



The International Association
for the Evaluation of
Educational Achievement

The Second IEA International Research Conference: Proceedings of the IRC-2006

VOLUME TWO



Proceedings of the IEA IRC-2006 (Vol. 2)

The Second IEA International Research Conference: Proceedings of the IRC-2006

**Volume 2: Civic Education Study (CivEd),
Progress in International Reading Literacy
Study (PIRLS), Second Information Technology
in Education Study (SITES)**



The International Association
for the Evaluation of
Educational Achievement

Copyright © 2007 International Association for the Evaluation of Educational Achievement (IEA)

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, electrostatic, magnetic tape, mechanical, photocopying, recording, or otherwise without permission in writing from the copyright holder.

ISBN: 978-90-806643-6-4

Copies of the *Proceedings of the IEA IRC-2006, Volumes 1 and 2*, can be obtained from:

IEA Secretariat

Herengracht 487

1017 BT Amsterdam, the Netherlands

Telephone: +31 20 625 3625

Fax: + 31 20 420 7136

Email: Department@IEA.nl

Website: www.iea.nl

Printing: Excel Digital Print, Freestyle Artworks, Wellington, New Zealand

Editing: Paula Wagemaker, Christchurch, New Zealand

Design and layout: Becky Bliss Design and Production, Wellington, New Zealand

Table of contents

Foreword	<i>vii</i>
<i>Hans Wagemaker</i>	
Civic knowledge of high-school students in Israel: Personal and contextual determinants	1
<i>Orit Ichilov</i>	
Factors that distinguish the most from the least effective schools in reading: A residual approach	15
<i>Constantinos Papanastasiou</i>	
Innovative pedagogical practices using technology: Diffusion patterns within schools	25
<i>Alona Forkush-Baruch and Dorit Tubin</i>	
Examining literacy, gender, and the home environment in PIRLS 2001 countries	39
<i>Kathleen Trong and Ann Kennedy</i>	
A comparison of fourth-graders' academic self-concept and attitudes toward reading, mathematics, and science in PIRLS and TIMSS countries	49
<i>Ann Kennedy and Kathleen Trong</i>	
Equating errors in international surveys in education	61
<i>C. Monseur, H. Sibberns, and D. Hastedt</i>	
Sociological perspectives on youth support for social movements	67
<i>Oren Pizmony-Levy</i>	
Online data collection in SITES 2006: Design and implementation	87
<i>Ralph Carstens, Falk Brese, and Barbara N. Brecko</i>	
Political attitudes and behaviors across adolescence and early adulthood: A comparison of IEA and European Social Survey findings	101
<i>Marc Hooghe and Britt Wilkenfeld</i>	
School climate for citizenship education: A comparison of England and the United States	115
<i>Gary Homana and Carolyn Barber</i>	
Home reading environments and children's reading performance: A comparative study of 25 countries	131
<i>Hyunjoon Park</i>	
Student government and voluntary organizations: A comparative study of Australia and the United States	145
<i>Gary Homana and Jeff Greene</i>	
Predicting the political involvement of European adolescents	157
<i>Rainer H. Lehmann</i>	
Reading achievements in urban and rural communities: A comparative analysis of equity in education	169
<i>Aisté Elígio</i>	
Secondary analysis of PIRLS 2001 Norwegian data	177
<i>Victor H. P. van Daal, A. Charlotte Begnum, Ragnar Gees Solheim, and Herman J. Adèr</i>	
The effect of multilingual policies on performance and progression in reading literacy in South African primary schools	191
<i>Sarah Howie, Elsie Venter, Surette van Staden, and Michelle van Gelder</i>	

Equity of achievement: A matter of education structures? <i>Ariane Baye and Christian Monseur</i>	199
Characteristics of United States Grade 4 language minority students in an international context: Findings from PIRLS 2001 <i>Laurence T. Ogle, David C. Miller, and Lydia B. Malley</i>	215
Teachers' practices in relation to students' civic engagement in three countries <i>Carolyn Barber</i>	233
Relationships between parental factors and children's reading behaviors and attitudes: Results from the PIRLS 2005 field test in Taiwan <i>Chia-Hui Chiu and Hwa-Wei Ko</i>	249
Online data collection in SITES 2006: Paper survey versus web survey—do they provide comparable results? <i>Barbara Nęza Brečko and Ralph Carstens</i>	261
Schools that exceed expectations: A cross-country comparison <i>Marjet Doupona Horvat and Alja Krevh</i>	271
Changes in reading variations and their relationship with socioeconomic status at school and individual levels in trend countries <i>Kajsa Yang-Hansen</i>	281
Analyzing trends in levels of reading literacy between 1970 and 2001 in Sweden <i>Monica Rosén</i>	295
A cross-country comparison of direct and indirect effects of parents' level of education on students' reading achievement <i>Eva Myrberg and Monica Rosén</i>	307

Foreword

As part of its mission, the International Association for the Evaluation of Educational Achievement is committed to the development of the community of researchers who work in the area of assessment both nationally and internationally. The association also has a commitment to provide policymakers with the types of data and analyses that will further their understanding of student achievement and the antecedent factors that are implicated in student learning.

As part of a larger strategy to achieve these broad goals, the IEA sponsors a research conference every two years as a means of providing opportunities for new researchers and more experienced scholars to meet, discuss, and present the findings of their work as it relates to the secondary analysis of IEA studies. The proceedings of the Second IEA International Research Conference, which was held in Washington DC, November 2006, and hosted by the Brookings Institution, are published here in two volumes.

The papers in Volume 1 of the proceedings have as their central focus the Trends in Mathematics and Science Study (TIMSS). Volume 2 brings together papers that focus on the Progress in International Reading Literacy Study (PIRLS), the Second Information on Technology in Education Study (SITES), and the Civic Education Study (CivEd).

IEA is grateful to everyone who participated in this conference and hopes that the papers provided here will interest those who work in the various areas of educational research represented in these pages.

We look forward to future contributions to our conferences, and hope that these papers not only contribute to our understanding of educational achievement but also lead to the development of the community of researchers involved in international and national assessment.

A handwritten signature in black ink, appearing to read 'Hans Wagemaker', with a stylized flourish extending to the right.

Hans Wagemaker PhD

EXECUTIVE DIRECTOR, INTERNATIONAL ASSOCIATION FOR THE
EVALUATION OF EDUCATIONAL ACHIEVEMENT

Civic knowledge of high-school students in Israel: Personal and contextual determinants¹

Orit Ichilov
Tel Aviv University
Tel Aviv, Israel

Introduction

Civic knowledge is gained from a great variety of sources (Ichilov, 1984; Niemi & Junn, 1998; Torney-Purta, Lehmann, Oswald, & Shulz, 2001; Torney-Purta, Schwille, & Amadeo, 1999). Past research on civic education suggests that students' performance is largely influenced by individual socioeconomic background and motivational factors. There has been little attention to the effects of school and classroom ideological and social attributes, such as the socioeconomic make-up or the political climate of the school or classroom. My point of departure is that opportunities and motivations to acquire civic knowledge may vary depending mainly on the ideological and socioeconomic texture of both the school and the classroom, as well as on students' background and personal characteristics. This paper, based on a multilevel analysis of information drawn from a sample of Israeli 11th-graders in 2000, examines the effects of a series of student-level and school-level independent variables on civic knowledge. Data for the study were collected in Israel as part of the International Association for the Evaluation of Educational Achievement (IEA)² comparative study of civic education in which 16 countries took part.

Schooling processes and citizenship education

A politically knowledgeable citizenry is a central goal in democracy. Knowledge about the basic democratic principles and procedures empowers citizens, and helps them make rational decisions and choices (Dewey, 1916/1966; Gutmann, 1987). Ample empirical evidence accumulated over several decades suggests that formal education explains more aspects of democratic citizenship than does any other factor (Almond & Verba, 1963; Barnes & Kaase, 1979;

Berelson, Lazarsfeld, & McPhee, 1954; Campbell, Converse, Stoke, & Miller, 1969; Campbell, Gurin, & Miller, 1954; Converse, 1972; Dahl, 1961; Dalton, 1988; Delli Caprini, & Keeter, 1996; Kamens, 1988; Nie, Powell, & Prewitt, 1969; Verba, Nie, & Kim, 1978; Wolfinger & Rosenstone, 1980). Yet how schooling does it remains an enigma, and the causal connection between various processes of formal education and democratic citizenship is pretty much an un-deciphered "black box" (Torney-Purta, 1997). The present study attempts to make a modest contribution to deciphering the puzzle, by focusing scholarly attention on the social and ideological attributes of the socializing contexts.

My basic assumption is that students' motivation and opportunities to acquire civic knowledge greatly depend on the social and ideological texture of both their school and their classroom. I argue that the nature and the content of educational influences concerning citizenship are structured by various social contexts, and that schooling effects cannot be envisioned solely as an encounter between individual students, possessing a variety of personal qualifications and resources, and educational institutions that vary in the quality of education they provide (Ichilov, 1991, 2002). While acknowledging the distinctiveness of each student, I consider individual students to be interdependent. This means that it is necessary to see individual behaviors and attitudes as partially contingent upon those of other individuals. Thus, civic orientations and knowledge of individual students can be understood within the larger social aggregates of which they are part. For instance, how interested in politics are a student's classmates? How frequently do they engage in political discussions? At the same time, however,

¹ This paper also appears in the August 2007 issue of *Political Psychology*.

² The International Association for the Evaluation of Educational Achievement (IEA) is an independent international co-operative of national research institutions and comparative studies of educational achievements. Since its inception in 1958, 18 cross-national studies have been done. The regular cycle of research projects encompasses learning in basic school subjects, as well as studies of particular interest to IEA member countries, such as civic education. The Israeli portion of the IEA study, on which this article is based, was financed by the Israeli Ministry of Education (Office of the Chief Scientist).

aggregated behavior can be understood as more than the simple accumulation of individually determined preferences. It represents the characteristics of the context in which students interact, and the climate of opinions, motivations, and behaviors within that context.

I consider the ideological and social texture of the school and classroom to be of central importance in determining students' motivations and opportunities to acquire civic orientations and knowledge. Schools convey to students a normative description of society and of their anticipated place in society as adults (Kamens, 1981). I argue that, even in school, students may experience being placed closer or further away from the center, where meaningful social assets are allocated. The development of status-related citizenship orientations therefore is linked to school and classroom contexts, and begins to emerge in the course of schooling. Thus, for example, motivation to acquire civic knowledge, opportunities to discuss politics with interested others, and aspirations for a lifestyle with active citizenship as an important component are closely linked to the ideological and socioeconomic texture of students' educational contexts.

Educational contexts greatly determine whom students associate with in school and out of school. Association with peers from a similar sociocultural background may solidify and reinforce status-related lifestyles that exist in the students' communities outside school (Ichilov, 1991). The concentration of minority and low socioeconomic students in low-prestige programs or in highly homogeneous classrooms may foster among these students feelings of alienation, rejection, and marginality in school, and of being discriminated against. Studies have shown that marginalized students are less engaged in school because they perceive fewer returns from education and limited opportunities for occupational mobility and for being politically effective and influential (Edwards & Foley, 1997; Fordhan & Ogbu, 1986; Johnson, Crosnoe, & Elder, 2001). Such feelings toward school as a social agency can be generalized and result in civic apathy or militancy, distrust for societal institutions, and a feeling that democracy is a game played chiefly by the elites.

Interaction among students in the upper tracks or in a classroom with a majority of well-to-do students may reinforce feelings of personal competence, acceptance,

success, and prospects for belonging to the elites in the future. These feelings can translate into greater motivation to acquire civic knowledge, and a stronger sense of political efficacy, reinforcing anticipation of political engagement in adulthood.

Israeli high schools as educational environments

Civic education is highly sensitive to the macro-political cultures of nations and the micro-political culture of schools within these nations (Mintrop, 2002; Schwille & Amadeo, 2002). Citizenship education in western democracies aims at inculcating a shared concept of citizenship that bridges ethnic, national, and socioeconomic rifts. It does so by incorporating particularistic identities and values, such as patriotism and national pride, and universalistic democratic codes, such as tolerance and respect for civil liberties. Current multiculturalism, however, often endorses social visions that do not necessarily value unity and cohesion. Minorities may resist and contest any attempt to bring multiplicity and heterogeneity into unity, arguing that such attempts entail a denial of difference and therefore are inherently oppressive (Young, 1990). Consequently, the kind of "civic knowledge" that schools impart is often considered part of the oppressive western hegemony, and it is advocated instead that "civic knowledge" should empower minorities in ways that enable them to dismantle their ideological scaffolding and to develop strategies and practices of resistance (Kanpol & McLaren, 1995; McLaren, 1997). Social rifts may thus radicalize minority groups, making them suspicious of appeals to certain ideals of "good citizenship," which they see as a demand that minorities should quietly learn to play by the rules of the majority (Kanpol & McLaren, 1995; McLaren, 1997; Samson, 1999; Young, 1990). Social divides may, likewise, nurture a decline in political participation and associational life, and encourage and entrench passive, inward-looking, and resentful forms of group identity that inhibit wider co-operation, dialogue, and solidarity (Kymlicka & Norman, 2000; McDonnell, Timpane, & Benjamin, 2000; Putnum, 1995; Schlozman, Verba, Brady, & Erkulwater, 1998). Under these circumstances, educating the younger generation for citizenship becomes an especially sensitive and difficult task (Byrne, 1997; Ichilov, 1999, 2004).

A major attribute of Israeli society is the wide and deep rifts between religious and non-religious groups, between Israeli Arabs and Jews, between the political left and the political right, and between the rich and the poor. These divides, that often intersect and overlap, represent contesting visions of Israel as a Jewish-democratic state and they profoundly shape Israel's political culture. I contend that these rifts permeate the schools in ways that may have an effect on both the implementation and the outcomes of civic education.

The very structure of the Israeli education system creates distinct ideological and educational climates in schools. Three types of school sub-systems exist within Israel's state system of education: Arab state schools, Hebrew religious state schools, and Hebrew regular (i.e., non-religious) state schools. The type of school that students attend is of central importance because these schools cater to distinct sectors within Israeli society (Ichilov, 1999). The present study focuses solely on academic (i.e., non-vocational) schools, thereby minimizing the effects of curricular differences. All schools follow the same civics curriculum and use the same textbook. A matriculation examination in civics is mandatory for all students at either the 11th or 12th grade, but unlike more central school subjects (such as mathematics, history, or English) that can be studied at an advanced level, civics is offered at a basic level only, and constitutes a very small fraction of the credits toward a matriculation certificate. The teaching of civics at the high school level occurs during the year in which students take the civics matriculation examination.

Non-religious Jewish students form the majority and attend mainly regular (non-religious) state schools. Originating from a great variety of socioeconomic and ethnic backgrounds, students' families share religiously non-observant lifestyles and ascribe the Jewishness of Israel mainly to national and cultural, rather than to theocratic, attributes (Ichilov, 1999). Religious state schools serve primarily Zionist orthodox Jews. The religious dimension overlaps to some extent with political and ideological controversies within Israeli society. After the Six Day War (in 1967) and the Israeli occupation of Judea and Samaria (the West Bank), which forms the heart of the biblical land of Israel, religious Jews became religiously and politically radicalized, and came to play a central role in political

movements and parties on the right side of the Israeli political map. Many settlements on the West Bank and the Gaza Strip were founded by religious nationalist groups, who form the spearhead of opposition to territorial concessions and the evacuation of settlements as part of a peace treaty with the Palestinian Authority (Don-Yehiya, 1994). Bar-Lev (1977) and Don-Yehiya (1994) claim that religious education at both the high school level and in higher education religious institutions (Yeshivot Leumiot) produces the next generations of religiously and politically radicalized individuals. In religious schools, students pray at school, and the dress codes are more conservative than in regular state schools. In some instances, classrooms or schools may be gender-segregated, with boys and girls studying separately (Schwarzwald, 1990).

Arab state schools serve mainly Israeli-Arab students. The term "Palestinian citizens of Israel" appears to express the self-definition of Israeli-Arabs and signifies a change in their collective identity. There is wide agreement among scholars that, since 1967, Israeli-Palestinians have undergone a radicalization process that involves a strengthening of the Palestinian national identity and a concomitant weakening of the Israeli civic identity (Landau, 1993; Rekhess, 1976, 1989). The 1987 Palestinian uprising greatly fostered the Palestinian identity of Israeli-Arabs (Schiff & Ya'ari, 1990). Islamic fundamentalism also gained strong hold among some segments of the Israeli-Arab population, as a form of collective identity. Israeli Palestinian-Arabs subscribe to the Palestinian historical narrative that clearly defies the claim of Jews to have a right to the land of Israel from days immemorial. Thus, it is not surprising that the national symbols of the State of Israel that represent Jewish themes are not an acceptable form of Israeli identity for the Palestinian-Arab minority. Israeli Palestinian-Arabs do not participate in the celebration of national holidays and memorial days, and they commemorate Israel's Independence Day as the day of the "Nakba," that is, the national disaster of the Palestinian people.

Arab educators and scholars claim that Israel has used the education system as a means of controlling its Palestinian-Arab citizens. The state-mandated curricula frequently are depicted as characterized by "the absence of any reference to Palestinian identity in history, literature, and social studies. Instead the curriculum offers a detailed Zionist narrative of history" (Rouhana,

1997, p. 86). The present state curriculum, however, compared to earlier versions, reflects some movement toward the inclusion of Arab and Palestinian identities (Al-Haj, 1995). Arab students take the matriculation examinations in their first language—Arabic. Arab schools are more conservative than are regular Hebrew state schools concerning school climate and teaching practices. Relationships between students and teachers tend to be more hierarchical and authoritarian (Al-Haj, 1993).

Existing rifts clearly carry over from the students' families and communities into the schools (Ichilov, 2000, 2003, 2004, 2005). Ichilov (2000) reports that students in Arab and in Hebrew religious schools seem more politicized than students in regular (non-religious) Hebrew state schools. For example, they are more likely than students in regular state schools to support the use of military power to retrieve occupied territories. Arab students are willing to engage in militant and illegal protest activities, as well as in legal active citizenship practices. The alienation of Arab students manifests itself in these students being less likely than students in Hebrew schools to support and show pride in Israel. Religious school students, in contrast, are the most patriotic students (Ichilov, 2005). In addition, Arab and religious school students participate more intensively than do regular state school students in out-of-school organizations. The greater politicization of Arab and religious Jewish students may result in stronger motivation to acquire civic knowledge. Similarly, radicalized students may consider the academic discussion of democratic principles sterile, irrelevant, and boring. Minority students may resent civic education as an attempt at "domesticating" them. These students may thus be unmotivated to acquire the kind of civic knowledge that schools impart. Furthermore, Arab and Hebrew religious schools provide a relatively conservative and "closed" school climate—one that is less conducive for inculcating civic knowledge and competences.

We must also take into account the fact that teachers are not neutral bystanders (Ichilov, 2002). Arab teachers attach lower importance than do teachers in Hebrew schools to classroom discussions of patriotism and national symbols, and to issues that are of greater concern for the Jewish population of Israel, such as global anti-Semitism. Arab teachers also assign greater importance to discussing views that

challenge the Zionist historical narrative. Teachers in religious state schools are the teachers most supportive of discussing patriotism and the national symbols of the state, and those least supportive of discussing the peace process, and Arab-Jewish relations in Israel (Ichilov, 2002). It is reasonable to assume that teachers may interpret events and curricular materials in the classroom based on their ideological stance, and that this, in turn, creates different climates of opinion in Hebrew and Arab schools, and in religious and non-religious Hebrew state schools. In summary, it seems that, through interactions with their teachers, school, and classmates, students can encounter political ideas mainly prevalent in their families and communities. This seems to be especially true in Arab and Hebrew religious state schools.

Having stressed the importance of the school as a socializing context, I should note that the home classroom is the major unit to which students belong. Most educational activities within Israeli high schools take place within it, with the exception of the instruction of several tracked school subjects such as mathematics. Students may thus experience more intensive exposure to their classmates than to other students. I should also mention that civics instruction takes place within untracked classrooms.

Individual characteristics of students

Studies of the effects of schooling on the acquisition of civic knowledge, attitudes, and competences demonstrate the importance of several socioeconomic and motivational characteristics of students. High levels of parental educational attainment, economic affluence, and substantial home literacy resources have a strong positive effect on students' civic knowledge (Ichilov, 2000; Niemi & Junn, 1998; Torney-Purta, Lehmann, Husfeldt, & Nikolova, 2002; Torney-Purta et al., 2001; Torney, Oppenheim, & Farnan, 1975). Also, students with high aspirations for further education attain high scores on civic knowledge assessments (Niemi & Junn, 1998; Torney-Purta, Lehmann, Oswald & Schulz, 2001; Ichilov, 1991). Citizenship orientations and behaviors are clearly an important component of status-related lifestyles, as indicated by studies that show a relationship between individuals' socioeconomic status and their voting behavior, their sense of political efficacy, and their involvement in public affairs (Easton, 1965;

Himmelweit, Humphreys, Jaeger, & Katz, 1981; Lane, 1959; Lipset, 1960; Milbrath, 1965; Stouffer, 1955).

These findings suggest that affluent and better-educated parents are more interested and involved than lesser-educated parents in politics, and that they are more likely to engage their children in political discussion and to motivate them to acquire knowledge about politics. The findings also suggest that anticipated lifestyles associated with further education motivate students to acquire civic knowledge. The importance of “anticipated future” for the formation of citizenship orientations and knowledge is evident in several studies. The report of the IEA International Civic Education Study (CivEd) (Torney-Purta et al., CivEd, 2001) reveals that “expected years of further education” had the strongest effect on civic knowledge among 14-year-old students in 28 countries, stronger than “home literacy resources” and “open classroom climate.” A similar variable was an important predictor of civic knowledge in the IEA Civic Education Study of 1971 (Torney et al., 1975) and in Niemi and Junn’s (1998) analysis of the National Assessment of Civics in the United States.

The evidence concerning gender differences in civic knowledge is less consistent. In a study conducted among Israeli high school students, Ichilov (1991) found that the positive effect of schooling was greater for males than for females. In other words, schooling reinforced the involvement, efficacy, and support of freedom of speech of males more than it did of females. A more recent study reported no significant differences between 14-year-old boys and girls in 27 countries in mean performance on a civic knowledge test (Torney-Purta et al., 2001). The disappearance of gender differences suggests that greater gender egalitarianism may have advanced the view that politics is not an exclusively male domain, encouraging female students to become interested and engaged.

Discussing and debating social and political issues in school is also an influential factor (Almond & Verba, 1963; Niemi & Junn, 1998). Apparently, political discussions allow for the transmission of information through active learning and can thus enrich students’ understanding of political issues. In addition, through political discussions in school, students become more skillful and thus perhaps more willing to discuss politics in other settings as well (Conover & Searing, 2000).

Based on the preceding discussion, my conclusion is that opportunities and motivations to acquire civic knowledge may vary depending on the ideological and socioeconomic texture of both the school and the classroom, and on students’ backgrounds and personal characteristics. I expect the contextual factors to have the greatest effect upon students’ civic knowledge.

Research questions and hypotheses

The present study examines the effects of a series of student-level and contextual independent variables on civic knowledge. The purpose of the analysis was to explore what percentage of the variance in students’ civic knowledge scores would be accounted for by student-level and contextual variables, and which variables, at each level of analysis, would account for such variance. It was expected that, overall, school-level variables, especially the type of school that students attend (Arab versus Hebrew), would have the greatest effect upon students’ civic knowledge test scores.

More specifically, the expectations were as follows:

- *Student-level variables:* Higher test scores were expected among those students who had better-educated parents, a large number of books at home, and high aspirations for further education, and who frequently engaged in political discussions. No gender differences in test scores were expected.
- *School-level variables:* Higher test scores were expected in Hebrew schools and in those schools with high percentages of students from socioeconomically well-off families. Higher test-scores were expected in classrooms with an open climate, and in classrooms where high proportions of classmates had well-educated parents and more books at home, frequently engaged in political discussions, and had high expectations for further education. The percentage of male students in the classroom was not expected to affect test scores.

Multilevel analysis of civic knowledge

Data and sample

The target population consisted of the age cohort of students ages 16 years to 16 years and 11 months at the time of the study and who were Grade 11 students within Israeli academic (i.e., non-vocational) state high schools. Data provided by the Central Bureau of Statistics (1996) estimated that the target population included about 80% of the relevant age cohort.

Uniform sampling procedures were set for all participating countries in the IEA Civic Education Study 2001, using a two-stage stratified cluster design. During the first stage, schools were sampled using a probability proportional to size (Foy, Rust, & Schleicher, 1996). During the second stage, the sample consisted of one intact classroom per school from the target grade (Grade 11). Sampled schools were drawn from a database provided by the Ministry of Education. The final sample consisted of 5,847 Grade 11 students in 157 Hebrew schools and 76 Arab schools. The total number of students assessed was 4,430, representing a participation rate of 83% (Amadeo, Torney-Purta, Lehmann, Husfeldt, & Nikolova, 2002). The sampling procedure made it possible to address phenomena at the individual student-level, as these were embedded within the school and classroom contexts.

The data were collected using a questionnaire that included a multiple-choice test of civic knowledge and skills, designed to have correct and incorrect answers. The items assessed knowledge of civic content as well as skills in interpreting material with civic or political content, including short-text passages and cartoons. Also included were items assessing economic literacy, but these were excluded from this study.

The questionnaire was developed by the International Steering Committee of the IEA in collaboration with the participating countries, and in Israel was translated from English into both Hebrew and Arabic. The Hebrew version was administered to students in Hebrew schools, and the Arabic version was administered to students in the Arab school system (Ichilov, 2000; Torney-Purta, 1996).³ The final version of the test was preceded by an elaborate process of piloting, and was adopted based primarily on confirmatory factor analysis and IRT modeling, indicating that the test meets high psychometric standards across countries (Amadeo et al., 2002; Torney-Purta & Klandl-Richardson, 2002). The IEA resolved that test items must remain confidential.

Method and variables

We assessed the effect of both individual-level and school-level variables on civic knowledge using hierarchical linear modeling (HLM). Multiple-level analysis suited both our theoretical assumptions concerning the acquisition of civic knowledge, and the structure of our data (Bryk & Raudenbush, 1987; Hoffman, 1997; Kreft & De Leeuw, 1998).⁴

The dependent variable

Civic knowledge test score: The outcome variable was the number of correct answers on a civics test consisting of 28 items. The items tested students' knowledge and understanding of the foundational principles and processes of democracy. The alpha score for the civic knowledge scale was .77.

Student-level independent variables

Two sets of variables were included:

a) *Demographic and socioeconomic variables*

- *Gender:* Coded 1 for males and 0 for females.
- *Parental education:* A composite index representing the mean number of years of mother's and father's education, ranging between 4 = partial elementary education and 15 = college education.
- *Home literacy resources:* Number of books at home, ranging from 1 = none to 6 = more than 200. Given the great variation among participating countries in income, standard of living, prestige grading of occupations, etc., in IEA international studies, number of books at home is considered a proxy for socioeconomic status (SES). This variable correlates highly with parental education (Beaton, Mullis, Martin, Gonzales, Kelly, & Smith, 1996).

b) *Motivation and activity variables*

- *Expected years of further education:* Ranging from 1 = 0 years to 7 = over 10 years.
- *Frequency of political discussion:* A composite index reflecting the mean tendency to discuss politics with parents, teachers, and peers. 1 = never, 2 = rarely, 3 = sometimes, 4 = frequently. Alpha = .65.

³ The participating countries in which the upper-secondary questionnaire was developed and applied were Chile, Colombia, Cyprus, Czech Republic, Estonia, Hong Kong SAR, Israel, Latvia, Norway, Poland, Portugal, Russian Federation, Slovenia, Sweden, and Switzerland.

⁴ This technique and its advantages are described in the cited literature. I mention here only a few of the key features of HLM that are of special relevance for this study. First, it explicitly takes into account the fact that students are clustered within schools, and are not statistically independent observations. Standard errors can be underestimated when this within-school clustering is not taken into account. Second, it involves distinguishing the variation that occurs among students within a school from the variation that occurs among schools.

School and classroom-level variables

Two types of variables were included: variables that apply to the entire school (such as school type and school SES), and aggregated classroom variables that characterized the context of each classroom. Empirically, we treated both types of variables as school-level variables, because only one classroom was randomly sampled within each sampled school.

- *School-type*: two dummy variables:
 - Arab school = 1, Hebrew state schools = 0.
 - Religious Hebrew schools = 1, Hebrew state schools = 0.
- *SES*: A composite index reflecting the socioeconomic composition of the student body. Schools were grouped into 10 categories ranging from the upper 10% of schools (those with the lowest proportion of disadvantaged students) to the lowest 10% of schools (those with the highest proportion of disadvantaged students). The index used was produced by the Ministry of Education.
- *Time of testing*: The grade at which students take the civics matriculation examination: 1 = at Grade 12, 0 = at Grade 11.
- *Aggregated school variables*: Scores as above, unless mentioned otherwise.
- *Gender*: Percentage of male students in a classroom.
- *Classroom climate*: An aggregated variable representing the mean score of items reflecting open classroom climate. Based on students' perceptions of classroom practices associated with an open classroom climate. Included, for example, were items concerning the presentation of different views on various issues in the classroom, teachers encouraging students to form their own views, and students' ability to freely express themselves and to disagree with their teachers. These were coded 1 = never, 2 = rarely, 3 = sometimes, 4 = frequently. Thus, the higher the mean score, the more "open" students perceived the classroom climate to be. Alpha = .82.
- *Parental education*: The average number of years of education for parents of students within each classroom.
- *Home literacy resources*: An aggregated variable reflecting the mean number of books at home for students within each classroom.

- *Expected years of further education*: An aggregated variable measuring the mean number of years of anticipated education for students within each classroom.
- *Discussing politics with parents, teachers, and peers*: Mean frequency of discussing politics with each of the following: parents, teachers, and classmates of students within each classroom.

Results

Tables 1 and 2 present descriptive statistics and correlations pertaining to student-level and school-level variables. The interesting statistics at both the student and school levels reveal that the mean scores of the civics test were fairly low (18.53 and 17.55 correct answers or 66.17% and 62.89% correct answers, respectively). At both levels, test scores positively correlated with similar variables (differentially defined as an individual or as a classroom attribute) and reflected socioeconomic, motivational, and behavioral components. These included number of books, expected years of further education, and parental education.

The study required two levels of analysis:

1. *The student level, which examines the effects of students' characteristics on civics test scores*: Students' characteristics were introduced into the model in two steps: first demographic and socioeconomic variables ("Model 1," column 1 in Table 1), followed by the inclusion of motivation and activity variables ("Model 2," column 2 in Table 1).
2. *The school level, which explores between-school differentiation in mean test scores*: School-level variables were introduced into the model in three steps: first school type ("Model 3," column 3 in Table 1), and then school SES ("Model 4," column 4 in Table 1), followed by aggregated contextual classroom variables ("Model 5," column 5 in Table 1).

Two equations were computed. At the student level, the test score of student i in school j is predicted as follows:

$$1. (\text{score})_{ij} = \beta_0j + \beta_1(\text{gender})_{ij} + \beta_2(\text{parental education})_{ij} + \beta_3(\text{number books})_{ij} + \beta_4(\text{expected years})_{ij} + \beta_5(\text{discussion})_{ij} + r_{ij}.$$

At the school level, only the intercept $[\beta_0j]$ was allowed to vary among schools. The slopes of all

Table 1: Descriptive Statistics and Correlations: Student-level Variables (N = 4,642) (All students)

	1	2	3	4	5	Mean	SD
1. Gender, male	–					0.50	0.50
2. No. of books	0.03*					4.57	1.30
3. Expected years of further education	-0.07**	0.26**				3.72	1.36
4. Political discussions	0.02	0.13**	0.13**			2.54	0.63
5. Parental education (years)	-0.06**	0.33**	0.28**	0.01		11.92	2.71
6. Civic knowledge score	0.01	0.21**	0.31**	0.05**	0.36**	18.53	6.98

Note: * $p < 0.05$; ** $p < 0.01$.

Table 2: Descriptive Statistics and Correlations: School-level Variables (N = 216) (All Schools)

	1	2	3	4	5	6	7	8	9	10	11	Mean	SD
1 Gender, male	–											0.50	0.25
2 No. of books	0.04											4.48	0.62
3 Expected years of further education	-0.12	0.61**										3.61	0.64
4 Political discussions	0.08	0.17*	0.21**									2.52	0.22
5 Parental education	-0.10	0.55**	0.59**	-0.15*								11.71	1.68
6 Hebrew state school	-0.15*	-0.12	0.13*	-0.38**	0.51**							0.52	0.50
7 Religious Hebrew school	0.03	0.24**	0.01	-0.10	0.15*	-0.42						0.14	0.35
8 Arab school	0.13	-0.05	-0.15*	0.47**	-0.65**	-0.74	-0.29**					0.34	0.47
9 SES	0.05	-0.30**	-0.36**	0.12	-0.58**	-0.34	0.14*	0.25**				5.03	2.75
10 Testing time	-0.14*	0.08	0.07	-0.43**	0.44**	0.44	0.24**	-0.65**	-0.21**			0.55	0.50
11 Classroom climate	0.01	0.22**	0.22**	0.44**	0.22**	0.00	0.12	-0.09	-0.05	0.01		2.95	0.24
12 Civic knowledge score	-0.05	0.55**	0.66**	-0.03	0.75**	0.36	0.16*	-0.51**	-0.41**	0.32**	0.32**	17.55	5.14

Note: * $p < 0.05$; ** $p < 0.01$.

independent variables were fixed because of a lack of theoretical rationale for the expectation that their impact on test scores would vary, and because their between-school variance did not reach statistical significance. The following equation defines the intercept as a function of the school-level variables and a random error:

$$2. \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{school: Arab})_j + \gamma_{02}(\text{school: religious})_j + \gamma_{03}(\text{school SES})_j + \gamma_{04}(\text{testing time})_j + \gamma_{05}(\% \text{ male})_j + \gamma_{06}(\text{mean parental education})_j + \gamma_{07}(\text{mean number books})_j + \gamma_{08}(\text{mean expected years})_j + \gamma_{09}(\text{mean discussion})_j + \gamma_{10}(\text{mean climate})_j + v_{0j}.$$

All variables at the student and school level (with the exception of the dummy variables) were centered around the grand mean. Thus, the intercept represents the average test score of a female student characterized by parents with elementary education, an average number of books at home, and average motivations and activities score, who studies in a Hebrew state school, with average school characteristics, and who takes the civics test at Grade 11.

We began the analyses of students' performance on the civics test by using a preliminary unconditional model to determine the amount of variation that occurs among students within schools and the amount of

variation that occurs between schools. The estimation of the variance components from the unconditional model for civic knowledge revealed that 46.70% of the variation in students' test scores occurred between schools. We then proceeded to examine the influence of student-level and school-level variables on students' test scores. The results of the analyses are presented in Table 3, and will be discussed step by step.

We began our analysis by examining whether demographic and socioeconomic differences were evident in students' civics test scores. The data ("Model 1," column 1 in Table 3) reaffirmed the findings of many previous studies concerning the effects of students' socioeconomic background, by revealing that students with more books at home scored higher than students with fewer books at home, and students whose parents were better educated scored higher than students with less-educated parents. In addition, contrary to our expectation, male students scored higher than female students. The effect of these variables remained significant throughout the next facets of our analysis (with the exception of "number of books," which became insignificant at the last phase of the analysis). However, while the introduction of the motivation and activity variables reinforced the observed effect of gender ("Model 2," column 2 in Table 3) and remained strong throughout the various phases of the analysis, the effect of parental education and number of books diminished. Test scores were positively affected by a student's educational aspirations (expected years of further education) and by a student's tendency to discuss politics with others.

As was reported previously, estimation of the variance components from the unconditional model for civic knowledge revealed that 46.70% of the variation in students' test scores occurred between schools. To determine the unique effect of each school variable, we compared each phase in Model 3 with each phase in Model 2, and each time, we deleted the percentage of reduction of the between-schools variance in the previous phase. The addition of school type (column 3) revealed that Arab schools scored lower than Hebrew state schools on the civic knowledge test. School type reduced 30.41% of the between-schools variance in civic knowledge. The direct effect of school type remained significant with the introduction of school SES (column 4). Lower SES related to lower test scores. The socioeconomic composition of the student body

in each school reduced another 8.20% of the between-schools variance in test scores. With the addition of classroom contextual variables (column 5), the direct effect of school SES became insignificant and some context effects became evident. This suggests that school SES did not directly affect civics test scores, and that classroom contextual variables mediated its effect. Furthermore, classroom contextual variables reduced an additional 28.69% of the between-schools variance in test scores. Open classroom climate made a positive contribution to students' test scores: the higher the average number of expected years for further education among the students in a classroom, the more likely they were to gain the higher scores on the test. The same relationship held for level of parental education: the higher the level of parental education among the students in a class, the more likely it was for these students to gain the higher scores on the test.

Overall, the contextual variables (introduced in columns 3 to 5) reduced 67.3% of the between-schools variance in test scores. School type and classroom contextual variables made the largest contribution to the reduction of between-schools variance.

Conclusion

My basic assumption was that students' motivation and opportunities to acquire civic knowledge would greatly depend on the social and ideological texture of both the school and the classroom. The results support my contention that contextual effects outweigh individual socioeconomic and motivational characteristics, and play a consequential role in determining students' civic knowledge scores. As expected, the type of school that students attend strongly affected civics test scores, with Arab students scoring lower than Hebrew school students. The direct effect of school type persisted when school SES and contextual classroom variables were introduced.

These results suggest that the lower civics scores among Arab students can be partially attributed to and may reflect the rift between Israeli-Jews and Israeli-Arabs. In other words, Arab students could be resentful of or indifferent to the kind of "civic knowledge" that schools are expected to impart. Arab teachers are also unenthusiastic about teaching their students how to become citizens in a Jewish-democratic state (Ichilov, 2003, 2004, 2005). Such findings may not be unique to Israeli society alone. Many western

Table 3: Gamma Coefficients from Hierarchical Linear Models Analysis of Civic Knowledge Scores (N = 4,642) (Standard Errors in Brackets)

Variable	Model 1: Students' Background	Model 2: Model 1+ Students' Motivations and Activities	Model 3: Model 1+ School Type	Model 3: Model 1+ School Type + SES	Model 3: Model 1+ School Type + SES + Classroom Composition
STUDENT-LEVEL VARIABLES					
Constant	13.64** (0.71)	(0.54) 15.08**	11.06** (0.08)	(0.62) -13.10**	12.84** (3.48)
Demographic and socioeconomic variables					
Gender, male	0.63** (0.20)	(0.20) 0.75**	0.74** (0.19)	(0.20) 0.75**	0.76** (0.20)
Parental education	0.21** (0.04)	(0.04) 0.13**	0.17** (0.04)	(0.04) 0.10*	0.14** (0.04)
No. of books	0.27** (0.07)	(0.07) 0.16*	0.15** (0.06)	(0.07) 0.12	0.17* (0.06)
Motivation and activity variables					
Expected years of further education		0.59** 0.60**	(0.07) (0.07)	0.61** 0.54**	(0.07) (0.08)
Political discussions		0.58** 0.64**	(0.14) (0.14)	0.62** 0.63**	(0.14) (0.14)
SCHOOL-LEVEL VARIABLES					
School type (Hebrew state school is reference)					
Arab			-5.20** (0.61)	(0.63) -2.74**	-4.29** (0.79)
Hebrew religious			0.07 (0.59)	(0.62) -0.17	0.9 (0.58)
School's SES: Student body composition				-0.51** -0.06	(0.09) (0.08)
Civics matriculation testing time: 1 = 12th, 0 = 11th grade					0.01 (0.48)
Percentage of male students					0.30 (0.84)
Mean no. years of parents' education					0.65* (0.32)
Mean no. of books					0.96 (0.53)
Mean expected years of further education					2.39** (0.52)
Mean frequency of political discussions					-1.45 (1.33)
Mean classroom climate score					3.15** (1.00)
Between-school variance components	21.02	19.76	13.75	12.13	6.46
Percentage reduction in between-schools variance components (comparing Model 2)	–	–	30.41	8.20	28.69

Note: * $p < 0.05$; ** $p < 0.01$.

societies are becoming increasingly multicultural. The unifying effect of citizenship traditionally founded on a shared collective memory, cultural togetherness, and nationality may be missing in multicultural societies, and it could be that students of citizenship education are neglecting to take into account the likelihood that citizenship and multiculturalism are forces pulling in diverse directions—unity versus disunity. Thus, minorities that are unable or do not wish to be absorbed and integrated into mainstream society could resent the type of “civic knowledge” that represents the ideas of the dominant majority.

School SES also affected test scores: the lower the SES, the lower the test scores. However, the direct effect of SES became insignificant when classroom contextual variables were introduced, suggesting that the effect of SES is mediated by classroom attributes. The results indicate that studying in a classroom that has an open classroom climate, with classmates who have high expectations for further education and who have well-educated parents, improves students’

test scores. This implication, in turn, suggests that when highly motivated students from affluent and educated homes set the tone, less motivated students and students from less advantaged backgrounds may follow. In these situations, it is possible that aspirations for a lifestyle with active citizenship as an important component could be induced, and likewise motivations and opportunities to acquire civic knowledge.

At the individual level, the results reaffirm findings of previous studies, namely that more expected years of further education, more books at home, and a greater tendency to discuss politics improve students’ test scores. While gender differences in civic knowledge are reportedly disappearing in many countries (Amadeo et al., 2002; Torney-Purta et al., 2001), Israeli male students scored higher than female students on the civics test. The persistence of gender differences among Israeli students should be explored, especially as politics might still be considered a predominantly male domain, with female students accordingly less motivated to be interested and engaged in this area.

References

- Al-Haj, M. (1993). Education for democracy in Arab schools: Problems and challenges. In O. Ichilov (Ed.), *Citizenship education in democracy* (in Hebrew) (pp. 23–43). Ramat-Gan: Massada and Tel-Aviv University, School of Education, Sociology of Education and the Community Unit.
- Al-Haj, M. (1995). *Education, empowerment, and control: The case of the Arabs in Israel*. Albany, NY: State University of New York Press.
- Almond, G. A., & Verba, S. (1963). *The civic culture: Political attitudes and democracy in five nations*. Princeton, NJ: Princeton University Press.
- Amadeo, J. A., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement: An IEA study of upper secondary students in sixteen countries*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Bar-Lev, M. (1977). *Graduates of yeshiva high schools in Israel: Between tradition and modernity* (in Hebrew). Unpublished doctoral dissertation, Bar-Ilan University, Ramat-Gan.
- Barnes, S. H., & Kaase, M. (1979). *Political action: Mass participation in five western democracies*. Beverly Hills, CA: Sage.
- Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematical achievements in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.
- Berelson, B. R., Lazarsfeld, P. F. & McPhee, W. N. (1954). *Voting: A study of opinion formation in a presidential campaign*. Chicago: University of Chicago Press.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101, 147–158.
- Byrne, S. (1997). *Growing up in a divided society: The influence of conflict on Belfast schoolchildren*. Teaneck, NJ: Fairleigh Dickinson University Press.
- Campbell, A., Converse, P. E., Stoke, D. E., & Miller, W. E. (1969). *The American voter*. New York: Wiley.
- Campbell, A., Gurin, G., & Miller, W. E. (1954). *The voter decides*. Evanston, IL: Row, Peterson.
- Central Bureau of Statistics. (1996). *Statistical abstracts*. Jerusalem: Author.
- Conover, P., & Searing, D. D. (2000). A political socialization perspective. In L. M. McDonnell, P. M. Timpane, & R. Benjamine (Eds.), *Rediscovering the democratic purposes of education* (pp. 91–127). Lawrence, KS: The University Press of Kansas.

- Converse, P. E. (1972). A change in the American electorate. In A. Campbell & P. E. Converse (Eds.), *The human meaning of social change* (pp. 263–337). New York: Russell Sage Foundation.
- Dahl, R. A. (1961). *Who governs? Democracy and power in an American city*. New Haven, CT: Yale University Press.
- Dalton, R. J. (1988). *Citizen politics in western democracies*. Chatham, NJ: Chatham House.
- Delli Carpini, M. X., & Keeter, S. (1996). *What Americans know about politics and why it matters*. New Haven, CT: Yale University Press.
- Dewey, J. (1916) (1966). *Democracy and education*. New York: Macmillan.
- Don-Yehiya, E. (1994). The book and the sword: The nationalist yeshivot and political radicalism in Israel. In M. Marty & S. Appleby (Eds.), *Accounting for fundamentalism: The dynamic character of movements* (pp. 264–302). Chicago: Chicago University Press.
- Easton, D. (1965). *A framework for political analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Edwards, B., & Foley, M. W. (1997). Social capital and the political economy of our discontent. *American Behavioral Scientist*, 40(5), 669–679.
- Fordhan, S., & Ogbu, J. U. (1986). Black students' school success: Coping with the burden of "acting white". *Urban Review*, 18, 176–206.
- Foy, P., Rust, K., & Schleicher, A. (1996). Sample design. In M. O. Martin & D. J. Kelly (Eds.), *Third International Mathematics and Science Study technical report* (Vol. 1) (pp. 4.1–4.17). Chestnut Hill, MA: Boston College.
- Gutmann, A. (1987). *Democratic education*. Princeton, NJ: Princeton University Press.
- Himmelweit, H. T., Humphreys, P., Jaeger, M., & Katz, M. (1981). *How voters decide*. London: Academic.
- Hoffman, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, 23(6), 723–744.
- Ichilov, O. (1984). *The political world of children and adolescents* (in Hebrew). Tel-Aviv: Yachdav.
- Ichilov, O. (1991). Political socialization and schooling effects among Israeli adolescents. *Comparative Education Review*, 35(3), 430–447.
- Ichilov, O. (1999). Citizenship education in a divided society: The case of Israel. In J. Torney-Purta, J. Schwille, & J. A. Amadeo (Eds.), *Civic education across countries: Twenty-four national case studies from the IEA Civic Education Project* (pp. 371–395). Amsterdam: IEA.
- Ichilov, O. (2000). *Citizenship orientations of 11th grade students and teachers in the Israeli Hebrew and Arab high schools* (Israeli national report).
- Ichilov, O. (2002). Differentiated civics curriculum and patterns of citizenship education: Vocational and academic programs in Israel. In D. Scott & H. Lawson (Eds.), *Citizenship education and the curriculum* (pp. 81–111) (Vol. 3 in the International Perspectives on Curriculum series). Westport CT: Greenwood.
- Ichilov, O. (2003). Teaching civics in a divided society: The case of Israel. *International Studies in Sociology of Education*, 13(3): 219–243.
- Ichilov, O. (2004). *Political learning and citizenship education under conflict*. New York and London: Routledge.
- Ichilov, O. (2005). Pride in one's country and citizenship orientations in a divided society. *Comparative Education Review*, 49(1), 44–62.
- Johnson, M. K., Corsnoe, R., & Elder, G. H, Jr. (2001). Students' attachment and academic engagement: The role of race and ethnicity. *American Sociological Review*, 74(4), 318–341.
- Kamens, D. H. (1981). Organizational and institutional socialization in education. *Research in Sociology of Education and Socialization*, 2, 208–219.
- Kamens, D. H. (1988). Education and democracy: A comparative institutional analysis. *Sociology of Education*, 61, 114–127.
- Kanpol, B., & McLaren, P. (Eds.). (1995). *Critical multiculturalism: Uncommon voices in a common struggle*. Westport, CT: Bergin & Garvey.
- Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.
- Kymlicka, W., & Norman, W. (Eds.). (2000). *Citizenship in diverse societies*. New York: Oxford University Press.
- Landau, J. (1993). *The Arab minority in Israel, 1967–1991: Political aspects*. Oxford: Clarendon Press.
- Lane, R. E. (1959). *Political life*. New York: Free Press.
- Lipset, S. M. (1960). *Political man*. New York: Doubleday.
- McDonnell, L. M., Timpane, P. M., & Benjamin, R. (Eds.). (2000). *Rediscovering the democratic purposes of education*. Lawrence, KS: The University Press of Kansas.
- McLaren, P. (1997). *Revolutionary multiculturalism: Pedagogies of dissent for the new millennium*. Boulder, CO: Westview.
- Milbrath, L. W. (1965). *Political participation*. Chicago: Rand McNally.
- Mintrop, H. (2002). Teachers and civic education instruction in cross-national comparisons. In G. Steiner-Khamsi, J. Torney-Purta, & J. Schwille (Eds.), *New paradigms and recurring paradoxes in education for citizenship: An international comparison* (pp. 61–85). Amsterdam: JAI Press, Elsevier Science.
- Nie, N. H., Powell, B., & Prewitt, K. (1969). Social structure and political participation. *American Political Science Review*, 63, 361–378.

- Niemi, R. G., & Junn, J. (1998). *Civic education: What makes students learn?* New Haven, CT: Yale University Press.
- Putnam, R. D. (1995). Bowling alone: America's declining social capital. *Journal of Democracy*, 6(1), 65–78.
- Rekhess, E. (1976). *The Arabs in Israel after 1967: The exacerbation of the orientation of problem* (in Hebrew). Tel-Aviv: Shiloach Center, Tel-Aviv University.
- Rekhess, E. (1989). Israeli Arabs and Arabs of the West Bank and the Gaza Strip: Political ties and national identification. *Hamizrah Hebadash (The New East)*, 32, 165–191.
- Rouhana, N. (1997). *Palestinian citizens in an ethnic Jewish state*. New Haven, CT: Yale University Press.
- Samson, C. (1999). The dispossession of the Inu and the colonial magic of Canadian liberalism. *Citizenship Studies*, 3(1), 5–25.
- Schiff, Z., & Ya'ari, E. (1990). *Intifada: The Palestinian Uprising—Israel's Third Front*. New York: Simon and Schuster.
- Schlozman, K. L., Verba, S., Brady, H., & Erkulwater, J. (1998). *Why can't they be like we are? Understanding the generation gap in participation*. Unpublished paper, Boston College.
- Schwarzwalz, J. (1990). *A research perspective on religious public education in Israel*. Ramat-Gan: Bar-Ilan University Press.
- Schwille, J., & Amadeo, J. A. (2002). The paradoxical situation of civic education in schools: Ubiquitous and yet elusive. In G. Steiner-Khamsi, J. Torney-Purta, & J. Schwille (Eds.), *New paradigms and recurring paradoxes in education for citizenship: An international comparison* (pp. 105–137). Amsterdam: JAI Press, Elsevier Science.
- Stouffer, S. A. (1955). *Communism, conformity and civil liberties*. Garden City, NJ: Doubleday.
- Torney, J. V., Oppenheim, A. N., & Farnen, R. F. (1975). *Civic education in ten countries: An empirical study*. New York: John Wiley and Sons.
- Torney-Purta, J. (1996). The Second IEA Civic Education Project: Development of content guidelines and items for a cross-national test and survey. *Canadian and International Education*, 25(2), 199–215.
- Torney-Purta, J. (1997). Links and missing links between education, political knowledge, and citizenship. *American Journal of Education*, 105, 447–457.
- Torney-Purta, J., & Klandl-Richardson, W. (2002). An assessment of what fourteen-year-olds know and believe about democracy in twenty-eight countries. In W. C. Parker (Ed.), *Education for democracy* (pp. 185–211). Greenwich, CT: IAP Publishing.
- Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement among upper-secondary students*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Torney-Purta, J., Lehmann, R., Oswald, H., & Schultz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Torney-Purta, J., Schwille, J., & Amadeo, J. A. (Eds.). (1999). *Civic education across countries: Twenty-four national case studies from the IEA Civic Education Project*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Verba, S., Nie, N., & Kim, J. (1978). *Participation and political equality: A seven nation comparison*. New York: Cambridge University Press.
- Wolfinger, R. E., & Rosenstone, S. J. (1980). *Who votes?* New Haven, CT: Yale University Press.
- Young, I. M. (1990). The ideal of community and the politics of difference. In L. Nicholson (Ed.), *Feminism/Postmodernism* (pp. 300–323). New York: Routledge.

Factors that distinguish the most from the least effective schools in reading: A residual approach

Constantinos Papanastasiou
University of Cyprus
Nicosia, Cyprus

Abstract

The universal importance of reading as one of the components of literacy was highlighted when UNESCO (the United Nations Educational, Scientific and Cultural Organization) declared 1990 the “International Year of Literacy.” This declaration confirmed the widely held belief that one of the main purposes of education is to ensure school students acquire the necessary skills to read with understanding. Worldwide, primary schools generally are responsible for teaching basic literacy skills. However, throughout the world, societies vary in terms of the degree of equality among their communities and among their schools. Given that, the question arises as to why some schools have higher average student achievement than would be expected and why some have lower average student achievement despite their

home and school circumstances. This study was based on data from the student questionnaire and tests of the Progress in International Reading Literacy Study (PIRLS), conducted under the auspices of IEA. The sample included 3,001 students whose average age was 9.7 at the time of testing. The statistical analysis used to distinguish the more effective and less effective schools was based on the residuals. From this analysis, six factors explaining school differences in reading achievement were identified. The first, and strongest, factor related to class activities subsequent to the reading class. The second most important factor related to reading outside school. The third factor related to time spent on homework, the fourth to attitudes toward reading, the fifth to activities during teaching, and the sixth to school climate.

Introduction

The universal importance of reading as one of the components of literacy was highlighted when UNESCO (the United Nations Educational, Scientific and Cultural Organization) declared 1990 the “International Year of Literacy.” Reading skills are fundamental to most everyday activities as well as necessary for personal intellectual growth. Reading is one of the most important abilities students acquire as they progress through their early school years. In a global society, a literate population is essential to a nation’s social and economic development. Therefore, one of the main purposes of education in all countries is to ensure that school students acquire the skills they need to read with understanding. Because reading is vital to every child’s development, the International Association for the Evaluation of Educational Achievement (IEA) conducted the Progress in International Reading Literacy Study (PIRLS), which took place in 2001. PIRLS defined reading literacy in a way that applies not only to young readers but across all ages: *Reading literacy is the ability to understand and use those written language forms required by society and/*

or valued by the individual.

Due to the continual struggle to improve education worldwide, the issue of school effectiveness has attracted considerable attention. School effectiveness as an area of study seeks to improve educational practice by studying how schools discharge their role as institutions of learning and instruction. More specifically, this area of interest looks to define what makes for a successful school. Although we might assume that effective schools are simply those with high average student achievement, recent literature makes it clear that high achievement often depends on the composition of the student intake (Martin, Mullis, Gregory, Hoyle, & Shen, 2000). It is also important to take into account the difficulty of the educational task when evaluating the effectiveness of a school. All societies vary in terms of the degree of equality among their schools: some schools are relatively rich and others are relatively poor; some schools are well equipped and others less well equipped. In extreme cases, we find schools located in affluent communities with children whose parents do everything possible to help their children’s

learning, and schools located in poor communities where parents believe their children's learning is the school's concern and not their own (Postlethwaite & Ross, 1992). But there are exceptions. Some schools serving poor communities perform well above the level we might expect given their circumstances, and some schools serving affluent communities perform well below the expected level. The schools that perform above the expected level are called *more effective*, and the schools that perform below their expected level are called *less effective* (Postlethwaite & Ross, 1992). The question that arises is: *What are the factors related to teaching that distinguish the more effective schools from the less effective schools?*

This inconsistency is mainly due to four reasons:

1. Some schools are located in affluent areas;
2. Schools with high achievement have more and better resourcing;
3. Schools with high average scores have principals who are good managers; and
4. Schools with high average scores have good teachers.

Good teachers are defined as teachers who get feedback systematically from the students, demand a lot from their students, know their subject matter, know how to structure the material to be learned, give help to those who are having problems mastering some objectives, and keep good order in the classroom (Postlethwaite & Ross, 1992).

PIRLS, conducted every five years by IEA, provides a tool to investigate both student achievement and school effectiveness, as it factors in the varying influences of instructional contexts and practices as well as home environment. Consistent with IEA's primary purpose of providing policy and instructionally relevant information, PIRLS includes a full range of context questionnaires. Since the home environment plays such an important role in providing children with early language and literacy experience, parents and caregivers are asked to complete a questionnaire about various aspects of fostering reading development. Additionally, students are asked about their reading, both within and outside of school (Mullis, Martin, Gonzalez, & Kennedy, 2003). Questionnaires are also administered to school principals and teachers. Gathering information about children's experiences in learning to read, together with reading achievement on the PIRLS tests, allows identification of factors or

a combination of factors associated with high levels of reading literacy.

PIRLS and the other IEA studies also offer objective information on student performance from different countries and cultures, thereby providing international and national policymakers and educators with an important data source (Mullis et al., 2000). These international studies, which focus on students' academic outcomes in relation to the school, teachers, student, and home environment (Gadeyne, Ghesquiere, & Onghena, 2006) not only give direction to policymakers but also enable them to develop improvement strategies more effectively (Brown, Duffield, & Riddell, 1995; Grobler, Grobler, & Esterhuyse, 2001; Nasser & Birenbaum, 2005; Secada, 1992).

Research studies on school effectiveness have attracted considerable political support in several countries (Luyten, Visscher, & Witziers, 2005). This kind of research aims to identify the factors that contribute to effective education and especially those that schools can implement (Creemers & Reezigt, 2005; Scheerens & Demeuse, 2005). One focal point of school effectiveness research is to highlight the characteristics of schools and classrooms that are associated with differences in school effectiveness. If we know the special characteristics of an effective school, especially those relating to the sphere of features that could be changed, then we are in a position to improve underperforming schools by encouraging them to adopt those characteristics (Luyten et al., 2005). Another objective of school effectiveness research is to identify what works in education and why (Creemers & Reezigt, 2005). Furthermore, the extent to which schools vary in effectiveness and the school factors that seem to promote effectiveness are academically interesting.

Many studies have focused on the fact that the composition of the student body has a substantial impact on achievement over and beyond the effects associated with students' individual abilities and social class. Some studies report that schools with low social class intakes have certain disadvantages associated with their context (Baumert, Stanat, & Watermann, 2005; Opdenakker & Van Damme, 2005; Van de Grift & Houtveen, 2006; Wilms, 1992). Other studies argue that among the factors most influencing performance are clear, well-organized teaching that keeps students

actively involved in the learning process; teaching that motivates students and connects learning to their extant knowledge; teachers who maintain students' active involvement in lessons; and classes that are efficiently organized and well structured. Recent research reveals that a major problem in underperforming schools is that their students are not given opportunity to attain the minimum objectives of the curriculum. This lack of opportunity can be a product of numerous factors, including textbooks that are inappropriate for attaining curriculum objectives, insufficient time allotted for learning and teaching, and poor teaching that does not stimulate students (Van de Grift & Houtveen, 2006). This same study found that underperforming schools advanced student performance when the teaching was improved, the class was better organized, and the students were kept actively involved.

As Stoll and Wikeley (1998) indicate, school improvement efforts in recent years have given greater attention to effectiveness issues such as the teaching and learning processes and student outcomes. For school improvement to be successful, certain characteristics of the school atmosphere must be favorable. For example, a school and its students must have common goals, mutual respect and support, and positive attitudes toward learning. Moreover, the school must feel responsible for its students' success (Creemers & Reezigt, 2005). Behavioral theorists agree that schools will not change if the staff within the schools—the teaching staff especially—do not change (Creemers, 1994).

Schools differ in terms of average student reading achievement. In general, a school's student intake will result in outcomes that are not caused by school processes (e.g., teaching methods, attitudes, etc). Thus, before comparing schools, we need to correct for student intake. A significant factor considered relevant in this respect is the socioeconomic status of students' families. The student populations of schools can differ considerably in relation to the home environment. The extent to which and how the home situation affects educational achievement has also received much attention (Luyten et al., 2005; Papanastasiou, 2000, 2002; Schreiber, 2002).

With its basis in school effectiveness research, this present study investigated achievement in schools in relation to the factors that enhance school performance.

The main research question was: *Why do students at some schools learn much less than would be expected on the basis of their family background and, conversely, why do students at some schools learn much more than would be expected on the basis of their family circumstances?* The aim of the study was to find out whether a set of characteristics taken from the student PIRLS questionnaire for Grade 4 of the primary school could be held responsible for a difference in reading literacy achievement.

Method

This study was based on Cyprus data from the student questionnaire and the reading tests of PIRLS. In this study, a two-stage, stratified, clustered sampling process was used to obtain the sample. In the first stage, a representative sample of 150 public schools was selected from a total of 299 schools (the 150 schools represented 50.2% of the total). During the second stage, one intact classroom of students in each sampled school was randomly selected. The resulting sample included 3,001 students, representing 29.7% of the total population of students. The mean age of the sampled Grade 4 students was 9.7 at the time of testing (April, 2001). Fifty-one percent of the students were male and 49% were female.

This present study analyzed data from both the student questionnaire and the reading literacy tests to identify the particular school characteristics/factors that differentiate between more effective and less effective schools. Scholars generally agree that comparing the extreme groups in the distribution of schools affords more reliable results, and that this approach offers some protection against imputing too much meaning to index scores for particular schools (Postlethwaite & Ross, 1992).

The statistical analysis to distinguish the more effective and less effective schools was based on the residuals, which are the differences between actual and predicted reading scores. In total, seven steps were followed in order to identify the factors distinguishing the more from the less effective schools (Postlethwaite & Ross, 1992).

Step 1: This step involved identifying the measures related to home characteristics, given these are thought to have an effect on student achievement. Three factors were identified from the PIRLS student questionnaire data: the economic status of the family, the size of the

home library, and the number of books the students had at home. For the first factor, 10 measures related to the economic status of the family. Four of these included certain items in the students' homes, namely, swimming pool, washing machine, air-conditioning, and computer. A fifth was ownership by the family of a country house. The second factor related to the size of the library at home. More specifically, students were asked about the number of books in their homes, excluding magazines and school books. The third factor related to the number of books each student had at home.

Step 2: A regression analysis was run with reading achievement as the dependent variable, and the three above-mentioned factors as the independent variables (economic status, the size of the home library, and the number of student books at home). The students placed above the regression line were those students whose reading scores were higher than would be expected, and the students placed below the regression line were those who had achieved lower scores than would be expected.

Step 3: In the third step, the residual scores were calculated. Students with positive residuals achieved higher scores than would be expected, and students with negative residuals achieved lower scores than would be expected. These residual scores were then averaged for all schools. The schools with positive mean residuals were considered the more effective schools, and the schools with negative mean residuals were deemed the less effective schools.

Step 4: In this step, the schools were ranked from the most effective to the least effective school. The schools with average residuals $>+14$ and the schools with average residuals <-17 were then selected. Our purpose was to select the schools at the extremes, as we considered these schools would give more reliable factors determining school effectiveness.

Step 5: Here, we tried to choose the variables that educational authorities have under their control and that influence student achievement. When selecting the indicators for further analysis, we accepted those where correlation of the residuals with all variables were statistically significant and excluded those variables not related to teaching reading. This process produced 18 variables.

Step 6: The 18 variables were then grouped in seven categories: reading outside of school, activities during

teaching, activities after reading in class, time spent on homework, attitudes toward reading, school climate, and hard work in class.

Step 7: In this seventh step, z -scores were calculated, and the t -test used for the final analysis. Our reason for calculating the z -score was that doing so allowed us to place all indicators on the same scale to facilitate the interpretation of the difference in mean scores between the more and the less effective schools. We then summed up the values of the indicators in each category to produce a composite value, and standardized the values of each of the seven factors to a mean of zero and standard deviation of one. Finally, we calculated the mean differences of the z -scores between the more effective and the less effective schools by using the t -test.

Results

This study aimed to determine the factors that distinguish schools as more effective or as less effective, specifically in terms of students' reading achievement. From Table 1, we can see that the composite measure of *economic status* of the family and the variables *home library* and *student books* correlated with achievement in reading. All three correlations were positive. The highest correlation was between student books and reading achievement ($r_{s-books_reading} = 0.24$), followed by the economic status ($r_{ES_reading} = 0.20$), and then the size of the home library ($r_{library_reading} = 0.12$). Thus, the greater the number of books the student had at home, the higher the economic status of that student's family, and the larger the size of the home library, the more likely it would be for that student to have a high level of reading achievement.

Table 2 presents the regression equation of the composite "economic status" and two variables as independent factors, and reading achievement as the dependent variable. The regression analysis was based on the hypothesis that reading achievement is a function of the number of student books at home, the economic status of the family, and the size of the home library. We can see from the equation that the most significant factor in predicting reading achievement was the number of student books. For all three independent factors, the contribution of variance to the prediction of reading achievement was statistically significant, although not high ($R = 0.282$, $R^2 = 0.080$).

Table 1: Correlations between Reading Achievement and Number of Student Books at Home, Economic Status of the Family, and Size of the Home Library

<i>r</i>	Student books	Economic status	Size of library
Reading achievement	0.24*	0.20*	0.12*

Note: * $p < 0.000$.

Table 2: Regression Equation for Predicting Reading Achievement

Predict reading achievement = .196 (s-books) + .138 (economic status) + .078 (size of library)

R = 0.282

R^2 = 0.080

Figure 1 illustrates the position of the 150 schools based on their achievement and their average residuals. With reference to this graph, we can compare schools that were more effective beyond their reading achievement. For example, we can see that some schools that had about the same average achievement in reading had positive average residuals while others had negative average residuals. Furthermore, some schools had lower reading achievement but were more effective than schools with higher reading achievement.

Figure 2 presents the schools after the exclusion of schools with small average positive or negative residuals. More specifically, the schools with average residuals $> +14$ and < -17 were excluded. In total, 87 schools out of the 150 were retained, 44 of which had positive residuals and 43 of which had negative residuals.

Table 3 presents the seven categories of factors that were selected for further analysis, and the corresponding variables and r values between variables and residuals. The significance of r was the criterion for variable selection. The seven factors were used to distinguish the more from the less effective schools. Table 4 presents the factors, the t -test values, the significance level, and the mean differences for the two groups of schools, that is, the more effective and the less effective.

Our analysis revealed six factors explaining school differences in reading achievement. The most influential factor was *things happening after teaching in class*. These “things” included students writing something about what they had read, talking to other students about what they had read, taking a written quiz or test about what they had read, doing a group

project with other students in the class about what they had read, acting in a play or drama about what they had read, and drawing pictures or doing an art project about what they had read. The second factor was *reading outside school*. With this factor, the students, when outside school, read aloud to someone at home, read newspapers, read directions or instructions, and/or listened to someone at home read aloud to them. The third factor distinguishing the two groups of schools was the *time students spent on homework*, and the fourth was *student attitudes toward reading*. The last two factors distinguishing the more effective and the less effective schools were *things happening during teaching* and *school climate*. The only factor that did not show a statistically significant difference between the more and the less effective schools was hard work in class.

Conclusion

The purpose of this study was to find the school characteristics that differentiate more effective from less effective schools in relation to reading achievement. For this reason, the analysis was based on the residuals, which highlight the differences between the actual reading scores and the predicted reading scores. From this analysis, six factors were found that explain school differences in reading achievement. The strongest factor was related to activities after teaching in class. The significant part of this factor concerns what students actually do after teaching in class, and not how the teacher conducts the class. In other words, it is important for teachers to include activities in their teaching that require the students to work and show that conceptualization took place. It seems that when

Figure 1: Position of Schools Based on the Average Reading Achievement and on the Average Residuals of their Students

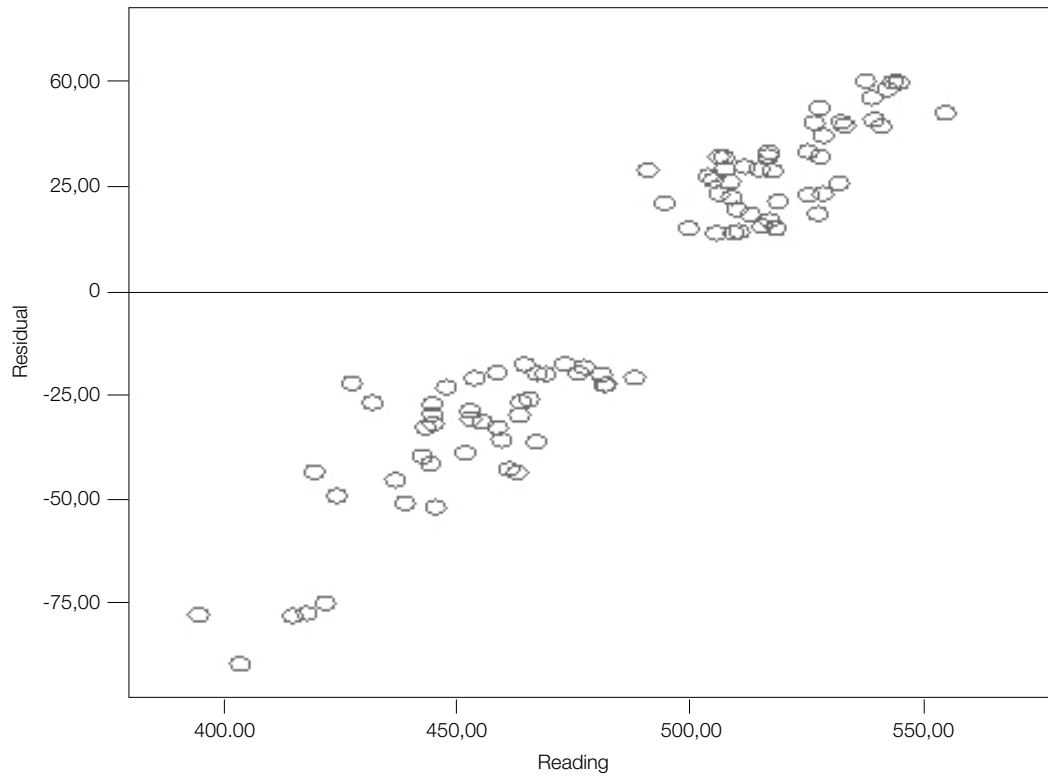


Figure 2: The Remaining Schools for Further Analysis

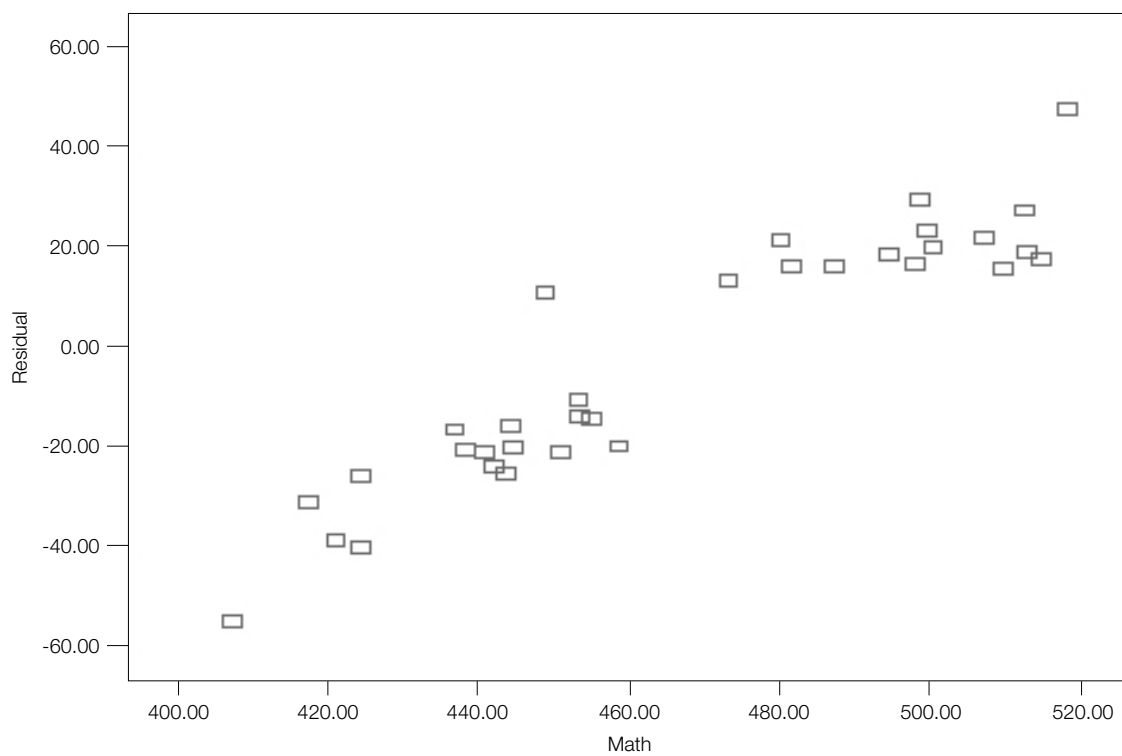


Table 3: Composites, Variables, and r Values, with Reading Achievement

Factors	Variables	r^*
1. Reading outside school	I read aloud to someone at home	0.122
	I read newspapers	0.120
	I read directions or instructions (to learn how to play a game)	0.122
	I listen to someone read aloud to me	0.185
2. Things happening during teaching	I read aloud to the whole class	0.105
	I read aloud to a small group of students in my class	0.144
	I read silently on my own	-0.142
3. Things happening after teaching in class	I write something about what I have read	0.110
	I talk to other students about what I have read	0.078
	I take a written quiz or test about what I have read	0.214
	I do group projects with other students in the class about what I have read	0.199
	I act in a play or drama about what I have read	0.215
	I draw pictures or do an art project about what I have read	0.172
4. Time for homework	When you have reading, how much time do you spend on this reading?	0.215
5. Attitudes toward reading	I think reading is not boring	0.251
	I read only if I have to	0.222
6. School climate	I think that teachers in my school care about me	0.188
7. Hard work in class	I think that students in my school work hard	0.269

Note: $p < 0.05$.

Table 4: Rank Order of Factors Distinguishing the More from the Less Effective Schools

Factors	t -test	p	Mean difference of z-scores
Things happening after teaching in class	6.34	0.00	0.31
Reading outside school	5.84	0.00	0.28
Time for homework	5.33	0.00	0.25
Attitudes toward reading	4.49	0.00	0.22
Things happening during teaching	2.96	0.003	0.14
School climate	2.96	0.003	0.14
Hard work in class	1.44	0.15	0.07

teachers include a component in their teaching that leads to students engaging in “after-teaching” activities, such as writing something about what they have read, taking a written quiz or being tested on what they have read, doing a group project in the class or home about what they have read, and so on, then both students and schools are more effective.

The second most important factor related to reading outside school. This factor meant that students read aloud to someone at home, and/or they read material additional to their text books, such as newspapers, directions, or instructions. These components all

relate primarily to the student and are secondary to the teaching as such. The teaching is important in this context in terms of encouraging students to work effectively or to engage in activities after class or after school.

The third factor related to time spent on homework. Again, this factor is in perfect combination with the above two factors in relation to distinguishing the more from the less effective schools. Students who spend time on homework belong to the effective group of students. Teaching is the main umbrella, but a vital part of it is homework.

The fourth factor related to attitudes toward reading. In PIRLS, student attitudes toward reading were positive in all participating countries, but students with the most positive attitudes had the highest reading achievement. Since attitudes can be taught, this factor is a significant consideration for teachers, and one they should not forget. The fifth factor, related to activities during teaching, also related to the factor reading outside school. Finally, the sixth factor related to school climate, which refers to students' beliefs about teacher attitudes toward them. The significance of this factor is that teacher behavior is important to student achievement, especially at this age. The only factor not statistically significant was student beliefs on working hard in school.

Taking part in PIRLS is of fundamental importance to the educational system of Cyprus. In addition to giving the educational authorities international achievement comparisons, the quality of these data enable in-depth analyses of the national results in an international context (Gonzales & Miles, 2001). The conceptual framework of the study was based on instructional practices applied in teaching reading, seen from the students' perspectives, together with some background factors. Such knowledge is often regarded as a potential foundation for school improvement interventions. If we know the features of effective

schools, we can improve the lower performing schools by encouraging them to adopt the characteristics of effective schools.

The contribution of this study is significant in that it was conducted in a country where all Grade 4 students follow the same reading curriculum. Our analysis revealed, however, two distinctly different learning environments. The findings that certain factors, that is, activities after teaching in class, reading outside school, time for homework, attitudes toward reading, activities during teaching, and school climate, distinguish the more effective schools, at least in terms of reading achievement, have major implications for language education since all these variables are amenable to change through instruction. For less effective schools to be more effective, educational interventions should take into consideration all of these factors underlying reading achievement.

The results of this analysis on school effectiveness contribute to a fuller understanding of the complicated issue of school improvement. However, the area of educational effectiveness still demands further theoretical and empirical research. Important issues that require further research are outcomes, inputs, and the learning process, as well as ideas on how we can promote an active learning environment, in the classroom and in schools.

References

- Baumert, J., Stanat, P., & Watermann, R. (2005, August). *Social and intellectual composition of schools' effects on student outcomes*. Paper presented at the EARLI conference in Nicosia, Cyprus.
- Brown S., Duffield J., & Riddell, S. (1995). School choice, social class and distinctions: The realization of social advantage in education. *Journal of Education Policy*, 11(1), 89–112.
- Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.
- Creemers, B. P. M., & Reezigt, G. J. (2005). Linking school effectiveness and school improvement: The background and outline of the project. *School Effectiveness and School Improvement*, 16(4), 359–371.
- Gadeyne, E., Ghesquiere, P., & Onghena, P. (2006). Psychological educational effectiveness criteria and their relation to teaching in primary education. *School Effectiveness and School Improvement*, 17(1), 63–85.
- Gonzales, E. J., & Miles, J. A. (Eds.). (2001). *TIMSS-1999 user guide for the international database. IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: Boston College.
- Grobler, A., Grobler, A. A., & Esterhuyse, F. G. (2001). Some predictors of mathematics achievement among black secondary school learners. *South African Journal of Psychology*, 31(4), 48–54.
- Luyten, H., Visscher, A., & Witziers, B. (2005). School effectiveness research: From a review of the criticism to recommendations for further development. *School Effectiveness and School Improvement*, 16(3), 249–379.
- Martin, M. O., Mullis, I. V. S., Gregory, K. D., Hoyle, C., & Shen, C. (2000). *Effective schools in science and mathematics*. Chestnut Hill, MA: Boston College.

- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J., & Smith, T. A. (2000). *TIMSS 1999: International mathematics report*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001: International report*. Chestnut Hill: Boston College.
- Nasser, F., & Birenbaum, M. (2005). Modeling mathematics achievement of Jewish and Arab eight grades in Israel: The effects of learner-related variables. *Educational Research and Evaluation, 11*(3), 277–302.
- Opdenakker, M. C., & Van Damme, J. (2005, August). *Are school structures and/or teaching processes responsible for the group composition effect?* Paper presented at the EARLI conference in Nicosia, Cyprus.
- Papanastasiou, C. (2000). Internal and external factors affecting achievement in mathematics: Some findings from TIMSS. *Studies in Educational Evaluation, 26*(1) 1–7.
- Papanastasiou, C. (2002). Effects of background and school factors on mathematics achievement. *Educational Research and Evaluation, 8*(1), 55–70.
- Postlethwaite, T. N., & Ross, K. N. (1992). *Effective schools in reading: Education for educational planners*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Scheerens, J., & Demeuse, M. (2005). The theoretical basis of the effective improvement model (ESI). *School Effectiveness and School Improvement, 16*(4), 373–385.
- Schreiber, B. J. (2002). Institutional and student factors and their influence on advanced mathematics achievement. *Journal of Educational Research, 95*(5), 274–286.
- Secada, G. W. (1992). The reform of school mathematics in the United States. *International Journal of Educational Research, 17*(5), 399–516.
- Stoll, L., & Wikeley, F. (1998). Issues on linking school effectiveness and school improvement. In W. Th. J. G. Hoebe (Ed.), *Effective school improvement: State of the art contribution to a discussion* (pp. 29–58). Groningen: GION, Institute for Educational Research, University of Groningen.
- Van de Grift, W. J. C. M., & Houtveen, A. A. M. (2006). Underperformance in primary schools. *School Effectiveness and School Improvement, 17*(3), 255–273.
- Wilms, J. D. (1992). *Monitoring school performance: A guide for educators*. London: Falmer Press.

Innovative pedagogical practices using technology: Diffusion patterns within schools

Alona Forkosh-Baruch
Tel Aviv University
Tel Aviv, Israel

Dorit Tubin
Ben-Gurion University
Beer-Sheva, Israel

Abstract

The study reported here is a secondary analysis of data collected in IEA's SITES Module 2, focusing on innovative pedagogical practices at the classroom level. It also presents an analysis of data in the parallel OECD/CERI international study, comprising case studies of ICT and organizational innovation, and focusing on ICT-related innovations at the school system level. In our analysis, we refer to 10 Israeli schools that participated in the study and in which novel initiatives took place across varied scopes. The scopes of innovation, referred to in this paper as two diffusion patterns of innovative pedagogical practices using technology, that is, "school-wide implementations" and "islands of innovation," are demonstrated and examined. The findings reported in

this paper focus on the two types of diffusion patterns with relation to (a) levels and domains of the ICT-based innovations within schools; (b) communication agents and school variables affecting the diffusion of the innovation; and (c) roles of factors affecting the diffusion and scalability of the innovation. In the discussion, we elaborate on the potential value of sustainable islands of innovation models as agents of innovation; the similarities and differences between both patterns of ICT implementation in schools; the question of continuity between the two types of diffusion patterns identified; and the unique value of each diffusion pattern for successful diffusion of ICT in schools.

Introduction

Information and communication technologies (ICT) have affected diverse aspects of our lives for some decades. However, schools maintain their image as traditional organizations with predispositions to a somewhat conservative structure. This traditionalism inhibits the adoption of change and slows down the process even in institutes willing to consider alterations in their teaching and learning processes. In addition, attempts to transform traditional structures usually encounter resistance, with the range of opposition dependent on several factors constituting the school milieu. Tyack and Cuban (1995) use the term "school grammar" (e.g., age-graded classrooms, rigid time units) to express the engraved standardized institutional template.

The organizational and designed implementation of ICT into whole educational systems began about a decade after the incorporation of computers in education. In the 1980s, computers in education were a relative novelty (Pogrow, 1983). However, during the

1990s, national plans for large-scale implementation of computers and internet connections were evident in several countries (Plomp, Anderson, Law, & Quale, 2003).

Change in school structure occurred in correspondence with transformations in the social and economic environment within which schools are inscribed and operate. Success in implementing this change required a certain degree of vision and willingness to "yearn" for and embrace change, a phrase coined by Papert (1993). Nowadays, given the fundamental impact of ICT on our lives, it is only natural that we anticipate implementation of ICT in schools will affect their grammar and lead to fundamental transformations in their structure (Watson, 2001).

This article focuses on such change, manifested in diffusion patterns of innovative pedagogical practices employing ICT in K–12 schools. The study recounted here draws on data from two parallel studies: IEA's

Second Information on Technology in Education Study (SITES) Module-2 (Kozma, 2003, 2005), and an Organisation for Economic Co-operation and Development/Centre for Educational Research and Innovation study (Venezky & Davis, 2002). Two main patterns in the diffusion process of ICT-based innovations were identified: “islands of innovation,” in which innovative pedagogical practices involved a minority (about 15%) of the teacher and/or student populations, and “school-wide innovations,” involving 50% or more of the teacher and/or student populations. In this article, we present an analysis, which draws on Rogers’ (1995) diffusion of innovation theory, of the characteristics of both ICT-implementation patterns.

Background

The increased emphasis on worldwide integration of ICT into education systems since the 1990s has facilitated significant efforts in several countries to pre-plan this implementation and to allocate adequate government funds (Hoffman, 2002). Consequently, in recent years, an increasing number of countries have endorsed ICT implementation as part of their national education policy, including aspects such as installation of computer infrastructure in schools, connecting computers to the internet, and teacher training (Pelgrum & Anderson, 1999; Venezky & Davis, 2002). Special attention has been given to diffusion of innovative ICT practices as an ongoing process in several aspects of school life, such as the curriculum, teaching and learning processes, timetabling, and planning learning space. However, one of the main difficulties concerning the diffusion of innovative ICT-based practices in schools is finding ways to engage teachers and students in using the new technologies effectively (Dodgson & Bessant, 1996).

Theories dealing with the diffusion of innovations have been developed in fields such as economics and marketing, in an attempt to identify factors affecting the release and marketing of new products (Amendola & Gaffard, 1988). Hitherto, most attempts to apply these theories to education have concentrated on higher education (Bennett & Bennett, 2003; Macchiusi & Trinidad, 2000), on teachers (Atkins & Vasu, 2000), and on nationwide diffusion of innovation processes (Porter, 2000). Attempts to apply these theories for examining ICT implementation at elementary and secondary school levels are scarce.

ICT implementation

The great potential of ICT to transform education continues to be acclaimed widely on different levels of education systems across the world. The obvious potential of ICT for enhancing teaching and learning processes was recognized as early as the 1980s, for example in the conceptualization, theory, and practice of Papert’s work related to constructivism (Papert, 1980). Stakeholders within education systems, specifically policymakers, acknowledge the advantages of using technology in ways that support individual growth, professional development, and systemic change in educational contexts (Gibson, 2002).

Schools around the world perceive integration of ICT into teaching and learning as a challenge, given its aim of promoting considerable change within the school structure as a whole, or as a lever for local change within one or more of its components. These components include, for example, creation of new learning configurations, formation of novel curricular solutions, broadening and alteration of teachers’ traditional roles, and generation of novel educational settings. ICT in schools has modified the role of teachers from that of instruction to guidance, assisted students in search of individual learning methods and evaluation of their learning processes and outcomes, and brought about change in the role of students by encouraging them to become active learners engaged in collaborative, authentic learning within the community context (Kozma, 2003). However, the complexity of processes and culture shifts required to realize the huge potential that ICT has for education, especially in terms of changes in teaching and learning, has yet to be fully recognized and appreciated.

Nonetheless, the prospective role of ICT in schools has led to high expectations concerning the opportunities and challenges that these novel technologies pose (Tearle, 2003). Schools have reacted in diverse ways to these prospects: some have embraced ICT in teaching and learning of several disciplines and multidisciplinary curricula; others have created novel didactic solutions and processes for their school or classroom settings; and yet others have focused on developing new assessment modes. Also, within the school setting, different models of ICT implementations have been incorporated within varying scopes. Some modes of ICT implementation include a school-wide context. Others focus on an

implementation of information technologies that is remarkably narrow in scale, emphasizing specific skills and knowledge, rather than an across-the-school approach.

Innovation in education and ICT roles

In the digital era, in which endless information is available at the push of a button and learning is ubiquitous, researchers are examining theoretical and empirical aspects of the impact of ICT on educational processes (Becker, 1994; Mioduser & Nachmias, 2002; Pelgrum & Anderson, 1999). Educational transformation processes and their evolution have become highly interesting research themes, with theoretical (e.g., cognitive or curricular issues) as well as practical (e.g., policy or planning issues) implications. However, innovation in itself is not a clearly defined concept, and educational innovation using ICT is a rather vague and multifaceted notion (Bamberger, 1991; Fullan, 2000).

In general, an educational innovation can be regarded as a shift in educational paradigm; schools assume the role of being the primary agents for preparing students to function in and become an integral part of the information society (Pelgrum, Brummelhuis, Collis, Plomp, & Janssen, 1997). This paradigm shift is oriented toward lifelong learning. Accordingly, the main goal of schools is to supply the skills and competencies required for living and working in a continually changing world (Fisher, 2000). ICT serves as a driving force behind the design, establishment, and evolution of this paradigm-shift, affecting both contents (new technology-related concepts and skills included in the curriculum, re-arranging of the curriculum) and general skills (e.g., learning how to learn, acquiring generic knowledge-manipulation skills and teamwork skills). Consequently, innovations can be defined in operational terms as the wide range of activities and means (e.g., curricular decisions, learning materials, learning configurations, lesson plans, tools and resources) that reflect the school's educational and philosophical orientation toward lifelong learning. Eisner (1994) unveils a relatively novel notion of education: *The ultimate aim of education is to enable individuals to become the architects of their own education and through that process to continually reinvent themselves.*

Numerous frameworks have been developed and offered with the aim of characterizing the ways ICT supports and promotes educational change. Technology-based innovations may facilitate transition from traditional to emerging pedagogical paradigms, leading to novel approaches to instruction in a number of dimensions constituting the school milieu, including the curriculum, time configuration, teacher and student practices and roles, grouping and collaboration (Means et al., 1993). These innovations can be specified within four main dimensions: curriculum content and goals, student practices, teacher practices, and how schools use ICT (Kozma, 2000). The North Central Regional Educational Laboratory (NCREL) offers another approach to the examination of ICT implementation in schools when it refers to "a framework for understanding the system wide factors that influence the effective use of educational technology" (enGauge, 2002).

ICT has the potential to affect schools irreversibly, and may contribute to teaching and learning processes and outcomes by meeting students' individual needs, providing a rich educational environment, and presenting educational materials in ways that motivate and stimulate meaningful learning, and arouse motivation in students (Abbot, 2001; Norton & Wilburg, 2002). ICT can also offer new possibilities for reorganizing educational processes, thereby creating innovative teaching and learning opportunities. However, the extent of this contribution has yet to be explored because of controversial evidence regarding this issue (Becker, 1998; Marsh, 2004).

ICT, when implemented in a school, is perceived as innovative in itself. However, ICT-supported innovations are defined herewith as pedagogical solutions and means of supporting a shift from traditional educational paradigms toward emerging pedagogical approaches based on current understanding of learning (e.g., learner-centered and constructivist acquisition of lifelong learning skills) (Mioduser, Nachmias, Tubin, & Forkosh-Baruch, 2003; Pelgrum et al., 1997). Therefore, an innovation is a qualitative educational shift toward a new educational paradigm as an enduring process (Mioduser, 2005).

Diffusion of innovations

Diffusion is the process by which an innovation is communicated through several channels over time and among members of a social system (Rogers,

1995). Rogers provides insight, through his diffusion of innovation theory, into diffusion patterns of innovative initiatives within schools. He defines this process as comprising four components:

1. The *innovation in itself*: an idea, practice, or object perceived as novel by an individual or a group of individuals;
2. *Communication channels*: the means by which communication passes from one individual to another;
3. *Time*: includes the decision process regarding the innovation, the adoption time by an individual and/or the group, and the adoption rate of the innovation; and
4. The *social system*: a set of interrelated units engaged in joint problem solving to achieve a common goal.

An innovation can therefore be defined as a thought or process that is new to a certain system, but not necessarily new altogether. When we examine the diffusion of innovations, it is vital that we look at the various types of innovations within a system. These can be categorized into three main types of innovations, each diffused in different ways (Rogers, 1995):

1. *Continuous innovation*: a simple changing or improving of an already existing product, in which adopters act in the same fashion as they had before, regardless of the novelty;
2. *Dynamically continuous innovation*: this can be the creation of a new product or a radical change to an existing one, in which the usage patterns of people are altered to some extent; and
3. *Discontinuous innovation*: this refers to a new and innovative product that brings overall change to consumers' acquisition and usage patterns.

The manner by which the actors involved in the innovation communicate among themselves can vary as well. Diffusion of innovations takes place through not only general or formal communication channels, but also—and possibly mainly—interpersonal communication. The information flow in this communication mode is characterized by processes such as knowledge transactions among individuals, leaders' influence, and/or peer pressure. The diffusion process at this stage is time consuming. According to Rogers (1995), adopters of an innovation undergo a five-stage process: awareness, interest, evaluation, trial, and adoption. This path, taken by the system as a

whole as well as by its individuals, must not be taken lightly, as the road to a novel educational paradigm is bumpy and not free of obstacles. As Bagehot (2004) observes, "One of the greatest pains to human nature is the pain of a new idea. It ... makes you think that, after all, your favorite notions may be wrong, your firmest beliefs ill-founded."

With regard to communication or diffusion channels of ICT-based innovations in schools, two factors have a major effect: key function-holders within the school, and the school background. In most schools, it seems that the principal, the computer coordinator, and computer experts serve as opinion leaders who mediate between the new technology and the teachers, and who promote diffusion of an innovation by encouraging their fellow teachers to embrace these initiatives (Cuban, 2002). Also, school background factors, such as school size, school level (elementary, high school), location within the country (central, periphery), and settlement type (urban, rural settlement, such as a kibbutz) affect interpersonal communication, solidarity among teachers, and possibly the effectiveness of peer pressure. This influence, in turn, affects the way the innovation is communicated, adopted, and shared by the teachers, and the time frame for the diffusion process.

Overall, research findings from international and national studies demonstrate that initiatives using ICT are mainly at the assimilation or transitional levels. However, many examples can be found of particular agents at the school level who put time, effort, and creative thinking into coping with the complex task of implementing ICT in novel pedagogical ways (Mioduser, 2005; Mioduser & Nachmias, 2002; Pelgrum & Anderson, 1999). Hence, the people included in the social milieu are of the utmost importance whenever consideration is given to integrating innovations in schools.

The social system refers to the group or groups of people among whom an innovation diffuses within their settings (Rogers, 1995). Research on educational change addresses several factors associated with the social systems involved in the diffusion process of a pedagogical innovation (Datnow & Stringfield, 2000; Fullan, 2001; Kinsler & Gamble, 2002; Tyack & Cuban, 1995). These factors can be classified into two major categories: internal factors and external factors. Internal factors are those that are located

within the school. They include the principal, the teachers, and the computer coordinator. However, they do not apply only to human factors, but also to the school's vision and history, teacher training, and ICT infrastructure and maintenance. External factors reside outside school boundaries, and include the government, the municipality, parents, experts, and intervening organizations, as well as national and regional policy and finance (Nachmias, Mioduser, Cohen, Tubin, & Forkosh-Baruch, 2004). This distinction between internal and external factors is fundamental with regards to the question of whether a school can generate innovations by changing the use and purpose of its inner resources (re-engineering), or whether there is need for change in decision-making policies (restructuring) within the school, followed by necessary systemic external action (Papagiannis, Easton, & Owens, 1998).

Emphasis on factors within the school is based on the assumption that the main barriers to change and diffusion of innovations are existing thinking patterns and human behavior. Therefore, conceptual change among school staff members is the first step toward adoption of organizational learning and innovation, a process that involves a bottom-up diffusion of the innovation process (Argyris & Schön, 1996; Sizer, 1992). In contrast, emphasis on factors outside the school stresses the role of decision-makers and top-down processes (e.g., placement of students, allocation of resources) as vital to any change taking place at the school level (Papagiannis et al., 1998; Tyack & Cuban, 1995). Both groups of factors, each in turn addressing separate sets of factors, seem to be essential for ICT-based innovation to diffuse into and improve school pedagogy (Cuban, 2002).

The aim of the present study, in light of the issues raised herein, was to examine differences between two patterns of ICT implementation in curricula within schools: *islands of innovation* and *school-wide implementation*. When observing these differences, we addressed the following:

1. Issues relating to differences in the levels of innovation in schools among each of the implementation-pattern groups, with an emphasis on curricular issues;
2. Communication agents and school variables affecting diffusion of an innovation within schools in each implementation-pattern group; and

3. The effects of internal and external factors related to diffusion of innovations in schools in each implementation-pattern group.

The diffusion of innovation theoretical considerations subsequently presented led to our secondary analysis of the data collected in Israeli schools. The questions we addressed were these:

1. What were the differences in the levels of innovation observed in schools in each of the implementation-pattern groups?
2. What were the major communication agents and school variables affecting the diffusion of innovations in schools in each of the implementation-pattern groups?
3. Which internal and external factors affected the diffusion of innovations in schools in each of the implementation-pattern groups?

Method

Data were collected via qualitative methodology from 10 Israeli schools that had implemented ICT-based successful innovative pedagogical practices. The research population included two elementary schools, one lower secondary school, three high schools, and four six-year secondary schools. These schools were similar in nature to most Israeli schools, so the results can shed light on patterns and processes in other schools. The selection of schools was based on indicators such as meaningful use of ICT, changes in teacher and student roles, curricular changes, and evidence of sustainability, scalability, and transferability (Kozma, 2000; OECD/CERI, 2000). These schools were chosen by a steering committee, the members of which selected the schools according to international and local indicators of innovative pedagogical practices using technology (see Tubin, Mioduser, Nachmias, & Forkosh-Baruch, 2003).

The data-collection tools included questionnaires and interviews (with the principal, computer coordinator, teachers involved in the innovation, teachers not involved in the innovation, student focus groups, parent focus groups, and agents external to the school), class observations, and documentation related to the ICT-based innovation. The study was conducted between February and July 2001. Researchers spent a five-day period in each school. All raw materials were transcribed and uploaded to the Israeli research website. The final Hebrew reports and

documentation, the final school research reports (in Hebrew and English), and the Israeli research report were also placed on the website.

After the data analysis stage of the international study, pursued according to the SITES Module 2 criteria and procedures (see Kozma, 2003), we conducted a secondary analysis of the data from the 10 participating Israeli schools in which the innovations took place. Two analysis tools were applied to assess the levels and domains of innovation in each school (Mioduser et al., 2003) as well as the factors, internal and external, involved in the diffusion of the innovations (Nachmias et al., 2004).

The two analysis tools applied in this study were constructed with reference to existing literature detailing the components of the school setting prone to change as a result of ICT implementation (Bransford, Brown, & Cocking, 1999; Pelgrum & Plomp, 2002; Wenglinski, 1998). Two axes defined the dimensions of the first tool—the levels and domains of innovation schema. The vertical axis consisted of a detailed list of domains of innovation, stemming from a summary of the literature and focusing on four main constituents of the school milieu. These were time/space configurations (consisting of three components—physical space, digital space, and time); students' role; teachers' role (consisting of two components—roles relating to students and roles relating to fellow teachers); and the curriculum (also consisting of three components—content, didactic solutions, and assessment methods). The horizontal axis represented levels of innovation, ranging from minor alterations of the school routine as a result of the implementation of ICT-based initiatives (1) to comprehensive transformation of pedagogical practices and teaching and learning processes (5). Three main levels were defined, as briefly mentioned in the background section. These were the assimilation level (1), the transition level (3), and the transformation level (5). Two additional in-between ratings were also defined. The levels of internal and external factors affecting the innovation were rated according to this analysis framework, which is detailed in Mioduser et al. (2003).

The second tool was composed of two axes: the vertical axis presented 21 factors assembled into seven categories, representing groups of factors involved

in the pedagogical practices using technology. These categories were roles within the school, roles outside the school, organization of learning, organizational climate, staff training and development, infrastructure and resources, and ICT policy. The horizontal axis specified the intensity of the factors' influence or involvement, also in a five-level scale, with 1 representing the lowest degree of influence and 5 representing the highest level of influence. Data, consisting of ratings using the above two tools, were evaluated by two independent judges, who reached an agreement rate of 83%. The remaining 17% of the ratings, in which there was disagreement, were discussed until full agreement was reached.

Finally, all schools were classified according to one of two diffusion-of-innovation patterns: *island-of-innovation schools* (IoIs), in which the innovation engaged only a particular group of students and/or teachers within the school setting, and *school-wide implementation schools* (SWs), in which the majority of a school's student and/or teacher populations were involved in the pedagogical innovation. In schools comprising the IoI group, 4% to 14% of the students and 2% to 28% of the teachers were involved in the activities. In SW schools, 64% to 100% of the students and 27% to 100% of the teachers were involved. The description of the content and scope of the innovations implemented in the participating schools is presented in Table 1; for a detailed description, see Forkosh-Baruch, Mioduser, Nachmias, and Tubin (2005).

These 10 innovations were not intended to represent ICT implementation, as defined by the SITES Module 2 methodology, within schools in Israel. The initiatives were chosen as remarkable examples of successful ICT implementation in innovative pedagogies. As a result, these cases are insufficient for the yielding of significant statistical conclusions. Nevertheless, as the selected innovations took place in schools similar in nature to most schools in Israel, the results do shed light on and enhance our understanding of similar educational patterns and processes in other schools via "naturalistic generalizations" (Stake, 1997). Therefore, despite (a) the small number of innovations analyzed, and (b) the non-representative nature of the chosen cases, we can still make assertions with reference to the uniqueness of the two diffusion patterns.

Table 1: Scope of Innovations Incorporating ICT in Two Diffusion Patterns: “Islands of Innovation” and “School-wide Implementation:” Percentage of Students and Teachers Involved

ID	Title of innovation	Description of innovation	Students involved	Teachers involved
<i>Islands of innovation</i>				
IL002	Computer trustees	Support of teachers during lessons, running of the school ICT support center, and coaching senior citizens as well as students with special needs	6%	24%
IL006	Computerized radio station	Studies toward their matriculations, theoretical and practical topics relating to mass communications, print, and broadcast media, operating a computerized studio, and preparing reports and broadcasts	6%	4%
IL009	Excellence center	Simulation of surroundings within a hi-tech factory, aiming to create a connection between education and industry; acquaintance with the hi-tech world while “connecting” industrialists to educational practice	14%	8%
IL010	Peace network	Making use of internet as a lever for facilitating tolerance, the changing of prejudice, bonding with peers from the Arab culture, and improving facility in English as a foreign language	4%	6%
IL015	Computerized greenhouse	An experimental site for planning and carrying out projects in biology, technology, ecology, and engineering, some being matriculation subjects; contact with research institutions, and coaching of academic experts	9%	2%
<i>School-wide implementation</i>				
IL001	“Beehive”: computerized projects	All students participate in ICT web-based projects via learning communities in subject matters such as literacy, geography, science, mathematics, history, and technology	100%	50%
IL003	ICT-saturated future school	All students implement ICT as a means of developing independent learning skills, adjusting to personal learning styles, applying several fields of interests, raising motivation, and involving parents	100%	100%
IL007	Website story	Students implement web-based learning in educational websites developed locally by teachers and students in over 20 subject matters as an integral part of the curriculum	64%	50%
IL008	“Aviv” virtual school	All students study courses in a virtual high school, develop ICT literacy, and interact with experts in project-based distant learning	100%	35%
IL013	Man and environment virtual learning space	All students develop independent learning skills by inquiry projects related to geographical surroundings, accompanied by a locally built virtual learning space developed mainly by students	100%	27%

Results

Question 1: What were the differences in the levels of innovation in schools between each of the implementation-pattern groups?

Our assumption was that there would be evident differences between the IoI and SW diffusion patterns. Theory indicates that we could have expected higher levels of innovation in IoI schools, since activities in these schools are generated and implemented by a specific group of highly motivated students and teachers, facilitating and accelerating the innovation adoption process, and not bound by the restrictions of the school system and setting. Findings, however, showed that the average level of innovation for domains in both groups of schools was identical, with all schools being located at the transition level. Table 2 describes the average levels of innovation for schools according to diffusion patterns, and presents the averages as well as the variance in levels of innovation within the domains for both diffusion patterns of innovation.

However, a closer look at the domain means unfolds interesting differences between the two diffusion patterns. In IoI schools, learning time and scheduling were defined more flexibly. It seems that ICT in IoIs is challenging the traditional definitions of rigid time slots, enabling ubiquitous learning. Also, teacher–student relationships were rated higher in IoIs, indicating more open and equal educational interaction than in the SW schools. The relatively small number of participants and the extraterritorial nature of some of the projects enabled greater flexibility of

time in relation to IoI projects, as opposed to SW implementations, which were normally embedded within the school timetable.

The findings also indicated different diffusion patterns for interactions between teachers and their peers and between teachers and their students. In SW schools, the nature of teachers' interactions with their peers changed dramatically, emphasizing collaboration and creativity due to their mutual interest in facilitating the diffusion of the innovation. Change and transformation in teachers' roles were supported and promoted by the same school policies that endorsed implementation of the innovation. This situation is indicated by the differences in mean values for teacher role with regards to fellow teachers ($M = 2.2$ for IoIs and $M = 3.6$ for SWs). In IoI initiatives, however, where teacher–student partnerships were a driving force in diffusion of the innovation and a major component in defining the nature of the innovation, a sense of confidence, efficacy, and mutual commitment to the initiative facilitated the emergence of strong, non-mediated, and sustainable tutor–tutee relationships. These relationships sometimes extended to an exchange of roles between teachers and students. In teacher–student roles, most IoI schools reached the transformation level of innovation ($M = 4.0$), stressing the nature of this pattern as a student-centered innovation.

The variance in levels of innovation within the domains for both diffusion patterns revealed intriguing information. The low variance observed in both

Table 2: Mean Levels, Standard Deviation, and Variance Levels of Domains of Innovation in 10 Israeli Initiatives Associated with Implementing ICT, by Diffusion Patterns (IoI and SW)

	Time and space configuration			Student role	Teacher role		Curriculum			
	Physical space	Digital space	Time	Student role	With students	With teachers	Content	Didactic solutions	Assessment methods	Mean
<i>Islands of innovation</i>										
Mean	2.4	3.6	3.8	3.8	4.0	2.2	4.0	3.2	3.0	3.3
St. dev.	1.67	1.14	0.83	0.83	0.70	0.83	1.22	1.09	1.67	
Variance	2.80	1.30	0.70	0.70	0.50	0.70	1.50	1.20	3.50	
<i>School-wide implementation</i>										
Mean	2.8	3.6	2.8	3.2	3.2	3.6	3.6	3.4	3.4	3.3
St. dev.	1.48	0.89	1.78	1.09	1.09	0.89	1.34	1.14	0.54	
Variance	2.20	0.80	3.20	1.20	1.20	0.80	1.80	1.30	0.30	
Difference	-0.4	0.0	1.0	0.6	0.8	-1.4	0.4	-0.2	-0.4	0.0

diffusion patterns was in the relationship patterns of teachers with fellow teachers. This finding strengthens this domain as a differentiator between both diffusion patterns, in light of the mean difference between the two patterns. In contrast, the assessment methods sub-domain varied from 1 to 5 in IoI projects ($SD = 3.5$), but from only 3 to 4 in SW implementation ($SD = .30$), and the difference between the means was small. Another sub-domain that showed a difference between variances was flexibility in terms of time. In IoI projects, the mean score was high ($M = 3.8$) and the variance was low ($SD = .70$), while in SW implementations the mean score was lower altogether, but the variance was much higher ($SD = 3.2$). This indicates the relative flexibility of time altogether in IoI diffusion patterns, and enhances the difference in time allocation in SW implementations, depending on additional intervening variables.

To conclude: although the means were identical in terms of the innovation, a closer examination of the means for each domain highlights the differences and unique emphases for each of the diffusion pattern. SWs focused mostly on systemic change, therefore allowing alterations throughout the fundamental school constituents, while IoIs concentrated on developing the innovation specifically, and not necessarily applying change to the school as a whole.

Question 2: Which communication agents and school variables affect the diffusion of innovations within schools in each implementation-pattern group?

Preliminary hypotheses suggested certain configurations of variables leading to the emergence of either IoI or SW diffusion patterns. The analysis of the innovations in the participating schools focused on the initiating agent (principal, leading teachers), the duration of the innovation, and school variables (such as, size, location, grade levels) that might affect diffusion patterns. For example, it can be assumed that if the principal is the initiator of the innovation, the school is relatively small and therefore facilitates peer relations and mutual influence, and the innovation has been sustainable for a long time, school-wide diffusion of innovation is more likely to occur. Overall, no noticeable differences between diffusion patterns for most variables emerged (see Table 3).

However, it seems that principals in SWs assumed more of a leading role in initiating the innovation and the diffusion process than did principals in IoI schools. In IoI schools, the leaders among the teaching staff (teacher, computer coordinator) usually took the initiative upon themselves. Also, the duration of the innovation was slightly longer in the IoI schools. Findings emphasize the pioneering nature of IoIs: highly motivated individuals initiate and operate innovations within a pedagogical and administrative context, unaware of the entire potential of ICT for teaching and learning. These initiatives persist as IoI, and in many instances do not grow into school-wide enterprises displaying scalability or sustainability

Table 3: Communication Agents and School Variables Affecting Initiatives Related to Implementing ICT in Israeli Schools, by Diffusion Patterns: IoI and SW

Diffusion pattern	Innovation initiator	Size (no. students)	Location	Settlement type	School level	Year of initiation
Islands of innovation	Principal-2 Leader-3	630–1,400	Center-3 Periphery-2	Urban-4 Rural-1	6-year secondary-3 Lower secondary-1 Higher secondary-1	1996–1998 With an exception: 1 since 1985
School-wide implementation	Principal-3 Leader-1 Both-1	380–1,260	Center 3 Periphery-2	Urban-4 Rural-1	6-year secondary-2 Higher secondary-1 Elementary-1	1995–1999

constraints. This situation once again highlights the different nature and traits of IoIs, and emphasizes the differentiation between the two diffusion patterns, rather than positioning them as the two ends of one continuum.

Question 3: What are the effects of internal and external factors related to diffusion of innovations in schools in each implementation-pattern group?

External factors refer to the people and the features outside the school boundaries that have a bearing on educational aspects of the school's functioning. These include the Ministry of Education, municipalities, academic supervisors and consultants, and private agents (e.g., software houses, educational services suppliers). Internal factors include the principal and teaching staff, as well as training within the school, teacher support, school history of innovation, and ICT vision and goals. Data are displayed in Table 4.

The intensity of the internal factors was higher in the SW than the IoI diffusion pattern, especially in terms of agents such as the school's history of innovations

($M = 4.8$ vs. 4.0), vision and goals (4.6 vs. 2.8), ICT coordinator (4.6 vs. 3.2), principal and leading teachers (4.8 vs. 3.8 in both), and accessibility of training (4.0 vs. 2.8). This finding suggests schools need to have a systemic predisposition for SW (i.e., a combination of highly rated internal factors) if the innovation is to be successfully adopted. In contrast, IoIs do not need to sustain these factors. The more prominent accessibility of training in SW schools suggests the need to create formal and systematic channels of information flow, ensuring diffusion of knowledge and practice with reference to the innovation.

The intensity of external factors, however, was similar in both diffusion patterns. This is interesting in light of the fact that IoIs have more outside-school financing resources, and more support from external intervening agents and local authorities. We can accordingly conclude that internal factors are those that differentiate between the two diffusion patterns, and that external factors are not noticeably different between the two.

Table 4: Mean Intensity of Internal versus External Factors Involved in the Innovations, Separated into the Two Diffusion Patterns—IoI and SW

	Factor involved	IoIs	SWs	Difference
External factors	Ministry of Education	2.8	3.4	0.6
	Municipalities	3.4	3.0	-0.4
	Parents	2.0	2.6	0.6
	Intervening factor	4.2	3.8	-0.4
	Expert teacher for students	2.2	2.8	0.6
	National ICT policy	3.0	3.4	0.4
	Financing of innovation	3.8	4.2	0.4
		4.4	3.8	-0.6
Mean		3.2	3.4	0.2
Internal factors	Principal	3.8	4.8	1.0
	Teaching staff	2.0	3.0	1.0
	ICT coordinator	3.2	4.6	1.4
	Leading teachers	3.8	4.8	1.0
	History of innovation	4.0	4.8	0.8
	Vision and ICT goals	2.8	4.6	1.8
	Relevancy of training	2.8	3.8	1.0
	Accessibility of training	2.8	4.0	1.2
	Computers and peripherals	3.4	4.0	0.6
	Technical support	4.4	4.4	0.0
Mean		3.3	4.3	1.0

Discussion and conclusions

ICT has the potential to enhance learning by offering students authentic problem-solving similar to that found in real-world adult settings, engaging them in virtual communities, and enabling them to use sophisticated ICT tools parallel to those used in high-tech companies. ICT can also enhance learning by facilitating inquiry-based projects that generate complex products; utilizing modeling and visualization for better understanding of abstractions; enhancing collaboration; facilitating sharing of experiences; and generating joint knowledge (Dede, 2000). The manner in which ICT is implemented in order to reach these goals can take the form of two distinctive patterns: islands of innovation (IoIs) and school-wide implementations (SWs). In this study, we examined the differences between these patterns of implementation.

IoIs appear to be sustainable, innovational configurations, clear in their goals and functions, their development and maintenance conditions, the roles of the participants, and their relationships to overall processes within school settings. Usually, they are initiated and sustained by a leader or leading group within the school. The goals and functions of IoIs are generally ambitious, aiming at depth rather than extension (e.g., in terms of number of students involved, number of topics covered, and characteristics of the activities). Though sustainable, they frequently do not prove to be scalable, as in the cases of the educational greenhouse or the computerized radio station (see Table 1). This is because the activities involved demand long-term learning processes, close and intense teacher–student interactions, and specific resources that also cannot be easily scaled (e.g., due to complexity of implementation or high costs).

Finally, factors external to the school were found to be important for the functioning of IoIs. These included intervening agencies supporting the initiation of activities (e.g., educational research and development institutions), regional high-tech industries providing support (e.g., expertise, equipment) and opportunities for students to participate in real-life projects, and administrative units in the structure of the educational system that promote novel pedagogical initiatives in schools.

In contrast, the SW implementations emphasized principals' vision and motivation, formal school policy,

and structured teacher training. These factors bring change in the nature of teacher–teacher relationships, based on collaboration and mutual support. Overall, in SWs, principals and decisionmakers face the challenge of defining an appropriate balance between the demands posed by the innovative practices and the features (e.g., structural, curricular, human) characterizing the regular functioning of the school. These points of strength and the potential contribution of SWs, assuming they are appropriately implemented, are evident when aiming to implement an innovation that is intended to change the school setting altogether.

When differentiating between the two types of diffusion of innovations, we have to ask whether they always depend on strongly situated and specific conditions. For example, is it possible to replicate islands of innovations on a larger scale without losing their fundamental traits? The reply to this question will help determine whether we should encourage transition processes of IoI to SW as a natural evolvement within a school, thereby creating a continuum of innovativeness. However, preliminary data do not support this idea.

Another interesting configuration would be schools in which SW and IoI initiatives coexist, in terms of issues such as policy emphases, allocation of resources to more than one innovation, conflicting time and space solutions, different (not necessarily, but possibly contradictory) curricular demands, and various curricular solutions meeting the needs of different innovations. This process may raise issues of priorities regarding not only the nature of the innovation, but also relating to scalability and transferability (which can tie into policy issues such as equity versus selective opportunities).

To conclude, in the analysis of the Israeli case studies, we presented the features and traits of two main innovation implementation patterns, and detailed the pros, cons, and trade-offs of each pattern. The strengths and potential contribution of SW initiatives, when implemented appropriately, are evident. However, the educational value of sustainable IoI initiatives as levers for scalable change over time and as models for varying forms of replication by others within the school and beyond it represents an intriguing ground of research that has theoretical as well as practical implications.

References

- Abbot, C. (2001). *ICT: Changing education*. London: Routledge-Taylor and Francis Group.
- Amendola, M., & Gaffard, J. (1988). *The innovative choice: An economic analysis of the dynamics of technology*. New York: Basil Blackwell.
- Argyris, C., & Schön, D. A. (1996). *Organizational learning II: Theory, method, and practice*. Reading, MA: Addison-Wesley.
- Atkins, N. E., & Vasu, E. S. (2000). Measuring knowledge of technology usage and stages of concern about computing: A study of middle school teachers. *Journal of Technology and Teacher Education*, 8, 279–302.
- Bagehot, W. (2004). *Physics and politics*. Whitefish, MT: Kessinger Publishing.
- Bamberger, P. (1991). Reinventing innovation theory: Critical issues in the conceptualization, measurement, and analysis of technological innovation. *Research in the Sociology of Organizations*, 9, 265–294.
- Becker, H. J. (1994). *Analysis and trends of school use of new information technologies*. Washington, DC: U.S. Congress Office of Technology Assessment/U.S. Government Printing Office.
- Becker, H. J. (1998). Running to catch a moving train: Schools and information technologies. *Theory into Practice*, 37(1), 2–30.
- Bennett, J., & Bennett L. (2003). A review of factors that influence the diffusion of innovation when structuring a faculty training program. *The Internet and Higher Education*, 6, 53–63.
- Bransford, J., Brown, A., & Cocking, R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academic Press.
- Cuban, L. (2002). *Undersold and underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.
- Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed at Risk*, 5, 183–204.
- Dede, C. (2000). Emerging influences of information technology on school curriculum. *Journal of Curriculum Studies*, 32(2), 281–303.
- Dodgson, M., & Bessant, J. (1996). *Effective innovation policy: A new approach*. London: International Thompson Business Press.
- Eisner, E. W. (1994). *The educational imagination: On the design and evaluation of school programs* (3rd ed.). NY: Macmillan.
- enGauge (2002). *A framework for effective technology use*. North Central Regional Educational Laboratory and the Metiri Group. Retrieved September 20, 2006 from <http://www.ncrel.org/engauge/framework/index.htm>
- Fisher, G. (2000). Lifelong learning: More than training. *Journal of Interactive Learning Research*, 11(3/4), 265–294.
- Forkosh-Baruch, A., Mioduser, D., Nachmias, R., & Tubin, D. (2005). “Islands of innovation” and “school-wide implementations”: Two patterns of ICT-based pedagogical innovations in schools. *Human Technology*, 1(2), 202–215.
- Fullan, M. (2000). The return of large-scale reform. *Journal of Educational Change*, 1(1), 1–23.
- Fullan, G. M. (2001). *The new meaning of educational change* (3rd ed.). London: Casswell.
- Gibson, I. W. (2002). Leadership, technology, and education: Achieving a balance in new school leader thinking and behavior in preparation for twenty-first century global learning environments. *Journal of Information Technology for Teacher Education*, 11(3), 315–334.
- Hoffman, E. S. (2002). *Can research improve technology planning policy?* Paper presented at American Education Research Association Annual Meeting, New Orleans, LA.
- Kinsler, K., & Gamble, M. (2002). *Reforming schools*. New York: Continuum.
- Kozma, R. (2000). *Qualitative studies of innovative pedagogical practices using technology: SITES M2 design document*. Enschede, the Netherlands: International Association for the Evaluation of Educational Achievement.
- Kozma, R. (Ed.). (2003). *Technology, innovation and educational change: A global perspective*. Eugene, OR: ISTE Publications.
- Kozma, R. (2005). National policies that connect ICT-based education reform to economic and social development. *Human Technology*, 1(2), 117–156.
- Macchiusi, L., & Trinidad, S. (2000). Implementing IT at an Australian university: Implications for university leaders. In A. Herrmann & M. M. Kulski (Eds.), *Flexible futures in tertiary teaching: Proceedings of the 9th Annual Teaching Learning Forum*. Perth, WA: Curtin University of Technology.
- Marsh, C. J. (2004). *Key concepts for understanding the curriculum* (3rd ed.). New York: RoutledgeFalmer.
- Means, B., Blando, J., Olson, K., Middleton, T., Cobb Morroco, C., Remz, A., & Zorfass, J. (1993). *Using technology to support education reform*. Washington DC: U.S. Department of Education/U.S. Government Printing Office.
- Mioduser, D. (2005). From real virtuality in Lascaux to virtual reality today: Cognitive processes with cognitive technologies. In T. Trabasso, J. Sabatini, D. Massaro, & R. C. Calfee (Eds.), *From orthography to pedagogy: Essays in honor of Richard L. Venezky* (pp. 173–192). Mahwah, NJ: Erlbaum.
- Mioduser, D., & Nachmias, R. (2002). WWW in education: An overview. In H. Adelsberger, B. Collis, & M. Pawlowsky (Eds.), *Handbook on information technologies for education and training* (pp. 23–43). Berlin/Heidelberg/New York: Springer-Verlag.

- Mioduser, D., Nachmias, R., Tubin, D., & Forkosh-Baruch, A. (2003). Analysis schema for the study of domains and levels of pedagogical innovation in schools using ICT. *Education and Information Technologies*, 8, 23–36.
- Nachmias, R., Mioduser, D., Cohen, A., Tubin, D., & Forkosh-Baruch, A. (2004). Factors involved in the implementation of pedagogical innovations using technology. *Education and Information Technologies*, 9, 291–308.
- Norton, P., & Wilburg, K. M. (2002). *Teaching with technology* (2nd ed.). Belmont, CA: Thomson Wadsworth Publishing.
- OECD/CERI (2000). *Schooling for tomorrow: Methodology for case studies of organizational change*. Paris: OECD.
- Papagiannis, J. G., Easton, A. P., & Owens, J. T. (1998). *The school restructuring movement in the USA: An analysis of major issues and policy implications*. Paris, France: UNESCO and IIEP.
- Papert, S. (1980). *Mindstorms: Children, computers and powerful ideas*. New York: Harvester Press.
- Papert, S. (1993). *The children's machine: Rethinking school in the age of the computer*. New York: Basic Books.
- Pelgrum, W. J., & Anderson, R. E. (Eds.). (1999). *ICT and the emerging paradigm for life long learning: A worldwide educational assessment of infrastructure, goals and practices*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Pelgrum, W. J., Brummelhuis, A., Collis, B., Plomp, T., & Janssen, I. (1997). *Technology assessment of multimedia systems for pre-primary and primary schools*. Luxembourg: European Parliament, Scientific and Technological Options Assessment Panel.
- Pelgrum, W. J., & Plomp, T. (2002). Indicators of ICT in mathematics: Status and covariation with achievement measures. In A. E. Beaton & D. F. Robitaille (Eds.), *Secondary analysis of the TIMSS data* (pp. 317–330). Dordrecht: Kluwer Academic Press.
- Plomp, T., Anderson, R. E., Law, N., & Quale, A. (Eds.). (2003). *Cross national policies and practices on information and communication technology in education*. Greenwich, CT: Information Age Publishing.
- Pogrow, S. (1983). *Education in the computer age*. Beverly Hills, CA: Sage.
- Porter, B. (2000). Measuring our progress with real tools: Why does technology work in some schools and not in others? *Proceedings of the Secretary's Conference on Educational Technology, 2000: Measuring impacts and shaping the future*, Alexandria, VA. (ED452838).
- Rogers, E. M. (1995). *Diffusion of innovations*. New York: The Free Press.
- Sizer, T. R. (1992). *Horace's school: Redesigning the American high school*. Boston, MA: Houghton-Mifflin.
- Stake, R. E. (1997). *Case study methods: Seeking sweet water*. In R. M. Jaeger (Ed.), *Complementary methods for research in education* (2nd ed., pp. 73–116). Washington, DC: American Educational Research Association.
- Stake, R. E. (2000). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 435–454). Thousand Oaks, CA: Sage.
- Tearle, P. (2003). ICT implementation: What makes the difference? *British Journal of Educational Technology*, 34(5), 567–583.
- Tubin, D., Mioduser, D., Nachmias, R., & Forkosh-Baruch, A. (2003). Domains and levels of pedagogical innovation in schools using ICT: Ten innovative schools in Israel. *Education and Information Technologies*, 8(2), 127–145.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward Utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
- Venezky, R. L., & Davis, C. (2002). *Quo vademus? The transformation of schooling in a networked world*. Paris, France: OECD/CERI.
- Watson, D. (2001). Pedagogy before technology: Re-thinking the relationship between ICT and teaching. *Education and Information Technologies*, 6(4), 251–266.
- Wenglinski, H. (1998). *Does it compute? The relationship between educational technology and student achievement in mathematics*. Princeton, NJ: ETS.

Examining literacy, gender, and the home environment in PIRLS 2001 countries

Kathleen Trong and Ann Kennedy
Boston College
Chestnut Hill, Massachusetts, USA

Abstract

Girls outperformed boys in reading achievement scale scores in all 35 countries that participated in PIRLS 2001. The PIRLS 2001 *Learning to Read Survey* provided information about early home literacy activities that indicates the onset of such gender differences in reading occurs before children enter school, as these activities were found to be significant predictors of reading achievement. This cross-cultural analysis investigated

the extent to which girls and boys differed in their participation in literacy activities and subsequent student attitudes toward reading, and explored these differences in the contexts of 32 PIRLS 2001 countries. Results of this study provide valuable information to parents and teachers in early childhood education in understanding the ways in which interactions with young children may shape the development of reading literacy.

Introduction

In 2001, the International Association for the Evaluation of Educational Achievement (IEA) conducted the first cycle of an international assessment of students' reading achievement at Grade 4, the Progress in International Reading Literacy Study (PIRLS). The PIRLS 2001 International Report (Mullis, Martin, Gonzalez, & Kennedy, 2003) revealed a gender gap favoring girls in all 35 participating countries, as well as the positive impact that early literacy activities have on later reading achievement. Using these two findings as a basis, this research investigated the relationship between gender and early childhood home environments and the influence of this environment on literacy and student attitudes. We hypothesized that substantially more girls than boys would be from home environments associated with high achievement and that more girls than boys would be participating in early literacy activities, thus contributing to the girls'

later success in reading. We also sought to determine if there were gender differences in Grade 4 students' attitudes toward reading, given that positive attitudes are associated with high achievement and may be influenced by early literacy activities. Thirty-two of the countries that participated in PIRLS 2001 explored these issues. Table 1 lists them.

It has been well documented in the literature that the home environment of a child before beginning school can strongly influence later reading achievement (Baker, Afflerbach, & Reinking, 1996; Martin, Mullis, & Gonzalez, 2005) and that the early literacy activities that parents do with their children can have a strong effect (Nord, Lennon, Liu, & Chandler, 2000). In addition, Mullis et al. (2003) found that, at Grade 4 level, girls have significantly higher reading achievement than boys internationally. The different treatment of children based on their gender from a very young age

Table 1: The 32 PIRLS 2001 Countries Used for Analysis

Argentina	France	Kuwait	Romania
Belize	Germany	Latvia	Russian Federation
Bulgaria	Greece	Lithuania	Scotland
Canada	Hong Kong SAR	Macedonia, Rep. of	Singapore
Colombia	Hungary	Moldova, Rep. of	Slovak Republic
Cyprus	Iceland	Netherlands	Slovenia
Czech Republic	Iran, Islamic Rep. of	New Zealand	Sweden
England	Italy	Norway	Turkey

has also been extensively researched (Statham, 1986; Yelland, 1998), demonstrating that a child's gender role is largely the result of environmental influences. Using the combination of these findings as a foundation, we considered it highly plausible that there is a difference in the way boys and girls are engaged in early literacy activities in the home.

Method

This research investigated parents' reports of a series of variables that play a part in early literacy and were provided by the PIRLS *Learning to Read Survey*, a questionnaire administered to parents of students who participated in PIRLS. Student attitudes toward reading were also examined, as reported by students in the *Student Questionnaire*. Our investigation focused on all of the participating countries that administered and had sufficient data for the *Learning to Read Survey*¹ and the *Student Questionnaire*. In these countries, approximately 136,500 students participated. Table 2 sets out the average reading achievement scores of these students. The table shows a significant difference in achievement between boys and girls internationally, ranging from 8 to 48 score points. In all participating countries, this difference favored girls, although there was not a direct relationship to the country's overall average scale score.

For our analyses, we examined variables that allowed us to gain a better grasp of the role of literacy in the students' home environments and of the students' attitudes toward reading once they reached Grade 4. We specifically chose home variables that involved interaction with the child or that he or she might influence in some manner. Our aim was to determine if such variables vary by gender in a meaningful way. These variables included early home literacy activities that parents and their children do together, the literacy skills that children had acquired before beginning school, parents' attitudes toward reading, and the number of children's books in the home. We investigated student variables measuring attitudes toward reading on the premise that these might also vary by gender and therefore be strong predictors of student achievement. In the PIRLS 2001 International Report (Mullis et al., 2003), many of

these home and student attitude variables were used to create indices that provided a broader depiction of literacy by combining items that were related by an underlying construct.

The Early Home Literacy Activities Index (EHLA) is a summary of six activities that foster literacy that parents may have done with their child before he or she began school. Parents were asked to respond if they did each of these activities "often," "sometimes," "never," or "almost never:"

Before your child began <ISCED Level 1>, how often did you or someone else in your home do the following activities with him or her?

- Read books
 - Tell stories
 - Sing songs
 - Play with alphabet toys (e.g., blocks with letters of the alphabet)
 - Play word games
 - Write letters or words
 - Read aloud signs and labels
-

Parents' reports of their child's early literacy skills when they began school were also used to create an index of Early Literacy Skills (ELS). Parents were asked to respond by indicating whether their child could do each of these five activities "very well," "moderately well," "not very well," or "not at all:"

How well could your child do the following when he/she began <ISCED Level 1>?

- Recognize most of the letters in the alphabet
 - Read some words
 - Read sentences
 - Write letters of the alphabet
 - Write some words
-

The index of Student Attitudes Toward Reading (SATR) was also analyzed, because it is reasonable to expect that students' early literacy environment influences their attitudes. The index was created using responses from the *Student Questionnaire*, in which students were asked to "agree a lot," "agree a little," "disagree a little," or "disagree a lot" with a series of statements:

¹ Morocco and the United States did not administer the *Learning to Read Survey*, and Israel did not have a response rate high enough to allow us to include their data.

Table 2: Gender Differences in Achievement of Students in 32 of the Countries that Participated in PIRLS 2001

Country	Girls			Boys		Achievement difference
	Percent	Average scale score		Percent	Average scale score	
Italy	48 (0.9)	545 (2.6)	▲	52 (0.9)	537 (2.7)	8 (2.5)
France	48 (0.9)	531 (2.7)	▲	52 (0.9)	520 (3.0)	11 (3.3)
Colombia	50 (1.2)	428 (5.1)	▲	50 (1.2)	416 (4.7)	12 (4.3)
Russian Federation	49 (0.9)	534 (4.0)	▲	51 (0.9)	522 (4.8)	12 (2.3)
Czech Republic	49 (1.0)	543 (2.8)	▲	51 (1.0)	531 (2.6)	12 (2.8)
Germany	50 (0.8)	545 (2.2)	▲	50 (0.8)	533 (2.5)	13 (2.7)
Romania	51 (1.0)	519 (4.2)	▲	49 (1.0)	504 (5.7)	14 (3.8)
Hungary	51 (1.0)	550 (2.4)	▲	49 (1.0)	536 (2.5)	14 (2.1)
Netherlands	50 (0.8)	562 (2.7)	▲	50 (0.8)	547 (2.8)	15 (2.2)
Slovak Republic	50 (0.9)	526 (3.0)	▲	50 (0.9)	510 (3.3)	16 (3.0)
Lithuania	51 (1.0)	552 (3.0)	▲	49 (1.0)	535 (2.7)	17 (2.7)
Scotland	52 (1.0)	537 (3.9)	▲	48 (1.0)	519 (4.2)	17 (4.0)
Canada (O,Q)	50 (0.2)	553 (2.6)	▲	50 (0.7)	536 (2.6)	17 (2.1)
Argentina	51 (1.1)	428 (6.2)	▲	49 (1.1)	410 (6.5)	18 (4.7)
Hong Kong SAR	50 (1.0)	538 (3.0)	▲	50 (1.0)	519 (3.5)	19 (2.9)
Iceland	50 (0.8)	522 (1.9)	▲	50 (0.8)	503 (1.5)	19 (2.0)
Turkey	48 (0.9)	459 (4.0)	▲	52 (0.9)	440 (3.7)	19 (3.1)
International Avg.	50 (0.2)	510 (0.7)	▲	50 (0.2)	490 (0.7)	20 (0.7)
Greece	50 (1.0)	535 (3.8)	▲	50 (1.0)	514 (4.0)	21 (3.9)
Macedonia, Rep of	49 (0.9)	452 (5.1)	▲	51 (0.9)	431 (4.8)	21 (3.6)
Norway	48 (1.0)	410 (3.5)	▲	52 (1.0)	489 (3.4)	21 (3.9)
Slovenia	50 (0.8)	512 (2.5)	▲	50 (1.1)	534 (2.6)	22 (2.8)
Latvia	48 (1.1)	556 (3.1)	▲	52 (1.1)	534 (2.6)	22 (3.4)
Sweden	49 (0.7)	572 (2.6)	▲	51 (0.7)	550 (2.5)	22 (2.6)
England	52 (1.1)	564 (3.9)	▲	48 (1.1)	541 (3.7)	22 (3.3)
Cyprus	49 (0.9)	506 (3.3)	▲	51 (0.9)	482 (3.6)	24 (3.5)
Bulgaria	51 (0.9)	562 (3.7)	▲	49 (0.9)	538 (4.7)	24 (3.6)
Singapore	48 (1.5)	540 (5.3)	▲	52 (1.5)	516 (5.7)	24 (4.1)
Moldova, Rep of	50 (1.0)	504 (4.7)	▲	50 (1.0)	479 (4.0)	25 (4.0)
New Zealand	49 (1.3)	542 (4.7)	▲	51 (1.3)	516 (4.2)	27 (5.4)
Iran	55 (3.6)	426 (5.7)	▲	45 (3.6)	399 (5.6)	27 (8.1)
Belize	50 (0.9)	341 (5.3)	▲	50 (0.9)	314 (5.3)	27 (4.8)
Kuwait	48 (0.3)	422 (5.6)	▲	52 (0.3)	373 (6.3)	48 (8.4)

Note: ▲ Significantly higher percentage of students than other gender.

Source: Mullis, I. V. S, Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools in 35 countries*. Chestnut Hill, MA: Boston College, p. 30.

What do you think about reading? Tell how much you agree with each of these statements.

- I read only if I have to (reverse coded)
- I like talking about books with other people
- I would be happy if someone gave me a book as a present
- I think reading is boring (reverse coded)
- I enjoy reading

Each of these indices was calculated using an average of the responses to the individual items, and

were grouped into categories of varying levels of literacy support. In our analyses, for all of the indices and individual variables, the percentage of girls and boys in the response categories that were most and least supportive of literacy were compared. We then tested these for statistically significant differences, which allowed us to identify if either gender was disproportionately engaged in particular home literacy activities or if the genders differed significantly in their attitudes.

Results

Early home literacy activities

The Early Home Literacy Activities Index (EHLA) discussed above is a summary of the home literacy environment, and we found a strong link between it and high literacy achievement for students in Grade 4. Table 3 shows significantly more girls in the high EHLA category (indicating that parents reported doing all of the activities frequently) and boys in the low EHLA category (indicating that parents reported doing all of

the activities never or almost never) in many countries, but that it was not a universal trend. Differences in the high EHLA category were significant in 19 countries, and 12 countries had significant differences in the low EHLA category.

The gender differences in relation to these activities become more apparent when examined individually. In 29 participating countries, significantly greater percentages of girls than boys sang songs often with their parents. However, only seven countries showed

Table 3: Percentages of Boys and Girls at Each Level of the Early Home Literacy Activities Index (EHLA) (32 countries)

Country	High EHLA		Medium EHLA		Low EHLA	
	Percent girls	Percent boys	Percent girls	Percent boys	Percent girls	Percent boys
Italy	65 (1.2) ▲	60 (1.4)	27 (1.2)	32 (1.3) ▲	7 (0.7)	8 (0.8)
France	58 (1.3)	55 (1.5)	34 (1.2)	35 (1.4)	8 (0.8)	10 (0.7)
Colombia	41 (2.0)	39 (2.2)	39 (1.6)	39 (1.7)	20 (1.5)	22 (1.5)
Russian Federation	69 (1.4) ▲	63 (1.6)	25 (1.0)	29 (1.6) ▲	7 (0.8)	8 (1.2)
Czech Republic	56 (1.6) ▲	48 (1.4)	38 (1.6)	44 (1.6) ▲	6 (0.8)	9 (0.9) ▲
Germany	47 (1.1) ▲	40 (1.1)	42 (1.0)	44 (1.1)	11 (0.6)	16 (1.0) ▲
Romania	55 (1.6)	55 (1.7)	30 (1.5)	31 (1.5)	15 (1.3)	14 (1.6)
Hungary	64 (1.4) ▲	58 (1.2)	30 (1.4)	34 (1.2) ▲	6 (0.6)	8 (0.7) ▲
Netherlands	59 (1.5) ▲	51 (1.8)	34 (1.4)	40 (1.5) ▲	7 (0.7)	9 (1.0) ▲
Slovak Republic	64 (1.4)	61 (1.5)	31 (1.4)	33 (1.4)	5 (0.5)	6 (0.7)
Lithuania	51 (1.5) ▲	46 (0.7)	37 (1.6)	41 (1.5)	13 (1.0)	13 (1.3)
Scotland	83 (1.4)	80 (1.5)	15 (1.3)	17 (1.4)	2 (0.5)	2 (0.6)
Canada (O,Q)	68 (1.0)	66 (1.0)	26 (0.9)	29 (1.1)	5 (0.5)	6 (0.6)
Argentina	51 (1.8)	48 (2.0)	34 (1.5)	35 (1.8)	14 (1.7)	17 (1.8)
Hong Kong SAR	18 (1.1) ▲	15 (1.0)	53 (1.0)	54 (1.4)	29 (1.3)	31 (1.6)
Iceland	55 (1.2) ▲	51 (1.4)	38 (1.2)	40 (1.4)	8 (0.7)	9 (0.8)
Turkey	29 (1.3) ▲	23 (1.1)	39 (1.6)	39 (1.2)	32 (7.2)	38 (1.5) ▲
International Avg.	54 (1.3) ▲	50 (0.3)	34 (1.2)	36 (0.3) ▲	12 (0.2)	15 (0.2) ▲
Greece	60 (1.9) ▲	53 (1.7)	31 (1.7)	35 (1.2)	9 (1.2)	12 (1.4) ▲
Macedonia, Rep of	62 (1.5)	61 (1.9)	31 (1.4)	30 (1.4)	6 (0.8)	8 (1.3)
Norway	50 (1.3) ▲	44 (1.7)	38 (1.3)	43 (1.6) ▲	12 (1.1)	13 (1.3)
Slovenia	60 (1.5) ▲	55 (1.7)	35 (1.4)	39 (1.6)	5 (0.7)	6 (0.7)
Latvia	60 (1.6)	57 (1.9)	33 (1.4)	35 (2.0)	7 (0.8)	8 (0.9)
Sweden	44 (1.2) ▲	39 (1.1)	44 (1.2)	46 (1.1)	12 (0.7)	15 (0.8) ▲
England	84 (1.7)	82 (1.5)	14 (1.5)	15 (1.5)	2 (0.5)	3 (0.5)
Cyprus	59 (1.7)	56 (1.7)	33 (1.4)	33 (1.6)	8 (0.8)	11 (0.9) ▲
Bulgaria	64 (1.9) ▲	60 (2.0)	25 (1.2)	26 (1.3)	11 (1.7)	14 (1.9)
Singapore	40 (1.3) ▲	35 (1.1)	42 (1.0)	41 (0.9)	18 (1.1)	24 (0.9) ▲
Moldova, Rep of	44 (1.6) ▲	39 (1.6)	38 (1.3)	39 (1.6)	18 (1.3)	21 (1.7) ▲
New Zealand	70 (1.5)	67 (1.6)	26 (1.5)	26 (1.5)	4 (0.7)	7 (1.0)
Iran, Islamic Rep. of	19 (1.4)	17 (1.5)	37 (1.5)	33 (1.6)	44 (2.3)	50 (2.5) ▲
Belize	43 (2.3) ▲	38 (2.3)	36 (1.5)	39 (2.0)	21 (2.0)	23 (2.4)
Kuwait	33 (1.0) ▲	27 (1.0)	48 (0.9)	48 (1.2)	19 (0.8)	25 (1.1) ▲

Note: ▲ Significantly higher percentage of students than other gender.

significant differences between boys and girls telling stories often with their parents. This range across variables suggests that some activities are more gender-neutral than others. More importantly, all of the significant differences for each early literacy activity followed the same pattern, with significantly more girls doing them frequently and significantly more boys never or almost never doing them. There were no instances where this pattern altered, even within individual countries.

Even more striking than the disproportionate participation in activities are the differences in the number of children's books in the home that emerged between boys and girls in some countries, based on parents' reports (Table 4). Thirteen countries showed significantly higher percentages of girls with more than 50 children's books; in 10 countries, significantly more boys had 25 or fewer children's books. Reading materials in the home are considered one of the strongest predictors of a child's reading achievement,

Table 4: Percentage of Boys and Girls Whose Parents Reported Having Children's Books at Home (32 Countries)

Country	0-25 children's books		26-50 children's books		More than 50 children's books	
	Percent girls	Percent boys	Percent girls	Percent boys	Percent girls	Percent boys
Italy	55 (1.6)	57 (1.6)	28 (1.2)	25 (1.1)	17 (1.0)	18 (1.1)
France	23 (1.3)	27 (1.3) ▲	31 (1.4)	30 (1.3)	46 (1.9)	43 (1.9)
Colombia	90 (1.1)	92 (0.9) ▲	7 (0.9)	6 (0.7)	3 (0.4)	2 (0.4)
Russian Federation	39 (1.8)	40 (1.9)	29 (1.2)	30 (1.4)	32 (1.6)	30 (1.2)
Czech Republic	17 (1.6)	17 (1.1)	35 (1.1)	38 (1.4) ▲	48 (1.8) ▲	44 (1.8)
Germany	22 (1.2)	27 (1.4) ▲	34 (1.0)	32 (1.1)	45 (1.5) ▲	40 (1.4)
Romania	67 (1.9)	65 (1.9)	20 (1.3)	23 (1.3) ▲	13 (1.3)	12 (1.0)
Hungary	25 (1.5)	27 (1.2)	29 (1.0)	30 (1.1)	46 (1.5) ▲	42 (1.5)
Netherlands	19 (1.7)	23 (1.5)	30 (1.5)	31 (1.8)	51 (1.8) ▲	46 (1.9)
Slovak Republic	33 (1.6)	38 (1.7) ▲	39 (1.3)	37 (1.3)	28 (1.6) ▲	24 (1.5)
Lithuania	51 (1.9)	53 (1.8)	28 (1.3)	29 (1.3)	21 (1.6)	18 (1.4)
Scotland	13 (1.3)	15 (1.3)	25 (1.6)	29 (2.8)	62 (1.9)	56 (2.9)
Canada	13 (1.0)	15 (0.9) ▲	26 (1.0)	25 (1.2)	62 (1.4)	59 (1.5)
Argentina	82 (2.0)	83 (1.7)	13 (1.4)	11 (1.5)	6 (1.0)	6 (0.9)
Hong Kong SAR	71 (1.9)	72 (1.4)	18 (1.0)	18 (1.0)	11 (1.2)	10 (0.8)
Iceland	4 (0.5)	5 (0.5)	24 (1.2)	22 (1.1)	72 (1.2)	73 (1.1)
Turkey	81 (1.4)	83 (1.3)	13 (1.1)	12 (0.9)	5 (0.8)	5 (0.7)
International Avg.	42 (1.3)	44 (0.3) ▲	25 (0.2)	25 (0.2)	33 (0.3) ▲	31 (0.3)
Greece	44 (2.2)	48 (2.2)	29 (1.7)	27 (1.7)	27 (2.4)	26 (2.2)
Macedonia, Rep of	60 (2.1)	61 (2.0)	26 (1.5)	27 (1.3)	14 (1.2)	12 (1.2)
Norway	8 (1.0)	11 (1.0)	23 (1.2)	26 (1.5)	69 (1.5) ▲	63 (1.6)
Slovenia	47 (1.6)	48 (1.5)	31 (1.4)	32 (1.2)	22 (1.3)	21 (1.3)
Latvia	27 (2.1)	32 (1.5) ▲	32 (1.5)	32 (1.5)	41 (1.8)	36 (1.6)
Sweden	12 (1.3)	12 (1.1)	20 (0.8)	25 (1.0) ▲	68 (1.4) ▲	63 (1.4)
England	10 (1.2)	13 (1.4) ▲	24 (1.4)	22 (1.7)	66 (1.8)	65 (2.2)
Cyprus	44 (1.9)	47 (1.7)	30 (1.5)	32 (1.6)	26 (1.5) ▲	21 (1.5)
Bulgaria	46 (2.4)	51 (1.8) ▲	29 (1.4)	28 (1.4)	24 (1.7) ▲	21 (1.4)
Singapore	23 (1.2)	28 (1.5) ▲	30 (0.9)	31 (0.8)	47 (1.5) ▲	41 (1.6)
Moldova, Rep of	76 (2.2)	79 (1.9)	16 (1.8)	14 (1.2)	7 (1.1)	7 (1.0)
New Zealand	12 (1.1)	15 (1.4) ▲	22 (1.6)	25 (1.5)	66 (1.8) ▲	60 (1.9)
Iran	82 (2.0)	86 (1.5)	11 (1.1)	9 (1.0)	7 (1.0) ▲	4 (0.6)
Belize	75 (2.0)	74 (1.9)	17 (1.5)	16 (1.6)	8 (1.0)	10 (1.6)
Kuwait	68 (1.5)	72 (1.3)	21 (0.9)	19 (1.9)	11 (0.9)	9 (0.8)

Note: ▲ Significantly higher percentage of students than other gender.

and are essential to early literacy development (Baker et al., 1996). While these data do not reveal whether these children's books were acquired because the children showed interest in early literacy activities or if the relationship was reversed, the presence of print materials has been shown to elicit conversations about reading in the home (ter Laak, 1994).

Early literacy skills

This participation in early literacy activities and possession of resources helps shape students' early literacy skills when they start school. Parents' reports of these skills were summarized in the Early Literacy Skills Index (ELS), which was positively associated with reading achievement. In line with the trend set in previous tables, Table 5 shows girls often over-represented in the "very well" category (indicating that they could do most tasks very well) and boys over-

Table 5: Percentages of Boys and Girls Whose Parents Reported Their Children Could Do Early Literacy Activities at the Time They Began School (32 Countries)

Country	Very well		Not at all	
	Percent girls	Percent boys	Percent girls	Percent boys
Italy	18 (1.1) ▲	15 (1.0)	13 (0.8)	18 (1.0) ▲
France	25 (1.2)	23 (1.2)	5 (0.7)	8 (0.7) ▲
Colombia	25 (1.5) ▲	21 (1.3)	16 (1.4)	15 (1.5)
Russian Federation	15 (1.0) ▲	12 (1.2)	21 (1.8)	27 (2.2) ▲
Czech Republic	9 (0.7)	7 (0.8)	34 (1.6)	41 (1.6) ▲
Germany	14 (0.7) ▲	9 (0.7)	20 (0.9)	29 (1.0) ▲
Romania	20 (1.6) ▲	15 (1.1)	20 (1.7)	22 (1.7)
Hungary	10 (0.8)	9 (0.9)	39 (1.1)	43 (1.4)
Netherlands	13 (1.0) ▲	8 (0.8)	19 (1.3)	27 (1.4) ▲
Slovak Republic	7 (0.6)	6 (0.6)	31 (1.4)	40 (1.3) ▲
Lithuania	28 (1.4) ▲	18 (1.2)	10 (0.9)	13 (1.1) ▲
Scotland	13 (1.2) ▲	9 (0.9)	8 (1.0)	18 (1.7) ▲
Canada (O,Q)	36 (1.1) ▲	27 (1.3)	4 (0.4)	7 (0.6) ▲
Argentina	21 (1.3) ▲	15 (1.3)	12 (1.7)	14 (2.2)
Hong Kong SAR				
Iceland	20 (1.1) ▲	13 (0.8)	15 (0.9)	22 (1.2) ▲
Turkey	25 (1.6) ▲	20 (1.1)	22 (1.6)	26 (1.5) ▲
International Avg.	24 (0.2) ▲	18 (0.2)	15 (0.2)	20 (0.3) ▲
Greece	41 (1.6)	37 (1.8)	8 (0.8)	13 (1.5) ▲
Macedonia, Rep of	42 (1.5) ▲	37 (1.6)	6 (1.0)	8 (1.3)
Norway	22 (1.3) ▲	16 (1.0)	9 (0.9)	18 (1.0) ▲
Slovenia	33 (1.5) ▲	20 (1.1)	7 (0.7)	14 (1.2) ▲
Latvia	29 (1.5) ▲	18 (1.2)	6 (1.0)	7 (1.0)
Sweden	34 (1.1) ▲	21 (1.0)	5 (0.7)	13 (0.7) ▲
England	33 (1.8) ▲	25 (1.6)	5 (0.9)	7 (1.0)
Cyprus	22 (1.3)	20 (1.1)	19 (1.3)	23 (1.5) ▲
Bulgaria	34 (1.5) ▲	23 (1.1)	13 (1.8)	19 (2.1) ▲
Singapore	54 (1.8) ▲	46 (1.7)	1 (0.2)	2 (0.3) ▲
Moldova, Rep of	15 (1.0) ▲	9 (0.9)	21 (1.7)	26 (2.0) ▲
New Zealand	26 (1.5) ▲	19 (1.4)	7 (0.9)	11 (1.2) ▲
Iran	30 (1.5) ▲	19 (1.2)	23 (2.3)	28 (2.2)
Belize	21 (1.7) ▲	14 (1.3)	16 (1.7)	22 (1.8) ▲
Kuwait	16 (0.7) ▲	11 (0.6)	26 (0.8)	31 (1.0) ▲

Note: ▲ Significantly higher percentage of students than other gender.

represented in the “not at all” category (indicating that they could not do most of the activities). These differences were significant for each category in 25 countries.

Several potential interpretations are possible for these data and the gender gaps they represent. It may be that girls have more advanced skills due to developmental differences at this age. However, it is also likely that these skills were enhanced by the early literacy activities that girls engaged in more frequently than did boys. One must also remember that these reports are parents’ perceptions and so possibly influenced by what parents think girls and boys should be capable of. Nonetheless, Table 5 suggests that girls enter schools with greater literacy abilities than do boys, and that they maintain this advantage through to Grade 4.

Parents’ reading attitudes and activities

In addition to asking about activities involving children, the *Learning to Read Survey* asked parents about their personal attitudes and activities in regards to reading. Positive parental attitudes have shown positive correlations with children’s reading achievement (Mullis et al., 2003; Spiegel, 1994). We examined the proportions of boys and girls in these categories of variable, but found few significant differences. This is as one might expect, given that parents’ attitudes develop independently of the gender of their child. However, one item asked parents to agree or disagree with the statement “Reading is an important activity in my home.” This variable differed from others in that its wording focused on the home environment, rather than parents as individuals. This item could be interpreted to mean that the child (and thus, his or her gender) in some way influenced the parent’s response. Perhaps because of this wording, we found significant gender differences in several countries (Table 6).

Students’ attitudes toward reading

The data thus far have demonstrated that home literacy environments tend to support girls more than boys, and that girls’ parents believe girls begin school with more literacy skills than do boys. Table 7 demonstrates that these differences persist through to Grade 4, and may influence the way that boys and girls feel about reading. The Student Attitudes Toward Reading Index (SATR) showed significantly higher percentages of girls than boys in the “high” category (indicating favorable

attitudes) and higher percentages of boys in the “low” category in every country. As with the other indices, the SATR was positively linked with achievement, and could help explain boys’ lower reading achievement.

Conclusions

These analyses provide strong evidence that girls are being raised in more supportive early literacy environments, and that they subsequently have stronger literacy skills when they begin school and more favorable attitudes toward reading when they reach Grade 4. This evidence reaffirms the idea that reading is a socially embedded activity that develops within a cultural context; thus, cultural roles play a part in the types of readers children become. This idea helps make sense of these data, as many cultures consider reading a feminine activity. Children most often engage in literacy activities with their mother, and many boys cite female family members as leading reading and writing activities in the home. In western cultures, particularly, this situation may contribute to boys’ disassociation from reading in an attempt to avoid a “feminine” activity (Millard, 2003).

The increasing support being given to early childhood education makes this information critical in terms of building a strong literacy foundation for both boys and girls. And while these data do not explain the reasons for such gender discrepancies, they do provide a strong impetus for further research in this arena. In particular, these findings beg the question: what activities are boys engaged in that are taking the place of these early literacy activities? This information could help us better understand the literacy background and needs of boys so that they can be addressed in a way that will encourage reading. Future international studies could explore this concern by expanding the information that is gathered regarding the early home environment. More specifically, future cycles of PIRLS could include more variables about home activities before students begin school in the Learning to Read Survey, and expand the Early Home Literacy Activities Index into a set of subscales that focuses on various facets of early literacy.

These data also show that, while the discrepancies between boys and girls are pervasive, they are not consistent. That is, countries often have highly significant differences for some variables and no differences on others, and there may be no discernible

Table 6: Percentages of Boys and Girls Whose Parents Reported Reading as an Important Activity in Their Homes (32 Countries)

Country	Strongly agree		Strongly disagree	
	Percent girls	Percent boys	Percent girls	Percent boys
Italy	44 (1.4)	43 (1.3)	3 (0.5)	4 (0.5)
France	45 (1.5)	42 (1.2)	2 (0.5)	3 (0.5) ▲
Colombia	75 (1.5)	75 (1.5)	4 (0.8)	4 (0.7)
Russian Federation	40 (1.4)	41 (1.4)	4 (0.5)	4 (0.5)
Czech Republic	34 (1.6)	33 (1.4)	4 (0.6)	5 (0.7)
Germany	43 (1.2) ▲	40 (1.1)	8 (0.6)	9 (0.6)
Romania	49 (1.6)	51 (1.8)	9 (1.1)	8 (1.1)
Hungary	60 (1.3)	58 (1.4)	3 (0.5)	4 (0.6)
Netherlands	34 (1.6) ▲	26 (1.7)	4 (0.5)	4 (0.6)
Slovak Republic	46 (1.7)	45 (1.6)	3 (0.5)	2 (0.4)
Lithuania	33 (1.5)	30 (1.4)	12 (1.0)	10 (0.9)
Scotland	62 (1.9)	58 (1.8)	2 (0.5)	4 (0.9)
Canada (O,Q)	59 (1.0)	57 (1.2)	1 (0.2)	2 (0.4) ▲
Argentina	51 (1.7)	48 (2.1)	3 (0.5)	3 (0.6)
Hong Kong SAR	35 (1.2) ▲	30 (1.2)	9 (0.9)	11 (0.7) ▲
Iceland	51 (1.3)	50 (1.4)	3 (0.4)	3 (0.4)
Turkey	56 (1.6) ▲	53 (1.4)	9 (1.3)	10 (1.0)
International Avg.	50 (0.3) ▲	48 (0.3)	5 (0.1)	5 (0.1)
Greece	56 (1.8)	58 (1.7)	7 (1.1)	7 (0.9)
Macedonia, Rep of	72 (1.5) ▲	64 (1.6)	5 (1.1)	6 (1.1)
Norway	53 (1.5)	52 (1.5)	4 (0.8)	4 (0.5)
Slovenia	30 (1.3)	29 (1.5)	4 (0.6)	4 (0.6)
Latvia	27 (1.5)	26 (2.1)	3 (0.8)	2 (0.3)
Sweden	54 (1.3)	54 (1.0)	2 (0.3)	2 (0.3)
England	61 (2.4)	65 (1.7)	3 (0.6)	3 (0.7)
Cyprus	53 (1.4)	51 (1.5)	5 (0.7)	5 (0.6)
Bulgaria	47 (1.7)	46 (1.9)	9 (1.4)	11 (1.8)
Singapore	56 (1.1) ▲	53 (1.0)	4 (0.5)	6 (0.5)
Moldova, Rep of	33 (2.1)	30 (1.6)	18 (1.6)	17 (1.2)
New Zealand	69 (1.8)	66 (2.0)	2 (0.5)	1 (0.4)
Iran	67 (2.1)	64 (1.7)	5 (0.9)	6 (1.1)
Belize	69 (2.2)	66 (2.3)	5 (0.9)	5 (1.0)
Kuwait	42 (1.0)	42 (1.1)	5 (0.3)	6 (0.6)

Note: ▲ Significantly higher percentage of students than other gender.

pattern within or across countries, and no direct relationship to the gender gap in achievement. We suspect that these incongruities are also indicative of the cultural context of literacy, demonstrating that reading literacy develops differently in different cultures. Future research focusing on the culture of reading in particular countries could delve further

into this topic, and would be useful in illuminating patterns that are not visible at this large scale. Such steps would give us a clearer understanding of boys' early childhood experiences and allow us to better explain and address these troubling gender differences in reading achievement.

Table 7: Percentages of Boys and Girls at Each Level of the Student Attitudes Toward Reading Index (SATR) (32 Countries)

Country	High SATR		Medium SATR		Low SATR	
	Percent girls	Percent boys	Percent girls	Percent boys	Percent girls	Percent boys
Italy	37 (1.5) ▲	46 (1.4)	30 (1.4)	45 (1.4) ▲	3 (0.5)	9 (0.8) ▲
France	65 (1.5) ▲	51 (1.4)	33 (1.6)	44 (1.4) ▲	2 (0.4)	5 (0.5) ▲
Colombia	52 (1.7) ▲	42 (2.2)	46 (1.6)	55 (2.0) ▲	2 (0.4)	3 (0.5) ▲
Russian Federation	63 (1.8) ▲	46 (2.0)	35 (1.8)	49 (1.8) ▲	2 (0.4)	5 (0.6) ▲
Czech Republic	49 (1.5) ▲	33 (1.6)	47 (1.4)	57 (1.7) ▲	4 (0.7)	10 (1.0) ▲
Germany	61 (1.2) ▲	40 (1.1)	35 (1.1)	45 (0.8) ▲	4 (0.4)	15 (0.8) ▲
Romania	67 (1.8) ▲	54 (2.1)	33 (1.7)	44 (2.1) ▲	1 (0.2)	2 (0.5) ▲
Hungary	59 (1.4) ▲	40 (1.6)	35 (1.2)	46 (1.4) ▲	6 (0.6)	14 (1.2) ▲
Netherlands	57 (1.6) ▲	30 (1.6)	37 (1.3)	48 (1.2) ▲	6 (0.6)	23 (1.3) ▲
Slovak Republic	54 (1.8) ▲	33 (1.7)	43 (1.7)	57 (1.5) ▲	4 (0.6)	9 (0.9) ▲
Lithuania	59 (1.7) ▲	33 (1.9)	38 (1.5)	58 (1.9) ▲	3 (0.6)	10 (0.9) ▲
Scotland	56 (1.6) ▲	38 (1.8)	39 (1.6)	45 (1.5) ▲	6 (0.7)	17 (1.5) ▲
Canada (O,Q)	65 (1.3) ▲	43 (1.2)	32 (1.2)	44 (1.2) ▲	4 (0.5)	13 (0.7) ▲
Argentina	50 (1.6) ▲	36 (1.5)	48 (1.5)	57 (1.3) ▲	2 (0.3)	7 (1.0) ▲
Hong Kong SAR	60 (1.3) ▲	38 (1.3)	39 (1.3)	54 (1.4) ▲	2 (0.3)	8 (0.6) ▲
Iceland	59 (1.5) ▲	39 (1.3)	39 (1.4)	51 (1.5) ▲	2 (0.4)	9 (0.8) ▲
Turkey	66 (2.0) ▲	52 (1.5)	34 (1.9)	46 (1.3) ▲	1 (0.1)	1 (0.3) ▲
International Avg.	61 (0.3) ▲	42 (0.3)	36 (0.3)	48 (0.3) ▲	3 (0.1)	9 (0.2) ▲
Greece	68 (2.1) ▲	55 (2.1)	29 (1.9)	37 (1.8) ▲	3 (0.6)	8 (0.9) ▲
Macedonia, Rep of	68 (1.7) ▲	56 (1.8)	31 (1.7)	43 (1.8) ▲	0 (0.1)	1 (0.4) ▲
Norway	55 (1.4) ▲	33 (1.6)	40 (1.3)	54 (1.5) ▲	5 (0.6)	13 (1.2) ▲
Slovenia	71 (1.7) ▲	46 (1.7)	27 (1.6)	43 (1.5) ▲	2 (0.4)	10 (1.2) ▲
Latvia	57 (1.9) ▲	29 (1.8)	39 (1.8)	58 (1.7) ▲	4 (0.6)	12 (1.0) ▲
Sweden	65 (1.4) ▲	43 (1.5)	32 (1.3)	46 (1.2) ▲	3 (0.4)	11 (0.9) ▲
England	53 (1.9) ▲	35 (1.5)	39 (1.6)	47 (1.5) ▲	8 (0.9)	18 (1.3) ▲
Cyprus	69 (1.8) ▲	45 (1.9)	28 (1.7)	45 (1.5) ▲	2 (0.5)	10 (1.1) ▲
Bulgaria	70 (1.6) ▲	49 (2.0)	28 (1.5)	45 (1.9) ▲	2 (0.4)	6 (0.8) ▲
Singapore	66 (1.4) ▲	42 (1.3)	33 (1.3)	51 (1.3) ▲	2 (0.3)	7 (0.5) ▲
Moldova, Rep of	72 (2.2) ▲	59 (2.0)	27 (2.1)	39 (2.0) ▲	1 (0.3)	3 (0.5) ▲
New Zealand	60 (1.7) ▲	43 (1.6)	36 (1.7)	46 (1.6) ▲	4 (0.7)	12 (1.2) ▲
Iran	74 (2.1) ▲	67 (2.6)	24 (2.0)	30 (2.4) ▲	2 (0.4)	3 (0.5)
Belize	36 (1.8) ▲	29 (1.6)	60 (2.2)	64 (1.6)	4 (0.8)	6 (0.8) ▲
Kuwait	55 (1.6) ▲	42 (2.0)	43 (1.6)	51 (2.0) ▲	2 (0.3)	6 (0.7) ▲

Note: ▲ Significantly higher percentage of students than other gender.

References

- Baker, L., Afflerbach, P., & Reinking, D. (1996). *Developing engaged readers in school and home communities*. Hillsdale, NJ: Earlbaum.
- Martin, M. O., Mullis, I. V. S., & Gonzalez, E. J. (2005). *Home environments fostering children's reading literacy: Results from the PIRLS 2001 study of reading literacy achievement in primary schools in 35 countries*. Paper presented at the International Research Conference of the International Association for the Evaluation of Educational Achievement, Lefkosia, Cyprus.
- Millard, E. (2003). Gender and early childhood literacy. In N. Hall, J. Larson, & J. Marsh (Eds.), *Handbook of early childhood literacy* (pp. 22–33). London: Sage Publications.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools in 35 countries*. Chestnut Hill, MA: Boston College.
- Nord, C. W., Lennon, J., Liu, B., & Chandler, K. (2000). Home literacy activities and signs of children's emerging literacy: 1993 and 1999. *Educational Statistics Quarterly*, 2(1). Retrieved May 10, 2006, from <http://nces.ed.gov/programs/quarterly/>
- Spiegel, D. (1994). A portrait of successful readers. In E. Cramer & M. Castle (Eds.), *Fostering the love of reading: The affective domain in reading education* (pp. 74–87). Newark, DE: International Reading Association.
- Statham, J. (1986). *Daughters and sons: Experiences of non-sexist childraising*. New York: Basil Blackwell.
- ter Laak, J. (1994). Literacy acquisition and social context: Approaches, emphases, and questions. In E. Assink (Ed.), *Literacy acquisition and social context* (pp. 220–237). Hemel Hempstead: Harvester Wheatsheaf.
- Yelland, N. (Ed.). (1998). *Gender in early childhood*. New York: Routledge.

A comparison of fourth-graders' academic self-concept and attitudes toward reading, mathematics, and science in PIRLS and TIMSS countries

Ann Kennedy and Kathleen Trong
Boston College
Chestnut Hill, Massachusetts, USA

Abstract

This study capitalized on the availability of data for Grade 4 students across subject areas to examine relationships between these students' attitudes and self-concepts and their achievement on the PIRLS 2001 reading achievement scale and the TIMSS 2003 mathematics and science achievement scales. The PIRLS 2001 and TIMSS 2003 student questionnaires included comparable items about student attitudes and self-perceived competencies in reading, mathematics, and science. Graphical displays

of the correlation between countries' average achievement scale scores and measures of attitude and self-concept were created overall and by gender across the three subject areas. These data plots were studied for emerging patterns in the groupings and positions of countries across the analyses for each subject area. Differences in patterns cross-culturally were explored further in terms of curricular characteristics.

Introduction

Achievement motivation has many dimensions, including affect and self-efficacy, and is believed to play a complex and important role in the development of positive attitudes and habits associated with high achievement. Furthermore, within the school curricula, characteristics related to motivation are themselves viewed as learning outcomes or goals. Results from the International Association for the Evaluation of Educational Achievement's (IEA) Trends in Mathematics and Science Study (TIMSS) revealed a positive association between mathematics and science achievement and both self-concept and attitudes within participating countries for Grade 8 students (Beaton, Martin, Mullis, Gonzalez, Smith, & Kelly, 1997; Beaton, Mullis, Martin, Gonzalez, Kelly, & Smith, 1997; Martin, Mullis, Gonzalez, & Chrostowski, 2004a, 2004b). Similar results were reported in PIRLS 2001 for relationships among the achievement motivation variables and reading achievement for Grade 4 students (Mullis, Martin, Gonzalez, & Kennedy, 2003).

The purpose of this research was to investigate patterns among the relationships between achievement

and those variables associated with achievement motivation, both within and across countries. Previous research on TIMSS 1999 data has focused on the phenomenon whereby within-country relationships are positive, but the direction of this relationship reverses when examined across countries. That is, when data are aggregated to the country level, many students in high-achieving countries demonstrate less favorable attitudes and lower confidence in their abilities than do students in lower-performing countries. The present study used the PIRLS 2001 and TIMSS 2003 data at Grade 4 to explore this phenomenon across the three curriculum areas of reading, mathematics, and science. We expected these relationships to differ by content in the context of curricular emphases, and that the results for Grade 4 students may reveal different patterns compared with prior research focusing on Grade 8 students. Exploring these variables has the potential to further our understanding of student background characteristics and achievement at the end of schooling in the primary grades, and how the relationship between the two is embedded in major school subject areas and cultural contexts.

Theoretical framework

Research in achievement motivation has expanded our knowledge and understanding of the complexities of students' motivation to learn and perform in an academic environment. Although there are many theories of achievement motivation, most consider motivation to include several interrelated constructs comprising both internal and external influences on student behaviors, attitudes, and achievement (Schunk, 1996). Students' self-concept, or perceptions of ability and task difficulty, as well as their interest or enjoyment of the subject are influenced in various ways through social interactions within their unique cultural contexts, in the home, at school, or in the community at large (Wigfield, Eccles, & Rodriguez, 1998).

Much research has been conducted on the relationship between self-concept and student achievement in terms of students' perceptions of ability and task difficulty, and beliefs of the value or utility of the subject (Valentine, DuBois, & Cooper, 2004). Results of many of these studies, focusing on the subject areas of mathematics and reading in secondary schools, show small but significant positive relationships in both subject areas (Marsh, Trautwein, Ludtke, Koller, & Baumert, 2005; Shen & Pedulla, 2000). Several studies confirm hypotheses that self-concept is a domain-specific measure, with mathematics and reading self-concept measures having little influence over one another (Marsh, 1990; Valentine et al., 2004).

Analyses of TIMSS 1995 and 1999 data for Grade 8 students suggested a directional shift in the relationship between self-concept and achievement when data are aggregated to the country level (Marsh & Hau, 2003; Shen, 2002; Shen & Pedulla, 2000; Wilkins, 2004). Researchers consider this phenomenon could be attributed to institutional differences in policies or curricula at the country level. They propose that countries with more demanding studies produce high-achieving students, but with a negative effect on students' self-concept. Such explanations align with Marsh and Hall's (2003) theory of the Big-Fish/Little-Pond effect, in which equally able students in higher-performing schools perceive themselves within the context of their environment and therefore are expected to have a lower self-concept. Thus, students with a more able peer group in a school with higher

expectations will see themselves as less able than those in a lower-performing school environment with less rigorous expectations. Skaalvik and Skaalvik (2002) suggest that more research should be done to explore students' frame of reference when responding to research questions about self-perception in order to better understand the response patterns for aggregated data.

Other researchers have examined the role of external factors, such as home and school contexts, on self-concept and attitudes. Studies of parental influences on self-concept and attitudes reveal that mothers' beliefs of self-efficacy, perceptions of their child's ability, and expectations for their child's future have a strong positive influence on their child's own self-perceptions and interest (Eccles Parsons, Adler, & Kaczala, 1982; Lynch, 2002). Hufton, Elliot, and Illushin (2002) explored the different ways that education is conceptualized in various cultures and the classroom contexts for learning. They found United States and English students more likely than Russian students to attribute academic success to innate ability; the Russian students were more likely to attribute success to hard work. Similarly, research by Li (2002, 2004) examined the different roles and purposes of education in Asian and western countries, attributing these differences to those found in the Confucian model of education in Asian countries and the Socratic model of their western counterparts. Asian countries are more likely to view education as a lifelong pursuit linked with one's moral and social status, whereas western cultures associate education with the attainment of a personal goal. Researchers from England examined possible cultural effects in England, a top-performing country in PIRLS 2001. They noted that fewer students in England and Scotland reported liking reading in comparison to other high-performing countries such as Sweden in PIRLS 2001 (Twist, Gnaldi, Schagen, & Morrison, 2004). However, Elliott and Bempechat (2002) and Lachman (1997) point out that attributing differences to culture often oversimplifies the complexities of contexts within cultures. They therefore encourage researchers to identify variables that seek to explain relationships within as well as between national contexts.

The link between student attitudes and achievement has been extensively studied and debated (Abu-Hilal, 2000), but it is generally agreed that there is a reciprocal

relationship (Marsh et al., 2005; Mullis, Kennedy, Martin, & Sainsbury, 2006; Valentine et al., 2004). Thus, within countries, students with more positive attitudes have higher achievement, which then feeds into more positive attitudes. Abu-Hilal's (2000) review of the research on attitudes toward achievement show inconclusive links between the two variables, although studies generally confirm that the association between attitudes and achievement is higher for reading than for mathematics and science. In one of the very few studies across the subject areas, Andre, Whigham, Hendrickson, and Chambers (1999) found that students reported liking both mathematics and reading better than science, with physical science reportedly liked the least, suggesting that relationships between attitudes and achievement may vary across subjects.

Data sources and methods

This study investigated comparable variables across the PIRLS 2001 and TIMSS 2003 studies that gauged students' attitudes, self-concept, and academic achievement in the subjects of reading, mathematics, and science. The questions provided in the background questionnaires administered to all participating students for each subject included ones focusing on students' enjoyment of the subject area as a measure of attitudes, and questions related to students' self-perceptions of subject competencies as a measure of students' self-concept. Academic achievement was measured on separate scales for each subject.

The PIRLS 2001 assessment was administered to Grade 4 students in 35 countries. The TIMSS 2003 assessment for Grade 4 was administered in 27 countries and benchmarking entities. The 19 countries and benchmarking participants that participated in both PIRLS 2001 and TIMSS 2003 (at Grade 4) were included in the analyses. The participants, listed in Table 1, represented a range of achievement, self-beliefs, and attitudes, as well as cultural contexts for learning.

Analyses were conducted using IEA's PIRLS 2001 and TIMSS 2003 international databases, each of which included data for approximately 80,000 Grade 4 students. This pair of studies represents the first time that IEA collected information on student attitudes and self-concept in Grade 4 reading, mathematics, and science using comparable variables. All countries included in these analyses were administered a student questionnaire—unique to each study—that asked students about various experiences and opinions relating to student achievement.

The PIRLS 2001 international database was used for the analyses of the reading achievement scale scores and the following student background questionnaire items, which measured student attitudes and self-concept, respectively:

- What do you think about reading? Tell how much you agree with the statement
"I enjoy reading."
- How well do you read? Tell how much you agree with the statement
"Reading is very easy for me."

Responses to each of these items were based on a four-point Likert scale, with the following options: agree a lot (4); agree a little (3); disagree a little (2); disagree a lot (1).

The TIMSS 2003 international database was used for analyses of the mathematics and science achievement scale scores and the following student background questionnaire items measuring student attitudes and self-concept, respectively:

- How much do you agree with these statements about learning mathematics/(science)
"I enjoy learning mathematics/(science)."
"I usually do well in mathematics/(science)."

Responses to each of these items were based on the same four-point scale as in PIRLS, with the following options: agree a lot (4); agree a little (3); disagree a little (2); disagree a lot (1).

Table 1: Countries that Participated in PIRLS 2001 and TIMSS 2003

Canada (O, Q)*	Iran, Islamic Rep. of	Morocco	Scotland
Cyprus	Italy	Netherlands	Singapore
England	Latvia	New Zealand	Slovenia
Hong Kong SAR	Lithuania	Norway	United States
Hungary	Moldova, Rep. of	Russian Federation	

Note: * Canada was represented by the provinces of Ontario and Quebec.

Within each country, a Pearson correlation coefficient (r) was calculated between each achievement scale score and the variables measuring attitudes and self-concept by subject area. To examine the relationships among variables between countries, scatterplots were constructed to show the relationships between country-level aggregates of average student achievement scores and average scores on the attitude items as well as separate scatterplots with average student achievement scores and average scores on self-concept items. Finally, to assist in the interpretation of the scatterplots across subjects, countries were organized in a table according to their position in one of four quadrants of the graphs based on average achievement scores and average scores on each variable:

- Above/Above, in which the average achievement score was above the international scale average and the average score on the attitude/(self-concept) variable was above the average score across the 19 participating countries.
- Below/Below, in which the average achievement score was below the international scale average and the average score on the attitude/(self-concept) variable was below the average score across the 19 participating countries.
- Above/Below, in which the average achievement score was above the international scale average and the average score on the attitude/(self-concept) variable was below the average score across the 19 participating countries.
- Below/Above, in which the average achievement score was below the international scale average and the average score on the attitude/(self-concept) variable was above the average score across the 19 participating countries.

Results

Student achievement

To put the results of the correlation analyses into context, it is important to note the distributions of achievement across the three subject areas. Table 2 shows the average scale scores for reading, mathematics, and science achievement arranged alphabetically by country. Each of the achievement scales has a scale average of 500 and a standard deviation of 100. Note also that while the same sample of students was measured on the mathematics and science scales, a separate sample of students was measured on the reading achievement scale.

Within each subject area, scale scores varied across countries, ranging from 304 to 565 on the science achievement scale, 347 to 594 on mathematics, and 350 to 554 on reading. The relative position of each country along the scales varied by subject. For example, while Singapore was the top performing country on the mathematics and science scales, it was outperformed in reading achievement by the Netherlands, England, Latvia, Canada (O, Q), Hungary, Lithuania, the United States, and Italy.

Within-country correlations

Tables 3 and 4 present the within-country correlations between achievement scale scores and scores on the self-concept and attitude variables. All correlations in the two tables were significant at the .01 level unless noted otherwise. Before discussing the results, we need to point out that while nearly all relationships were significant due to the large sample sizes in the data, the magnitude of the relationships was small or, in some cases, negligible. Generally, we refer in the discussion to the strength of the relationships in terms of the coefficient of determination, calculated by squaring the correlation coefficient, which is interpreted as the percent of variation in achievement scores that can be explained by the different responses to the background variables.

Within-country correlations: Self-concept

The relationship between achievement and self-concept (see Table 3) differed according to subject area. The relationship appeared to be strongest for the mathematics scale, with nearly 80% of the countries having correlation coefficients greater than .22. In other words, these countries were able to explain at least 5% of the variance in mathematics achievement scores by the students' measures of self-concept. In eight of these countries—Canada, Cyprus, Hong Kong SAR, Hungary, Latvia, the Netherlands, Singapore, and the United States—the percent of variance explained was greater than 9%. With regard to the reading scale, more than one-third of the countries demonstrated a small positive relationship between achievement and self-concept, with $r > .22$. Countries in which the percent of variance explained was greater than 9% were England and Norway. The weakest relationship between self-concept and achievement across countries was for the science achievement scale. Only one country, Hungary, showed a correlation coefficient greater than .22.

Table 2: Average Achievement Scores for Reading, Mathematics, and Science Achievement Scales Arranged Alphabetically by Country

Country	Reading achievement			Math achievement			Science achievement		
Canada (O,Q)	544 (2.4)	▲		509 (2.5)	▲		525 (2.7)	▲	
Cyprus	494 (3.0)	▼		510 (2.4)	▲		480 (2.7)	▼	
England	553 (3.4)	▲		531 (3.7)	▲		540 (3.6)	▲	
Hong Kong SAR	528 (3.1)	▲		575 (3.2)	▲		542 (3.1)	▲	
Hungary	543 (2.2)	▲		529 (3.1)	▲		530 (3.0)	▲	
Iran	414 (4.2)	▼		389 (4.2)	▼		414 (4.1)	▼	
Italy	541 (2.4)	▲		503 (3.7)			516 (3.8)	▲	
Latvia	545 (2.3)	▲		536 (2.8)	▲		532 (2.5)	▲	
Lithuania	543 (2.6)	▲		534 (2.8)	▲		512 (2.6)	▲	
Moldova, Rep. of	492 (4.0)	▼		504 (4.9)			496 (4.6)		
Morocco	350 (9.7)	▼		347 (5.1)	▼		304 (6.7)	▼	
Netherlands	554 (2.5)	▲		540 (2.1)	▲		525 (2.0)	▲	
New Zealand	529 (3.6)	▼		493 (2.2)	▼		520 (2.5)	▲	
Norway	499 (2.9)			451 (2.3)	▼		466 (2.6)	▼	
Russian Federation	528 (4.4)	▲		532 (4.8)	▲		526 (5.2)	▲	
Scotland	528 (3.6)	▲		490 (3.3)	▼		502 (2.9)		
Singapore	528 (5.2)	▲		594 (5.6)	▲		565 (5.6)	▲	
Slovenia	502 (2.0)			479 (2.6)	▼		490 (2.5)	▼	
United States	542 (3.8)	▲		518 (2.4)	▲		536 (2.5)	▲	
International Scale Avg.	500 (0.0)			500 (0.0)			500 (0.0)		

Notes: ▲ Significantly higher than international scale average.

▼ Significantly lower than international scale average.

Table 3: Within-country Correlations: Achievement and Self-concept

Country	Correlation with "Reading is very easy for me"	Correlation with "I usually do well in math"	Correlation with "I usually do well in science"
Canada (O,Q)	0.29	0.39	0.22
Cyprus	0.16	0.30	0.14
England	0.33	0.25	0.17
Hong Kong SAR	0.18	0.31	0.17
Hungary	0.18	0.39	0.25
Iran	0.16	0.24	0.18
Italy	0.16	0.25	0.13
Latvia	0.29	0.37	0.19
Lithuania	0.24	0.25	0.14
Moldova, Rep. of	0.19	0.19	0.18
Morocco	-0.01*	0.11	0.00*
Netherlands	0.22	0.33	0.19
New Zealand	0.29	0.27	0.13
Norway	0.34	0.22	0.11
Russian Federation	0.10	0.23	0.19
Scotland	0.21	0.21	0.06
Singapore	0.17	0.39	0.13
Slovenia	0.28	0.23	0.11
United States	0.29	0.31	0.19

Note: * Not significant.

Table 4: Within-country Correlations: Achievement and Attitude

Country	Correlation with "I enjoy reading"	Correlation with "I enjoy learning math"	Correlation with "I enjoy learning science"
Canada (O,Q)	0.27	0.08	-0.03
Cyprus	0.10	0.11	0.01*
England	0.29	0.03	-0.04
Hong Kong, SAR	0.23	0.20	0.12
Hungary	0.23	0.16	0.08
Iran	0.24	0.20	0.17
Italy	0.18	0.12	0.03
Latvia	0.17	0.15	0.00*
Lithuania	0.13	0.13	-0.03
Moldova, Rep. of	0.12	0.15	0.11
Morocco	0.14	0.17	0.15
Netherlands	0.29	0.06	-0.03
New Zealand	0.22	0.03	0.01
Norway	0.26	0.06	0.03
Russian Federation	0.07	0.14	0.03
Scotland	0.21	-0.01**	0.01*
Singapore	0.28	0.17	-0.02
Slovenia	0.26	0.07	-0.01*
United States	0.24	0.05	0.01

Note: * Not significant; ** significant at .05.

Within-country correlations: Attitudes

The relationship between achievement and attitudes (see Table 4) also differed according to subject area. In the case of student attitudes, the relationship appeared to be strongest for the reading scale, with students' measures of self-concept explaining between 5% and 9% of the variance in reading achievement scores in more than half of the participating countries. With regard to the other subject scales, the relationships between achievement and Grade 4 students' attitudes were negligible. No countries had correlation coefficients greater than .2 for mathematics or science.

The relationship between reading and attitudes was perhaps stronger because of the nature of the subject area. Many students are able to read independently by Grade 4 and can choose books that suit their interests. However, students may not reach this level of autonomy in mathematics and science until the later grades. Also, reading tends to transcend the context of school, as children often develop reading habits for purposes other than school-related tasks.

Cross-country correlations

Scatterplots for cross-country correlations between aggregated average scale scores and average scores for the background variables are displayed in Figures 1 (self-concept) and 2 (attitudes). When we aggregated the data to the country level, the results revealed negative correlation patterns similar to those reported in the existing TIMSS research at Grade 8 (Shen, 2002; Shen & Pedulla, 2000). Across the three subject areas, higher achieving countries reportedly had lower values of self-concept. The same pattern appeared to hold true for attitudes as well, with students in higher achieving countries on average having poorer attitudes toward the subject areas. This negative relationship for both self-concept and attitudes was most pronounced for mathematics.

Dotted lines divide each scatterplot into four quadrants. The upper-right quadrant (Above/Above) holds those countries with average scale scores higher than the scale mean of 500 and with average scores on the background variables higher than the average across the 19 participating countries. These countries have response patterns for the motivation variables

that align with their within-country associations. The lower-right quadrant (Below/Above) contains those countries performing below the achievement scale average but with average scores on the motivation variables. The countries in the upper-left quadrant (Above/Below) are those scoring above the achievement scale average but

lower-left quadrant (Below/Below) are those countries scoring below the achievement scale average and with below-average scores on the motivation variables. The countries in the upper-left quadrant (Above/Below) are those scoring above the achievement scale average but

Figure 1: Cross-country Scatterplots: Achievement and Self-concept

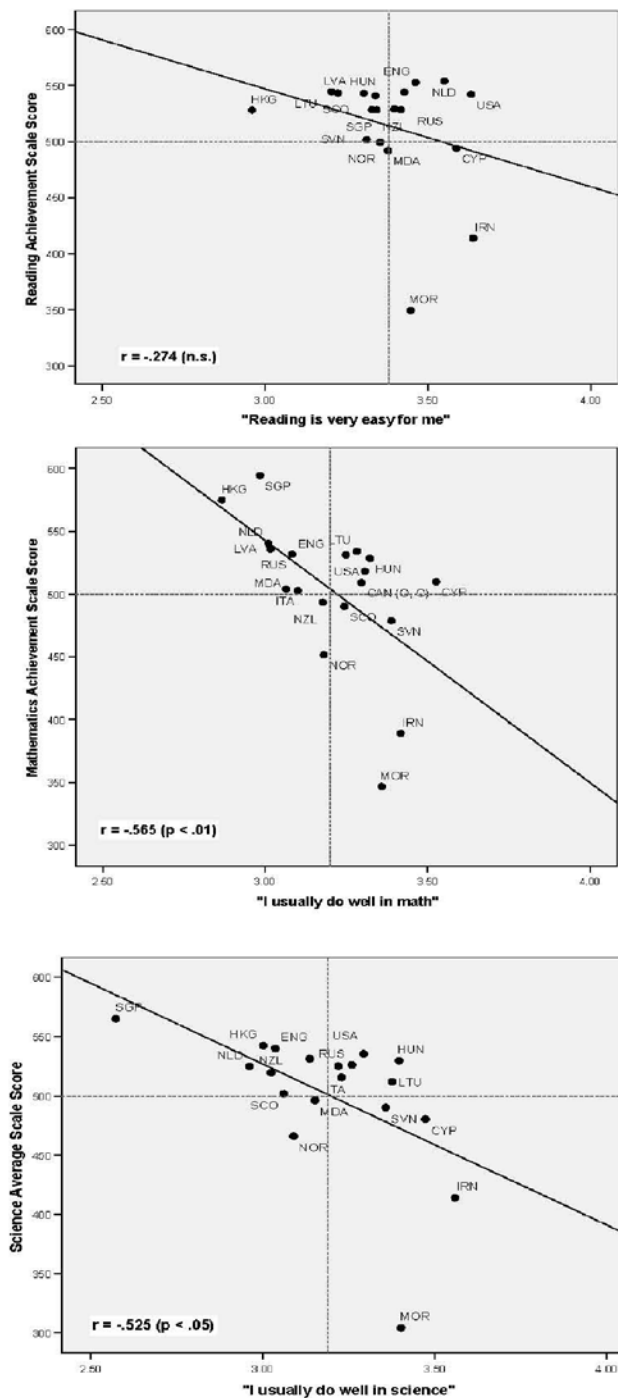
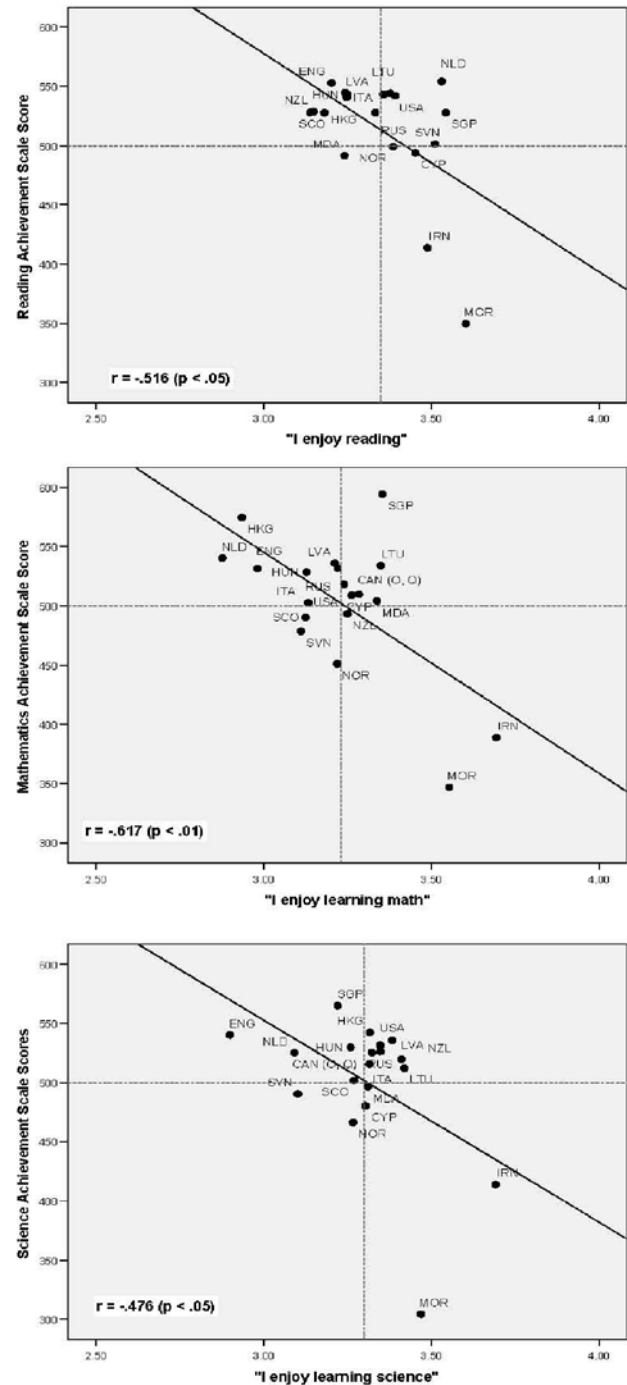


Figure 2: Cross-country Scatterplots: Achievement and Attitudes



with below-average scores on the motivation variables. The countries in the Above/Below and the Below/Above quadrants display response patterns at the aggregate level that are inconsistent with expectations, given the positive correlations within countries. Thus, students in the countries in these quadrants had a lower self-concept and more negative attitudes despite their good performance, or they had a self-concept and attitudes that were inflated relative to their poorer performance.

Tables 5 and 6 show which quadrant each country appeared in across the subject areas for self-concept and attitudes, respectively. Our discussion of these tables centers on those countries within the shaded areas, that is, those countries with high achievement and low self-concept or attitudes, and those countries with low achievement and high self-concept or attitudes. We chose to focus on these countries because they were the contributors to the inverse relationship between achievement and motivation. Furthermore, these relationships tend to contradict expectations based on within-country observations as well as theoretical frameworks for achievement motivation and learning.

Self-concept

Above/Below

Students in the countries in the Above/Below quadrant for self-concept had low self-concept in relation to their overall performance on the achievement scales (Table 5). Note that three countries, Hong Kong SAR, Latvia, and Singapore, fall into this category for all subject areas. Three additional countries appear for two subject areas: Italy (reading and mathematics), The Netherlands (mathematics and science), and Scotland (reading and science). This pattern suggests a systemic influence on self-concept within individual countries. Exploration of curriculum-related variables from the TIMSS 2003 international report, such as content coverage and instructional time, did not reveal any clear patterns across the countries. Within-country analyses are therefore necessary to allow us to make solid conclusions rather than anecdotal ones.

The countries within this quadrant are culturally diverse, although the two Asian countries, Singapore and Hong Kong SAR, are included in this group. The diversity leads us to conclude that while there may be a set of variables that explain a country's position within this quadrant, this set of variables may be unique to

each country and its social and educational contexts. The appearance of both participating Asian countries in this quadrant supports Shen's (2002) claims that curricula in Asian countries tend to be rigorous, leading to lower self-concept.

Below/Above

Countries in the Below/Above quadrant for self-concept had students with high self-concept in relation to their overall performance on the achievement scales. Iran and Morocco can be found in this category for all three subjects. Investigation of the percentage of TIMSS science and mathematics topics covered up to and including Grade 4 showed Morocco having covered 3% of the science topics and 21% of the mathematics topics for all students (Martin et al., 2004a, 2004b). These low percentages may indicate a curriculum that is less demanding than the mathematics curricula in other countries, thereby creating an effect opposite to that found in the Asian countries.

Attitudes

When examining the attitudes table (Table 6), it is important to keep in mind the lack of a noteworthy relationship between attitudes and achievement within the subject areas of mathematics and science, as reported in the literature.

Above/Below

Students in the countries in the Above/Below quadrant had low self-concept in relation to their overall performance on the achievement scales. Note that countries within this quadrant differ slightly from those in the self-concept table (i.e., Table 5). Three countries fall within this category for all subject areas: England, Hungary, and Scotland. Other countries appearing for more than one subject area include Hong Kong SAR, Italy, Latvia, and the Russian Federation (reading and mathematics), and the Netherlands (mathematics and science). Again, the patterns suggest some systemic influences on student attitudes within countries.

The location of England and Scotland in this quadrant is not surprising for reading, given Twist et al.'s (2004) research on the PIRLS results for these countries in comparison with the results for other high-performing countries. Explanation for England and Scotland in this location also comes from Sainsbury and Schagen's (2004) analysis of negative trends in attitudes occurring over the five-year period between

Table 5: Countries Arranged by Quadrants for Achievement and Self-concept

Reading	Above/Below Mathematics	Science	Reading	Above/Above Mathematics	Science
Hong Kong SAR	Hong Kong SAR	England	Canada (O,Q)	Canada (O,Q)	Canada (O,Q)
Hungary	Italy	Hong Kong SAR	England	Cyprus	Hungary
Italy	Latvia	Latvia	Netherlands	England	Italy
Latvia	Moldova, Rep. of	Netherlands	New Zealand	Hungary	Lithuania
Lithuania	Netherlands	New Zealand	Russian Fed.	Lithuania	Russian Fed.
Singapore	Russian Fed.	Scotland	United States	Scotland	United States
Scotland	Singapore	Singapore		United States	
Slovenia					
Below/Below			Below/Above		
Reading	Mathematics	Science	Reading	Mathematics	Science
Norway	New Zealand	Moldova, Rep. of	Cyprus	Iran, Islamic Rep. of	Cyprus
	Norway	Norway	Iran, Islamic Rep. of	Morocco	Iran, Islamic Rep. of
			Moldova, Rep. of	Slovenia	Morocco
			Morocco		Slovenia

Table 6: Countries Arranged by Quadrants for Achievement and Attitudes

Reading	Above/Below Mathematics	Science	Reading	Above/Above Mathematics	Science
England	England	England	Canada (O,Q)	Canada (O,Q)	Canada (O,Q)
Hong Kong SAR	Hong Kong SAR	Hungary	Lithuania	Cyprus	Hong Kong SAR
Hungary	Hungary	Netherlands	Netherlands	Lithuania	Italy
Italy	Italy	Scotland	Singapore	Moldova, Rep. of	Latvia
Latvia	Latvia	Singapore	Slovenia	Singapore	Lithuania
New Zealand	Netherlands		United States	United States	New Zealand
Russian Fed.	Russian Fed.				Russian Fed.
Scotland	Scotland				United States
Reading	Above/Below Mathematics	Science	Reading	Above/Above Mathematics	Science
Moldova, Rep. of	Norway	Norway	Cyprus	Iran, Islamic Rep. of	Cyprus
	Slovenia	Slovenia	Iran, Islamic Rep. of	Morocco	Iran, Islamic Rep. of
			Morocco	New Zealand	Moldova, Rep. of
			Norway		Morocco

1998 and 2003, during which the British Government implemented a major initiative, the National Literacy Strategy (NLS). Further investigations of the national contexts are necessary to understand the complexities of these relationships not only for England and Scotland, but also for the other countries within this category.

Below/Above

Students in the countries in the Below/Above quadrant for attitudes had positive attitudes in relation to their overall performance on the achievement scales. Similar to the patterns for self-concept, both Iran and Morocco sit within this category across subject areas. One explanation for Iran's positive attitude toward reading may be the explicit goals of enjoying and valuing the utility of reading within the country's Farsi Language and Literature Curriculum for Primary Education

(Mullis, Martin, Kennedy, & Flaherty, 2002). Additional within-country analyses are necessary for more detailed conclusions.

Conclusions

The findings of this study support the existing research on both positive relationships within countries and negative relationships when aggregated to the country level between self-concept and achievement in the three subject areas. The results at Grade 4 broaden our understanding of these relationships beyond Grade 8. Also consistent with previous research is the discovery of varying relationships between achievement and motivation across subjects. While it is not uncommon to find research on both mathematics and reading, research involving science self-concepts and attitudes is not as prevalent.

Another key finding of this study is evidence of systemic relationships within countries across subject areas. For example, several countries exhibited consistent patterns of high achievement and low motivation across at least two subject areas. Worth mentioning is the absence of a pattern of countries within the quadrants. Although the two Asian countries appear in the same quadrant for self-concept, the quadrant also includes countries from Western and Eastern Europe. It is clear that explanations focusing solely on the differences between Confucian and Socratic

models of education are inappropriate. Certainly, more information is needed to understand these complex relationships, including deliberate survey development plans in further studies, and qualitative exploration of social and educational contexts at various levels (e.g., classroom, schools, national initiatives).

The results of this study will contribute to the ongoing discussion of achievement motivation in education. Policymakers can use these findings to inform the implementation of programs that will best enable students to learn and feel positive about education. Additionally, these results can be used to promote further research using TIMSS and PIRLS data and to inform questionnaire development for future study cycles in terms of including variables that will enhance our understanding of attitudes and self-concept across countries and subject areas.

The intention of this research was exploratory in nature. We recognize that results and conclusions seem to open additional discussion on the theories of achievement motivation and the identification of contributing variables. Further analyses will be enhanced by publication of the PIRLS 2006 and TIMSS 2007 international databases, which will include, for the first time, information about reading, mathematics, and science curricula in Grade 4, as well as encyclopedias describing the education systems in the primary grades across the subject areas.

References

- Abu-Hilal, M. M. (2000). A structural model of attitudes towards school subjects, academic aspiration, and achievement. *Educational Psychology, 20*(1), 75–84.
- Andre, T., Whigham, M., Hendrickson, A., & Chambers, S. (1999). Competency beliefs, positive affect, and gender stereotypes of elementary students and their parents about science versus other school subjects. *Journal of Research in Science Teaching, 36*(6), 719–747.
- Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A., & Kelly, D. L. (1997). *Science achievement in the middle school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.
- Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1997). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.
- Eccles Parsons, J., Adler, T. F., & Kaczala, C. M. (1982). Socialization of achievement attitudes and beliefs: Parental influences. *Child Development, 53*(2), 310–321.
- Elliott, J. G., & Bempechat, J. (2002). The culture and contexts of achievement motivation. *New Directions for Child and Adolescent Development, 96*, 7–26.
- Huften, N., Elliott, J. G., & Illushin, L. (2002). Achievement motivation across cultures: Some puzzles and their implications for future research. *New Directions for Child and Adolescent Development, 96*, 65–85.
- Lachman, R. (1997). Taking another look at the elephant: Are we still (half) blind? Comments on the cross-cultural analysis of achievement motivation by Sagie et al. (1996). *Journal of Organizational Behavior, 18*, 317–321.
- Li, J. (2002). Learning models in different cultures. *New Directions for Child and Adolescent Development, 96*, 45–63.
- Li, J. (2004). "I learn and I grow big": Chinese preschoolers' purposes for learning. *International Journal of Behavioral Development, 28*(2), 116–128.
- Lynch, J. (2002). Parents' self-efficacy beliefs, parents' gender, children's reader self-perceptions, reading achievement and gender. *Journal of Research in Reading, 25*(1), 54–67.

- Marsh, H. W. (1990). Influences of internal and external frames of reference on the formation of math and English self-concepts. *Journal of Educational Psychology*, 82(1), 107–116.
- Marsh, H. W., & Hau, K. (2003). Big-fish-little-pond effect on academic self-concept: A cross-cultural (26 countries) test of the negative effect of academically selective schools. *American Psychologist*, 58(5), 364–376.
- Marsh, H. W., Trautwein, U., Ludtke, O., Koller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, 76(2), 397–416.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004a). *TIMSS 2003 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004b). *TIMSS 2003 international science report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Kennedy, A. M., Martin, M. O., & Sainsbury, M. (2006). *PIRLS 2006 assessment framework and specifications* (2nd ed.). Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools in 35 countries*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Flaherty, C. L. (2002). *PIRLS 2001 encyclopedia: A reference guide to reading education in the countries participating in IEA's Progress in International Reading Literacy Study (PIRLS)*. Chestnut Hill, MA: Boston College.
- Sainsbury, M., & Schagen, I. (2004). Attitudes to reading at ages nine and eleven. *Journal of Research in Reading*, 27(4), 373–386.
- Schunk, D. H. (1996). *Learning theories: An educational perspective* (2nd ed.). Englewood Cliffs, NJ: Merrill.
- Shen, C., (2002). Revisiting the relationship between students' achievement and their self-perceptions: A cross-national analysis based on TIMSS 1999 data. *Assessment in Education*, 9(2), 161–184.
- Shen, C., & Pedulla, J. J. (2000). The relationship between students' achievement and their self-perception of competence and rigour of mathematics and science: A cross-national analysis. *Assessment in Education*, 7(2), 237–253.
- Skaalvik, E. M., & Skaalvik, S. (2002). Internal and external frames of reference for academic self-concept. *Educational Psychologist*, 37(4), 233–244.
- Twist, L., Gnaldi, M., Schagen, I., & Morrison, J. (2004). Good readers but at a cost? Attitudes to reading in England. *Journal of Research in Reading*, 27(4), 387–400.
- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, 39(2), 111–133.
- Wigfield, A., Eccles, J. S., & Rodriguez, D. (1998). The development of children's motivation in school contexts. *Review of Research in Education*, 23, 77–118.
- Wilkins, J. L. M. (2004). Mathematics and science self-concept: An international investigation. *Journal of Experimental Education*, 72(4), 331–346.

Equating errors in international surveys in education

C. Monseur
University of Liège
Liège, Belgium

H. Sibberns
IEA Data Processing Center
Hamburg, Germany

D. Hastedt
IEA Data Processing Center
Hamburg, Germany

Abstract

For a decade, more or less, one of the major objectives of international surveys in education has been to report trends in achievement. For that purpose, a subset of items from previous data collections has been included in a new assessment test and the equating process (i.e., reporting the cognitive data of different data collections into a single scale) implemented through Item Response Theory models. Under IRT assumptions, the same equating function is obtained regardless of which common items are used because item-specific properties are fully accounted for by the item's IRT parameters. However, model misspecifications always occur, such as small changes in the items, position effects, and curriculum effects. Therefore, other sets of linked items

can generate other equating transformations, even with very large examinee samples. According to Michaelides and Haertel (2004), error due to the common-item sampling does not depend on the size of the examinee sample, but rather on the number of common items used. As such, this error could constitute the dominant source of error for summary scores. During its history, the International Association for the Evaluation of Educational Achievement (IEA) has reported trends in achievement through TIMSS 1999, TIMSS 2003, and PIRLS 2001, but has not added equating errors to the usual sampling and imputation errors, leading to an increase in Type I errors. It is for this reason that this study analyzes the variability of the trends estimate.

Introduction

Policymakers' interest in the monitoring of education systems and in measuring the effects of educational reforms has contributed to an increased emphasis on trend indicators in the design of recent surveys of educational achievement. Trends over time provide policymakers with information not only on how the achievement level of students in their country changes in comparison with the achievement levels of students in other countries, but also on how within-country differences, such as gender gaps in achievement, evolve over time. The progressive emphasis on trend indicators constitutes a major change in international surveys of education over the past decade. The names of two current IEA surveys reflect this growing interest: the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS).

Under IRT assumptions, the same equating function is obtained regardless of which common items are used because item-specific properties are fully accounted for by the item's IRT parameters. However, model misspecifications always occur. These include small changes in the items, position effects, and curriculum effects. This means other sets of linked items can generate other equating transformations, even with very large examinee samples. According to Michaelides and Haertel (2004), error due to common-item sampling depends not on the size of the examinee sample but on the number of common items used. As such, common error due to the common-item selection could constitute the dominant source of error for summary scores.

Although IEA reports trends indicators for achievement in its current studies, it bases the standard error on the trends estimates only on the standard

errors associated with the two mean achievement estimates used to compute the trends. The standard error has two components: the sampling uncertainty and the measurement uncertainty.¹ The PISA 2003 initial report also reports trends indicators in reading. As described in the PISA 2003 technical report, the standard error on the trends estimates adds a third error component, denoted linking error. This error reflects model misspecifications between the two data collections. However, the PISA 2003 linking error appears to be unsatisfactory because:

1. It assumes item independency, which is inconsistent with the embedded structure of items into units;
2. It considers partial credit items as dichotomous items; and
3. It takes only the international misspecifications between the two data collections into account.

Not recognizing the uncertainty due to the linking process leads to an underestimation of the linking errors and thus increases the Type I error, thereby resulting in the reporting of significant changes in achievement when, in fact, these are not significant. Furthermore, results are usually interpreted and published without regard to the test used. In other words, IEA reports achievement results in terms of reading literacy, mathematics, and science in general and not in terms of, for example, reading literacy on a specific test, such as with the PIRLS test. It is also very likely that an achievement trend will be interpreted in terms of change in the student performance and not in terms of changes in achievement on the anchoring items. In this context, the political importance of trends in achievement should not be underestimated. Also, if scholars suggest educational reforms based on the significant shifts, they may actually end up offering inappropriate policy recommendations.

Throughout the history of international surveys of achievement in education, the IEA Reading Literacy Study has offered a unique opportunity to study the equating error. This is because the achievement test used in 2003 is exactly the same achievement test used by the IEA Reading Literacy Study in 1991. Indeed, in other surveys, instruments are different, changes in the test design can occur, or, as is the case in PISA, the relative importance of the domains can vary from one data collection to another.

Method

Nine countries participated in both the IEA Reading Literacy Study of 1991 and the Reading Literacy Repeat Study of 2001. However, the data from only eight countries were re-analyzed (Greece, Hungary, Iceland, Italy, New Zealand, Slovenia, Sweden, and the United States). For timing reasons, it was not possible to include the Singapore data.

The Reading Literacy Study 1991 performance instrument consisted of 108 items administered to all students, without any rotation (Wolf, 1995). The first 40 items, which assessed “*word recognition*,” were not included in our study because they showed a severe ceiling effect. Of the 68 remaining items, we deleted three from the database because they had been recoded “*not applicable*” for all students. We therefore had a pool of 65 items from which we could randomly select particular numbers of items.

Simulations were used to empirically compute the linking error. Two factors that might have had an impact on the linking distribution for the IEA Reading Literacy Study were (i) the number of anchor items, and (ii) the importance of the shifts in the item parameters between the two data collections. This present study therefore analyzed the variability of the linking error depending on the number of anchor items by using replication methods. Let us suppose that 20 items of the 65 were used in the IEA Reading Literacy Repeat Study. This would have resulted in about 28 millions of billions of possible tests of 20 items out of the pool of 65.

For this study, 50 tests of 20 items randomly selected from the item pool were constructed. The same method was used to construct 50 tests of 30 items, 50 tests of 40 items, and 50 tests of 50 items. Each of the data sets (i.e., eight countries by two data collections by 50 tests by four types of tests or 3,200 data sets) was submitted to ConQuest (Wu, Adams, & Wilson, 1997) for drawing plausible values. Note that no conditioning variable was used.

Before generation of the plausible value, random samples of 500 students per country and per data collection were drawn, and a joint calibration of the whole item pool performed to obtain the item parameters. The plausible values on the *logit* scale were

¹ Because student performance estimates are reported through plausible values, the measurement uncertainty corresponds to the imputation variance.

then transformed on a new scale with a mean of 500 and a standard deviation of 100 by using *senat* weight per test,² whatever the number of items included in the test. Thus, the distribution of the eight countries and the two data collections had a mean of 500 and a standard deviation of 100. The achievement trend was then computed per test by comparing the country mean at Time 1 (1991) and the country mean at Time 2 (2001). Finally, the mean and the standard deviation of the 50 trends estimated were computed per type of test.

Results

The average trends of the test all correlated with the reported trends in the international report (Martin, Mullis, Gonzalez, & Kennedy, 2003). A perfect correlation could not be expected because one country was not included in the analyses. Also, the scaling model in this approach (1PL) was different from the model used in the 10-year trend study (3PL).

Table 1 and Figure 1 present the linking error, that is, the standard deviation across the 50 trends estimate per type of test. As the table and figure show, the variability of the trends increases as the number of items decreases. These results clearly demonstrate the impact of the item selection on the trend estimates and advocate the use of a linking error for testing the significance level of a particular trend. Because, in international surveys, the link between two data collections usually is based on fewer than 40 items, the linking error is quite substantial, as it has more or less the same size as the sampling error. For instance, the standard errors on the achievement trend estimates in PIRLS Repeat (Martin et al., 2003) ranged from 3.7 to 7.4. No doubt, the outcomes of the test would differ for countries with low trend estimates.

Figure 1: Linking Error per Country and per Type of Test

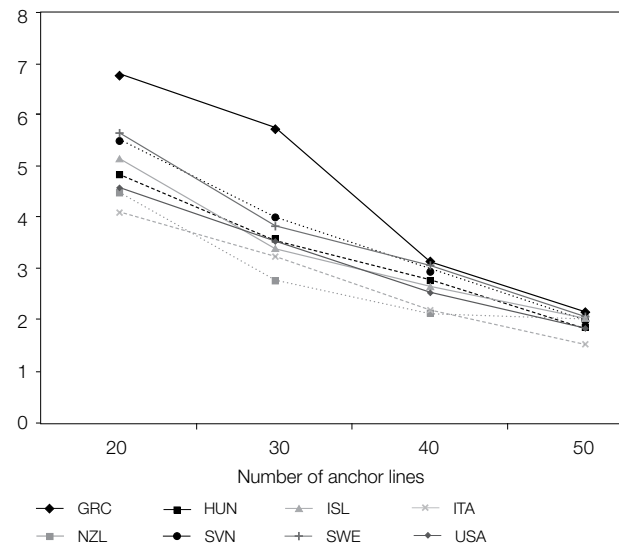


Table 1 and Figure 1 also show the variability of the linking error from one country to another for a particular test type. This observation implies that a single linking error for all countries is not as accurate as it should be. For example, the linking error is 6.78 for Greece but only 4.11 for Italy. Different analyses therefore were implemented in order to understand the outlying linking error for Greece.

First, the variability of the shifts in the national item parameters between 1991 and 2001 was computed for the eight countries. As expected, the variance of the national item parameters correlated at 0.49 with the linking error. In other words, the larger the shifts in the national item parameters, the larger the linking error. However, the factor that seemed to contribute mainly to the size of the linking error was the trends estimate. Table 2 provides the correlation between the *absolute* value of the trend estimates and the linking

Table 1: Linking Error per Country and per Type of Test

	GRC	HUN	ISL	ITA	NZL	SVN	SWE	USA
Test of 20 items	6.78	4.88	5.16	4.11	4.51	5.52	5.64	4.60
Test of 30 items	5.74	3.57	3.41	3.24	2.79	4.00	3.83	3.54
Test of 40 items	3.15	2.76	2.67	2.21	2.13	2.97	3.07	2.56
Test of 50 items	2.15	1.85	2.05	1.53	2.00	2.00	2.08	1.84

² Here, the sum of the student weights per country and per data collection is a constant, which means that each country contributed equally to the linear transformation.

errors per type of test. The table shows that as the trend estimate increased, the linking error increased. Finally, the linking error was computed for each country and gender. Table 3 and Figure 2 present the overall linking error, as well as the linking errors for gender. In three countries (Iceland, New Zealand, and Sweden), there was nearly no difference between the overall linking error and the linking error for each gender. For Hungary, the linking error for girls was actually higher than the overall linking error. Note, however, that for all countries, the linking error was higher for girls than for boys. Further research is necessary to explain these differences. It is possible that the item format was the main cause for the observed differences.

Table 2: Correlation between the Trend Estimate (Expressed in Absolute Value) and Its Linking Error

Type of test	Correlation
20	0.91
30	0.88
40	0.82
50	0.66

Figure 2: Overall Linking Error and Linking Error per Gender

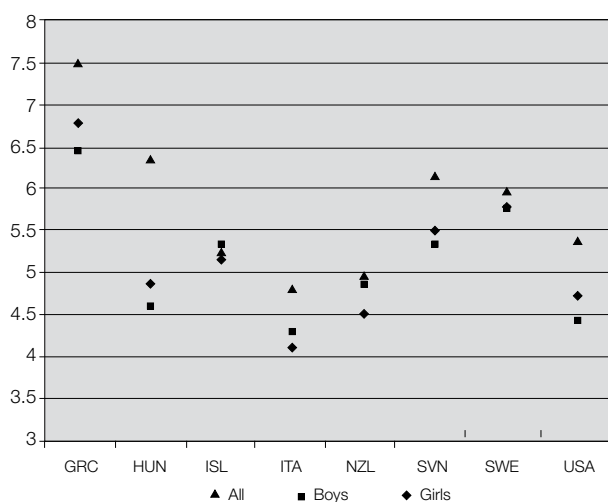


Table 3: Overall Linking Error and Linking Error for Gender

	GRC	HUN	ISL	ITA	NZL	SVN	SWE	USA
All	6.78	4.88	5.16	4.11	4.51	5.52	5.64	4.60
Boys	6.43	4.60	5.33	4.29	4.85	5.32	5.63	4.30
Girls	7.50	6.36	5.24	4.82	4.98	6.15	5.86	5.26

Conclusion

In 2004, the OECD PISA 2003 initial report (Organisation for Economic Co-operation and Development/OECD, 2004) also reported trends. However, as described in the OECD PISA 2003 technical report (OECD, 2005), the standard error of the trend estimate included a linking error. As discussed in Monseur and Berezner (2006), while the addition of a linking component in the standard error constituted a methodological improvement, it did raise several issues. In particular, the PISA 2003 linking error appears to be unsatisfactory because:

1. It made the assumption of item independency, which is inconsistent with the embedded structure of items into units;
2. It considered partial credit items as dichotomous items; and
3. It took into account only international misspecifications between the two data collections.

The results of the simulations presented in this study highlight the relationship between the number of items and the linking error and (more importantly) the variability of the linking error from one country to another. The linking error also correlated highly with the achievement trend estimates. The results also highlight the increase of the linking error for within-country analyses as shown by the gender example.

Further analyses should now be devoted to computation of the linking error on the final set of anchoring items. Replication methods like jackknifing and bootstrapping usually used in the sampling area might be of interest.

If policymakers and international report readers limited their interpretation of the trend estimates to the anchoring items, it would not be necessary to recommend the addition of a linking error. However, an improvement in student performance based on several dozen anchor-items is currently interpreted as an improvement for the students for the whole

domain assessed by the study. As such, the inclusion of a linking error in reporting trends would be consistent with how trends are presently interpreted.

According to Michaelides and Haertel (2004), common items should be considered as being chosen from a hypothetical infinite pool of potential items. Cronbach, Linn, Brennan, and Haertel (1997) also adhere to this point of view. Remember that a test score is based on an examinee's performance on a particular test form consisting of certain items. What is of most interest here is not how well the examinee did on those particular items at that particular occasion.

Rather it is the inference drawn from that example of performance to what the examinee could do across many other tasks requiring the application of the same skills and knowledge.

The interpretations of the trends indicators by policymakers and the arguments presented by scholars like Michaelides and Haertel (2004) and Cronbach et al. (1997) advocate for hypothetical infinite populations. In other words, even if a new international test did include all items from a previous survey, a linking error would still need to be reported. This linking error would reflect the model misspecifications.

References

- Cronbach, L. J., Linn, R. L., Brennan, R. L., & Haertel, E. H. (1997). Generalizability analysis for performance assessments of student achievement or school effectiveness. *Educational and Psychological Measurement*, 5(7), 373–399.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS trends in children's reading literacy achievement 1991–2001*. Chestnut Hill, MA: Boston College.
- Michaelides, M. P., & Haertel, E. H. (2004). *Sampling of common items: An unrecognized source of error in test equating*. Los Angeles: Center for the Study of Evaluation (CSE), University of California.
- Monseur, C., & Berezner, A. (2006). *The computation of linking error*. Paper presented at the AERA annual convention's symposium on measuring trends in international comparative research: Results from the first two cycles of the OECD/PISA study. San Francisco.
- Organisation for Economic Co-operation and Development (OECD). (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: Author.
- Organisation for Economic Co-operation and Development (OECD). (2005). *PISA 2003 technical report*. Paris: Author.
- Wolf, R. M. (1995). *The IEA Reading Literacy Study: Technical report*. The Hague: International Association for the Evaluation of Educational Achievement.
- Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Multi-Aspect test software* [computer program]. Camberwell, VIC: Australian Council for Educational Research.

Sociological perspectives on youth support for social movements

Oren Pizmony-Levy
Indiana University¹
Bloomington, Indiana, USA

Introduction

In the last four decades, social movements have expanded dramatically worldwide. Social movements have become an important vehicle through which collectivities give voice to their concerns about contested issues of our time (Snow, Soule, & Kriesi, 2004). Prominent social movements are the environmental movement (Frank, 1997; Frank, Hironaka, & Schofer, 2000), the peace movement (Kriesi, 1995), and the human-rights movement (Smith, 1995).

The development of social movements is accompanied by significant processes of institutionalization and professionalization, which have altered the way we conceptualize social movements (Snow et al., 2004). Thus, the common definition of the term is: *“collectivities acting with some degree of organization and continuity outside of institutional or organizational channels for the purpose of challenging or defending extant authority, whether it is institutionally or culturally based, in the group, organization, society, culture or world order of which they are a part”* (Snow et al., 2004, p. 11). The environmental and human-rights movements are prime examples of this definition because we can find in both of them local, national, and international organizations (e.g., Greenpeace, Amnesty International) that mobilize resources and serve as stable structures for continuity and success of the movements. These movements promote their agendas by engaging in various types of collective actions such as staging protests, collecting signatures on petitions and pressure-letters, lobbying governmental officials, and even initiating educational programs.

Another example of the institutionalization of social movements is found in the Civic Education Study (hereafter CivEd) carried out by the International Association for the Evaluation of Educational Achievement (hereafter IEA). The findings of the study show that, world-wide, 14-year-olds are “gravitating to

affiliation and action connected to social-movement groups and not to ... formal relations with political parties” (Torney-Purta, Lehmann, Oswald, & Schulz, 2001, p. 81). More specifically, 14-year-olds in all 28 countries sampled express readiness to vote in national elections as well as to engage in charity work (collecting money for a social cause, volunteering) and legal social-movement activities (collecting signatures, participating in non-violent demonstration).

Although the CivEd findings give us a notion about the civic engagement of future generations, I suggest we can interpret world-wide youth support as an immense achievement of contemporary movements. Within only a short time, social movements have gained remarkable legitimacy as an avenue for political participation. Torney-Purta et al. (2001) frame this process as the construction of “conventional citizenship” and “social-movement citizenship.” The latter term is similar to concepts suggested by other scholars. For example, according to the taxonomy of dimensions and role patterns of citizenship developed by Ichilov (1990), social-movement citizenship fits the broad definition of the citizen role in democracy where there is a balance between universalistic and particularistic orientations and where expression of disagreement with current situation is included. In addition, the term “social-movement citizenship” overlaps parts of the concept “global citizenship” (Falk, 1994; Noddings, 2005).

The growing importance of social movements in local, national, and international politics (Edwards & Gaventa, 2001; Meyer, Boli, Thomas, & Ramirez, 1997) together with the declining participation in conventional politics (for the United-States, see Verba, Scholzman, & Brady, 1995; for other western democracies, see Dalton & Kuechler, 1990; Gray & Caul, 2000) undermine the representativeness of

¹ Direct correspondence to Oren Pizmony-Levy, Department of Sociology, Indiana University, Bloomington 47401, Indiana (email: opizmony@indiana.edu). I thank Ivan Furre, Brian Powell and Christi Smith for their insights and advice. Support for the research was provided by the Discipline Based Scholarship in Education (DBSE) at Indiana University.

democracies. These factors also raise questions, such as those relating to who is participating and who is expected to participate and to become involved in politics.

The present study aimed to examine how series of individual characteristics combine with certain factors in national social/institutional contexts to mobilize youth support for social movements. The model suggested here is based on sociological literature regarding micro-level and macro-level determinants of participation in social movements (Guerin, Crete, & Mercier, 2001; Mertig & Dunlop, 2001; Schofer & Fourcade-Gourinchas, 2001), voluntary associations (Cornwell & Harrison, 2004; Curtis, Baer, & Grabb, 2001) and other altruistic behavior (Healy, 2000). In line with the CivEd theoretical framework (Torney-Purta et al., 2001), I will argue that young people's support for social movements is embedded within larger social contexts.

Individual-level explanations

Previous research has established a strong correlation between support and participation in social movements on the one hand, and social class, social capital, motivations, and values on the other (Klandermans, 2004; Kriesi, 1989; McAdam, 1988). Studies have found that the *new middle class* (e.g., highly educated people employed in the teaching, creative, welfare, and caring professions) is over-represented in social movements (Maheu, 1995; Rootes, 2004). Review of the literature points out four different theoretical explanations for this pattern. First, it is argued that the new middle class is relatively separated from the productive sector and the market. Its limited influence on important economic issues is therefore fertile soil for radical or critical views of the dominating social order. The second argument is that social movements give priority to issues such as identities and lifestyles and thus serve the needs of the new middle class (Klandermans, 2004; Melucci, 1996). Third, from a cultural standpoint, there is the argument that people from the new middle class hold the *cultural toolkit* of competencies, skills, and funds of knowledge that enable them to participate in social movements (Swidler, 1986). Another cultural explanation argues that participation in social movements is a type of "cultural capital" and, as such, signals recognition to fellow social-class members and excludes others

(Bourdieu, 1984; Strandbu & Krange, 2003).

In relation to *social capital*, studies have found that social networks may increase individuals' chances to become involved in politics and in social movements (Diani, 2004; McFarland & Thomas, 2006). Social networks and involvement in different organizations facilitate a process of social learning and diffusion of crucial information necessary for political involvement (Putnam, 2000; Verba et al., 1995). Studies on values orientations have found that holding *post-materialist* values (e.g., values of well-being and quality of life) explain high levels of support and participation in social movements (Inglehart, 1990, 1997; Inglehart & Baker, 2000).

Studies have found mixed results regarding gender differences in support for social movements (Barkan, 2004; Rootes, 2004). In the environmental movement, for example, women have been found to be more supportive and more engaged in local activities, yet less engaged at the national level. One theory is that women are more likely to have values systems consistent with the goals of social movements and voluntary associations. However, this relationship is not fully translated into action since women suffer from lack of biographical availability (McAdam, 1986; Rootes, 2004). McAdam (1986, p. 70) defines biographical availability as the "absence of personal constraints that may increase the costs and risks of movement participation, such as full-time employment, marriage and family responsibilities."

Contextual explanations

The importance of the social context to the ebb and the flow of social movements is well recognized (McAdam, McCarthy, & Zald, 1988; Snow et al., 2004). Review of the literature gleaned three different theoretical explanations for differences between countries in social-movement activities.

The first concerns level of economic development. This approach asserts that industrialization leads to occupational specialization, as well as to rising education and income levels. In turn, these changes increase social status distinctions in society, around which a diverse range of collective actions will form (Lipset, 1960, 1994). Following on from this argument, Inglehart (1990) found that as societies industrialize, individuals get more education and become wealthier and develop post-materialist values.

In more recent research, Inglehart and Baker (2000) show that national economic development is positively correlated with shifts toward participatory norms.

The second explanation relates to political opportunity structure. This approach argues that exogenous factors enhance or inhibit social-movement activities and mobilization into social movements (McAdam et al., 1988; Meyer & Minkoff, 2004). For example, democracy provides the political freedoms to engage in civic activities and is thought to foster a participatory political culture that leads to the formation of social movements. Another characteristic of the political structure is the openness of the political system. In the parliamentary system, for example, the executive branch of government depends on the direct or indirect support of the parliament. Hence, citizens have a greater potential influence on the government through participation in conventional politics. When the electoral system is based on proportional representation, this potential influence is even greater (International Institute for Democracy and Electoral Assistance/IDEA, 2005).²

The welfare regime approach is the third theoretical explanation for differences between countries in social-movement activities. This approach is rooted in the neo-institutionalism theory and contends that individual behavior is shaped or “scripted” by institutionalized cultural frames (Berger & Luckman, 1966; Meyer et al., 1997). In other words, this approach assumes that welfare regimes—institutionalized scripts—affect the level of support for social movements and expected participation in these movements.

According to Esping-Andersen (1990), the emergence of the welfare state is a product of complex interrelationships among social classes and social institutions. Based on historical analysis of western capitalist countries, Esping-Andersen identified three distinct types of welfare regimes:

1. The “liberal” welfare state common in Anglo-Saxon countries and characterized by limited, means-tested assistance with strict entitlement rules;
2. The “conservative/corporatist” welfare state more common on the continent of Europe in which the

state supplies welfare assistance but preserves many of the status differences of pre-modern society; and

3. The “social-democratic” welfare state in the Nordic countries, where welfare provision is separated from the market system.

Studies have found that welfare regimes are reflected in different facets of citizens’ attitudes and behavior (cf., Fraile & Ferrer, 2005; Scheepers & Grotenhuis, 2005). Salamon and Anheier (1998) found that the establishment of non-profit organizations (also known as third-sector organizations) is associated with particular welfare regimes. Hence, in countries characterized by a modest welfare regime (e.g., liberal) more non-profit organizations are established. In contrast, in countries characterized by a generous welfare regime (e.g., social-democratic), fewer non-profit organizations are established. The analysis by Schofer and Fourcade-Gourinchas (2001) of individual associational activity (as measured in the 1991 World Values Survey) generated similar findings.

Other studies point out different effects of the welfare regime on political behavior (Curtis et al., 2001; Curtis, Grabb, & Baer, 1992; Janoski, 1998). Drawing on work by Esping-Andersen (1990), Janoski identified three kinds of democracies—liberal, social-democratic, and traditional-corporatist. Janoski argued that due to specific institutional arrangements, traditional-corporatist democracies tend to lag behind both liberal and social democracies in forming voluntary associations. He also argued that, despite the similarity between liberal and social democracies, the creation of voluntary associations in each of them differs. In liberal countries, voluntary associations are a “substitute” for a strong welfare regime; in social democracies, the establishment of voluntary associations is encouraged by the government and in some cases is even state-funded (Janoski, 1998).

In sum, the theory and the empirical evidence indicate that support and participation in social movements is determined by individual-level characteristics and by contextual-level factors. Thus, the first objective of this study was to evaluate the extent to which individual characteristics influence

2 Research shows that electoral systems have psychological effects (IDEA, 2005). Voters who wish to support a minor party are often faced with a dilemma as how best to avoid “wasting” their vote. In these terms, proportional representation systems are much more open, and the pressure to vote strategically is reduced. In contrast, plural majority systems are less open and people vote without expressing their sincere choice but rather vote for the candidate they believe has a realistic chance of winning a seat.

youth support for social movements and the expected participation of young people in social-movement activities. The second objective was to estimate the impact of country-level attributes on these individual outcomes.

Data, measures, and method

I analyzed data from IEA's CivEd, which is the largest and most rigorous study of civic education ever conducted internationally (Torney-Purta et al., 2001). The study was designed to assess students' civic knowledge, their skills in interpreting political communication, and their civic attitudes and engagement. Hence, the research instrument encompassed items measuring students' concepts of democracy and citizenship; their attitudes related to trust in institutions, their nation, opportunities for immigrants, the political rights of women; and their expectations for future participation in civic-related activities. CivEd surveyed two populations. The standard population included all students enrolled on a full-time basis in Grade 8 (the grade in which most of the participating students were age 14). The optional population included students enrolled in upper secondary school (where the students' ages ranged from 16.6 to 19.4 years).

The CivEd data set is suited to the model proposed earlier because it provides comprehensive information on a variety of issues and concepts related to social movements, such as perceptions of governmental responsibilities (a proxy for post-materialist values) and membership in organizations (a proxy for potential social-capital and recruitment). Note, however, that the CivEd data are limited, since they relate to a cross-sectional and not a longitudinal design. Hence, the outcome variables examined here are restricted to attitudes and expected behavior (versus actual behavior). The standard population data set is particularly suited to the model because it is drawn from 28 countries (compared with 16 in the optional population), which are varied in terms of economic development, welfare regime, and level of democracy.

Measures

The analyses included two dependent variables, 11 independent variables at the individual level (social background and knowledge, attitudes, and behavior), two control variables at the individual level (parallel to the two dependent variables), and 10 independent

variables at the country level. Table 1 presents the basic descriptive statistics for these variables.

Dependent variables

1. *Support for Social-Movement Citizenship (SMC)* is a composite measure of four items describing two facets of social movements: content (environment, human rights) and means (general activities, peaceful protest). The respondents were asked about the importance of statements offering explanations for "what a good adult citizen is or does." The statements analyzed were: "takes part in activities to protect the environment;" "takes part in activities promoting human rights;" "participates in activities to benefit people in the community;" and "would participate in a peaceful protest against a law believed to be unjust."
2. *Expected Participation in Social-Movement Activities (SMA)* is a composite measure of four items describing common social-movement (and civil-society) activities. The respondents were asked how likely it would be for them to undertake different types of political action. The questions distinguished between "action that adults could take" and "actions that young people could take during the next few years." The actions analyzed here were: "volunteer time to help [benefit] [poor or elderly] people in the community;" "collect money for a social cause;" "collect signatures for a petition;" and "participate in a non-violent [peaceful] protest march or rally."

Independent variables (individual level)

1. *Gender*, coded 1 for females and 0 for males.
2. *Number of Books at Home*, ranging from 0 = none to 5 = more than 200 books. This variable is considered a proxy for socioeconomic status (Beaton, Martin, Mullis, Gonzales, Smith, & Kelly, 1996) and a proxy for a "scholarly culture" background (Kelley, Evans, & Sikora, forthcoming).
3. *Expected Years of Further Education*, ranging from 1 = 0 years to 7 = 10 plus years. This variable is considered a proxy for general motivation.
4. *Parental Education* is a composite index representing the mean number of years of mother's and father's education, ranging between 4 = partial elementary education to 15 = college education. This variable is considered a proxy for socioeconomic status and social class.

Table 1: Definitions and Descriptive Statistics for Variables Used in the Analyses: 28 Countries from the IEA Civic Education Study, Standard Population, 1999

Variable	Range	Mean	SD
<i>Dependent variables</i>			
Social-Movement Citizenship (SMC)	1–4	3.11	0.62
Expected Participation in Social-Movement Activities (SMA)	1–4	2.55	0.59
<i>Individual-level variables</i>			
Gender	0–1	0.52	0.50
Books	0–5	3.25	1.35
Expected Education	0–6	3.30	1.44
Parental Education	0–15	9.72	4.37
Civic Knowledge	0–38	24.77	7.76
Membership in Organizations	0–15	2.82	2.32
Discuss National Politics	1–4	2.21	0.74
Discuss International Politics	1–4	2.21	0.74
Attention to News	1–4	2.93	0.63
Perceived Governmental Role as Related to Economy	1–4	3.14	0.49
Perceived Governmental Role as Related to Society	1–4	3.43	0.46
Conventional Citizenship	1–4	2.73	0.53
Expected Participation in Conventional Politics Activities	1–4	2.36	0.55
<i>Country-level variables</i>			
Nation Size (in millions)	0.8–274	30.98	57.26
National Economic Development	4,809–29,605	15,307	7,661
Electoral System, Proportional/Representative	0–1	0.64	0.48
Governing System, Parliament-based	0–1	0.57	0.50
Women in National Parliament	5.6–42.7	18.05	10.67
<i>Welfare Regime</i>			
Liberal (reference group)	0–1	0.11	0.31
Social Democracy	0–1	0.14	0.36
Conservative	0–1	0.11	0.31
Former Socialist	0–1	0.39	0.50
Latin America	0–1	0.07	0.26
Southern Europe	0–1	0.14	0.36
Hong Kong SAR	0–1	0.04	0.19
<i>Freedom of Press</i>			
Free	0–1	0.86	0.36
Partly free	0–1	0.11	0.32
Not at all free	0–1	0.04	0.19
<i>Freedom in the World</i>			
Level at 1990	2–14	6.36	4.40
Level at 2000	2–9	3.39	1.81
Changes in Freedom	(0.50)–0.75	0.23	0.42

5. *Civic Knowledge* is the number of correct answers on the civics test, which consisted of 38 items. The items were designed as multiple-choice with correct and incorrect answers.³
6. *Membership in Organizations* is the number of organizations in which respondents were participating, for example, "student council/student government," "group which prepares a school newspaper," "sports organization or sports team." This variable is a proxy for the respondents' social network and affiliation with extra-curricular organizations.

In addition to the background variables and characteristics, the analyses included independent variables relating to attitudes and behavior. Discussion of politics involved two composite indexes reflecting the mean tendency to discuss politics with parents, teachers, and peers:

7. *Discussion of National Politics*; and
8. *Discussion of International Politics*.
9. *Attention to News* is a composite index reflecting the mean tendency to experience exposure to news by reading newspapers and by listening to news broadcasts on television/radio.
Respondents answered questions relating to these above three measures by using a scale comprising 1 = never, 2 = rarely, 3 = sometimes, 4 = frequently.
The last two independent variables in the analyses reflect youth perceptions of governmental role and responsibilities. Respondents were asked to decide whether different statements/descriptions "should or should not be the government's [state's] responsibility."
10. *Economy-Related Responsibilities* is a composite measure of five items: "guarantee a job for everyone who wants one;" "keep prices under control;" "provide industries with the support they need to grow;" "provide an adequate standard of living for the unemployed;" and "reduce differences in income and wealth among people."
11. *Society-Related Responsibilities* is a composite measure of seven items: "provide basic health care for everyone;" "provide an adequate standard of living for old people;" "provide free basic education for all;" "ensure equal political opportunities

for men and women;" "control pollution of the environment;" "guarantee peace and order within the country;" and "promote honesty and moral behavior among people in the country."

The answer range for these two measures was 1 = definitely should not be the government's responsibility, 2 = probably should not be, 3 = probably should be, and 4 = definitely should be the government's responsibility. These variables provide a proxy for post-materialist values (Torney-Purta et al., 2001).

Control variables (individual level)

I decided to control for the strong correlations between variables concerning conventional politics and the two dependent variables.

1. *Support for Conventional Citizenship* is a composite measure of six items explaining what a good adult citizen is or does: "votes in every election;" "joins a political party;" "knows about the country's history;" "follows political issues in the newspaper, on the radio, or on TV;" "shows respect for government representatives;" and "engages in political discussions."
2. *Expected Participation in Conventional Politics Activities* is a composite measure of five items describing common (and conventional) political actions: "vote in national elections;" "get information about candidates before voting in an election;" "join a political party;" "write letters to a newspaper about social or political concerns;" and "be a candidate for a local or city office."

Independent variables (country level)

After referring to previous research relating to contextual effects on participation in social movements and voluntary associations, I selected eight variables that capture differences between countries.⁴

1. I based *Nation Size*, as measured by population size in millions (UNDP, 2000), on Katzenstein's (2003) distinction between small states and big states.
2. For *National Economic Development*, I used the measurement of gross domestic product (GDP) per capita in US\$ (United Nations Development Program/UNDP, 2000).

3 Of the 38 items, only 16 items have been chosen for release. The 22 non-released items are being retained for possible use in future IEA studies in this area. Further information on the content categories to which the different items relate may be found in Torney-Purta et al. (2001, pp. 191–194). The full text of the 16 released items may be found online at <http://www.wam.umd.edu/~jtpurta/studentQ.htm>

4 Due to the scarce availability of data, I could not use theoretically justified variables such as size of civil society and adult engagement in social movements.

For testing the effect of political opportunity structure, I used common measures that reflect the level of democracy (Freedom House, 2006a, 2006b).⁵

3. I measured *Freedom of Press* as a series of three dichotomies: “free,” which included most of the countries; “partly free,” which included Colombia, Romania, and the Russian Federation; and “not free,” which included Hong Kong SAR.
4. I measured *Political Freedom* by referring to a number of “political rights” and “civil liberties” at two time-points: 1990 and 2000. To capture the amount of change between these time-points, I calculated another variable that reflected percentage change.

I also used three variables that reflect the structure of a country’s political system. Taken together, these variables are a good proxy for how open or closed a political system is (Inter-Parliamentary Union, 2005).

5. *Electoral System* (IDEA, 2005), coded 1 for countries with a proportional representation electoral system (Belgium, Bulgaria, Chile, Colombia, Cyprus, Czech Republic, Slovak Republic, Denmark, Estonia, Finland, Greece, Latvia, Norway, Poland, Portugal, Romania, Sweden, Slovenia) and 0 for all the rest (such as plurality/majority or mixed system).
6. *Governing System* (IDEA, 2005), coded 1 for countries with a parliamentary system (Bulgaria, Czech Republic, Estonia, Finland, Germany, Greece, Hungary, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovak Republic, Switzerland) and 0 for all the rest (such as presidential or constitutional monarchies).
7. *Women in National Parliament*, measured as the percentage of seats occupied by women.
8. Finally, I measured *Welfare Regime* as a series of seven dichotomies. Based on Esping-Andersen’s (1999) typology, I created three categories: “liberal,” which included Australia, the United Kingdom, and the United States; “social democracy,” which included Belgium, Denmark, Norway, and Sweden; and “conservative,” which included Finland, Germany, and Switzerland. In response to critiques and revision of this typology

(Esping-Andersen, 1999; Leibfried, 1992), I created a fourth category for Southern European regimes, namely Cyprus, Italy, Greece, and Portugal. Since former socialist countries share a similar legacy of welfare provision and are still in transition toward new welfare regimes (Deacon, 2000; Esping-Andersen, 1996), I created a fifth category—“former socialist” regimes. The relevant countries were Bulgaria, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, the Russian Federation, Slovakia, and Slovenia. According to the same rationale, I created a sixth category for Latin America regimes (Chile and Colombia). The last category had only one country, Hong Kong SAR, included here because of its unique welfare regime (Gough, 2001).

Method and analytic strategy

The first step in the empirical stage included two series of OLS regressions in which 12 individual-level variables were regressed on two dependent variables: Support for Social-Movement Citizenship (SMC) and Expected Participation in Social-Movement Activities (SMA). Here, I tested existing explanations and theories regarding individual-level characteristics of social-movement supporters. The OLS findings signified the range of cross-national variation in the social-movement supporters’ profile.

In the second step, I employed two series of hierarchical models to analyze the same dependent variables: SMC and SMA (Raudenbush & Bryk, 2000). Multi-level models are appropriate in this case because I am interested in an individual-level outcome that is affected by both individual-level and country-level variables. Using hierarchical models, I estimated equations at two levels: an individual-level or Level 1 equation (within-country), and a country-level or Level 2 equation (between-countries). The Level 1 equation takes the form

$$Y_{ij} = \beta_0j + \beta_1X_{1ij} + \beta_2X_{2ij} + \dots \beta_kX_{kij} + r_{ij}$$

and where subscript *i* refers to individual cases and *j* refers to countries. All independent variables are centered at the group mean and have random slopes.

5 Freedom House is an independent non-governmental organization that has surveyed and tracked trends in global political freedom and press freedom since 1980. Freedom House reports are widely utilized by social-science scholars. Further information regarding the methodology may be found online:

- Freedom in the world: <http://www.freedomhouse.org/template.cfm?page=35&year=2006>
- Freedom of press: <http://www.freedomhouse.org/template.cfm?page=56&year=2006>

Four Level 1 coefficients— β_0 , $\beta_{\text{(economy)}}$, $\beta_{\text{(society)}}$ and $\beta_{\text{(control)}}$ —are specified for each of the countries (j) and are, in turn, modeled as a function of country-level variables in a Level 2 equation that takes the form

$$\begin{aligned}\beta_0 &= \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}W_{2j} + \dots \gamma_{0m}W_{mj} + \mu_{0j} \\ \beta_1 &= \gamma_{10} + \gamma_{11}W_{1j} + \gamma_{12}W_{2j} + \dots \gamma_{1m}W_{mj} + \mu_{1j}\end{aligned}$$

and where W_{mj} refers to the value of country-level variable for any j country. Due to the centering method, I introduced compensation for lost variance in the intercept (β_0) equation (Raudenbush & Bryk, 2000; Snijders & Bosker, 1999).

Before presenting the findings, I would like to comment about the process of specifying the multi-level model (Snijders & Bosker, 1999, pp. 94–95). I began the specification by constructing a model for Level 1 that included all 12 independent variables. Using a deviance test, I trimmed the model and excluded the non-significant slopes from the model. Next, I specified a Level 2 model for the prediction of the intercept. The preliminary model included 10 country-level variables. Again, using a deviance test, I trimmed the model and excluded the non-significant slopes. Last, I specified a Level 2 model for the prediction of the three slopes: governmental responsibly related to economy/society and the control variable. Since the deviance test was significant, I did not exclude non-significant slopes. I used the same process for each of the dependent variables (i.e., SMC and SMA).

Findings

Individual variation by country

Table 2 shows the result from a series of OLS regressions for Attitudes Toward Social-Movement Citizenship (SMC) for each country in the sample. The findings indicated that females had stronger support than males for SMC in half of the countries (14 out of 28), most of them western democracies. This pattern was reversed in Belgium and Estonia, where males had stronger support for SMC. In other countries, most of them Eastern European, there was no significant gender gap in support for SMC.

According to social-movements literature, the level of support for SMC should increase with level of affluence. However, the findings presented here do not support this claim. In most of the countries, the coefficients of scholarly culture (number of books)

and parental education were not significant. In a few countries—Chile, Cyprus, Germany, Hungary, Italy, and Slovakia—these coefficients were actually negative and significant. Respondents from these countries who came from an affluent social background had lower support for SMC.

General motivation, as measured by the variable “expected years of education,” had a small and inconsistent effect on the support for SMC. In Belgium, Italy, Portugal, Romania, and Switzerland, the coefficient was positive and significant, but in the Czech Republic, Finland, and Latvia the coefficient was negative and significant. Furthermore, civic knowledge and membership in various organizations increased the level of support for SMC, but the size of these coefficients was nonetheless rather small (0.00–0.01).

Discussing national and international politics had no significant effect in the majority of countries. In Cyprus, young people who frequently discussed national politics had a higher level of support for SMC, while others who frequently discussed international politics had a lower level of support for SMC. In Estonia and Lithuania, young people who frequently discussed international politics had higher support for SMC. Similarly, attention to news had a positive and significant effect in only a few countries: Hong Kong SAR, Hungary, Romania, Russia, Slovakia, Sweden, and the United States.

According to social-movements literature, support for SMC is highly correlated with post-materialist values. Support for this claim was evident in the majority of countries (26 out of 28). The coefficients of the two variables that capture perceptions of governmental responsibilities were positive and significant. However, the coefficient of “responsibilities related to society” was stronger than the coefficient of “responsibilities related to economy.” On average, the ratio between these two coefficients was 1:2.

Last, as previously explained, I included Support for Conventional Citizenship as a control variable. The coefficients of this variable differed from one country to another. In some countries, respondents tended to separate the two modes of citizenship (small coefficients, such as Portugal $B = 0.24$); in other countries, respondents tended to combine them and to connect them (large coefficients, such as Hong Kong SAR $B = 0.64$).

Table 3 shows the results from a series of OLS regressions for Expected Participation in Social-Movement Activities (SMA) for each country in the sample. In general, the findings were similar to those described beforehand. In most countries, the likelihood of participation in SMA was positively correlated with gender (females), membership in organizations, discussion of national and/or international politics, and attention to news. The likelihood of participation in SMA also positively correlated with perception of governmental responsibilities, but more in relation to societal responsibility than to economic. In contrast to previous findings, the coefficients here were stronger and the pattern was more consistent among countries. As for the control variable, Expected Participation in Conventional Politics Activities, respondents in some countries tended to separate the two modes of political participation (small coefficients, such as Portugal and Germany $B = 0.24$), and in other countries to combine them and to connect them (large coefficients, such as Hong Kong SAR $B = 0.44$).

Support for social movements and the social context

The second step in the empirical stage encompassed two hierarchal models in which 11 individual-level variables and country-level variables were regressed on the two dependent variables.

Table 4 presents the results of hierarchal linear models predicting the Support for Social-Movement Citizenship (SMC). The fully unconditional model, Model 1, shows that differences between countries accounted for 11% of the variance in the variable. This figure was appropriate for employing hierarchical level modeling (Raudenbush & Bryk, 2000). The average level of support for SMC was high and equal to 3.09.

Model 2 included socio-demographic variables. Females were more supportive than males of SMC, as were young people with higher levels of social capital (membership in organization), higher general motivation (expected years of education), and higher scores on the civic knowledge test. In contrast to what is evident in social-movements literature, the social-class variables correlated negatively with the outcome variable. The association with scholarly culture slope (number of books) was negative and significant, while parental education was negative and marginally significant ($p < 0.1$).

Model 3 included individual-level attitude variables. All the variables had a positive and significant effect on support for SMC. Thus, the respondents who discussed international politics, paid attention to news, and supported government role as related to economy and society were those who had the higher levels of support for SMC. Consistent with the theory, holding post-materialist values associates more with the outcome variable than does holding materialist values. When we look at the control variable, Support for Conventional Citizenship, we notice the strong correlation ($\text{Gamma} = 0.40$): on average, the young people who participated in CivEd tended to support both methods of political participation.

Country-level variables also had effects on the dependent variable (Model 4). Adding these variables to the model explained 91% of the variance among countries. The welfare regime variables had an effect on support for SMC. Youth in social-democratic states and in Latin American states had more support for SMC than did youth in liberal states (reference group). In contrast, youth in conservative countries and in former socialist countries had less support for SMC than did youth in other countries. Parliamentary system and intensive process of democratization (changes between 1990 and 2000) also had positive and significant effects on support for SMC. Taking all the variables together, we can see that the level of youth support for SMC was higher in countries characterized by an open and dynamic political system. When we look at the control variable, nation size, we can see that the bigger nations had the higher levels of support for SMC.

Finally, I examined differential effects of individual characteristics across countries (i.e., across welfare regimes). In Southern European countries, the effect of materialist values (government role related to economy) on the dependent variable was stronger than in other countries; in Hong Kong SAR, the effect of materialist values was weaker. In conservative and Southern European countries, the effect of post-materialist values (government role related to society) on the dependent variable was stronger than in other countries. These findings suggest that support for SMC had a different meaning in the different countries. In the Southern European countries, youth support for SMC was associated with demand for more governmental intervention. In the conservative

Table 2: OLS Regression Coefficients on Attitudes toward Social-Movement Citizenship

Welfare regime	Country	Intercept	Female	Number of books at home	Parents' education	Expected education	Civic knowledge	Membership in organizations
Liberal	Australia	0.68 **	0.04 *	0.00	0.00	0.00	0.00 **	0.01 **
	United Kingdom	0.80 **	0.03	-0.01	0.00	-0.01	0.01 **	0.01
	United States	0.59 **	0.12 **	-0.02 **	0.00	0.01	0.00 **	0.01
Social Democracy	Belgium	0.67 **	-0.05 *	0.00	-0.01 *	0.03 *	0.01 **	0.01
	Denmark	1.11 **	0.07 **	0.00	0.00	0.00	0.00 **	0.02 **
	Norway	0.83 **	0.07 **	0.00	0.00	-0.01	0.01 **	0.01
	Sweden	0.73 **	0.09 **	-0.01	0.00 *	-0.01	0.00	0.01 *
Conservative	Finland	0.54 **	0.12 **	0.01	0.00	0.02 *	0.01 **	0.02 **
	Germany	0.95 **	0.08 **	0.00	-0.01 **	0.01	0.00	0.01
	Switzerland	0.89 **	0.08 **	-0.03 **	0.00	0.02 *	0.00	0.01 *
Southern Europe	Cyprus	0.83 **	0.03	0.01	-0.01 **	0.00	0.00	0.01
	Greece	0.89 **	0.07 **	-0.01 *	0.00	0.01	0.01 **	0.00
	Italy	0.65 **	0.04 *	0.00	-0.01 *	0.01 *	0.01 **	0.01
	Portugal	0.90 **	0.03	-0.01	-0.01	0.01 **	0.01 **	0.00
	Bulgaria	0.69 **	0.03	0.01	0.00	-0.01	0.01 **	0.00
Former Socialists	Czech Republic	0.84 **	0.04 **	-0.01	0.00	-0.02 **	0.01 **	0.02 **
	Estonia	1.26 **	-0.07 **	-0.01	0.00	0.00	0.00 *	-0.01
	Hungary	0.82 **	0.00	0.00	-0.01 *	0.00	0.00	0.01 **
	Latvia	1.03 **	-0.01	0.00	0.00	-0.02 **	0.00 **	0.00
	Lithuania	0.96 **	-0.01	-0.01	0.00	0.00	0.00	0.01
	Poland	0.30 **	0.00	0.00	0.00	-0.01	0.00	0.01 *
	Romania	0.98 **	0.02	-0.01	0.00	0.02 **	0.01 **	-0.01
	Russian Federation	1.22 **	-0.01	0.00	0.00	0.00	0.00 **	0.01
	Slovakia	0.95 **	-0.01	-0.01	-0.01 **	-0.01	0.00	0.01
	Slovenia	0.50 **	0.01	-0.01	0.00	0.00	0.01 **	0.01 *
Latin America	Chile	0.77 **	0.04 **	-0.03 **	0.00	0.00	0.00 **	0.01 **
	Colombia	1.75 **	0.04 **	-0.01	0.00	0.00	0.01 **	0.01 *
	Hong Kong SAR	0.45 **	0.03 *	-0.01	0.00	0.01	0.00 **	0.01 **

Table 2 (contd.): OLS Regression Coefficients on Attitudes toward Social-Movement Citizenship

Welfare regime	Country	Discuss national politics	Discuss international politics	Attention to news	Government responsibility		Conventional citizenship	N	R ²
					Economy	Society			
Liberal	Australia	0.00	-0.01	0.03	0.10 **	0.14 **	0.46 **	2,881	0.24
	United Kingdom	0.03	-0.02	0.00	0.10 **	0.17 **	0.41 **	2,614	0.21
	United States	0.01	0.01	0.05 **	0.11 **	0.17 **	0.47 **	2,454	0.35
Social Democracy	Belgium	0.03	-0.01	0.00	0.11 **	0.18 **	0.38 **	1,728	0.23
	Denmark	0.01	0.00	0.00	0.15 **	0.13 **	0.31 **	2,874	0.17
	Norway	0.01	0.02	0.02	0.15 **	0.19 **	0.34 **	2,998	0.27
	Sweden	-0.02	0.03	0.07 **	0.08 **	0.18 **	0.46 **	2,710	0.27
Conservative	Finland	0.03	-0.01	-0.02	0.06 *	0.21 **	0.43 **	2,597	0.24
	Germany	0.01	-0.02	0.01	0.08 **	0.28 **	0.34 **	3,254	0.22
	Switzerland	0.03	0.01	0.01	0.09 **	0.19 **	0.37 **	2,813	0.21
Southern Europe	Cyprus	0.04 *	-0.05 **	0.02	0.16 **	0.20 **	0.42 **	2,864	0.27
	Greece	0.01	0.00	0.01	0.16 **	0.25 **	0.30 **	3,138	0.31
	Italy	0.03	-0.01	0.00	0.16 **	0.24 **	0.34 **	3,673	0.25
	Portugal	0.03	0.02	-0.02	0.12 **	0.31 **	0.24 **	2,973	0.22
Former Socialists	Bulgaria	0.03	-0.02	0.03	0.06 *	0.16 **	0.51 **	2,278	0.37
	Czech Republic	-0.03	0.02	0.01	0.08 **	0.24 **	0.39 **	3,441	0.21
	Estonia	-0.01	0.06 **	-0.03	0.10 **	0.12 **	0.41 **	3,101	0.18
	Hungary	0.02	0.00	0.03 *	0.11 **	0.18 **	0.42 **	3,011	0.26
	Latvia	-0.03	0.03	0.01	0.09 **	0.19 **	0.43 **	3,107	0.27
	Lithuania	0.02	0.05 *	-0.03	0.09 **	0.16 **	0.38 **	2,290	0.20
	Poland	0.01	0.01	0.00	0.20 **	0.22 **	0.44 **	3,135	0.32
	Romania	0.00	0.00	0.03 *	0.12 **	0.09 **	0.40 **	2,737	0.28
	Russian Federation	-0.03	0.01	0.05 **	0.08 **	0.18 **	0.40 **	2,041	0.19
	Slovakia	-0.02	0.01	0.04 **	0.10 **	0.24 **	0.41 **	3,311	0.27
	Slovenia	-0.01	0.00	0.02	0.12 **	0.24 **	0.42 **	2,864	0.28
Latin America	Chile	0.01	-0.01	0.00	0.17 **	0.22 **	0.36 **	5,226	0.23
	Colombia	-0.02	0.03	0.02	0.10 **	0.12 **	0.26 **	4,128	0.17
	Hong Kong SAR	-0.01	0.00	0.02 *	0.02	0.15 **	0.64 **	4,435	0.42

Note: * $p < 0.05$; ** $p < 0.01$.

Table 3: OLS Regression Coefficients on Expected Participation in Social-Movement Activities

Welfare regime	Country	Intercept	Female	Number of books at home	Parents' education	Expected education	Civic knowledge	Membership in organizations
Liberal	Australia	0.89 **	0.18 **	0.00	0.00	-0.01	-0.01 **	0.04 **
	United Kingdom	0.77 **	0.21 **	-0.01	0.00 *	-0.01	0.00	0.03 **
	United States	0.63 **	0.17 **	-0.02 *	-0.01 **	0.02 *	0.00 **	0.03 **
Social Democracy	Belgium	0.55 **	0.21 **	-0.01	0.00	-0.02	0.00	0.04 **
	Denmark	0.67 **	0.21 **	0.00	0.00 *	0.00	0.00 **	0.05 **
	Norway	0.70 **	0.16 **	0.01	0.00	-0.01	-0.01 **	0.03 **
	Sweden	0.60 **	0.14 **	-0.02 **	0.00 *	0.00	-0.01 **	0.03 **
Conservative	Finland	0.87 **	0.20 **	-0.01	0.00	-0.03 **	0.00	0.04 **
	Germany	1.11 **	0.21 **	-0.01	0.00 *	0.01	-0.01 **	0.03 **
	Switzerland	1.16 **	0.23 **	-0.02 **	-0.01 **	0.00	-0.01 **	0.03 **
Southern Europe	Cyprus	1.19 **	0.13 **	-0.01	-0.01	0.00	0.00 **	0.02 **
	Greece	0.95 **	0.09 **	-0.01	-0.01 *	0.00	0.00 **	0.02 **
	Italy	0.83 **	0.20 **	0.00	-0.01 *	0.00	0.00	0.02 **
	Portugal	0.69 **	0.07 **	0.00	-0.01 *	0.01 *	0.00 *	0.02 **
Former Socialists	Bulgaria	1.25 **	0.06 **	-0.02 *	-0.01 **	-0.01	0.00 *	0.00
	Czech Republic	0.83 **	0.18 **	0.01	-0.01 **	-0.03 **	0.00	0.04 **
	Estonia	1.11 **	0.07 **	-0.01	0.00	0.00	-0.01 **	0.02 **
	Hungary	0.90 **	0.12 **	-0.01	0.00	-0.03 **	-0.01 **	0.03 **
	Latvia	0.94 **	0.09 **	-0.01	0.00	-0.01	-0.01 **	0.02 **
	Lithuania	1.23 **	0.08 **	0.01	-0.01 **	-0.01	-0.01 **	0.02 **
	Poland	0.74 **	0.11 **	-0.01	0.00	-0.02 *	-0.01 **	0.02 **
	Romania	1.11 **	0.00	-0.01	0.00	-0.02 *	0.00	0.02 **
	Russian Federation	0.92 **	0.00	-0.02 *	0.00	-0.02 *	-0.01 **	0.03 **
	Slovakia	0.69 **	0.12 **	0.00	-0.01 **	-0.02 *	-0.01 **	0.03 **
	Slovenia	1.35 **	0.06 **	-0.01	0.00	-0.01	0.00 **	0.03 **
Latin America	Chile	1.35 **	0.11 **	-0.02 **	0.00 *	0.00	-0.01 **	0.02 **
	Colombia	0.98 **	0.12 **	-0.02 **	0.00	0.00	0.00 **	0.01 *
	Hong Kong SAR	0.71 **	0.10 **	-0.01 *	-0.01 **	0.00	0.00	0.02 **

Table 3 (contd.): OLS Regression Coefficients on Expected Participation in Social-Movement Activities

Welfare regime	Country	Discuss national politics	Discuss international politics	Attention to news	Government responsibility		Conventional citizenship	N	R ²
					Economy	Society			
Liberal	Australia	0.02	0.02	0.08 **	0.03	0.08 **	0.38 **	2,881	0.28
	United Kingdom	0.06 **	0.06 **	0.11 **	0.09 **	-0.01	0.33 **	2,614	0.30
	United States	0.07 **	0.02	0.11 **	0.06 **	0.11 **	0.36 **	2,454	0.34
Social Democracy	Belgium	0.03	0.01	0.18 **	0.10 **	0.14 **	0.26 **	1,728	0.24
	Denmark	0.06 **	0.04 *	0.10 **	0.08 **	0.05 *	0.33 **	2,874	0.26
	Norway	0.04 *	0.06 **	0.08 **	0.04 *	0.07 **	0.35 **	2,998	0.25
	Sweden	0.03	0.09 **	0.08 **	0.04	0.11 **	0.37 **	2,710	0.26
Conservative	Finland	0.04	0.01	0.05 **	0.01	0.06 **	0.37 **	2,597	0.25
	Germany	0.04	0.08 **	0.10 **	0.01	0.12 **	0.21 **	3,254	0.19
	Switzerland	0.01	0.10 **	0.08 **	0.08 **	0.09 **	0.21 **	2,813	0.18
Southern Europe	Cyprus	0.04 *	0.04 *	0.10 **	0.11 **	0.11 **	0.23 **	2,864	0.20
	Greece	0.04 *	0.05 **	0.09 **	0.07 **	0.20 **	0.24 **	3,138	0.19
	Italy	0.06 **	0.05 **	0.10 **	0.07 **	0.12 **	0.27 **	3,673	0.22
	Portugal	0.02	0.06 **	0.12 **	0.12 **	0.17 **	0.21 **	2,973	0.20
Former Socialists	Bulgaria	0.05 *	0.05 *	0.08 **	0.07 **	0.01	0.31 **	2,278	0.19
	Czech Republic	0.00	0.05 **	0.06 **	0.05 **	0.05 *	0.30 **	3,441	0.22
	Estonia	0.09 **	0.08 **	0.03	0.06 **	0.04	0.26 **	3,101	0.16
	Hungary	0.00	0.08 **	0.08 **	0.06 **	0.08 **	0.39 **	3,011	0.26
	Latvia	0.05 **	0.06 **	0.06 **	0.03	0.06 **	0.41 **	3,107	0.23
	Lithuania	0.09 **	-0.01	0.05 *	0.05 *	0.04	0.33 **	2,290	0.20
	Poland	0.06 **	0.01	0.09 **	0.11 **	0.06 *	0.40 **	3,135	0.24
	Romania	0.05 **	0.03	0.10 **	0.03	0.07 **	0.31 **	2,737	0.18
	Russian Federation	0.05 *	0.02	0.10 **	0.02	0.08 **	0.42 **	2,041	0.30
	Slovakia	0.05 **	0.01	0.07 **	0.03	0.10 **	0.43 **	3,311	0.21
	Slovenia	0.02	0.04 **	0.05 **	0.07 **	0.06 **	0.24 **	2,864	0.14
Latin America	Chile	0.04 **	0.02	0.10 **	0.08 **	0.14 **	0.23 **	5,226	0.17
	Colombia	0.05 **	0.02	0.11 **	0.08 **	0.10 **	0.34 **	4,128	0.25
	Hong Kong SAR	0.06 **	0.01	0.11 **	0.02	0.10 **	0.44 **	4,435	0.34

Note: * $p < 0.05$; ** $p < 0.01$.

Table 4: Hierarchical Linear Models for Support for Social-Movement Citizenship

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	3.09 **	3.09 **	3.09 **	1.27 **
Level 1 Variables				
Gender	—	0.04 **	0.03 **	0.03 **
Books (x10)	—	-0.05 *	-0.06 **	-0.06 **
Expected Education (x10)	—	0.01 **	0.01	0.01
Parental Education (x10)	—	-0.01 +	-0.02 **	-0.02 **
Civic Knowledge (x10)	—	0.06 **	0.04 **	0.04 **
Membership in Organizations (x10)	—	0.01 **	0.07 **	0.07 **
Discuss International Politics (x10)	—	—	0.09 *	0.09 *
Attention to News	—	—	0.01 **	0.01 **
Governmental Role Economy	—	—	0.11 **	0.10 **
Social Democracy	—	—	—	0.03
Conservative	—	—	—	-0.04 +
Former Socialist	—	—	—	0.01
Latin America	—	—	—	0.03
Southern Europe	—	—	—	0.04 *
Hong Kong SAR	—	—	—	-0.08 **
Governmental Role Society	—	—	0.19 **	0.16 **
Social Democracy	—	—	—	0.00
Conservative	—	—	-0.06 *	—
Former Socialist	—	—	—	0.02
Latin America	—	—	—	-0.01
Southern Europe	—	—	—	0.08 **
Hong Kong SAR	—	—	—	-0.00
Conventional Citizenship	—	—	0.40 **	0.43 **
Social Democracy	—	—	—	-0.07 *
Conservative	—	—	—	-0.07 *
Former Socialist	—	—	—	-0.01
Latin America	—	—	—	-0.10 **
Southern Europe	—	—	—	-0.10 **
Hong Kong SAR	—	—	—	0.21 **
Level 2 Variables				
Nation Size (x100)	—	—	—	0.04 *
Welfare Regime (reference group: Liberal)	—	—	—	—
Social Democracy	—	—	—	0.12 *
Conservative	—	—	—	-0.16 *
Former Socialist	—	—	—	-0.15 *
Latin America	—	—	—	0.12 *
Southern Europe	—	—	—	-0.00
Hong Kong SAR	—	—	—	-0.00
Changes in Freedom 1990–2000	—	—	—	0.10 **
Governing System	—	—	—	0.08 *
Variance Component				
Intercept	0.034 **	0.034 **	0.034 **	0.003 **
Level 1	0.273 **	0.266 **	0.202 **	0.202 **
Change in Intercept VC a	—	—	—	91%
Change in Level 1 VC a	—	3%	24%	—

Notes: Level 2 also includes compensation for group centering of Level 1 variables (see text).

+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

countries, youth support for SMC was associated with demand for more governmental intervention in terms of well-being and other post-materialist values.

When we refer to the control variable, support for conventional citizenship, we notice the effect that welfare regime variables had on the correlation of the control variable and the outcome variable. In all of the countries, except for former socialist, the correlation between support for conventional citizenship and support for SMC became weaker. In Hong Kong SAR, however, the correlation became stronger. These findings suggest that in the liberal countries as well as in the former socialist countries and Hong Kong SAR, youth support both methods of participation.

Table 5 presents the results of hierarchical linear models predicting Expected Participation in Social-Movement Activities (SMA). The fully unconditional model, Model 1, indicated that differences between countries accounted for 13% of the variance in the variable. These figures were appropriate for employing hierarchical level modeling (Raudenbush & Bryk, 2000). As we can see, the findings in Table 5 are very similar to the findings discussed in relation to Table 4. Due to this similarity, I discuss here only the differences between the two analyses. First, the gender gap was much more salient ($\text{Gamma} = 0.13$ vs. $\text{Gamma} = 0.03$). Females in the CivEd sample not only supported social-movement citizenship more than men did, but they also expected to participate more in social-movement activities. Second, general motivation (expected years of education) and score on the civic knowledge test had a negative and significant effect on the outcome. These findings align with the pattern pointed out earlier wherein social movements are supported by disadvantaged youth and not necessarily by youth from affluent backgrounds.

Several differences were also evident in terms of the country-level variables. First, youth in Latin American states as well as in Southern European countries and in Hong Kong SAR were more likely to participate in CMA than were youth in liberal countries. Second, the variable “women representation in parliament” negatively and significantly correlated with CMA. In countries (generally social democratic) where women were highly represented in conventional politics, youth were more likely to participate in CMA.

Discussion

In this study, I have explored, through a comparative approach, the individual and contextual determinants of youth support for social movements. Previous research on youth engagement in social movements is scarce in the fields of political socialization and social movements. This empirical gap is counter-intuitive because social movements are one of the few available methods for gauging the political participation of youth under voting age. More surprisingly, this empirical gap coexists with the accumulating evidence that the adherents and activists of social movements come mainly from the younger segment of the population (McAdam et al., 1988; Mertig & Dunlop, 2001).

The first objective of this study was to evaluate the extent to which individual characteristics influence youth support for social movements. The findings showed that females had stronger support than males for social-movement citizenship and had a higher likelihood of participating in social-movement activities. In addition, support for social movements was present among youth from disadvantaged backgrounds. The negative relationship between social class and support for social movements contradicts previous research about social movements. One possible explanation is that the young people in the 28 countries sampled in the CivEd did not associate social movements with social class interests or identity. The findings also showed that those young people who discussed international politics, paid attention to news, and held post-materialist values were those who had stronger support for social movements. However, the variable “discussing national politics” had no significant effect on the outcome variable, a finding which suggests that the concept “social movement” relates more to global than local matters. In other words, this finding might reflect the “Think Global, Act Local” slogan used intensively by the environmental movement since the 1970s.

The second objective was to estimate the impact of country-level attributes on youth support for social movements. The findings did not support the national economic development approach proposed in previous research. Rather, they showed significant differences among types of welfare regimes and among different political systems. The social democracies and Latin American countries participating in CivEd produced the highest level of support for social-

Table 5: Hierarchical Linear Models for Expected Participation in Social-Movement Activities

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	2.53 **	2.53 **	2.53 **	1.75 **
Level 1 Variables				
Gender	–	0.13 **	0.13 **	0.13 **
Books (x10)	–	-0.00	-0.09 **	-0.09 **
Expected Education (x10)	–	0.01 **	-0.07	-0.06 **
Parental Education (x10)	–	-0.00	-0.03 **	-0.03 **
Civic Knowledge (x10)	–	0.00	-0.05 **	-0.05 **
Membership in Organizations (x10)	–	0.04 **	0.03 **	0.03 **
Discuss International Politics	–	–	0.07 **	0.07 **
Attention to News	–	–	0.09 **	0.09 **
Governmental Role Economy	–	–	0.05 **	0.04 **
Social Democracy	–	–	–	0.01
Conservative	–	–	–	-0.03
Former Socialist	–	–	–	0.03
Latin America	–	–	–	0.03
Southern Europe	–	–	–	0.04 *
Hong Kong SAR	–	–	–	-0.04
Governmental Role Society	–	–	0.09 **	0.06 **
Social Democracy	–	–	–	0.02
Conservative	–	–	–	0.04
Former Socialist	–	–	–	-0.01
Latin America	–	–	–	0.05 *
Southern Europe	–	–	–	0.08 **
Hong Kong (SAR)	–	–	–	0.04
Expected Conventional Politics	–	–	0.32 **	0.39 **
Social Democracy	–	–	–	-0.03
Conservative	–	–	–	-0.15 **
Former Socialist	–	–	–	-0.07 *
Latin America	–	–	–	-0.12 **
Southern Europe	–	–	–	-0.10 *
Hong Kong SAR	–	–	–	0.10 *
Variable				
Level-2 Variables				
Nation Size (x100)	–	–	–	–
Welfare Regime (reference group: Liberal)	–	–	–	–
Social Democracy	–	–	–	0.11 +
Conservative	–	–	–	-0.03
Former Socialist	–	–	–	-0.12 **
Latin America	–	–	–	0.19 **
Southern Europe	–	–	–	0.12 *
Hong Kong SAR	–	–	–	0.22 **
Women in Parliament (x10)	–	–	–	-0.08 **
Variance Component				
Intercept	0.043 **	0.043 **	0.043 **	0.005 **
Level 1	0.297 **	0.280 **	0.228 **	0.228 **
Change in Intercept VC a	–	–	–	88%
Change in Level 1 VC a	–	6%	19%	–

Notes: Level 2 also includes compensation for group centering of Level 1 variables (see text).

+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

movement citizenship and the highest level of expected participation in social-movement activities. One possible explanation is that the young people in these countries supported social movements for different reasons (Janoski, 1998). Furthermore, in conservative countries and in Southern European countries, youth holding post-materialist values stood out in their support of social movements.

Finally, the findings drawn from controlling the variables regarding conventional politics indicated that

the young people participating in CivEd tended to support both methods of political participation. This tendency was more prominent in liberal countries, in former socialist countries, and in Hong Kong SAR. It seems that the young people in these countries had created a “bricolage” of political participation methods and were innovatively recombining different elements in a manner constituting a new way of configuring political participation (Campbell, 2005).

References

- Barkan, S. (2004). Explaining public support for the environmental movement: A civic voluntarism model. *Social Science Quarterly*, 85(4), 913–937.
- Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzales, E. J., Smith, T. A., & Kelly, D. L. (1996). *Science achievement in the middle school years: IEA's TIMSS*. Chestnut Hill, MA: Boston College.
- Berger, P. L., & Luckmann, T. (1966). *The social construction of reality*. Garden City, NY: Anchor Books.
- Bourdieu, P. (1984). *Distinction: A social critique of the judgment of taste*. Cambridge, MA: Harvard University Press.
- Campbell, J. L. (2005). Where do we stand? Common mechanisms in organizations and social movements research. In G. F. Davis, D. McAdam, W. R. Scott, & M. N. Zald (Eds.), *Social movements and organization theory*. Cambridge: Cambridge University Press.
- Cornwell, B., & Harrison, J. A. (2004). Union members and voluntary associations: Membership overlap as a case of organizational embeddedness. *American Sociological Review*, 69(6), 862–881.
- Curtis, J., Baer, D. E., & Grabb, E. G. (2001). Nations of joiners: Explaining voluntary association membership in democratic societies. *American Sociological Review*, 66(6), 783–805.
- Curtis J., Grabb, E. G., & Baer, D. E. (1992). Voluntary association membership in fifteen countries: A comparative analysis. *American Sociological Review*, 57(2), 139–152.
- Dalton, R., & Kuechler, M. (Eds.). (1990). *Challenging the political order: New social and political movements in western democracies*. New York: Oxford.
- Deacon, B. (2000). Eastern European welfare states: The impact of the politics of globalization. *Journal of European Social Policy*, 10(2), 146–161.
- Diani, M. (2004). Networks and participation. In D. Snow, S. Soule, & H. Kriesi (Eds.), *The Blackwell companion to social movements* (339–359). Malden, MA: Blackwell Publishing.
- Edwards, M., & Gaventa, J. (Eds.). (2001). *Global citizen action: Perspective and challenges*. Boulder, CO: Lynne Rienner.
- Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Cambridge: Polity Press.
- Esping-Andersen, G. (1996). *Welfare states in transition*. Thousand Oaks, CA: Sage Publications.
- Esping-Andersen, G. (1999). *Social foundations of postindustrial economies*. Oxford: Oxford University Press.
- Falk, R. (1994). The making of global citizenship. In B. van Steenbergen (Ed.), *The condition of citizenship* (pp. 1–9). Thousand Oaks, CA: Sage Publications.
- Fraile, M., & Ferrer, M. (2005). Explaining the determinants of public support for cuts in unemployment benefits spending across OECD countries. *International Sociology*, 20(4), 459–481.
- Frank D. J. (1997). Science, nature, and the globalization of the environment, 1870–1990. *Social Forces*, 76(2), 4409–4435.
- Frank, D. J., Hironaka, A., & Schofer, E. (2000). The nation-state and the natural environment over the twentieth century. *American Sociological Review*, 65(1), 96–116.
- Freedom House. (2006a). *Report on freedom in the world*. Available online at <http://www.freedomhouse.org/template.cfm?page=15>
- Freedom House. (2006b). *Report on freedom of the press*. Available online at <http://www.freedomhouse.org/template.cfm?page=16>
- Gough, I. (2001). Globalization and regional welfare regimes: The East Asian case. *Global Social Policy*, 1(2), 163–189.
- Gray, M., & Caul, M. (2000). Declining voter turnout in advanced industrial democracies, 1950 to 1997: The effects of declining group mobilization. *Comparative Political Studies*, 33(9), 1091–1122.
- Guerin, D., Crete, J., & Mercier, J. (2001). A multilevel analysis of the determinants of recycling behavior in the European countries. *Social Science Research*, 30(2), 195–218.

- Healy, K. (2000). Embedded altruism: Blood collection regimes and the European Union's donor population. *American Journal of Sociology*, 105(6), 1633–1657.
- Ichilov, O. (Ed.). (1990). *Political socialization, citizenship education, and democracy*. New York: Teachers College Press.
- Inglehart, R. (1990). *Culture shift in advanced industrial society*. Princeton, NJ: Princeton University Press.
- Inglehart, R. (1997). *Modernization and postmodernization: Cultural, economic, and political change in 43 societies*. Princeton, NJ: Princeton University Press.
- Inglehart, R., & Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review*, 65(1), 19–51.
- International Institute for Democracy and Electoral Assistance (IDEA). (2005). *IDEA handbook for electoral system design*. Stockholm: Author.
- Inter-Parliamentary Union. (2005). *Report on percentage of women in national parliaments*. Available online at <http://www.ipu.org/wmn-e/world.htm>
- Janoski, T. (1998). *Citizenship and civil society: A framework of rights and obligations*. Cambridge: Cambridge University Press.
- Katzenstein, P. J. (2003). Small states and small states revisited. *New Political Economy*, 8(1), 9–31.
- Kelley, J., Evans, M., & Sikora, J. (forthcoming). *Scholarly culture and academic performance in 43 nations*. Available online at: <http://paa2006.princeton.edu/download.aspx?submissionId=61840>
- Klandermans, B. (2004). The demand and supply of participation: Social psychological correlates of participation in social movements. In D. Snow, S. Soule, & H. Kriesi (Eds.), *The Blackwell companion to social movements* (pp. 360–379). Malden, MA: Blackwell Publishing.
- Kriesi, H. (1989). New social movements and the new class in the Netherlands. *American Journal of Sociology*, 94(5), 1078–1116.
- Kriesi, H. (1995). *New social movements in Western Europe: A comparative analysis*. Minneapolis, MN: University of Minnesota Press.
- Leibfried, S. (1992). Towards a European welfare state? On integrating poverty regimes into the European community. In Z. Ferge & J. Kolberg (Eds.), *Social policy in a changing Europe*. Frankfurt am Main: Campus Verlag.
- Lipset, S. M. (1960). *Political man: The social bases of politics*. Baltimore, MD: Johns Hopkins University Press.
- Lipset, S. M. (1994). The social requisites of democracy revisited. *American Sociological Review*, 59(1), 1–20.
- Maheu, L. (Ed.). (1995). *Social movements and social classes: The future of collective action*. London: Sage Publications.
- McAdam, D. (1986). Recruitment to high-risk activism: The case of freedom summer. *American Journal of Sociology*, 92, 64–90.
- McAdam, D. (1988). Micro-mobilization contexts and recruitment to activism. In B. Klandermans, H. Kriesi, & S. Tarrow (Eds.), *From structure to action: Social movement participation across cultures* (Vol. 1) (pp. 125–154). Greenwich, London: JAI Press Inc.
- McAdam, D., McCarthy, J., & Zald, M. (1988). Social movements. In N. Smelser (Ed.), *Handbook of sociology* (pp. 695–738). Beverly Hills, CA: Sage Publications.
- McFarland, D. A., & Thomas, R. J. (2006). Bowling young: How youth voluntary associations influence adult political participation. *American Sociological Review*, 71(3), 401–425.
- Melucci, A. (1996). *Challenging codes: Collective action in the Information Age*. Cambridge: Cambridge University Press.
- Mertig, A. G., & Dunlop, R. E. (2001). Environmentalism, new social movements, and the new class: A cross-national investigation. *Rural Sociology*, 66(1), 113–136.
- Meyer, D. S., & Minkoff, D. (2004). Conceptualizing political opportunity. *Social Forces*, 82(4), 1475–1492.
- Meyer, J. W., Boli, J., Thomas, G. M., & Ramirez, F. O. (1997). World society and the nation-state. *American Journal of Sociology*, 103(1), 144–181.
- Noddings, N. (2005). Global citizenship: Promises and problems. In N. Noddings (Ed.), *Educating citizens for global awareness* (pp. 1–21). New York: Teachers College Press.
- Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.
- Raudenbush, S., & Bryk, A. (2000). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.
- Rootes, C. (2004). Environmental movements. In D. Snow, S. Soule, & H. Kriesi (Eds.), *The Blackwell companion to social movements* (pp. 608–640). Malden, MA: Blackwell Publishing.
- Salamon, L. M., & Anheier, H. K. (1998). Social origins of civil society: Explaining the nonprofit sector cross nationality. *Voluntas: International Journal of Voluntary and Nonprofit*, 9(3), 213–248.
- Scheepers, P., & Grotenhuis, M.T. (2005). Who cares for the poor in Europe? Micro and macro determinants for alleviating poverty in 15 European countries. *European Sociological Review*, 21(5), 453–465.
- Schofer, E., & Fourcade-Gourinchas, M. (2001). The structural contexts of civic engagement: Voluntary association membership in comparative perspective. *American Sociological Review*, 66(6), 806–828.

- Smith, J. (1995). Transnational political processes and the human rights movement. In L. Kriesberg, M. Dobkowski, & I. Walliman (Eds.), *Research in social movements, conflict and change* (pp. 185–219). Greenwood CT: JAI.
- Snijders, A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage Publications.
- Snow, D., Soule, S., & Kriesi, H. (2004). Mapping the terrain. In D. Snow, S. Soule, & H. Kriesi (Eds.), *The Blackwell companion to social movements* (pp. 3–16). Malden, MA: Blackwell Publishing.
- Strandbu, A., & Krange, O. (2003). Youth and the environmental movement: Symbolic inclusions and exclusions. *The Sociological Review*, 51(2), 177–197.
- Swidler, A. (1986). Culture in action: Symbols and strategies. *American Sociological Review*, 51(2), 273–286.
- Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- United Nations Development Program. (2000). *Human development report*. Available online at <http://hdr.undp.org/>
- Verba, S., Schlozman, K. L., & Brady, H. E. (1995). *Voice and equality: Civic voluntarism in American politics*. Cambridge, MA: Harvard University Press.

Online data collection in SITES 2006: Design and implementation

Ralph Carstens and Falk Brese

IEA Data Processing Center

Hamburg, Germany

Barbara N. Brecko

Educational Research Institute (ERI)

Ljubljana, Slovenia

Abstract

The thematic background of the Second Information Technology in Education Study (SITES) 2006 made it a very good candidate to explore the feasibility of collecting data over the internet. SITES 2006 is the first study in the history of international comparative educational assessments to apply such methodology, with the school and teacher questionnaires offered online as an international option. With the additional prospect of lowering response burden and costs, while at the same time allowing for more flexibility and efficiency, the Online Data Collection (ODC) component in SITES was gradually launched in a series of carefully monitored steps in close cooperation with participating systems and experts from late 2004 onwards. This paper describes the key design requirements and implementation principles.

These, to a large extent, mirror the routines and procedures successfully applied in earlier and current IEA studies, but at the same time take into account existing findings and approaches in the academic and commercial world with respect to online research. Consequently, several methodological issues, such as reliability, validity, and comparability with paper-and-pencil questionnaires, had to be addressed, including but not limited to languages and scripts, translation, national adaptation, quality control, technical limitations, equipment of respondents, security and access, usability, and (eventually) data processing to allow for seamless integration with the general IEA study framework and the highest possible participation of respondents.

Background and rationale

During planning for the third module (M3) of the Second Information Technology in Education Study (SITES), the study's research consortium proposed (see Anderson & Plomp, 2001) that computer and internet technology could be employed to assess ICT-related¹ indicators and measures via two methodologically innovative components: (i) a web-based performance assessment administered to a sub-sample of students in addition to a "core" paper-and-pencil test; and (ii) web-based background questionnaires. Although the student assessment component was eventually dropped from the survey (a result of operational concerns and obstacles), the then appointed SITES study consortium, consisting of the University of Twente, the University of Hong Kong, and the IEA Data Processing Center (DPC), confirmed their willingness to explore the feasibility of online data

collection (ODC) in their proposal to the IEA General Assembly. The consortium stressed that the thematic background of SITES 2006, with its questions of to what extent and how ICT is used in education and how it supports and enhances pedagogical practices, made the study a very good candidate to employ survey technology similar to the technology and tools used in pedagogical contexts under focus in SITES. The relative affinity of the surveyed Grade 8 mathematics and science teachers to using ICT in and outside of the classroom allowed the members of the International Coordinating Consortium (ICC) to be even more confident about the general feasibility of ODC, as they could anticipate a reasonable level of acceptance of ODC amongst these teachers. Thus, SITES became the first IEA study—and the first in the history of international comparative educational

¹ ICT stands for information and communication technologies.

assessments—to make ODC an integral framework component and to apply this methodology as an international option.

The ICC also expected that, in addition to fostering innovative methodology, the online mode would offer certain operational benefits and added value, most importantly, the ability to provide an alternative and potentially more efficient response channel for respondents, which they might ideally perceive as more convenient, interesting, or simply motivating. The consortium also assumed that the online mode would lower the overall response burden and significantly reduce, or even eliminate, costs and workload for the national study centers in the areas of printing, preparation, shipping, collection, and data entry. The consortium further assumed that in relation to the international data processing, cleaning, and analysis, the online mode would yield more accurate and reliable data that, in turn, would be more rapidly available for reporting.

With these considerations in mind, the consortium saw ODC not only as a means to conduct and support research, but as an object of research itself. This paper accordingly addresses the question: what are the characteristics of a feasible ODC system as used in an international comparative survey? The paper also overviews the conceptual considerations, the procedural integration (e.g., the sharing of work between international and national centers), and the resulting staged technical implementation. A second important concern discussed in this paper is whether the different data-collection modes introduce effects and biases into the data that could influence (adversely or otherwise) the making of meaningful statistical analyses.

Conceptualization and integration with survey operations

Since the emergence of the world wide web, the number of people using the internet has constantly grown (see, for example, Bandilla, 2002). Similarly, ODC is being increasingly used for both commercial and scientific purposes. This is also true for educational research. However, most educational surveys using ODC have been conducted at national level only. Several of the countries participating in SITES

2006 had experienced the use of ODC within their countries, but until the time of this study, large-scale educational surveys at the international level were based entirely on paper questionnaires. The SITES research consortium thus had to carefully consider the usability and comparability of ODC procedures and means applied simultaneously in multiple countries.

The most important requirements when using ODC within an IEA study are that IEA standards and guidelines are followed and that survey operation procedures are similar to the tried-and-true procedures used in previous IEA studies, whenever possible.² One key standard regarding survey procedures is to minimize the response burden and non-response (Martin, Rust, & Adams, 1999, p. 57f.). Even at the development stage, the study design should aim at clarity, ease of use, and simplicity. SITES tried to address these considerations through several arrangements discussed in this chapter.

General approach

To meet the goal of simplicity and to ensure comparability of data, the paper versions of the three questionnaire types had to be finalized first. From these final paper versions, the questionnaires were converted for the online mode. Respondents using the online mode could only fill in the questionnaire via the internet. No other options were allowed, such as sending PDF documents via email or printing out the online questionnaires and mailing them to the national center. However, as was the case with the paper mode, respondents did not have to answer the whole questionnaire in one session. From the time they received login information to the end of the data-collection period, respondents could log in and out as many times as needed and resume answering the questionnaire at the last question they had responded to before logging out. All answers were automatically saved whenever respondents moved to another question. Respondents could also change any given answer at any time while logged in.

To lower the burden for respondents during the SITES main study, the decision on whether to assign the online or the paper questionnaires was taken at the national center. In most of the countries, the default mode was set at the school level. Every respondent of

² IEA provides standards and guidelines for conducting its studies (cf. Martin et al., 1999).

a school—the principal, the technical coordinator, and the sampled teachers—were assigned to the same mode. In other countries, the default mode was alternatively set at the questionnaire level, so that all respondents of the same group (principals, technical coordinators, or teachers) received the questionnaire in the same mode. The assignment of the default mode within each participating country took into account what mode a specific school preferred. Furthermore, each country had to assure that every respondent assigned to the online mode by default had the option of answering the questionnaire on paper regardless of the reasons for not being willing or able to answer online.

In addition to procedural considerations, the design had to address technical issues regarding response burden. To ensure privacy, every respondent received individual login information. The only requirements for the respondent were an internet connection and a standard internet browser. No additional software was required to fill in the questionnaire (see the section on technical implementation below). The online system was designed to reduce the traffic load as much as possible so that the connection time could be spent on reading through and answering the questions rather than waiting for the website to load. This facility was also tested to work for low-speed connections (e.g., 33/56K modems). Because the focus of SITES 2006 was on pedagogical usage of ICT and not on computer literacy, terminology and technical hurdles were carefully considered and implemented in a way that reduced, to the very minimum, the computer skills a respondent needed to have to access and answer the questions.

Challenges using mixed mode

SITES was conducted using a true mixed-mode design. Data from different data-collection modes were planned to be merged to a single data set. Consequently, potential measurement error originating from the use of the two different modes had to be controlled for and reduced as much as possible to ensure comparability. In line with standards and practices from other surveys, comparable and similar conditions across modes and countries were desired because respondents resort to context to make sense of questions. To address this consideration, the SITES survey design established several general similarities

between the paper and the online questionnaires with regard to potential measurement error. Questionnaires in both modes were self-administered and equally situated in the visual domain, in contrast to mixed-mode surveys that simultaneously employ self-administered questionnaires and telephone or face-to-face interviewers. Moreover, data from both modes were collected over the same time period.

The navigational paradigm for the online questionnaire was designed to be similar to that of the paper questionnaires. Respondents could use “next” and “previous” buttons to navigate to an adjacent page, similar to flipping physical pages. In addition, a “table of contents” allowed respondents to go directly to a selected page of a paper questionnaire, instead of having to click the “next” or “previous” buttons multiple times. Both features allowed respondents to skip or omit questions just as they could in a self-administered paper questionnaire.

However, there were differences in terms of representation between the two modes. While a page in the paper questionnaire could contain several questions, assuming they all fitted together on one page, there was only one question per page in the online mode because of the length and complexity of a large number of questions in SITES and because of expectation of a sequencing (primacy/recency) effect. Furthermore, if a question exceeded a certain length, only parts of it could be seen on screen, which meant respondents had to scroll up and down to reach other parts of the question. In the paper mode, every part of a page was visible at once, with the exception of questions spanning multiple pages.

Another difference in terms of representation is the overall impression of the length of the questionnaire at first glance. This impression can be visual or sensory (if weight is taken into account), and is gained from simply looking at the thickness of the paper questionnaire or skimming the document. In the SITES online questionnaires, the length could be estimated by looking at the “table of contents,” which listed each individual question. But this estimate required an assumption about the average length of questions. Having answered half of the questions, respondents could not assume that the time needed to answer the second half would be the same. The same is true for the implemented question counter displaying the current question versus the total number of

questions (e.g., 6/22). This is an unresolved problem in mixed-mode designs in general. Page or question counters, tables of contents, and progress indicators that indicate the percentage of completed questions divide a questionnaire on the basis of the number of questions, not on the basis of time or response burden. Filter questions constitute an additional complication in this sense. A possible solution is to base progress indicators or the table of contents on the relative measure of estimated time (a sort of “weight”) for each individual question.

To further ensure the similarity of the two sets of instrumentation, responses to the online questionnaires were not made mandatory and they were not evaluated in detail, even though corresponding provisions already existed in the adopted software.

Overall, an isomorphic, near-identical representation between modes (Denscombe, 2006) was achieved, which was able to yield identical and comparable data. But smaller differences remained. For example, multiple-choice question formats that required respondents to pick one of several options were implemented with standard HTML “radio buttons” in the online mode. While it was possible for respondents to change the answer to any other option, it was not possible for them to uncheck the answer completely. In the paper questionnaires, however, respondents could cross out (cancel) a given answer in tick boxes or even clearly state that they did not want to give an answer. While the ICC members acknowledged this possibility to add extra “don’t know,” “not applicable,” or “don’t want to answer” categories to all such questions, they took a balanced decision against it.

Implications for national centers

Countries were asked whether they intended to use ODC for the SITES main data collection. If so, those countries were obliged to also field test the ODC procedures. The first step for national centers after receiving the international version of the paper questionnaires was to translate them to the language(s) used for the SITES survey within that country. Since there could be more than one language, the preparation of online instrumentation was based on the concept of “cultures.” A “culture” was referred to as a certain language within a certain cultural context.

For example, instrumentation for “Estonian in Estonia” and “Russian in Estonia” had to be prepared separately. Because adaptations (e.g., ISCED levels) to the questionnaire could differ for the same language used in different countries, “English in Singapore” and “English in South Africa” were treated as distinct cultures.

The translation of the paper questionnaires was verified by the IEA Secretariat. Afterwards, the layout of the translated questionnaires was checked by the International Study Center (ISC) against deviations from the international version. Only after these check points were countries allowed to commence conversion to the online versions of the questionnaires. For this task, the DPC provided an application to national centers, called the *IEA SurveySystem Designer* (see the section on technical implementation below). The DPC additionally provided country representatives with training for the application as part of a data management seminar. National centers received, along with the Designer, the international template files with the structure of the questionnaire already built into these definitions.

National centers used the *Designer* to translate all text and to adapt the questionnaire where necessary. Because translation had already been done for the paper questionnaire and because the online translations and adaptations had to be identical to the corresponding paper questionnaires, this task was largely a copy-and-paste procedure. However, in addition to the paper text passages, certain text passages were needed exclusively for online purposes. These included text passages on the welcome and help screen, on navigation buttons, for error and validation messages, and for other information related to the online mode.³ The integrated “web preview” component of the *Designer* enabled national centers to check what the online version would look like at any stage of the translation and adaptation process. Once the centers had completed translation and adaptation, all they needed to do was email an exported archive of files to the DPC. These files, which contained all structural information, all translations, and all national adaptations, were the only electronic resources needed to set the survey “online” for that particular culture.

³ Examples for online-related passages are “Login,” “Username,” “Password,” “Help,” “You are logged in as ...,” and text on navigation buttons like “Next” and “Previous.”

Once questionnaires had been set online, national centers sent information to respondents on how to access the online questionnaire. This information was sent in the form of so-called “cover letters,” which were essentially the first two pages of the respective paper questionnaire. The letters contained general information about the survey, a culture-specific internet address (URL), as well as a label with a username and a password. In line with the process for the paper questionnaires, the cover letters were sent to school coordinators, who distributed them to the designated individuals within their school.⁴ If needing support, respondents could address their national center, which, in turn, could contact the DPC if unable to solve the problem locally.

During the data-collection period, national centers were able to monitor progress and responses to the online questionnaires. National centers were given access to the monitor site where each respondent’s identification code, additional tracking information (e.g., first and last login), and responses to the questionnaire were displayed, thereby enabling national centers to retrieve a real-time status report and to send reminders to those people who had not responded in the expected period of time.

After data collection, national centers were required to enter data from the paper questionnaires (if any) via IEA’s Windows Data Entry Manager (WinDEM) software that used codebooks matching each center’s adapted instrument versions. The DPC provided the international version of the codebooks. Countries that used only the paper mode had to adapt the codebooks manually to reflect any adaptations made. Countries that used ODC did not have to adapt the codebooks manually since the *Designer* featured an export of codebooks for use in WinDEM, including all national adaptations.

In summary, participants using ODC in SITES faced additional, parallel workloads and complexity before and during the data-collection period. Specifically, they had to prepare online questionnaires in addition to the paper versions, learn how to use the *Designer*, prepare cover letters, and monitor progress. However, the centers had the benefit of a reduction in workload after data collection. Because answers to online questionnaires were already in electronic format,

and responses were stored on servers maintained by the DPC, there was no need for separate data entry. Once the data-collection period ended, national centers had no further obligations regarding the data collected online.

Additional procedures at the DPC

The use of ODC had an impact for the DPC even at the stage of questionnaire development. Because the *SurveySystem* was developed under a time constraint, the software could not “manage” every kind of layout of the questions evident within the original paper questionnaire. In these cases, mostly involving complex multi-matrix formats, the question was redesigned to be displayed similarly in both paper and online modes. In this respect, the use of ODC had a backwash effect on the paper questionnaires.

Working procedures at the DPC were also affected. DPC staff had to set up servers through which respondents could access the online questionnaires. For data collection, staff created a database for storing all answers. In the event of server failures or crashes or problems accessing the database, staff developed a backup strategy involving additional servers. They also defined an automated routine check of the permanent availability of the online questionnaires and a semi-automatic procedure to restore availability if needed. During the data-collection period, DPC staff constantly monitored the servers and the database.

After finalizing the field-trial paper instruments, the SITES consortium converted these to the online mode using the *Designer* at the DPC to ensure this mode reflected the structure of the paper questionnaire and all text passages. The resulting international definition templates were sent to countries and served as the basis for translation and adaptation. Another additional task for the DPC was to extend the training for national data managers to include use of the *Designer*. This training, a new feature of the data-management seminar held in Hamburg in June 2005, saw the DPC giving a special presentation and hands-on training on both the functionality and use of the *SurveySystem*.

When national centers finished translation and adaptation of the online questionnaire, they returned the culture-specific definition files to the DPC. As an additional quality control measure prior to uploading

⁴ “School coordinators” were the in-school liaison people for the national centers and were responsible for listing eligible teachers as well as distributing and collecting materials.

a country's survey, DPC staff made comprehensive visual question-by-question checks for differences between online and paper versions, and also employed triangulation by using the national adaptation forms that countries provided for documenting deviations from the international version of the questionnaires. Deviations were reported back to national centers, and online questionnaires were released and made accessible only after any remaining issues were resolved satisfactorily, ensuring an isomorphic representation of questions in both modes. The described procedure was done for all cultures that used ODC.

At the end of the data-collection period, the DPC took the questionnaires offline and merged and subsequently verified the data from the online and paper modes. Problems in the structure of the data, for instance out-of-range values, were reported back to countries. With respect to problems with data from paper questionnaires, countries were asked to check for mistakes that may have occurred during manual data entry. For data from the online mode, the only way to verify suspect values was to contact the respondent directly.

Technical implementation

After the procedural requirements and methodological necessities and constraints outlined in the previous section had been addressed, the ICC had to determine a suitable plan for technical implementation of the ODC.

The first step was to look at existing solutions, technologies, and tools in the field of electronic forms that might also serve the needs of SITES. For instance, the consortium reviewed Adobe's Acrobat (PDF) technology as an alternative for creating and collecting electronic forms. However, under the agreed strategy to avoid any prerequisites on the side of respondents' or schools' computers, the consortium discarded this and similar possibilities. The second step was to identify and evaluate approximately 15 commercial and open-source software systems for internet-based questionnaires, whether especially designed for this purpose or as an add-on for paper-, telephone-, or interviewer-based surveys. The reviewed systems were checked against the list of mandatory features described

in the previous section during test-driving, personally liaising with vendors, consulting with service providers, and attending subject-related conferences.⁵ The general conclusion was that no single "out-of-the-box" solution would integrate with the pre-agreed survey operations, most importantly in the areas of (i) decentralized translation, adaptation, and documentation, and (ii) mixed-mode data collection and subsequent data processing. In mid 2004, the consortium decided to delegate the task of developing a suitable self-owned and self-managed software product (later coined the "IEA SurveySystem") to the IEA DPC's software unit. This delegation was supervised and supported by the authors of this present paper.⁶

Essentially, the IEA *SurveySystem* is a hierarchical, structure-oriented model of a survey that stores and manages all questionnaire-related information, including text passages, translations, and adaptations, as well as verification rules, variable names, and labels for data management (see Figure 1). In contrast, data sets and data-entry software that were or are used in previous and current educational surveys, such as TIMSS and PIRLS,⁷ are "flat," in the sense that they store data on variable level only with limited contextual information in, for instance, variable names and value labels. The SurveySystem's consolidation of metadata in a single set of files that the SITES national and international centers could easily send to each other over the internet allowed for a consistent way to manage the localized versions of questionnaires, including all information that was later necessary for data entry. This approach also provided potential for use beyond the context of ODC (see the section below titled "Roadmap").

To serve the different usage scenarios, three distinct, but interconnected, components of the SurveySystem were developed: the *Designer*, the *Web*, and the *Monitor*. All three utilize the same core library, which provides the necessary logic for the particular usage context, for instance side-by-side translation for translators using the Designer, login validation and navigation for the actual respondents in the Web application, or active surveys and their current participation using the *Monitor* application to be used by the national research coordinators (NRCs).⁸

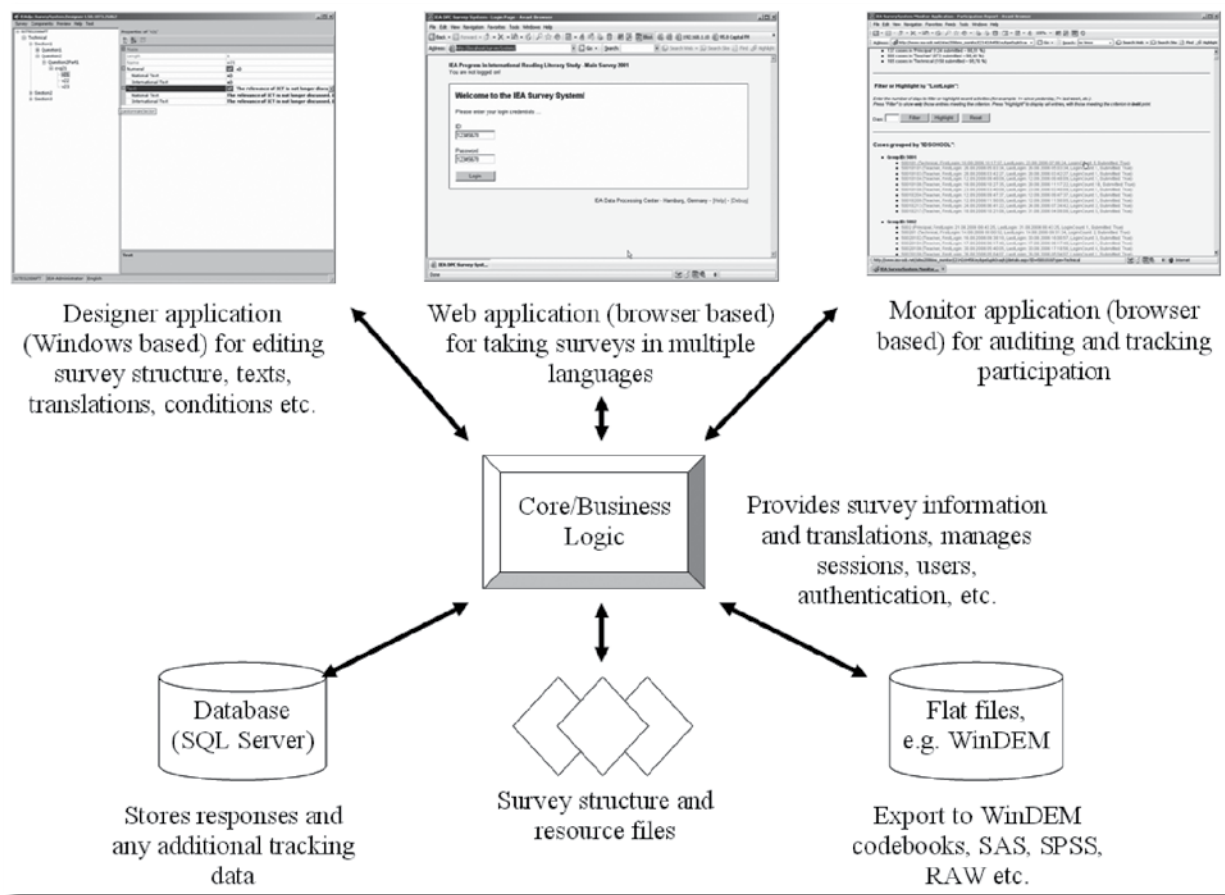
⁵ Most importantly, the General Online Research Conference (GOR, <http://www.gor.de>).

⁶ The authors would like to acknowledge the creative work of Stephan Petzchen, director of the DPC's software unit.

⁷ TIMSS stands for the Third International Mathematics and Science Study; PIRLS for the Progress in International Reading Literacy Study.

⁸ NRCs are country representatives at the national center in charge of conducting the study within a country and serve as the main contact person for their country with the International Study Center.

Figure 1: Architectural Overview of the IEA SurveySystem (Core, Components, Data Storage)



The *Designer* facilitated the flexible definition of survey elements, for example, questionnaires, sections within a questionnaire, questions, etc., and the translation/conversion of text passages.

- The *Designer* was used to create, delete, disable, and edit survey components and their properties and allowed for translation of all text passages in the existing national paper-and-pencil questionnaires and additional "system resources," for example, the "next" button or validation messages.
- Usage was based on three roles: administrator, manager, and translator. Whereas an administrator was responsible for setting up an international master template for a survey, managers were given the right to edit only the structure and text passage of their national definition. Translators were not allowed to make changes to the structure of a survey, for instance, to remove a question, and

had access only to the national part of the text passages.

- For the purpose of verification and test-driving, the *Designer* included a complete web server to drive the integrated preview system that delivers the survey exactly as under live conditions.
- The *Designer* supported the export of codebooks to IEA's generic data-entry software WinDEM to allow for structurally isomorphic data entry of online and paper questionnaires.
- Technically, the *Designer* was implemented as an offline Windows application to be run on computers in national centers to avoid the need for a permanent internet connection.

The *Web* component was a compiled web application that served the personalized HTML pages to the actual respondents around the world for completion from within standard internet browsers.

Given the overall goal of securing the maximal possible coverage, no respondents were to be excluded because of incompatible browsers or disabled features. In addition, computer literacy was expected to vary greatly among respondents and users. Therefore, the design of the application was balanced between minimally desirable capabilities on the one hand and simplicity on the other so that users could expect the application to work intuitively and under minimal conditions. This was especially important, as requirements in terms of connection speed, available software (browsers), and modem and processing speed were identified as highly critical obstacles during initial discussion and the review of literature (for example, Reips, 2002). In this sense, the approach in SITES (selected aspects below) was similar to the “respondent-friendly design” (Dillman, Tortora, & Bowker, 1998).

- In detail, the output was designed and tested to ensure (near) identical representation in all supported browsers. The screen size was set to 800 by 600 pixels with 256 colors. With the exception of the welcome screen, no logos or graphics were used.
- As a minimum, the web application required HTML 4.0 (an internet standard since 1998) to reliably support tables, the “dir” attribute to support bi-directional unicode layout, and cascading style sheets, version 1, for formatting.
- Given that a growing number of people (presumably especially those with a “higher” education) are aware of potential security and privacy threats associated with the use of the internet and of unwanted pop-ups and advertisements, it was likely that the corresponding browser features would be turned off or would work unreliably. The web application thus made use of plain HTML controls and did not require any “fancy” technologies, such as cookies, JavaScript, Flash, drop-down boxes, or pop-ups to work.
- Users with browsers that did not meet these minimal standards were presented with a localized lists of supported browsers, which were Internet Explorer 5.0 and above, Netscape 5.0 and above, Mozilla/Firefox 1.0 and above, Apple Safari/WebKit 1.0 and above, Opera 7.0 and above, and

Konqueror (all versions regardless of platform), where applicable.

Finally, the *Monitor* component, another compiled web application, allowed national centers to audit participation in real-time and to follow-up schools in the case of incomplete or not-returned questionnaires in a similar way to that used for administration of the paper questionnaires.

- After a secured login, the *Monitor* listed and grouped (by school) all participants who had logged in to the system at least once. In addition, the report included the number of logins of each respondent, the time of the first and last login, the questionnaire’s submission status, the browser used, and the reported screen resolution.
- To support the day-to-day monitoring, highlight and filter features were included that limited the group lists to the latest activities only, for example, all completed questionnaires in the last three days.

As a technical and programming basis for all components and modules, the Microsoft .NET Framework, which is similar to the Java programming and application framework, was used not only because of its proven robustness and excellent suitability for multi-lingual-based (Unicode) and internet-based applications in general but also because of the high level of existing expertise available within the DPC. The live systems were hosted on dedicated and redundant high-performance servers rented from a reliable and experienced solution provider based in Germany. Appropriate measures were taken to secure the data, and these were further strengthened by a security audit conducted by an external organization.

Staged development, implementation, and evaluation

The introduction of ODC technology and methodology in SITES, although warmly welcomed and much expected by NRCs, had to take place in a gradual and careful way. Provisions had to be made to rule out the risk that the ODC would fail in general or render data incomparable and therefore make them useless. In ongoing cooperation with the NRCs⁹ and experts, consortium staff launched the ODC

⁹ The authors would like to acknowledge the critical input and valuable contribution made by Anne Larson (Denmark), Juan Enrique Hinostroza (Chile), and Alona Forkosh-Baruch (Israel).

component in three main phases—a technical try-out, the field trial, and the main survey.

The first phase (i), the technical try-out with 12 participants in May 2005, sought to collect basic experience and to locate and eliminate basic technical problems with respect to languages, the conversion of instruments from paper to online, scripts and script direction (e.g., Hebrew), and browser compatibility. In general, this phase was highly successful in identifying key requirements and defining further development.

The second phase (ii), the field trial in autumn 2005 with 16 out of the 18 participants, made use of a feature-complete ODC system developed on the basis of the technical try-out findings. A split-sample design to identify, investigate, and statistically control for possible problems in relation to the data-collection mode, such as response bias and non-response at variable and questionnaire levels, was employed. The consortium planned to make a final go/no-go decision after reviewing these results.

The largest part of the evaluation therefore was based on the findings and outcomes of the field trial, especially in terms of the new procedures developed for SITES. Although a volume of authoritative literature is available on the general question and specific aspects of using ODC methodology in small- or large-scale surveys, mainly at regional or national level, very little of this work relates to past and present projects in the educational domain, and even less relates to projects in which instrumentation is administered in two parallel (mixed) modes and simultaneously, internationally. In addition, the degree to which outcomes and observations from one such survey are actually transferable to another was difficult to predict due to varying populations, sample designs, research domains, and perceived importance and relevance of the respective survey within the particular population. While producing a working technical and procedural solution was, of course, a key prerequisite, the evaluation of the online approach in SITES was thus based on existing evidence (where applicable and appropriate in general), but required separate investigation to reflect the unique SITES characteristics by addressing two key questions:

- *Key Question 1—Feasibility:* Investigate whether ODC methodology and procedures can work in the context of an international IEA study.

- *Key Question 2—Validity:* Analyze whether the two modes—online and paper and pencil—yield comparable data, thus allowing the implementation of both modes in and across countries.

In terms of the key question of feasibility in an international context within the IEA survey framework and with respect to the aspects described in the previous sections, especially “Conceptualization and Integration with Survey Operations,” the following list contains selected, yet central, aspects, which were largely based on experiences and observations reported by NRCs and national center staff.

- Participants found using the software package designed to transform paper questionnaires for the online mode and to serve the live survey (i.e., the IEA *SurveySystem*) to be straight-forward. However, the DPC received valuable suggestions for improvement during the initial development (try-out and field trial), mostly with respect to usability and time saving.
- There were no observed or reported problems with respect to translation, conversion, and representation of advanced scripts (such as Thai, Chinese, and Hebrew) or script-direction (right-to-left writing) during preparation and administration of either the field trial or the main run.
- The majority of localized definition files for the online administration were submitted for review within the proposed timeframe. On average, participants reported a required conversion and verification time of two to three person days, equivalent to one day per questionnaire, with additional time needed to communicate and resolve possible differences.
- The thorough comparison of the two sets of instruments discovered only minor mistakes introduced during the copy-and-paste conversion. In this respect, the conversion to the online mode did not interfere with the preceding translation and layout verification activities and check points. However, the time needed for this work at the international level was underestimated because of the need to conduct full reviews instead of simple spot checks.
- During administration of the field trial and main data-collection run, no major problematic behavior of the servers or users’ browsers was observed or

reported to the IEA DPC from national centers. The ongoing monitoring of server log files allowed the identification and resolution of what were “few and far between” technical glitches that, after fixing, had no measurable influence on respondents’ sessions and participation rates. No unplanned server down-time was recorded.

- The SITES procedural design required paper versions of questionnaires to be used in case respondents refused or simply were not in the position to participate online. Unfortunately, during the field trial, the distribution of these fall-back paper questionnaires seemed to work for only some of the ODC participants. The need to reliably record the demand for paper instead of online questionnaires was regarded as highly important by the ICC and the NRCs, who firmly monitored the successful application during the main data collection.
- The review of server log files showed that in relation to all online-administered questionnaires, less than 0.1 percent of users attempted to access the web application with browsers (e.g., Internet Explorer 4) that were intentionally excluded because of poor support for basic technologies (i.e., style sheets). The same magnitude was also observed during the main study.
- National centers reported making extensive use of the provided *Monitor* application to access the return status of all questionnaires in real-time in order to facilitate following up non-respondents. In parallel to the main study, additional features were introduced to filter and highlight recent activities.
- Eventually, the integration and subsequent data-cleaning and data-processing strategies adopted from PIRLS were adjusted to reflect this second set of data, but did not constitute a major obstacle with respect to the quality of the data or communication with NRCs. In all cases, manually entered data sets and data from the online mode structurally matched 100%. Thus, both modes used the exact same variable and coding scheme within a participating system.

Despite the increased workload and the added complexity during the initial phase, the group of ODC participants apparently managed the preparation and subsequent administration of the online mode well.

While both the SITES field trial and the main study in some of the participating systems were carried out in an atmosphere of “survey” or “educational reform fatigue,” there were no indications that subsequent reluctance or even resistance of schools to participate in the survey or the willingness of individual respondents to fill in the questionnaire related to the mode of administration. All in all, the majority of procedures and tasks, supported by manuals and direct support via email, went as intended, although for some areas, more attention and allocation of time and resources were required than originally expected, for instance, for cross-checking questionnaire wording and structure.

Evaluating the second key question (i.e., whether the paper and online modes yielded comparable data) naturally was a more complex process than that required for the first key question and involved statistical analyses. The existing literature on the question of mode effects allows one to identify issues that are repeatedly (yet inconsistently) observed in online surveys, for instance, drop-out increasing with the length of questionnaires and non-response in complex matrix questions. In SITES, these were expected to occur (if at all) in this measurement area, but not in the areas of population coverage or sampling. Unlike some surveys that have to rely on self-selection of respondents, and thus have no clearly defined, authenticated, and documented sample with known probabilities of selection, SITES employed a sound and defensible sampling design in both the field trial and the main data collection.

For the purpose of comparison, SITES implemented a split-sample design at school level during the field trial (with a typical sample size of 25 schools combined), thereby simulating the same survey with two independent yet near-identical samples assigned to uniform modes, using near-identical questionnaires within each. Although serving the analytical needs, this created a dilemma of sorts because the split design did not match the proposed implementation options for the main data collection, where schools and individual participants within schools were expected to have greater freedom in regard to preferred mode of administration. Other analytical limitations originated from the two-stage sample design, the limited effective sample sizes, and the partially experimental instrumentation. The mode comparison, however, had to be based on data coming from these preliminary

instruments, especially with respect to factors and scales that would allow the go/no-go decision before the start of the main run.

In short, the SITES consortium had to show that the ODC instruments could yield results comparable to paper-and-pencil instruments if both were to be used in the study. Analyses for this were conducted in November 2005 under the auspices of the DPC. The major conclusion drawn, based on various approaches, was that there were no substantial differences between the data derived from paper and that derived from the online mode that would reduce the ability to join these sets of data and to make joint analyses. Unfortunately, issues were identified in terms of participation and drop-out, which were partially rooted in the preliminary character of the field-trial instruments. For example, the level of drop-out was negligible or fairly limited in the two school-level questionnaires, which had fewer questions overall, but substantial, although not critical, in the significantly longer teacher questionnaire. Adequate measures, in this case the reduction of overall length, were implemented for the main run.

The SITES consortium, after reporting to and requesting support from the group of NRCs and the IEA's technical expert group (TEG) in December 2005, thus decided to recommend the further use of ODC during the main study where possible and desired.¹⁰ Accordingly, in this third and final phase (iii) in April/May 2006 (Northern Hemisphere) and August/September 2006 (Southern Hemisphere), ODC was offered as an international option. While nearly all of the participating systems expressed at least some interest in using the online mode as either the default for data collection (with paper-and-pencil questionnaires as a mandatory fall-back) or for selected schools, regions, or individuals only, 17 of the 22 participating systems eventually implemented ODC, usually as the default mode for collecting data, after taking into account the suitability for their local context.¹¹

A thorough review of the main survey is presently in progress. At the time of writing, there were no

major observed or reported problems with respect to the approximately 20,000 questionnaires administered online.

Conclusions

IEA SITES 2006 explored the possibility of administering an educational survey over the internet in addition to the traditional paper-and-pencil response channel. While the advantages of collecting the large amount of data in SITES over the internet were evident and substantial in terms of costs and time, these factors did not provide grounds in their own right to implement online questionnaires. Maintaining quality, securing the highest possible participation rates, and arriving at comparable data, whereby the results of the survey could be generalized to the population of Grade 8 mathematics and science teachers, were the key goals and criteria of the SITES approach, but unfortunately not automatic (see Couper, 2000).

The procedures that were necessary to support ODC alongside the conventional paper-and-pencil track, which will continue to be the dominant mode for some time, and certainly for participants with low internet penetration, usage, or affinity, were designed in such a way that little or no changes were required with respect to the tried-and-true survey operations employed in other IEA surveys (currently PIRLS, TIMSS, ICCS, and TEDS-M¹²), especially with respect to the fact that ODC was not mandatory, but offered as an international option only. Although the majority of participating systems that opted for ODC implemented it as the default mode and, hence, administered only a small number of paper questionnaires, the main challenges were to cater for isomorphic versions of the instrumentation in both modes, to reliably administer the resulting mixed-mode survey within the existing survey framework and procedures, and to subsequently integrate and process the two sets of data. Given the relatively short period of time that was available to determine and then implement an ODC strategy, the overall conclusion is that SITES 2006 was successful in achieving this. The obstacles experienced during the actual administration

10 For details, please refer to the accompanying paper on the comparison of this split-sample data (Brecko & Carstens, this volume), which covers the following aspects: comparison of response rates, individual variables, missing rates between the data gathered by the two modes, in particular the level of not-reached questions (drop-out), and the reliability of the data/scales gathered by the two modes.

11 Two of the 17 participants administered the survey in two languages each, resulting in a total of 19 sets of instrumentation. For one participant, a recommendation was issued to refrain from using ODC. The remaining four participants decided not to use ODC because of feasibility concerns.

12 ICCS is the International Civic and Citizenship Education Study; TEDS-M is the Teacher Education and Development Study-Mathematics.

of SITES, primarily difficulties in securing school participation, reportedly were attributable to factors such as general “survey fatigue” and not connected to the online mode itself.

Nonetheless, certain aspects of the migration from paper-based to web-based questionnaires were identified as being methodologically problematic during the evaluation activity. The length of questionnaire as well as the type and complexity of question types appeared to be factors that increased the level of non-response, especially toward the end of the questionnaire. The recommendation made for the main survey in order to ensure high response rates was to reduce the overall length as well as the number of matrix questions, given that these factors seemed to have been the main contributors to the observed level of drop-out. Another issue was accurate documentation of the necessary survey mode within schools. National centers were therefore requested to check with schools whether the online mode was acceptable and feasible for them before sending out materials and providing fall-back paper questionnaires for those individuals without internet access and/or required equipment, or who simply refused.

Provided that the above two main prerequisites (concise instrumentation and accurate allocation of mode) have been met, we, along with the ICC, think that the proposal to implement ODC in SITES was sound and defensible. However, the decision and responsibility for using ODC rested with the NRCs, who were free to decide on the type and scope of implementation, that is, as the default mode, for types of questionnaires, and for specific schools or individuals only. As a guideline, it was recommended that the decision made accorded with the level of confidence that a particular participant had in regard to using the online mode. The degree of confidence could be based, for example, on within-country computer/internet penetration, poor online response rates in previous surveys, and (more importantly) the outcomes of the mandatory field trial.

Roadmap

If we acknowledge that, in the foreseeable future, the use of conducting surveys online will continue to

grow, the question we need to consider is not whether ODC will be applied in international assessments, but rather under which conditions it can be efficiently integrated with or will eventually replace paper-and-pencil approaches altogether. More importantly, we also need to consider how IEA might be able to exploit such benefits. SITES 2006 constitutes only one specific instance of collecting data over the internet. Nevertheless, because the underlying procedures and artifacts were designed in general terms, future development of the methodology can be seen as having two tracks: (i) technical and procedural advancement; and (ii) application to other contexts.

Technically, the next iterations of the IEA *SurveySystem* may include a number of (presently discussed) advanced features, such as print-outs, the generation of questionnaires in Microsoft Word or other formats, the integrated documentation of national adaptations, the integration of translation and translation verification (partially automated), advanced monitoring (i.e., direct access from schools), improved creative flexibility for layout and question types, and more granular demand tracking for paper/online questionnaires. Other additions, such as multimedia elements, are equally possible within the created framework.

In general, future development of the procedures and artifacts used for ODC in IEA studies will involve attempts to further decrease the overall workload and complexity for national and international centers. Conversely, requests or recommendations to transfer more or all of the operations in IEA surveys to the online mode, for instance, listing and sampling, participation tracking, data entry and verification, are interesting ideas, but they are unlikely to be feasible in the near future given the range of national contexts that IEA has to cater for. Presently, we consider the online mode an efficient addition to paper-based surveys with regard to data collection only.

In relation to other contexts, the approach and software are currently being used for administrative data collection in PIRLS¹³ and SITES.¹⁴ Similar implementations are likely to be used for a variety of scenarios in all IEA surveys, including collection of feedback from test administrators, administrative forms, and post-hoc documentation. With respect to

13 “Survey Activities Questionnaire” and “Curriculum Questionnaire.”

14 “National Context Questionnaire.”

content survey data, an experimental implementation was done for the questionnaires in a small-scale regional survey on citizenship in Germany (BLK Demokratie), for which the DPC served as a fieldwork contractor. The next large-scale instances of applying ODC methodology similar to SITES, that is, for background and context information, will be the OECD's Teaching and Learning International Survey (TALIS 2008), for which the DPC is serving as international contractor and study center, and the IEA International Civic and Citizenship Education Study (ICCS 2009). We have yet to see, of course, if the ODC methodology will work for these surveys, where a much broader variety in terms of the subject matters under focus and the surveyed adult populations can be expected.

Furthermore, initial concepts to extend the usage of ODC to proficiency/performance assessments of, for the most part, students, are currently being developed

on an experimental basis. Due to quality standards, plans to offer instant online reporting/feedback to schools or countries were dropped from the original SITES 2006 conceptual framework and proposal.

Overall, there appears to be a strong interest for alternative methods of delivering background questionnaires or even assessment materials to respondents,¹⁵ assuming that these offer added value by, for instance, enriching the experience for respondents, opening options unavailable with paper-and-pencil methodology, reducing printing and/or data entry costs, and shortening the time until analysis. We therefore consider that it is important to share existing expertise and the relatively successful implementation in SITES 2006 with partners within and outside of IEA in order to further improve and extend the current technology and knowledge-base for use in future studies and contexts.

References

- Anderson, R. E., & Plomp, T. (2001). *Student assessment of ICT-supported knowledge management competencies: A plan for IEA SITES M3*. Report to the IEA Standing Committee. Minneapolis, MN: University of Minnesota.
- Bandilla, W. (2002). *Web surveys: An appropriate mode of data collection for the social sciences?* In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online social science* (pp. 1–6). Seattle, Toronto, Bern, Göttingen: Hofgrete & Huber.
- Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64, 464–494.
- Denscombe, M. (2006). Web-based questionnaires and the mode effect: An evaluation based on completion rates and data contents of near-identical questionnaires delivered in different modes. *Social Science Computer Review*, 24(2), 246–254.
- Dillman, D. A., Tortora, R. D., & Bowker, D. (1998). *Principles for constructing web surveys: An initial statement* (Technical Report No. 98–50). Pullman, WA: Washington State University Social and Economic Sciences Research Center.
- Martin, M. O., Rust, K., & Adams, R. J. (1999). *Technical standards for IEA studies*. Delft: Eburon.
- Reips, U.-D. (2002). Standards for internet-based experimenting. *Experimental Psychology*, 49(4), 243–256.

15 See Computer-Based Assessment of Science (CBAS) in PISA 2006, tentative ICT assessment components 1A/1B in PISA 2009 (PISA 2009—Call for Tender, OCDE/EXD/PCM/EDU/IA/05/148), and Anderson & Plomp (2001).

Political attitudes and behaviors across adolescence and early adulthood: A comparison of IEA and European Social Survey findings

Marc Hooghe

Catholic University of Leuven, Belgium

Britt Wilkenfeld

Catholic University of Leuven, Belgium

University of Maryland

College Park, Maryland, USA

Abstract

The International Association for the Evaluation of Educational Achievement (IEA) Civic Education Study (CivEd) measured important aspects of adolescents' civic attitudes and behavior. Two age cohorts were examined: 14-year olds in 1999 and upper-secondary students (with a mean age of 17.9 years) in 2000. The European Social Survey (ESS), conducted in 2002, examined similar constructs and included an age cohort of 18- to 30-year-olds. The current study compares attitudes and behaviors across adolescence and early adulthood by conducting secondary analysis on data from the two IEA studies and

the ESS. Patterns of continuity and change are examined within nations across the three age groups, as well as among nations. Countries examined include Denmark, Norway, Sweden (Scandinavian countries), the Czech Republic, Poland, Slovenia (Eastern European countries), Portugal, and Switzerland. Congruence across the three age groups was found for political trust, but not for voting behavior or positive attitudes toward immigrant rights. Explanations of continuity and discontinuity are explored.

Introduction

Is it meaningful to ask 14-year-olds if they intend to vote, believe immigrants are entitled to certain rights, or trust their nation's leaders? Although such questions, particularly those pertaining to voting, are regularly used in research on political socialization, their external validity is rarely studied in a systematic manner (Denver & Hands, 1990; Metz, McLellan, & Youniss, 2003; Sapiro, 2004; Torney-Purta & Amadeo, 2003). Reported intentions to vote and participate in other civic activities may be genuine. However, answers to these questions may be plagued by social desirability given that the most acceptable answers will be apparent in most societies and education systems. Although the issue of social desirability poses a challenge for the validity of all survey methodology, the Civic Education (CivEd) studies of the International Association for the Evaluation of Educational Achievement (IEA) may be more susceptible to social desirability than other IEA studies. Specifically, the CivEd studies may have a vulnerability to false positives that merely reflect the students' awareness of the value pattern prevalent in the school system. The same can be said for reports

of intended participation in other civic activities, such as community participation and political party membership, or civic attitudes, such as trust and tolerance.

Additional issues arise when examining adolescents' civic attitudes and behavior. Specifically, can an adolescent's attitude toward a particular issue predict his or her attitude as a young adult? Studies often ask adolescents about their faith in institutions as diverse as the courts, policy, or parliament. Fourteen-year-olds may have indirect experience with these institutions, through school and parents, but generally will not have had any direct experience. Therefore the grounds on which they form their opinions are unclear. Similarly, do reports of intended behavior at a young age have any predictive value for actual behavior at a later age? Adult voter turnout is consistently shaped by idiosyncratic motives, resulting in significant individual differences in participation from one election to another (Blais, 2006). This leads one to question whether citizens' intentions to vote are predictive of actual voter turnout.

Such questions bring us back to the discussion of the utility of describing the political and social attitudes of 14-year-olds. If values and attitudes are still developing during adolescence, perhaps their current attitudes are merely a snapshot with limited predictive value with regard to young adult attitudes. Conversely, if basic attitudes are well developed by age 14, then such fundamental beliefs should remain relatively stable throughout the life cycle. This consideration validates questioning 14-year-olds about their political attitudes because it offers a glimpse of their political future.

Ideally, these questions would be addressed using a panel design. The adolescents that participated in the IEA study could be sampled a second time to ascertain whether they actually voted, signed a petition, participated in a protest, and so on.¹ Indeed, the 14-year-old respondents of the 1999 IEA survey are now 21 years old and therefore past voting age in all examined countries. However, the anonymity inherent in the CivEd data collection makes it impossible to conduct a follow-up study.

This paper will explore a different route to assess the external validity of the IEA data on civic education and engagement, fully utilizing the study's comparative and international nature. The basic assumption is that if attitudes are stable, and if they are measured in a reliable manner, it is possible to predict stability across various phases in the lifecycle. Given the current data limitations, we cannot test this assumption on the exact same individuals, but we can test it across societies. For example, if 14-year-olds in society A report significantly higher trust levels than their counterparts in society B, we can assume that the differences in 18-year-olds' trust levels in societies A and B will reflect a similar pattern. The comparison can be taken a step further by examining the responses of young adults in a general population survey (in this case, the European Social Survey) that offers a reliable measurement of civic attitudes and behaviors for various European societies. Given the same assumption of attitude stability, observations among 14-year-olds should have some predictive value with regard to the observations among young adults. The current analysis will examine the continuity of civic attitudes and behaviors across adolescence and into early adulthood, including variation by country and gender.

Theoretical framework and literature review

Continuity versus discontinuity, or stability versus change, is a recurring topic of debate in developmental psychology. Some theorists assert that development is necessarily discontinuous or stage-like (Erikson, 1968), resulting in distinctly different attitudes and behavior throughout the lifespan. Changes in cognitive structures that occur during each stage facilitate changes in thinking. Theorists on the other side of the continuum assert that experiences at all points of development influence a person's cognition and behavior (Bandura, 1986). According to such theories, attitudes and behavior would not be distinctly different at different ages because of the cumulative and continuous nature of cognition.

Attitude and behavior stability

Research findings vary on the development and stability of attitudes in adolescence and adulthood. The first issue in attitude research is clarification of what can be justified as an attitude. Batista-Foguet and Saris (1997) assert that judgments must be stable over time, and consistent with related judgments, to be considered an attitude. Everything else is just an opinion and therefore subject to fluctuation. A second issue is whether development is even related to stability because of the proposition that attitude development and attitude change are not the same thing. Attitude development requires change in the quality of thinking, rather than merely change in thinking (Lao & Kuhn, 2002). Finally, when measuring social attitudes, how the issue is framed can manipulate cognition and attitudes, though specific attitudes such as affirmative action are more susceptible to framing than are general or core attitudes such as racism (Sibley, Liu, & Kirkwood, 2006). As these issues indicate, measuring attitude stability versus change is a complicated undertaking.

When examining the stability of political attitudes of different age cohorts in the 1950s (from 1956 to 1960) and the 1970s (from 1972 to 1976), Alwin and Krosnick (1991) found that the 18 to 25 age group (the youngest in the study) showed the least stability. Stability generally increased with age until late adulthood where it again became unstable. The

¹ However, social desirability can also be a problem when measuring actual behavior. In this kind of survey research, it is common for more people to report they have voted than is reflected in actual voter turnout.

researchers attribute attitude instability during youth and late adulthood to general instability at this point in the life cycle, as this is when major life events often occur. In a similar study, stability in cultural orientations (e.g., tolerance and equality) did not occur until late adolescence, defined as 21 to 24 years old (Vollebergh, Iedema, & Raaijmakers, 2001). The researchers posited that older adolescents' advanced organization of attitudes, as well as general decline in parental influence, contributed to generating attitude stability.

There has been less research on the stability of behavior, although one panel study found that adolescent reports of civic engagement were associated with civic engagement at a later age (Jennings & Stoker, 2004). Specifically, high school seniors' reports of volunteer work and other organizational involvement was strongly linked to involvement levels in their 40s. Granberg and Holmberg (1990) reported that intended voting behavior was consistent with actual voting behavior, although prior behavior and self-identity had a moderating influence.

Given the nature of the CivEd study, it is not possible to look at continuity or stability of attitudes and behavior within individuals. Rather, this paper examines continuity within individual nations for similar age cohorts. Historically, patterns of political attitudes and behaviors have shown persistent differences across European societies. These differences are quite apparent and stable among adults, so it can be assumed that similar patterns exist among adolescents. For instance, overwhelming differences exist between the political cultures of Scandinavian countries and the countries of Central and Eastern Europe. We anticipate from our look at political trust, attitudes toward immigrants' rights, and intention to vote, to find similar patterns across all three age cohorts in the selected countries.

Gender differences

In addition to attitude stability, we also examine attitude differences, specifically gender differences. Women have been found to be more trusting in governmental institutions than men, both in adolescence (Hooghe & Stolle, 2004) and adulthood (Johnson, 2005). In relation to attitudes, women tend to be more liberal on political issues and voting (Atkeson & Rapoport, 2003; Studlar, McAllister, & Hayes, 1998). This difference

may start in adolescence, as was found in a study that asked youth whether social and political rights (e.g., social benefits and voting) should be extended to "controversial" groups such as immigrants. Adolescent girls responded with more tolerance on 10 of the 12 items (Sotelo, 1999).

Differences between men and women are also found in regard to political participation. Some studies report that civic participation is higher among men (Hooghe & Stolle, 2004); others report that women have essentially closed the gender gap (Atkeson & Rapoport, 2003). Findings appear to vary depending upon the type of political activity. For instance, in the United States, women now have higher voter turnout than men (Jamieson, Shin, & Day, 2002), are equally involved in protests, and are less likely to contact a public official (Verba, Scholzman, & Brady, 1995). Atkeson and Rapoport (2003) expressed concern that although women are voting in higher numbers, they are less likely to publicly voice their political beliefs. This claim was supported by their finding that men, while not necessarily stronger in their political convictions, were more likely to openly express their political attitudes in questionnaires, on both closed- and open-ended items. Women had lower levels of political communication (for instance, regarding political parties or candidates they liked or disliked) and were more likely to choose "don't know" as a response (Atkeson & Rapoport, 2003).

Based on the findings in the literature, we cannot hypothesize as to the stability of attitudes and behavior by gender, but we can hypothesize that gender differences will exist on the examined items. Specifically, we anticipate that females will report higher trust than males. We also expect females to report more positive attitudes toward immigrant rights and higher intentions to vote, in accordance with current trends.

Country differences

The IEA CivEd study contains various items that assess respondents' trust in political institutions. Although it is generally assumed that such trust is essential for maintaining the legitimacy of democracy (Easton, 1965), some authors have expressed concern about a decline in political trust in various western societies (e.g., Putnam, 2000). Comparative research has shown distinct patterns of trust, with high trust levels

in Scandinavian countries, lower levels in Western European countries, and the lowest levels in Central and Eastern Europe (Delhey & Newton, 2005). Despite the fact that most countries in Central and Eastern Europe embarked on their transition to liberal democracy more than a decade ago, figures on political trust remain very low in comparison to Western Europe and Scandinavia (Mishler & Rose, 2001, 2002). Liberal attitudes, such as support for gender equality or the rights of ethnic minorities, tend to follow the same stable pattern. Liberal values are most prevalent in the Scandinavian countries, followed by Western European countries with a Protestant tradition. The European countries with a Catholic tradition tend to be less liberal, with the least support for liberal values evident in the countries of Central and Eastern Europe (Welzel, Inglehart, & Klingemann, 2003).

Differences among countries are quite large and persistent over time. Given this information, as well as Uslaner's (2002) assertion that trust is a very stable attitude, we assume that such differences will also be found among adolescents. In addition to the expectation that trust in governmental institutions will be stable, we anticipate that the trust levels of adolescents in Scandinavian countries (Denmark, Norway, and Sweden) will be significantly higher than in countries with traditionally lower levels of trust, namely Eastern European countries (Czech Republic, Poland, and Slovenia). Since patterns are quite distinct and appear to be stable across generations (Franklin, 2004), we follow the same logic with regard to the spread of liberal values and patterns of voter turnout. We expect adolescents in Scandinavian countries to report more liberal attitudes and higher intentions to vote, and adolescents from countries in Eastern Europe to show the opposite. We do not have specific hypotheses for Switzerland or Portugal, as they cannot be classified as either Scandinavian or Eastern European.

Method

To compare attitudes and behaviors across adolescence and early adulthood, it is necessary to employ three datasets—two from the IEA Civic Education Study and one from the European Social Survey. In 1999, the IEA study examined civic knowledge and engagement of approximately 90,000 14-year-olds in 28 countries (Torney-Purta, Lehmann, Oswald, & Schulz, 2001).

In 2000, a similar study was conducted in which nearly 50,000 upper secondary students in 16 countries were assessed on similar outcomes (Amadeo, Torney-Purta, Lehmann, Husfeldt, & Nikolova, 2002). The upper secondary students had a mean age of 17.9 years and are therefore referred to in this paper as 18-year-olds. Questionnaire items pertaining to civic concepts and attitudes are the focus of analysis because, in contrast to the civic knowledge items, these items were identical for both groups of students.

The European Social Survey (ESS) is a major comparative survey, conducted in 20 countries with approximately 42,000 respondents. The ESS is supported by the European Science Foundation and adheres to rigorous methodological norms (European Social Survey, n. d.). As such, it can be considered the most reliable measurement of political attitudes available for European adults. The data used here were collected in 2002. To make the comparison as close as possible, we limited the analysis to ESS respondents between 18 and 30 years of age. Our assumption here is that these young adults were likely to share the same political culture as the adolescents involved in the CivEd studies.

Given the comparative nature of our project, we can only include information on the countries included in both the CivEd studies and the ESS. Only eight countries took part in all three of these data-gathering efforts: the Czech Republic, Denmark, Norway, Poland, Portugal, Slovenia, Sweden, and Switzerland. Limiting the analysis to these eight countries allows us to look at data from 26,018 14-year-olds (IEA 1999), 22,761 18-year-olds (IEA 2000), and 2,652 18- to 30-year-olds (ESS 2002). For clarification of explanation, we refer here to 14-year-olds as adolescents, 18-year-olds as late adolescents, and 18- to 30-year-olds as young adults.

In addition to looking at stability within countries, we will look at gender differences and compare across countries to see if patterns exist for specified country groups. Specifically, six of the countries will be categorized into a Scandinavian group (Denmark, Norway, and Sweden) and an Eastern European group (Czech Republic, Poland, and Slovenia). The dependent variables examined will be adolescent, late adolescent, and young adult reports of political trust, positive attitudes toward immigrants' rights, and voting behavior.

Results

Political trust

In the IEA CivEd surveys, trust in governmental institutions was assessed with the question, "How much of the time can you trust each of the following institutions?" Respondents were instructed to answer on a four-point Likert scale (1 signified "never"; 4 signified "always"). We constructed a political trust scale on their responses for the following institutions: national/federal government, local government, courts, police, political parties, national parliament/Congress, and the United Nations. The composite scale was reliable for both 14-year-olds (Cronbach $\alpha = .795$) and 18-year-olds (Cronbach $\alpha = .814$). The reliability of this scale is rather remarkable, given that respondents of this age usually do not have personal experience with any of these government institutions. A likely explanation is that adolescents' perceptions and evaluations of such institutions are based on political conversations with adults, particularly their parents and teachers (Torney-Purta, Barber, & Richardson, 2004).

The ESS questionnaire inquired into political trust with a comparable question: "Please tell me how much you personally trust each of these institutions." Participants responded on a scale of 0 to 10, with 0 indicating a complete lack of trust and 10 indicating complete trust. Responses were examined for the following institutions: country's parliament, legal system, police, politicians, European parliament, and the United Nations.² The composite ESS scale was also found to be reliable with a Cronbach $\alpha = .855$. Since the IEA scale had seven items and the ESS scale had six items, both were rescaled to have an equivalent range (0 to 100).

Mean scores for the political trust scale were then examined across the three age groups. Through use of correlation coefficients, we could determine the strength of the relations between country means at the three age points, and therefore infer patterns of stability within countries. The country means with regard to political trust were very stable between 14 and 18 years ($r = .926, p < .01$) and were just as stable when we compared 18-year-olds with young adults

between 18 and 30 years of age ($r = .936, p < .01$). When comparing 14-year-olds with young adults, we found the patterns and country differences continued to show continuity ($r = .826, p < .05$). Essentially, in countries where young adults express high levels of trust in political institutions (such as the Scandinavian countries), adolescents and late adolescents also can be expected to express high levels of trust.

For females, trust was stable between adolescents and late adolescents ($r = .935, p < .01$) and late adolescents and early adults ($r = .893, p < .01$), but not between adolescents and early adults ($r = .686$). Males showed similar correlations between age groups. Trust was stable from adolescence to late adolescence ($r = .870, p < .05$) and from late adolescence to early adulthood ($r = .949, p < .01$), but not from adolescence to early adulthood ($r = .692$). Comparison of males and females showed males generally reported higher levels of trust, although the difference was significant only for the 18 to 30 age group ($F(1, 2198) = 6.86, p < .01$). Means for females, males, and total countries can be seen in Table 1, while means for countries are depicted in Figure 1.

Figure 1 depicts clear country patterns in trust levels for each age group. The patterns confirm that respondents from Denmark, Norway, and Sweden consistently reported higher trust, while Central and Eastern European respondents generally had lower trust figures. The only exception to this trend was Poland, with relatively high trust levels among 14-year-olds, and low trust levels among young adults. Overall, citizens from Scandinavian countries had more trust in political institutions than did citizens from Eastern European countries at all three examined ages. An ANOVA showed that this difference was significant at 18 years of age ($F(1, 4) = 10.88, p < .05$) and for young adults ($F(1, 4) = 67.38, p < .01$). The distinct drop in the means between 18-year-olds and young adults should not be interpreted at face value since the measuring method was different in the two surveys (four-point scale versus 11-point scale). Of more importance is the ranking between the various countries, and the finding that this ranking was quite continuous across the three age groups.

2 The fact that we also included the United Nations here, which is not a national political institution, did not affect results. The entire scale proved to be one-dimensional, indicating that trust for national political institutions and for the United Nations loads on the same factor. The fact that the United Nations loaded lower indicates that 14-year-olds clearly differentiate between institutions within the national political system and international organizations.

Table 1 and Figure 1 convey that country patterns of political trust were stable across the three data sources. Although this simple correlation should not be used to arrive at any conclusions with regard to individual stability, it is clear that distinct country patterns are already evident at the age of 14.

Positive attitudes toward immigrants' rights

The IEA CivEd studies (for both 14- and 18-year-olds) asked adolescents to respond to several questions regarding their feelings about immigrant rights in their country. We selected four of the items to construct a positive attitudes toward immigrants' rights scale. Items were selected based on their use in prior CivEd research (e.g., Amadeo et al., 2002; Torney-Purta et al., 2001)

and their association with ESS items. Students were asked to respond to the following questions on a four-point Likert scale (1 indicating strong disagreement and 4 indicating strong agreement):

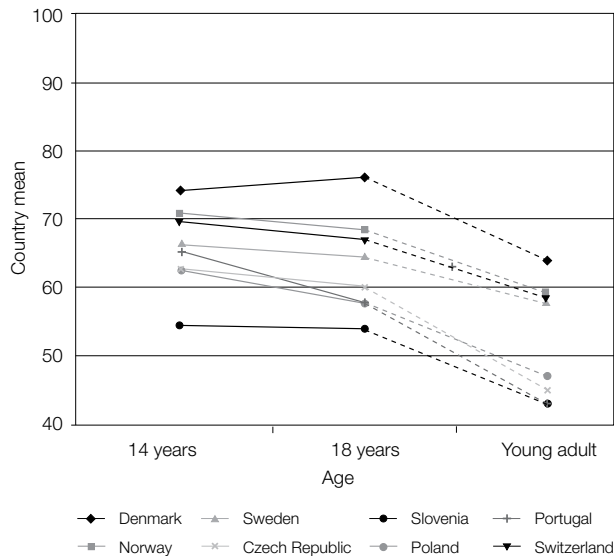
1. Immigrants should have the opportunity to keep their own language.
2. Immigrants' children should have the same opportunities for education that other children in the country have.
3. Immigrants should have the opportunity to keep their own customs and lifestyle.
4. Immigrants should have all the same rights that everyone else in a country has.

Table 1: Political Trust: Country Means across the Three Age Groups

Country	14-year-olds (IEA)	18-year-olds (IEA)	18- to 30-year-olds (ESS)
Female			
Denmark	–	–	64.09 (14.54)
Norway	71.36 (11.18)	68.95 (9.52)	58.70 (13.15)
Sweden	67.19 (12.10)	65.56 (10.44)	57.48 (15.47)
Czech Republic	63.26 (11.13)	60.51 (10.77)	43.79 (17.19)
Poland	65.44 (12.03)	58.76 (11.61)	41.83 (15.69)
Slovenia	54.33 (12.31)	53.45 (11.90)	44.11 (17.51)
Portugal	61.36 (11.05)	57.41 (8.68)	44.92 (15.19)
Switzerland	–	67.79 (10.21)	56.10 (12.53)
Mean across countries	63.73 (12.67)	60.09 (11.77)	50.74 (17.27)
Male			
Denmark	–	–	64.24 (16.31)
Norway	70.47 (13.40)	68.19 (11.50)	59.79 (14.61)
Sweden	66.43 (14.09)	64.92 (13.03)	59.51 (15.60)
Czech Republic	63.64 (12.47)	60.21 (11.81)	47.61 (17.04)
Poland	65.84 (13.78)	57.62 (12.88)	43.94 (15.14)
Slovenia	54.80 (13.61)	53.84 (13.21)	43.96 (18.42)
Portugal	63.30 (11.97)	58.77 (9.97)	46.98 (15.45)
Switzerland	–	65.37 (12.08)	58.67 (15.32)
Mean across countries	64.04 (14.05)	60.00 (13.04)	52.70 (17.73)
Total sample			
Denmark	74.04 (11.73)	76.28 (8.66)	64.16 (15.40)
Norway	70.88 (12.42)	68.56 (10.59)	59.12 (14.33)
Sweden	66.76 (13.19)	65.22 (11.83)	58.41 (15.89)
Czech Republic	63.44 (11.78)	60.36 (11.31)	45.69 (17.56)
Poland	65.65 (12.94)	58.18 (12.29)	42.71 (15.66)
Slovenia	54.56 (12.99)	53.62 (12.58)	43.06 (18.90)
Portugal	62.35 (11.57)	58.04 (9.32)	45.66 (15.57)
Switzerland	69.80 (12.07)	66.33 (11.40)	57.47 (14.12)
Mean across countries	65.51 (13.52)	61.81 (13.08)	51.46 (17.92)

Note: All entries are scored on an equivalent (0–100) measurement scale. Standard deviations are noted in parentheses.

Figure 1: Political Trust: Country Means across the Three Age Groups



A composite scale was constructed from these four items and then rescaled to a range of 0 to 100. For 14-year-olds, the scale had a Cronbach $\alpha = .814$, and for 18-year-olds a Cronbach $\alpha = .791$. The scale was therefore deemed reliable for both age groups.

Matching an ESS scale with the IEA scale proved to be quite difficult. Individual ESS items that matched in terms of content (such as language or education) did not actually correspond to what was being asked in the IEA items. For instance, the ESS asked participants their degree of agreement with the statement "Communities of people who have come to live here should be allowed to educate their children in their own separate schools if they wish." Agreement with this statement could indicate that the participant felt that immigrant children would have more academic success in a separate school and therefore should be entitled to attend such a school. However, agreement could also indicate segregationist attitudes, and therefore not a positive attitude toward immigrants. Therefore, an ESS scale that best measured an overall attitude toward immigrant rights in the country was constructed from the following items:

1. People who have come to live here should be given the same rights as everyone else.
2. Would you say it is generally bad or good for [country]'s economy that people come to live here from other countries?

3. Would you say that [country]'s cultural life is generally undermined or enriched by people coming to live here from other countries?
4. Is [country] made a worse or a better place to live by people coming to live here from other countries?
5. It is better for a country if almost everyone shares the same customs and traditions.
6. If a country wants to reduce tensions, it should stop immigration.

Item 1 was responded to on a five-point Likert scale, with 1 indicating strong agreement and 5 indicating strong disagreement. Since the high score was perceived to be a negative response, this item was reverse coded. Items 2, 3, and 4 were responded to on a 10-point Likert scale with 0 indicating a negative response (e.g., country is a worse place to live) and 10 indicating a positive response (e.g., country is a better place to live). Responses were recoded in order to have a five-point response scale instead of a 10. Items 5 and 6 were responded to with a five-point Likert scale, with 1 indicating that the respondents strongly agreed with the statement and 5 indicating that they strongly disagreed. Since strong disagreement with these statements indicated more positive attitudes toward immigrants' rights (and thus higher scores), both items used the original coding. After the relevant items had been recoded, the composite was rescaled to have range of 0 to 100. The ESS positive attitudes toward immigrant rights scale had a Cronbach $\alpha = .781$ and was therefore deemed reliable.

After establishing reliable scales for both the IEA and ESS data, we compared means across the three age groups. Overall, the means for the 14-year-olds were significantly related to the means for the 18-year-old ($r = .823, p < .05$), but were not correlated with the means for the young adult. Similarly, the means for 18-year-olds and for young adults were not correlated.

A similar pattern was apparent for females and males, with significant correlations between 14- and 18-year-old females ($r = .938, p < .01$) and males ($r = .962, p < .01$). However, when comparing males with females, we see significant differences across all three age groups. Females consistently reported higher positive attitudes toward immigrants' rights at 14 years of age ($F(1, 16,678) = 733.31, p < .001$), 18 years of age ($F(1, 17,477) = 690.42, p < .001$), and 18 to 30 years of age ($F(1, 2327) = 4.59, p < .05$). Table 2

Table 2: Immigrants' Rights: Country Means across the Three Age Groups

Country	14-year-olds (IEA)	18-year-olds (IEA)	18- to 30-year-olds (ESS)
Female			
Denmark	–	–	63.43 (14.54)
Norway	83.93 (13.82)	79.46 (13.45)	62.53 (12.35)
Sweden	87.09 (14.43)	87.12 (12.21)	71.13 (14.14)
Czech Republic	80.31 (11.09)	76.99 (10.58)	53.28 (14.32)
Poland	84.26 (12.66)	84.75 (12.76)	57.31 (10.96)
Slovenia	75.97 (13.72)	72.58 (13.32)	53.79 (11.29)
Portugal	81.39 (10.80)	80.97 (10.00)	54.21 (13.99)
Switzerland	–	70.62 (15.16)	60.60 (13.66)
Mean across countries	82.16 (13.16)	79.75 (13.35)	59.87 (14.06)
Male			
Denmark	–	–	59.60 (15.22)
Norway	73.69 (20.00)	70.57 (18.52)	57.79 (15.91)
Sweden	79.38 (19.74)	78.23 (18.07)	67.88 (13.54)
Czech Republic	74.92 (14.11)	71.93 (13.83)	52.51 (12.25)
Poland	78.02 (16.40)	78.70 (16.44)	56.42 (12.87)
Slovenia	68.91 (16.10)	67.75 (15.56)	53.72 (13.10)
Portugal	79.96 (12.15)	79.01 (11.64)	55.42 (13.73)
Switzerland	–	66.52 (19.07)	62.27 (12.00)
Mean across countries	75.81 (16.98)	73.77 (16.67)	58.60 (14.51)
Total sample			
Denmark	71.27 (18.59)	76.45 (14.60)	61.57 (14.97)
Norway	78.86 (17.90)	75.20 (16.68)	59.98 (14.55)
Sweden	83.27 (17.67)	83.13 (15.82)	69.36 (13.89)
Czech Republic	77.79 (12.87)	74.47 (12.56)	52.89 (13.26)
Poland	81.30 (14.88)	81.89 (14.92)	56.86 (11.96)
Slovenia	72.39 (15.33)	70.29 (14.62)	53.75 (12.25)
Portugal	80.69 (11.48)	80.13 (10.79)	54.74 (13.86)
Switzerland	72.30 (19.31)	68.37 (17.62)	61.43 (12.86)
Mean across countries	77.41 (16.61)	76.89 (15.25)	59.22 (14.30)

Note: All entries are scored on an equivalent 0–100 measurement scale. Standard deviations are noted in parentheses.

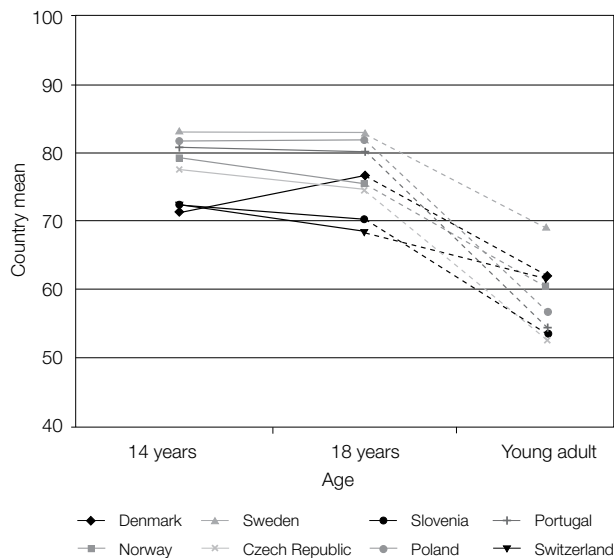
reports means for positive attitudes toward immigrants' rights by both country and gender. Country means are illustrated in Figure 2.

Examining patterns for groups of countries was more difficult because individual countries showed much variance within (e.g., Denmark 14-year-olds showed the lowest means, 18-year-olds showed the fourth highest, and young adults the second highest). However, differences between the two country groups were significant for young adults ($F(1, 4) = 8.472$, $p < .05$), with Scandinavian countries showing more positive attitudes toward immigrant rights.

The figures regarding attitudes toward immigrant rights are much less conclusive than are those pertaining to political trust. Within the IEA studies, we observe

strong correlations between 14-year-olds and 18-year-olds, which hint at continuity across adolescence. The finding of a non-significant correlation between the adolescents and the young adults could indicate many things. There may be an actual difference that can be attributed to the adults' increase in knowledge or experience related to immigrant issues, including both positive and negative encounters and exposure. It is possible that the discontinuity is indicative of a measurement issue. Although both the IEA and ESS scales reliably assessed one factor, there is no way to know that it was actually the same factor. Upon further investigation, it appears that the IEA specifically measured attitudes toward immigrants' rights, while the ESS measured attitudes toward immigrants.

Figure 2: Immigrants' Rights: Country Means across the Three Age Groups



Though similar in subject matter, these attitudes may be distinct from each other.

Expected and actual voting behavior

To assess student intentions for civic participation, the IEA questionnaire asked this question: "When you are an adult, what do you expect that you will do?" Several political actions were listed, but unfortunately many of them did not correspond to the ESS, and it was therefore not possible to construct an entire scale. However, the item "vote in national elections" had a clear counterpart. Students responded to this item on a four-point Likert scale (1 signifying that they would certainly not vote and 4 signifying that they would certainly vote). Responses were standardized to 100.

In the ESS, young adults were asked, "Did you vote in the last national election in [date]?" Respondents answered yes, no, or not eligible. "Not eligible" respondents were excluded from analyses. To enhance comparisons, responses were also standardized to 100. Again, seemingly large drops in young adults' voting behavior should not be interpreted at face value because of the difference in measurement.

Mean scores for voting were then examined across the three age groups. Significant correlations existed between 14- and 18-year-olds ($r = .922, p < .01$) and between 18- to 30-year-olds ($r = .707, p < .05$), but not between 14-year-olds and young adults. Females reported slightly higher voting intentions and actual behavior than males. The differences were significant

for adolescents at 14 years of age ($F(1, 18,115) = 27.01, p < .001$) and 18 years of age ($F(1, 18,979) = 13.62, p < .001$). There were no significant differences between Scandinavian and Eastern European countries. Means of intended and actual voting behavior by country and gender are reported in Table 3 and illustrated in Figure 3.

As is evident in Figure 3, the results regarding continuity of voting intentions at ages 14 and 18 with actual voting behavior as a young adult are not conclusive. Switzerland was a clear outlier, with low voting intentions and equally low turnout. However, all other countries showed no consistent pattern and therefore no clear ranking order. For instance, while Swedish 14-year-olds reported relatively low voting intentions, the young adults had among the highest turnout figures. Again, there is little evidence for stability between the three age groups, except in Switzerland, which is a country with a very distinct political culture (e.g., holding frequent elections). However, it cannot be concluded that these figures indicate genuine instability within countries because, again, we are dealing with different items. This time the wording was not what distinguished the IEA question from the ESS question. Rather, the difference exists in the aspect of voting being measured. The IEA measured adolescents' intended behavior while the ESS measured actual behavior. We also examined voting with a single item rather than a solid and reliable measurement scale.

Reliability of scales

Our results indicated that stability across the three age groups, and thus the three data sources, could only be established when nearly identical measurements were utilized. The attitude toward immigrants' rights scale and the voting intention item did not predict young adults' attitudes and behaviors in the same country, which may be attributed to inadequate measurement equivalence. Therefore, the political trust scale became the focus of further investigation into internal cohesion across the three age groups. A factor analysis was conducted to explore whether 14- year-olds, 18-year-olds, and young adults reporting on the political trust items were actually referring to the same latent variable. Results are reported in Table 4.

The factor analysis indicated that we were indeed measuring a one-dimensional attitude. For all 24

Table 3: Voting Behavior: Country Means across the Three Age Groups

Country	14-year-olds (IEA)	18-year-olds (IEA)	18- to 30-year-olds (ESS)
Female			
Denmark	–	–	89.14 (26.49)
Norway	82.46 (19.67)	88.01 (16.24)	73.09 (36.08)
Sweden	75.81 (22.69)	87.30 (17.40)	87.50 (28.04)
Czech Republic	72.23 (22.37)	77.13 (21.60)	67.35 (37.41)
Poland	85.79 (17.74)	89.38 (16.68)	64.34 (37.53)
Slovenia	79.46 (19.05)	80.56 (19.38)	77.35 (34.55)
Portugal	79.09 (19.20)	86.41 (16.59)	66.49 (37.42)
Switzerland	–	50.44 (25.04)	54.21 (36.77)
Mean across countries	79.14 (20.66)	82.87 (20.55)	73.05 (36.00)
Male			
Denmark	–	–	90.30 (25.27)
Norway	81.56 (22.58)	87.63 (17.89)	74.73 (35.54)
Sweden	71.82 (25.54)	85.23 (20.92)	83.55 (31.12)
Czech Republic	74.25 (23.38)	77.88 (21.97)	69.83 (36.99)
Poland	80.71 (23.37)	86.92 (19.80)	64.00 (37.55)
Slovenia	75.89 (22.61)	80.65 (21.15)	73.18 (36.06)
Portugal	79.89 (19.96)	84.73 (17.78)	70.43 (36.83)
Switzerland	–	57.80 (27.21)	56.25 (37.17)
Mean across countries	77.44 (23.19)	81.73 (22.02)	73.02 (36.01)
Total sample			
Denmark	81.98 (18.19)	91.83 (13.90)	89.66 (25.91)
Norway	82.02 (21.15)	87.84 (17.03)	73.95 (35.76)
Sweden	73.83 (24.26)	86.25 (19.41)	85.34 (29.79)
Czech Republic	73.17 (22.86)	77.50 (21.78)	68.79 (37.08)
Poland	83.39 (20.75)	88.22 (18.26)	64.17 (37.50)
Slovenia	77.61 (21.02)	80.60 (20.24)	75.16 (35.35)
Portugal	79.43 (19.57)	85.67 (17.16)	68.16 (37.15)
Switzerland	63.11 (25.63)	54.33 (26.46)	55.24 (36.89)
Mean across countries	77.08 (22.51)	83.45 (20.80)	73.05 (35.99)

Note: All entries are scored on an equivalent (0–100) measurement scale. Standard deviations are noted in parentheses.

observations in Table 4, one clear factor emerged, which explained an average of 47% of the variance. This is quite remarkable, given that 14-year-olds typically have not had direct experience with these institutions. Additionally, despite the fact that differences existed between the factor loads, the structures generally remained the same. For instance, the national parliament generally had the highest factor loading, confirming that respondents saw parliaments as the most important institution with regard to political legitimacy and representation.

The most apparent difference between the age groups was that the police and the courts usually received low factor loadings for the late adolescents, while these tended to be higher for the adults. The factor analysis therefore suggests it is reasonable to ask

14-year-olds about their trust in political institutions. Their answers followed the same pattern and structure as the answers of young adults in their respective countries.

Discussion

In this paper, we have compared the aggregate level scores for a number of political attitudes and behaviors in the IEA 1999 and 2000 CivEd studies, and in the ESS 2002. Our goal was to use this comparison to shed new light on the enduring question of the stability or continuity of attitudes and behavior. This comparison turned out to be more difficult than initially expected because of inadequate measurement equivalence across data sources. However, the items pertaining to trust in political institutions were nearly identical, and

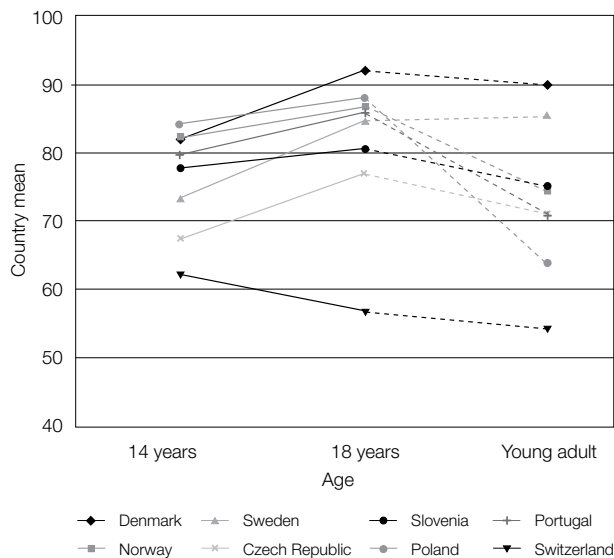
Table 4: Factor Analysis Results for Political Trust Scale

Item	Denmark			Norway			Sweden		
	IEA99	IEA00	ESS02	IEA99	IEA00	ESS02	IEA99	IEA00	ESS02
Natl. govt.	.75	.71	–	.75	.72	–	.75	.79	–
Local govt.	.65	.65	–	.64	.67	–	.71	.70	–
Courts	.63	.60	.78	.67	.70	.80	.62	.68	.82
Police	.58	.56	.65	.58	.63	.72	.59	.63	.70
Politicians	–	–	.80	–	–	.77	–	–	.81
Parties	.65	.54	–	.63	.55	–	.71	.67	–
UN	.63	.63	.64	.57	.51	.60	.51	.58	.65
Parliament	.74	.72	.80	.75	.75	.79	.77	.80	.81
Eigenvalue	3.08	2.80	2.73	3.04	2.99	2.73	3.16	3.40	2.90
Variance (%)	44.0	40.3	54.6	43.4	43.7	54.5	45.2	48.5	57.9
Item	Czech Republic			Poland			Slovenia		
	IEA99	IEA00	ESS02	IEA99	IEA00	ESS02	IEA99	IEA00	ESS02
Natl. govt.	.65	.72	–	.72	.74	–	.69	.72	–
Local govt.	.60	.56	–	.59	.63	–	.71	.66	–
Courts	.65	.63	.80	.69	.67	.80	.67	.62	.82
Police	.59	.62	.72	.58	.62	.70	.58	.57	.72
Politicians	–	–	.81	–	–	.79	–	–	.79
Parties	.71	.69	–	.61	.66	–	.66	.64	–
UN	.53	.53	.60	.50	.52	.63	.55	.56	.73
Parliament	.77	.80	.84	.75	.77	.78	.70	.74	.83
Eigenvalue	2.93	3.00	2.89	2.87	3.10	2.76	2.99	2.94	3.04
Variance (%)	41.8	42.8	57.8	44.0	43.8	55.1	42.7	42.0	60.7
Item	Portugal			Switzerland					
	IEA99	IEA00	ESS02	IEA99	IEA00	ESS02			
Natl. govt.	.72	.75	–	.71	.78	–			
Local govt.	.65	.68	–	.63	.57	–			
Courts	.55	.53	.81	.60	.65	.81			
Police	.48	.48	.71	.59	.60	.68			
Politicians	–	–	.81	–	–	.78			
Parties	.59	.63	–	.59	.51	–			
UN	.55	.52	.67	.55	.55	.62			
Parliament	.71	.71	.82	.70	.74	.83			
Eigenvalue	2.64	2.60	2.90	2.73	2.82	2.84			
Variance (%)	37.7	37.1	58.1	39.0	40.3	56.9			

Notes: Factor loadings are for 14- and 18-year-olds (from IEA99 and IEA00) and young adults (from ESS02).

– Indicates not administered.

Figure 3: Voting Behavior: Country Means across the Three Age Groups



strong continuity was found from early adolescence to late adolescence to adulthood. This finding supports Bandura's (1986) assertion that attitudes are relatively stable and indicates that distinct political cultures are well established at the age of 14. Furthermore, gender differences were slight and the structure of the latent attitude seemed to be equivalent. We therefore conclude that political trust is a stable attitude at the aggregate level, which strengthens Uslaner's (2002) argument that trust is stable, particularly at the country level. Longitudinal panel studies can help elucidate whether this argument is also valid at the individual level.

The evidence regarding other civic attitudes and behaviors is mixed, and it remains unclear whether certain attitudes are more stable than others. In the literature on political socialization and political attitudes, trust is often portrayed as the most stable attitude. In the current study, the inconclusive findings regarding positive attitudes toward immigrants' rights are likely attributable to measurement issues, specifically that the two scales are not comparable. However, the discontinuity between intended and actual voting behavior cannot be as easily explained. Within nations, there appears to be a noticeable difference between adolescents' intentions to vote and adults' actual voting behavior. Possible explanations for this discontinuity include political and economic conditions of which adults may be more cognizant or complicated voter registration processes. The finding

does support some theorists' expectation that attitude and behavioral changes are indicative of development and new cognitive structures. Regardless of the explanation for the finding, one should not conclude that questions pertaining to voting intentions should be avoided, as it is quite possibly a meaningful measurement of a sincere intention of the respondent. More research is needed on the stability of intended and actual behaviors, both of individuals and age groups within nations.

In summary, the congruence of political trust across the three examined age groups indeed strengthens the status of IEA findings and demonstrates the usefulness of questioning 14-year-olds on their civic attitudes. If patterns between countries persist into adulthood, then we can assume that early experiences, in the home and school, influence civic attitudes later in life. The educational implication of this finding is that adolescents are gaining a solid foundational knowledge regarding governmental institutions, and are forming their beliefs and attitudes at a young age. It is often expected that civic education and other elements of the school curriculum or climate influence students' civic engagement (Galston, 2001; Niemi & Junn, 1998; Torney-Purta, Barber, & Richardson, 2004).

Education systems have also been found to have an effect on students' voting intentions, although findings from the current study cannot contribute to such literature. Perhaps 14- and 18-year-olds do indeed prefer to give socially desirable answers even though they do not intend to vote. Or perhaps more applied civic experiences are needed in schools, such as mock elections, examination of prototypes of candidates' materials, and student governments, to help students translate their behavioral intentions into actions as adults. Certainly, schools cannot change the political and economic climates of their societies, but they can do their best to effectively prepare students for active civic engagement upon adulthood.

References

- Alwin, D. E., & Krosnick, J. A. (1991). Aging, cohorts, and the stability of sociopolitical orientations over the life span. *American Journal of Sociology*, 97, 169–195.
- Amadeo, J. A., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement: An IEA study of upper secondary students in sixteen countries*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Atkeson, L. R., & Rapoport, R. B. (2003). The more things change the more they stay the same: Examining gender differences in political attitude expression, 1952–2000. *Public Opinion Quarterly*, 67, 495–521.
- Bandura, A. (1986). *Social foundations of thoughts and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Batista-Foguet, J. M., & Saris, W. E. (1997). Tests of stability in attitude research. *Quality & Quantity*, 31, 269–285.
- Blais, A. (2006). What affects voter turnout? *Annual Review of Political Science*, 9, 111–125.
- Delhey, J., & Newton, K. (2005). Predicting cross-national levels of social trust. Global pattern or Nordic exceptionalism? *European Sociological Review*, 21(4), 311–327.
- Denver, D., & Hands, G. (1990). Does studying politics make a difference? The political knowledge, attitudes and perceptions of school students. *British Journal of Political Science*, 20(2), 263–279.
- Easton, D. (1965). *A systems analysis of political life*. New York: Wiley.
- Erikson, E. (1968). *Identity: Youth and crisis*. New York: W. W. Norton.
- European Social Survey. (n. d.). Retrieved June 23, 2006 from www.europeansocialsurvey.org
- Franklin, M. (2004). *Voter turnout and the dynamics of electoral competition in established democracies since 1945*. Cambridge: Cambridge University Press.
- Galston, W. A. (2001). Political knowledge, political engagement, and civic education. *Annual Review of Political Science*, 4, 217–234.
- Granberg, D., & Holmberg, S. (1990). The intention-behavior relationship among U.S. and Swedish voters. *Social Psychology Quarterly*, 53, 44–54.
- Hooghe, M., & Stolle, D. (2004). Good girls go to the polling booth, bad boys go everywhere: Gender differences in anticipated political participation among American fourteen-year-olds. *Women & Politics*, 26, 1–23.
- Jamieson, A., Shin, H. B., & Day, J. (2002). Voting and registration in the election of November 2000. *Current population reports* (pp. 20–542). Washington, DC: US Census Bureau.
- Jennings, M. K., & Stoker, L. (2004). Social trust and civic engagement across time and generations. *Acta Politica*, 39(4), 342–379.
- Johnson, I. (2005). Political trust in societies under transformation. *International Journal of Sociology*, 35, 63–84.
- Lao, J., & Kuhn, D. (2002). Cognitive engagement and attitude change. *Cognitive Development*, 17, 1203–1217.
- Metz, E., McLellan, J., & Youniss, J. (2003). Types of voluntary service and adolescents' civic development. *Journal of Adolescent Research*, 18(2), 188–203.
- Mishler, W., & Rose, R. (2001). Political support for incomplete democracies: Realist vs. idealist theories and measures. *International Political Science Review*, 22(4), 303–320.
- Mishler, W., & Rose, R. (2002). Learning and re-learning regime support: The dynamics of post-communist regimes. *European Journal of Political Research*, 41(1), 5–35.
- Niemi, R., & Junn, J. (1998). *What makes students learn?* New Haven: Yale University Press.
- Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster.
- Sapiro, V. (2004). Not your parents' political socialization: Introduction for a new generation. *Annual Review of Political Science*, 7, 1–23.
- Sibley, C. G., Liu, J. H., & Kirkwood, S. (2006). Toward a social representations theory of attitude change: The effect of message framing on general and specific attitudes toward equality and entitlement. *New Zealand Journal of Psychology*, 35, 3–13.
- Sotelo, M. J. (1999). Gender differences in political tolerance among adolescents. *Journal of Gender Studies*, 8, 211–217.
- Studlar, D. T., McAllister, I., & Hayes, B. C. (1998). Explaining the gender gap in voting: A cross-national analysis. *Social Science Quarterly*, 79, 779–798.
- Torney-Purta, J., & Amadeo, J.-A. (2003). A cross-national analysis of political and civic involvement among adolescents. *Political Science and Politics*, 36(2), 269–274.
- Torney-Purta, J., Barber, C. H., & Richardson, W. K. (2004). Trust in government-related institutions and political engagement among adolescents in six countries. *Acta Politica*, 39(4), 380–406.
- Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Uslaner, E. (2002). *The moral foundations of trust*. Cambridge: Cambridge University Press.

Verba, S., Schlozman, K. L., & Brady, H. E. (1995). *Voice and equality: Civic volunteerism in American politics*. Cambridge, MA: Harvard University Press.

Vollebergh, W. A. M., Iedema, J., & Raaijmakers, Q. A. W. (2001). Intergenerational transmission and the formation of cultural orientations in adolescence and young adulthood. *Journal of Marriage and Family*, 63, 1185–1198.

Welzel, C., Inglehart, R., & Klingemann, H. D. (2003). The theory of human development: A cross-cultural analysis. *European Journal of Political Research*, 42(2), 341–380.

School climate for citizenship education: A comparison of England and the United States¹

Gary Homana and Carolyn Barber
University of Maryland
College Park, Maryland, USA

Introduction

Historically, schools have served multiple purposes in society. While schools have played a critical role in the development of academic abilities of young people, they have also served as places that assist students in developing an understanding of society and commitment to political and civic engagement. This process is sometimes referred to as the civic education or civic learning of students, but it is advantageous to frame these concepts broadly in ways that go beyond courses labeled “civics.” Traditionally, citizenship education has focused on transmission of civic knowledge. However, recent work has advanced a broader notion of citizenship education (see Homana, Barber, & Torney-Purta, 2006; Kerr, 2005). Renewed interest in citizenship education has been fueled by the release of several influential publications, including two reports associated with the International Association for the Evaluation of Educational Achievement’s (IEA) Civic Education Study (CivEd) (Torney-Purta, Lehmann, Oswald, & Schultz, 2001; Torney-Purta, Schwille, & Amadeo, 1999) and one featuring the final report of the advisory group on citizenship in England (Crick, 1998).

At the same time, positive school climate is crucial for the development of competent and capable citizens. Along with the creation of pathways for academic development, school and classroom climates have the potential to assist in citizenship development. In this role, schools can help foster the knowledge, skills, and dispositions that young people need to develop into politically aware and socially responsible individuals (Torney-Purta & Vermeer, 2004). More recent work on school climate (Homana et al., 2006) has conceptualized and measured the connection between school climate and citizenship education.

The purpose of this paper is to examine the relationship between school climate and citizenship

education in England and the United States using items from the 1999 IEA Civic Education Study (Torney-Purta et al., 2001) to understand how both school characteristics and students’ school experiences can foster civic knowledge, skills, and dispositions in 14-year-olds in England and the United States. While some work has been done examining the relationship of student civic outcomes to an open classroom climate (Campbell, 2004; Hahn, 1999; Kerr, Ireland, Lopes, Craig, & Cleaver, 2004; Torney-Purta, 2002; Torney-Purta et al., 2001), this study addresses an expanded conceptualization of school climate and civic learning based on key dimensions from the School Citizenship Education Climate Assessment (Homana et al., 2006). Given the central role that items in the IEA CivEd played in formulating the School Citizenship Education Climate Assessment, an analysis of these dimensions using IEA data is a natural extension of previous school climate work. From a research perspective, the study provides a fruitful endeavor because it can lead to understanding the similarities and differences in both countries, as well as improve understanding of how researchers, practitioners, and policymakers can enhance students’ academic, political, social, and civic responsibilities.

England and the United States: Citizenship education policy reforms and challenges

Both England and the United States have approached citizenship education by implementing policies at various levels of government, albeit through different ways. For example, in a sweeping move, England enacted a comprehensive compulsory national curriculum (see Qualifications and Curriculum Authority, 2006) for secondary students, which began in 2002 and was based on the Crick Report (1998). Three key aspects of this national curriculum focus

¹ The comments and assistance of Judith Torney-Purta are gratefully acknowledged.

on citizenship education: progress toward becoming informed citizens; skills of inquiry and communication; and participation and responsible action (Nelson & Kerr, 2005). In the United States, although citizenship education is not a national policy (and despite, as some argue, the No Child Left Behind initiative substantially reduced the coverage of citizenship education in social studies classes), 41 states currently have statutes that address acquisition of citizenship knowledge and skills (Education Commission of the States, 2006). However, citizenship education often occurs in limited school and classroom climates.

In many ways, similar challenges exist for creating a positive school climate for citizenship education in both England and the United States. One challenge, for example, concerns ensuring opportunities for meaningful learning of civic-related knowledge that builds on and enhances academic and participation skills. Although England provides guidance to its citizenship curriculum, recent evaluation of citizenship education in the country suggests that “in order for active citizenship to be developed effectively, programmes need to have a clear learning aim, purpose and intention” (Nelson & Kerr, 2005, p. 15). Furthermore, the Office for Standards in Education (Ofsted) inspections in England (2005) revealed that citizenship education tends to be unsatisfactory when it (i) is not identified with subjects that can provide the greatest citizenship opportunities; (ii) lacks expertise to support subject development; and (iii) consists of inequality of opportunities among students. These limitations manifest themselves in terms of limited opportunities for students to think critically and deliberately about civic-related issues. At the time of IEA CivEd, open discussion and deliberation of social and political topics in the English classroom context was rare. Evidence collected for a “case study” of the state of civic education in England indicated that teachers avoided discussion of politics with their students for fear of transmitting “bias and indoctrination,” which could influence both how often teachers engage in discussions with students and how open teachers allow their classes to be for such discussion (Kerr, 1999). These findings have also more recently been supported in England by Cleaver, Ireland, Kerr, and Lopes (2004), who found frequent use of traditional teaching and learning in schools and moderate opportunities for more active student activities. Such discussion,

however, could have potential benefits for English students’ civic development. As Kerr and colleagues reported when summarizing England’s results from CivEd (Kerr, Lines, Blenkinsop, & Schagen, 2002), students who discuss political and social issues with others begin to think about how the events brought up relate back to what they have been taught in class.

In the United States, empirical evidence similarly indicates that teachers, more often than not, continue to utilize traditional teaching methods such as lecturing and worksheets, and the effect on student civic knowledge and engagement (civic outcomes) is noticeable. For example, in a study of 135 Grade 8 through Grade 10 social studies classrooms in Chicago, Kahne, Rodriguez, Smith, and Thiede (2000) found that a typical social studies classroom primarily engaged students in lower-level thinking, provided a thin and fragmented knowledge base, offered few substantive opportunities to experience democracy, rarely linked content to understanding and respecting diversity, and provided limited opportunities to examine and respond to social problems. The problem was made worse when teachers were required to teach to state civics goals. However, they also found that “when teachers provided students with more and varied opportunities to develop as citizens, they simultaneously provided significantly more opportunities for higher-order thinking and deep inquiry” (p. 311). Given these types of issues, it is important and interesting for researchers and policymakers to consider how changes in policy will change school climate toward citizenship education.

The study of England and the United States is also important in considering the climate of citizenship education in an international context. CivEd (Torney-Purta et al., 2001) found that despite the effectiveness of an open and participatory climate in promoting civic knowledge and engagement, the approach is not the norm in most countries. While about one-quarter of the students indicated they were often encouraged to voice opinions during discussion in their classrooms, an equal number said this rarely or never occurred. Teachers confirmed these findings: teacher-centered methods, such as the use of textbooks, recitation, and worksheets, dominate in civic-related classrooms in most countries, although there are some opportunities for discussion of issues (Torney-Purta, 2002). At the same time, some students lack these experiences

more than do others. Further analysis of the United States sample of students found this proportion was even lower among immigrant and Hispanic students (Torney-Purta, Barber, & Wilkenfeld, 2006).

Concerns for citizenship education in both England and the United States existed at the time of CivEd, and still exist today. The findings of this paper contribute to understanding the relationship between school climate and citizenship education and provide a foundation for future research. We begin by considering key dimensions of a positive school climate (Homana et al., 2006) and their expected implications for citizenship education among adolescents in England and the United States.

Theoretical framework for school climate for citizenship education

This theoretical framework was originally developed as part of a broader school climate for citizenship education assessment (Homana et al., 2006) for the National Center on Civic Learning and Engagement at the Education Commission of the States. The framework outlines key dimensions related to an education for responsible citizenship, serves to deepen understanding of the characteristics and their purpose, and establishes a basis for the assessment items used in this study. The framework was developed through a process that utilized a variety of school climate and culture literature.

Dimension 1: Official recognition and community acceptance of the civic purpose of education

A positive school climate for citizenship requires a substantial agreement among members of the school community on a philosophy of education committed to the goals and objectives of a common civic purpose. This common understanding provides a reflection of the collective will of the school community through strong instructional content and pedagogy that supports the development of civic learning and promotes understanding and commitment to political and civic engagement. As a result, positive school climate can create an environment across the school that fosters the well-being, academic achievement, and civic development of students. Underlying this relationship are an explicit school mission statement and supportive policies that promote conscious commitment to citizenship education. At the same

time, policy impediments are identified and reduced so citizenship education is successfully sustained. England provides an example of how schools can be supported to meet these goals. As part of its national civic work, England developed a framework for inspection to help guide schools regarding standards, made available citizenship education resources, and provided professional opportunities for development of communities of practice in citizenship through teacher-training courses to help schools reach these goals (Kerr, 2003).

Dimension 2: Meaningful learning of civic-related knowledge

Classroom instruction that explicitly focuses on meaningful civic content is a critical element for students' citizenship and enhanced learning. Instructional methods and approaches that foster civic-related knowledge engage students in activities that promote a range of academic competencies. Using data from the United States sample of the 1999 IEA Citizenship Education Study, Torney-Purta and Richardson (2003) concluded that meaningful civic knowledge builds upon students' past understanding, and that this understanding is made authentic by connections to current issues and concerns. A positive classroom climate can be promoted to enhance academic performance and cultivate development of the knowledge, skills, and dispositions necessary in a democratic society. Activities within the classroom such as debating, role-playing, and mock trials can encourage an individual's active construction of knowledge through participation in activities that are meaningful to a democratic society (Torney-Purta, 2002; Vosniadou, 2001). These activities are most successful in fostering citizenship education when they touch upon a variety of topics, including both national and international concerns (National Council for the Social Studies, 2006; Pigozzi, 2006; Qualifications and Curriculum Authority, 2006).

Dimension 3: Cooperation and collaboration in approaching civic-related learning

In a positive school climate for citizenship, members of the school community engage in cooperative and collaborative experiences that enhance and support learning and problem-solving connected to citizenship development. These cooperative activities engage students in meaningful learning tasks and problem-

solving by requiring students to co-construct new knowledge while building upon the prior experiences that each of them brings to the group (Vosniadou, 2001). At the same time, collaborative experiences can create spaces for students and faculty to collaborate as a team, share in the decision-making process, and promote meaningful educational experiences that actively engage students in their learning. Finally, a truly collaborative environment also encourages cooperation among teachers, which creates an environment of support and the creation of common goals and increased efficacy for improving students' cognitive and social development (Freiberg, 1999). As such, these schools create powerful learning environments that promote cooperative learning, group cohesion, respect, and mutual trust (Finnan, Schnepel, & Anderson, 2003; Ghaith, 2003; Ireland, Kerr, Lopes, Nelson, & Cleaver, 2006; Kerr et al., 2004)—all important characteristics for the development of a sense of school belonging.

Dimension 4: Mutual trust and positive interactions

Positive school climate for citizenship fosters a supportive environment that merits mutual trust and positive interactions among all its members. Group openness, collaboration, cooperation, and supportive atmosphere are consistent with schools' missions and goals. While quality collaboration and cooperation have positive influences on a school's citizenship education climate, these qualities are further enhanced when they occur in environments where members of the school community are respectful and trusting of one another. Cohen's (2001) work in the field of social and emotional education suggests that improved classroom and school climate can create a safe, responsive and caring environment where learning optimally takes place. Indeed, the Search Institute (2000) identifies a caring school community as one of several "external developmental assets," or characteristics of a student's environment, that encourage positive development both academically and socially. Supporting these contentions, Ireland and colleagues (2006) found that a strong sense of school community positively influenced students' trust and knowledge of people in their local communities. The authors also argue that a sense of community provides opportunities for social responsibility, awareness of the social and political environment, including cultural awareness,

and positive interactions among different groups of people.

Dimension 5: Students' input and participatory problem-solving skills are valued

McLaughlin (2004) suggests that organizations such as schools should provide opportunities for youth to engage as active learners in ways that lead to confidence in the value of participatory problem-solving. Positive school climates foster environments in which members of the school community respect, value, and promote students' abilities to shape their own learning and participate in solving school and community problems. In these schools, students feel a sense of freedom to express their ideas and respect the ideas of others, both in the classroom and through school-wide activities that bring students together to address issues related to school life. Members of schools with a positive climate ask students for their input regarding a range of issues, including school policies, and this input is taken into account by teachers and administrators. Through these experiences, students become active participants in the democratic process and meaningfully contribute to school life.

Dimension 6: Thoughtful and respectful dialogue about issues

The development of citizenship education through thoughtful and respectful deliberation and dialogue is supported by policies and practices that are student-centered, recognize and encourage civic-related professional development, and create genuine opportunities for collaboration, cooperation, and communication across the school. Torney-Purta (2002) found that civic knowledge, skills, and dispositions are fruitful in a rich classroom climate based on mutual engagement. In such an environment, trying out one's knowledge in interpersonal situations makes the learning experience more meaningful because it involves discussion with others. Schools with a positive climate support opportunities for deliberation and dialogue that cultivate the ability to listen to others' views, to acknowledge those views, and to build on them when expressing one's own ideas. In these schools, teachers encourage students to discuss political or social issues about which people have different opinions. They ensure there is enough time to talk about a topic, and they create a safe and welcoming environment where many students are comfortable expressing their

opinions (Cleaver et al., 2004; Hess, 2004) on social and controversial issues. In this way, students can develop the ability to critique, analyze, and formulate possibilities for action that are critical for responsive citizenship (Westheimer & Kahne, 2003).

Dimension 7: Engagement and interaction in service to the community

Citizenship education has an important relationship to active participation within the broader community. Beginning through involvement in student councils and other types of in-school civic participation, citizenship education continues by building relationships between the school and the community. Democratic concepts such as inequality, injustice, and an obligation to the broader community are elements that link real community issues, from the local to the international level, to co-curricular activities in the school.

A positive school climate also promotes a shared commitment to policies that encourage more and deeper collaborative interaction between the school and the community. Ongoing involvement in active student participation activities linked to the community, through experiences like high-quality service-learning, can help students identify community problems, create solutions to address those problems, and reflect on the process through multiple points of view.

In addition, beginning through in-school engagement, formal civic education serves as a catalyst to extend the understanding and application of citizenship education into the community. Democratic understandings can be successfully linked to practice through real issues, creating opportunities for students to become active and contributing members in their communities and leading to increased political and civic engagement (Cleaver et al., 2004; Homana & Greene, 2006; Ireland et al., 2006; Morgan & Streb, 1999). Beneficial to the community, students, and schools, these experiences provide opportunities for achievement and recognition and participation in meaningful peer interactions. They also help students work toward collective goals with one another and with adult members. These opportunities are all consistent with the goals of citizenship education.

Summary

This analysis addresses how a positive school citizenship education climate influences outcomes relating to each of the three civic-related competencies outlined by the Education Commission of the States: knowledge, skills, and dispositions. In particular, we examine how each of the seven dimensions of a positive citizenship education climate, as measured by a survey of teachers, students, and administrators as part of the IEA Civic Education Study, predict students' civic knowledge, their internal political efficacy, and their conceptions of what is good for democracy. This multilevel analysis addresses how both characteristics of the school itself and how students' experiences in and perceptions of school climate can foster civic competencies in 14-year-old students in two countries.

Method

In 1994, IEA began planning a study of civic education (CivEd). A case study phase from 1994 to 1998 used qualitative data to craft an instrument to measure political knowledge, attitudes (see Torney-Purta et al., 1999), and civic engagement of almost 90,000 14-year-olds in 28 countries (Torney-Purta et al., 2001). More specifically for this analysis, we were interested in measuring students in three competencies—civic knowledge, participatory skills, and dispositions toward democracy—and analyzing if these student competencies could be predicted by the seven dimensions of school citizenship education climate outlined above. CivEd is a particularly useful survey to use for these purposes. Knowledge, skills, and dispositions toward democracy were three of the competencies that CivEd explicitly listed as of interest (along with civic behaviors and social attitudes: Schulz & Sibberns, 2005). In addition, the IEA study also had an explicit emphasis on the role that schools and teachers play in civic education (Torney-Purta et al., 2001).

Sampling and variables chosen for this analysis

Nationally representative samples of students in England and the United States were selected as data sources for the current investigation. A total of 3,043 students in 145 schools from England and 2,811 students in 124 schools from the United States participated in the study. Because students are nested

within schools, multilevel modeling is an appropriate methodology for analyzing these data.

The analysis used two-level hierarchical linear (HLM) models. The first level of analysis utilized student-level data from CivEd's 14-year-old student data sets. In addition to the items and scales available in the international data sets, the study also incorporated various IRT scales developed by the Civic Education Data and Researcher Services (CEDARS) at the University of Maryland (see Husfeldt, Barber, & Torney-Purta, 2005). The dependent variables within this level included total civic knowledge, internal political efficacy, and concepts of democracy. We chose independent variables that related to the seven dimensions of school citizenship education and, in the majority of cases, were included as scales and items in the School Citizenship Education Climate Assessment available through the Education Commission of the States (indicated in Table 1).

The student-level variables corresponding to Dimension 2 were based on national option items from the United States, and were not included in the analysis of English students and schools. However, we believed it was important to consider how these items contributed to civic outcomes in the United States because these items appear in the School Citizenship Education Climate Assessment as administered by the Education Commission of the States (Homana et al., 2006). As a result, two versions of the United States analysis are presented for each outcome: one that does not include the national-option variables (so allowing us to compare the results with those from England), and one that includes the extra variables to determine how student perceptions of classroom activities contribute to the model above and beyond the other variables.

In contrast, the School Citizenship Education Climate Assessment does not include items from IEA CivEd in the first and fourth dimensions of the assessment. We did, however, include items from the teacher survey that we believed best captured these two dimensions at the school level (i.e., view of the importance of schools for civic education and racial/religious intolerance). Similarly, the scales relating to classroom activities taken from items on the teacher survey are not included in the Education Commission of the States' assessment. We included these items in

order to have a measure of meaningful learning in classrooms in both countries.

Data adequacy

It was important to choose data that included sufficient numbers of both students and schools in order to limit the possibility of the analysis suffering due to small sample size issues. To prevent this from happening and to ensure that missing cases were accounted for in this analysis, we imputed missing data for predictors at each level (using the mean for continuous variables, and the median of similar cases for dichotomous variables). Table 2 provides descriptive statistics for both the predictor and the outcome variables.

School and student variance

Data adequacy can also be observed by calculating the intraclass correlations (ICCs) and reliability estimates (λ) for each of the three outcomes of interest. The purpose of calculating these statistics is to determine the amount of variability at each level by partitioning the variance of the dependent variable (in this analysis, our three outcomes of citizenship education) into school variance (the between-group variance) and student variance (the within-group variance).

This analysis of ICCs indicated good variability of citizenship at the school level for civic knowledge and satisfactory variability for democracy in both England and the United States. In England, approximately 22% and 8% of the variance in civic knowledge and democracy, respectively, was between schools. In the United States, the variance in these same outcomes was approximately 30% and 12%, respectively. In addition, the reliability estimates for England (civic knowledge $\lambda = 0.890$; democracy $\lambda = 0.734$) suggested that the data available within schools are strong enough to detect effects at the school level. Reliability estimates for the United States (civic knowledge $\lambda = 0.798$; and democracy $\lambda = 0.674$) suggested the same.

At the same time, the low ICCs for efficacy outcomes in both countries suggested that HLM would not be appropriate to use in the analysis because there is not sufficient between-school data to be modeled (ICC = 5%, $\lambda = 0.542$ in England; ICC = 3%, $\lambda = 0.365$ in the United States). Ordinary least-squares regression was used for this outcome. The software we employed for this purpose allowed us to take into account the design effects in the data.

Table 1: Description of Variables

Student-level dependent variables	
TOTCG	<i>Total Civic Knowledge</i> IRT score (Torney-Purta et al., 2001)
EFFIC	<i>Internal Political Efficacy</i> IRT score (Husfeldt et al., 2005)
DEMOC	<i>Concepts of Democracy</i> IRT score, from eight items
Student-level predictors	
<i>Dimension</i>	<i>Variable</i>
Dimension 2: Meaningful Learning of Civic-Related Knowledge (US only)	* <i>Interactive Class Activities</i> composite scale (Homana et al., 2006) $\alpha = .72$
	* <i>Traditional Class Activities</i> composite scale (Homana et al., 2006) $\alpha = .58$
	* <i>Media Activities</i> composite scale (Homana et al., 2006) Two-item scale: split-half reliability = .76
	* <i>Write Long Answers to Questions</i> , single-item
	* <i>Read Extra Material</i> , single-item
Dimension 3: Cooperation and Collaboration	* <i>Learning about the Value of Cooperation</i> composite scale. Three-item composite of learning to cooperate with others to help the community (Homana et al., 2006). α in US = 0.73; in England = 0.46
Dimension 5: Students' Input is Valued	* <i>Confidence in Participation at School</i> IRT score (Torney-Purta et al., 2001)
Dimension 6: Thoughtful and Respectful Dialogue	* <i>Open Classroom Climate for Discussion</i> IRT score (Torney-Purta et al., 2001)
Dimension 7: Engagement in Service to Community	* <i>Average Engagement in Civic Activities</i> : Six-item composite of participation in civic-related extracurricular activities (Homana et al., 2006). α in U.S.= 0.58; England, 0.65 Recoded 0 = fewer than two activities; 1= two or more activities
School-level Predictors	
<i>Name (Dimension)</i>	<i>Description</i>
Dimension 1: Official Recognition of the Civic Purpose of Education	Average of teachers who <i>strongly agree</i> schools are important for teaching civic education. Recoded as dichotomous variable, teachers do not strongly agree = 0; teachers who strongly agree = 1
Dimension 2: Meaningful Learning of Civic-Related Knowledge	Average use of <i>interactive activities</i> , as reported by teachers. α in US and England = 0.73
	Average use of <i>teacher-centered activities</i> , as reported by teachers. α in US = 0.68; in England = 0.69
Dimension 3: Cooperation and Collaboration	*Principal report of <i>learning the value of cooperation</i> (Homana et al., 2006) α in United States = 0.55; in England = 0.83
Dimension 4: Mutual Trust and Positive Interactions	Principal report of whether <i>racial and religious intolerance</i> is an issue in the school Recoded as dichotomous variable (median-split), only one or neither is an issue = 0; both are an issue = 1
Dimension 5: Students' Input is Valued	* <i>Average Confidence in Participation at School</i> IRT score (Torney-Purta et al., 2001)
Dimension 6: Thoughtful and Respectful Dialogue	* <i>Average Open Classroom Climate for Discussion</i> IRT score (Torney-Purta et al., 2001), aggregated from student responses
Dimension 7: Engagement in Service to the Community	* <i>Average Engagement in Civic Activities</i> : Six-item composite (Homana et al., 2006), aggregated from student responses
	*Principal's report of <i>activity availability</i> (from six in climate assessment: Homana et al., 2006) Recoded as dichotomous variable: 0 = fewer than 5 activities; 1 = 5 or 6 activities

Note: * This scale appears in the School Citizenship Education Climate Assessment (Homana et al., 2006).

Table 2: Descriptive Statistics

	England			United States		
Dependent variables	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>
Total Civic Knowledge	3,011	99.41	18.81	2,786	105.01	22.18
Conception of Democracy	2,879	9.84	2.00	2,737	10.08	2.04
Internal Political Efficacy	2,743	9.35	2.22	2,611	10.25	2.10
<i>Independent variables: Student</i>						
Confidence in Value of School Participation	3,043	9.92	1.91	2,811	10.05	2.06
Perception of Open Classroom Climate for Discussion	3,043	10.02	1.93	2,811	10.44	2.15
Learn about the Value of Cooperation	3,043	3.08	0.52	2,811	3.03	0.56
High Engagement in Civic Activities	3,043	1.34	1.22	2,811	1.55	1.36
<i>Independent variables: School</i>						
Schools are Important for Civic Ed.	128	0.31	0.47	124	0.42	0.50
Average Confidence in Participation	128	9.88	0.64	124	10.04	0.62
Average Perception of Classroom Discussion Climate	128	9.99	0.83	124	10.47	0.74
Teacher Use of Interactive Activities	128	2.39	0.39	124	2.57	0.58
Principal Rating of Learning about Value of Cooperation	128	3.48	0.34	124	3.26	0.45
Teacher Use of Traditional Activities	128	2.50	0.34	124	2.77	0.50
Student Activity Participation	128	1.35	0.37	124	1.55	0.47
High Number of Activities Available in School	128	0.59	0.49	124	0.49	0.50

Models

Three models are reported for each of the three outcomes of interest: one for England; one for the United States that excluded the classroom activity national options; and one for the United States that included the classroom activity national options. The purpose of the first United States model was to allow comparison of what influenced civic outcomes in this country with what influenced the outcomes in England. The purpose of the second United States model was to determine how students' perceptions of classroom activities influenced civic outcomes over and above other factors.

In the models for civic knowledge and conceptions of democracy, which employ multilevel analyses, all predictors were entered into each model simultaneously. Because student-level variables were centered on their grand mean, we could interpret the school-level coefficients as the effects of these characteristics of a school after controlling for individual students (i.e., they could be interpreted as effects of the school

context itself, over and above the sum of student effects). The model for efficacy similarly considered all predictors simultaneously, but because of the low ICC, only student-level predictors were employed.

As predicted, all variables included in these analyses correlated with one another to various extents (maximum $r = .59$) at the student level, between learning about cooperation and confidence in the value of participation. In considering these correlated variables simultaneously as predictors of the same civic-related outcome, we knew that only those with the strongest relationships to the outcomes would be statistically significant in the model. Many variables that do relate to the outcome may not be significant because of the other variables included in the model. Therefore, it is important to remember that the purpose of this analysis was to determine which aspects of school climate are most important to various civic-related outcomes according to statistical tests, and not to determine whether certain aspects of climate are unimportant for civic education.

Results

Knowledge analyses

Our first set of analyses considered the extent to which characteristics of a school citizenship education climate, as well as individual students' perceptions of and experiences in that school climate, predicted students' civic knowledge. Results from the three models are reported in Table 3. In each country, students scored higher on CivEd's test of civic knowledge when they perceived their classroom climate to be an open one and when they were highly engaged in extracurricular activities that benefited the community. Students in both England and the United States also had higher civic knowledge scores when they attended schools with an open discussion climate and with high activity

participation on the part of students, regardless of individual participation in these activities. In this sense, not only is it important that students are in a school context that encourages open discussion and active participation outside of school, but also that they themselves take advantage of such participation.

There were also several differences between England and the United States in the manner in which school citizenship education climate predicted civic knowledge. We found that in England, students performed better on the test of civic knowledge in schools that offered more community-related extracurricular activities, even after we had controlled for the extent of student participation in these activities. In other words, simply *offering* additional activities in which students can become involved contributed to the citizenship

Table 3: Results of Civic Knowledge Analysis

	England	United States	United States (class activities)
TOTCG Constant	99.37 (0.7)	105.0 (1.0)	105.0 (1.1)
<i>School-level predictors</i>			
Schools are Important for Civic Ed.	n.s.	n.s.	n.s.
Teacher Use of Interactive Activities	n.s.	3.7* (1.6)	3.7* (1.6)
Teacher Use of Traditional Activities	n.s.	n.s.	n.s.
Principal Rating of Value of Cooperation	n.s.	n.s.	n.s.
Race and Religious Intolerance Happens	n.s.	n.s.	n.s.
Average Confidence in School Participation	n.s.	n.s.	n.s.
Average Classroom Discussion Climate	2.9* (1.3)	3.9* (1.8)	3.9* (1.8)
Student Engagement in Extra Activities	4.3* (2.1)	6.9** (2.7)	6.9** (2.5)
Number of Activities Offered	4.0** (1.5)	n.s.	n.s.
<i>Student-level predictors</i>			
Student has Learned about Value of Cooperation	-1.6* (0.7)	n.s.	n.s.
Student Confidence in the Value of School Participation	1.2** (0.2)	n.s.	n.s.
Student Perception of Classroom Climate	1.1** (0.2)	0.9** (0.2)	0.9** (0.2)
Student is Highly Engaged in Activities	0.8** (0.3)	1.0** (0.3)	1.0** (0.3)
High Use of Interactive Activities			n.s.
High Use of Traditional Activities			n.s.
High Use of Media Activities			3.6** (1.0)
Write Long Answers to Questions			n.s.
Read Extra Material			n.s.
% Reduction in School Variance	37%	35%	36%
% Reduction in Student Variance	4%	2%	3%

Note: n. s. indicates not significant; * indicates significant at the .05 level; ** significant at the .01 level.

education climate in a way that had a positive influence on students' knowledge. In addition, English students who perceived that their participation in school decisionmaking was valued were more knowledgeable. Interestingly, students who reported that they learned about the value of cooperation in their classrooms had significantly *lower* civic knowledge. However, it is important to remember that the relationship between learning about cooperation and knowledge only showed as negative after we had controlled for other characteristics of in-class learning, including discussion climate and teacher activity use.

Students in the United States had higher civic knowledge in schools where teachers reported using more interactive activities such as role-playing and group projects. In considering the student reports of classroom activities, we also found that students who reported that they watched and discussed television and other visual media as part of their social studies classes (or other civic-related classes) were more knowledgeable than those who did not.

All together, the predictors in this model reduced school-level variance in civic knowledge by over a third in both countries. A look at student-level (or within-school variance), however, showed that such characteristics reduced the variance by only 3% to 4%. The addition of students' reports of classroom activities in the United States reduced the variance by only an additional percentage point. Thus, while a positive school citizenship education climate can contribute to differences in average levels of civic knowledge between countries, it is less effective in explaining differences in civic knowledge between students attending the same school.

Concepts of democracy analyses

A positive school citizenship education climate also appears to influence students' own dispositions toward democracy, including their conceptualizations of what activities are good or bad for democracy. Table 4 reports the multilevel analysis of how climate influenced the concepts of democracy scale formed from CivEd.

Table 4: Results of Democracy Analysis

	England	United States	United States (class activities)
DEMOC Constant	9.8 (0.1)	10.1 (0.1)	10.1 (0.1)
School-level predictors			
Schools are Important for Civic Ed.	n.s.	n.s.	n.s.
Teacher Use of Interactive Activities			
Teacher Use of Traditional Activities	n.s.	n.s.	n.s.
Principal Rating of Value of Cooperation	n.s.	n.s.	n.s.
Race and Religious Intolerance Happens	n.s.	n.s.	n.s.
Average Confidence in School Part.	n.s.	n.s.	n.s.
Average Classroom Discussion Climate	0.4* (0.1)	0.4** (0.1)	0.4** (0.1)
Student Engagement in Extra Activities	n.s.	n.s.	n.s.
High Number of Activities Offered	n.s.	n.s.	n.s.
Student-level predictors			
Student has Learned about Value of Cooperation	n.s.	0.3** (0.1)	0.3** (0.1)
Student Confidence in the Value of School Participation	0.2** (0.03)	0.1** (0.02)	0.1** (0.02)
Student Perception of Classroom Climate	0.1** (0.02)	0.1** (0.02)	0.1** (0.02)
Student is Highly Engaged in Activities	n.s.	0.1* (0.03)	0.1* (0.03)
High Use of Interactive Activities			-0.2* (0.9)
High Use of Traditional Activities			n.s.
High Use of Media Activities			0.2** (0.1)
Write Long Answers to Questions			n.s.
Read Extra Material			n.s.
% Reduction in School Variance	29%	38%	38%
% Reduction in Student Variance	4%	7%	7%

Note: n. s. indicates not significant; * indicates significant at the .05 level; ** significant at the .01 level.

Of the school context variables analyzed, only one significantly predicted students' concepts of democracy. Regardless of individual perception of the openness of classroom climate, students in both England and the United States who attended schools that received higher average student ratings of openness of discussion climate had stronger concepts of what is important for democracy. Despite the few context effects seen, many characteristics of students' own perceptions of school citizenship education climate related positively to their conceptualization of democracy. Across both countries, students who perceived their classroom climates to be open and who felt that student participation in the school decisionmaking processes was valued had stronger concepts of democracy.

Additional aspects of students' experiences in their school further predicted conceptions of democracy in the United States. Students who learned about the value of cooperation in their classrooms and who were highly engaged in extracurricular activities intended to benefit the community had stronger concepts of democracy than those who did not. In regard to student perceptions of their classroom activities, students who reported using news media when learning about citizenship education had stronger conceptualizations of democracy than students who did not.

In England, school citizenship education climate variables reduced the school-level variance by close to 30%, but only reduced the student-level variance by 4%. In the United States, the variables reduced the student-level variance by 7%, and reduced the

school-level variance by 36%. The model that also included predictors relating to students' perceptions of meaningful learning did not notably further reduce the variability evident for the United States sample.

Efficacy analyses

In the third and final analysis, we considered how students' reports of their school climate influenced their internal political efficacy, considered a necessary civic participation skill according to the Education Commission of the States (2005). The results of the analyses for England, the United States, and the United States with class activities are summarized in Table 5.

Students who were high participators in community-related extracurricular activities were more efficacious than low participators in both England and the United States. In England, however, the perception of an open classroom climate for discussion had an additional significant effect on students' efficacy. In contrast, students in the United States were more efficacious when explicitly taught about the importance of cooperating and collaborating with others, and when they thought that student voice was valued in school. However, taken together, these variables explained only between 5% and 6% of the variance in students' political efficacy.

The addition of the classroom activities variables to the United States analyses showed that students who reported high use of interactive activities (i.e., working with others), students who regularly wrote long answers to questions, and students who

Table 5: Results of Efficacy Analysis

	England	United States	United States (class activities)
EFFIC Constant	7.3 (0.4)	7.5 (0.4)	7.3 (0.4)
<i>Student-level predictors</i>			
Student Has Learned about Value of Cooperation	n.s.	0.5** (0.1)	0.4** (0.1)
Student Confidence in the Value of School Participation	n.s.	0.1* (0.03)	n.s.
Student Perception of Classroom Climate	0.2** (0.02)	n.s.	n.s.
Student Highly Engaged in Activities	0.3** (0.04)	0.2** (0.03)	0.2** (0.03)
High Use of Interactive Activities			0.5** (0.1)
High Use of Traditional Activities			n.s.
High Use of Media Activities			n.s.
Write Long Answers to Questions			0.2* (0.2)
Read Extra Material			0.5** (0.1)
% Student Variance Explained	5%	6%	9%

Note: n. s. indicates not significant; * indicates significant at the .05 level; ** significant at the .01 level.

read extra material were more efficacious, as were students who agreed that they were explicitly taught to cooperate and collaborate with others. Adding the classroom activities variables, however, rendered the independent effect of confidence in student voice non-significant. Interestingly, after controlling for other aspects of students' in-class experiences, we found that students who reported regularly participating in activities such as role-play or group projects had lower conceptualizations of democracy than students who reported irregular or little participation in these activities.

Discussion

Our analyses produced several key findings important for citizenship education in England and the United States. Results from analysis of the three outcomes confirm that students' citizenship education is significantly affected by what occurs in their schools. It is particularly interesting that so many characteristics of schools (as reported by principals, teachers, and students) were significant predictors of student outcomes over and above students' individual attitudes toward and behaviors in school. This consideration suggests that the effects of a school are greater than the aggregate of students. In other words, a school's climate has its own, unique effect on students' development.

More specifically, it is striking how experiences offered in a school both inside and outside the classroom affect civic knowledge. In the classroom, the findings of our analysis for both countries suggest that promoting an open classroom climate for discussion is a key dimension for schools seeking to improve overall civic knowledge among students. In the United States, we also found an additional positive effect of a school's use of classroom activities encouraging interaction facilitative of civic knowledge. Building on the work of Hess (2004), Westheimer and Kahne (2003), and Torney-Purta (2002), the analysis highlights the importance of schools utilizing their opportunities for discussion to help students develop their abilities of critical analysis in order to promote action for responsive citizenship. The significant effect of students' use of media in the classroom on civic knowledge in the United States is another important finding. Given that a component of the civic knowledge tests assessed students' skills in interpreting political information (Torney-Purta et al., 2001), we should

not be surprised that exposing students to meaningful, real-world political communication in the form of television news increased students' knowledge about the political world.

Extracurricular activities have an additional significant effect on knowledge. Those students in our samples who were highly engaged in extracurricular activities based on civic involvement had higher civic knowledge scores. In addition, attending a school where students were highly engaged also related to higher civic knowledge, regardless of individual participation. This effect appeared to be especially pronounced in England, where additional significant and positive effects were found for attending a school that simply offered many extracurricular activities (regardless of participation) and for individual perceptions that student participation is valued. Perhaps the ambivalence that English teachers may have felt about formal civic education prior to the Crick Report (expressed in Kerr, 1999) resulted in a situation where students learned most about citizenship through their own experiences in the community and in school decisionmaking. Indeed, no additional positive effects of classrooms on knowledge were found in England. In fact, we found, after controlling for other aspects of climate, that learning about the value of cooperation had a negative effect on knowledge, suggesting that coverage of this issue was not fully integrated into civic education at the time of the study.

A similar combination of in-class and out-of-class influences on students' conceptions of democracy was evident in each of the two countries. In particular, having an open classroom climate for discussion in a school overall and perceiving that openness individually each had a significant and positive effect on the strength of concepts of democracy. These findings support both research and our theoretical framework regarding the value of creating a positive school climate where students can engage, listen, and share their ideas and opinions with one another in a safe and supportive environment and so gain an increased understanding of democracy. A look across countries at out-of-school influences suggests that the strongest such influence on conceptualization of democracy is individual perception that student participation is valued. In other words, unlike the situation with civic knowledge, it is not enough in both England and the United States for students just to participate in relevant

extracurricular activities. Students must believe that their voice is valued if they are to gain a sense of what democracy is about.

When comparing the findings for the United States and for England, we considered of particular interest the finding that there is not only more school-level variance in democracy concepts in the United States, but also that more of that variance can be explained in the United States than in England. Although confidence in the value of school participation was extremely important in predicting concepts of democracy among United States students, there was an additional significant effect of simply participating in extracurricular activities. In relation to in-class effects, students who learned how to cooperate with others in solving community problems and who were exposed to the media had stronger democracy conceptualizations. These in-class predictors provide contrasting examples of how a school citizenship education climate can influence civic development by connecting to real-world experiences. In learning about the value of cooperation, students learn to cooperate and collaborate with diverse members of the community in order to affect positive change in their community and to become more active members of a democratic society. As with civic knowledge, the media provides students with real-world political messages about democracy that they must interpret.

However, in the United States, there is a negative relationship between students' participation in interactive activities such as role-playing or letter-writing and their conceptions of democracy. Assuming that this finding is a result of active student reflection on their experience, then perhaps it is through these types of activities that students begin to challenge what it means to live in a democracy. Alternatively, the finding could reflect the quality of the experience, suggesting that students require better preparation and guidance so that through deliberate practice they acquire the conceptions of democracy that will lead to positive and active civic engagement. Indeed, this negative relationship is found only in the model that controls for other beneficial aspects of school climate, such as openness for political discussion.

The analysis of internal political efficacy, an important civic knowledge skill, provided an opportunity to explore how in-school experiences could influence students' individual perceptions of

their ability to participate in political discussions. Overall, the results suggest that a positive school citizenship education climate can enhance students' sense of efficacy by providing opportunities for students to use their political knowledge in learning, in discussion, and in solving community problems. In both England and the United States, students who participated in civic-related extracurricular activities to a greater extent had more internal political efficacy than students who engaged to a lesser extent, although confidence in value of school participation did not have as strong an effect. This finding lies somewhat in contrast to analysis of internal political efficacy among upper-secondary students in Scandinavian countries (Barber & Amadeo, 2006), which found a significant relationship between confidence in the value of school participation and efficacy when participation itself was not taken into account. It appears from the current analysis that it is actual participation in and experience with these activities, rather than a general sense that student participation in such activities is valued, which influences efficacy. In other words, students must have hands-on experience in their communities and schools in order to gain confidence in their abilities.

Within classrooms, students in England were more efficacious when they perceived their climates as open to discussion. In the United States, significant effects were found in relation to what students reported learning and doing in the classroom. Students who learned explicitly to cooperate with others, who were exposed to interactive classroom activities, and who were required to read and write extensively had higher internal political efficacy. Regardless of whether the experience came from discussion or from formal instruction, it appears again as if students must experience discussing and thinking about citizenship in order to increase their confidence in it.

Finally, several non-significant findings raise important issues about how school citizenship can best be measured in a school. The belief that teachers have in the importance of civic education was not a significant predictor for any of the three citizenship-related outcomes. However, we need to remember that this question was asked of only civic-related teachers, and we were in essence only comparing those teachers who "agreed" that this was important to those who "strongly agreed" that it was important. We also need to acknowledge, though, that a completely-recognized

civic mission requires the buy-in of all teachers and administrators in a school, not only those most directly responsible for teaching about citizenship. The School Citizenship Education Climate Assessment (Homana et al., 2006) is designed to include input from many adult stakeholders in a school. This inclusion provides a more complete sense of the civic mission and results in more variability in the average importance of civics in any given school. In addition, principals' report of racial and religious intolerance was not a significant predictor of civic outcomes.

This finding raises two issues. First, principals might not be the most honest reporters of intolerance, as they may not realize (or may not wish to report) the true issues that a school has. This speaks to the importance of considering student input in measuring school climate, as students may be better able to report whether they have felt intolerance among their peers. In addition, this item only captured negative relationships among students in a school. To positively affect citizenship education, interactions among students and between students and adults must be positive, rather than simply non-negative. The School Citizenship Climate Assessment includes a series of questions on the extent to which relationships among students and between students and teachers build trust (from Brand, Felner, Shim, Seitsinger, & Dumas, 2003). This measure may be a better indicator of how interactions can truly influence citizenship in a positive manner.

Implications for teachers and administrators

Instructional practice is clearly how most students are expected to achieve greater learning outcomes; however, this analysis demonstrates the importance of considering what students learn outside of formal classroom contexts as well. Classrooms, and the schools in which they are situated, can provide rich opportunities for students to learn and to engage in interactions with other students and adults, forge a

community based on common values, respect, trust and acceptance of human diversity, and participate in activities that encourage a wide range of learning (Homana et al., 2006). Research can be conducted that examines if and how schools provide better avenues for students to become involved in meaningful learning opportunities within the classroom, throughout the school, and extending into the greater community. These evaluations can examine how schools develop vibrant and positive learning communities within classrooms that encourage students to actively participate with others in their own learning, while still addressing important community and world issues. Research can also investigate how schools provide opportunities for improved civic learning and engagement in out-of-classroom settings. Through an exchange of shared decisionmaking, such activities enhance student motivation in ways that promote group learning and, by extension, encourage the development of social capital. The findings also suggest investigation into how schools can create engaged communities of practice in which the dimensions used to frame this current study are enmeshed into a school's social context, allowing the benefits of a positive school climate for citizenship education to be embraced and practiced by all school members.

Conclusion

Particularly fruitful endeavors for future analysis are those focused on student responses to the host of school climate variables ranging from open classroom climate for discussion, engagement in learning with the community, expectations for learning, and issues of cultural diversity. These endeavors will help us gain a more nuanced understanding of the relationship of these variables to increased civic learning and engagement, and this understanding, in turn, will help ensure that all students have opportunity to develop into politically and socially responsible individuals.

References

- Barber, C., & Amadeo, J. (2006, July). Sense of political efficacy among upper-secondary students in three Nordic countries. In I. Martin (Chair), *University and upper-secondary students' civic empowerment across countries*. Paper symposium presented at the International Society for Political Psychology Annual Meeting, Barcelona, Spain.
- Brand, S., Felner, R., Shim, M., Seitsinger, A., & Dumas, T. (2003). Middle school improvement and reform: Development and validation of a school-level assessment of climate, cultural pluralism, and school safety. *Journal of Educational Psychology*, 95, 570–588.
- Campbell, D. (2004). *Voice in the classroom: How an open classroom environment facilitates adolescents' civic development*. University of Maryland: Center for Information and Research on Civic Learning and Engagement.
- Cleaver, E., Ireland, E., Kerr, D., & Lopes, J. (2004). *Citizenship longitudinal study: Second cross-sectional survey 2004: Listening to young people in England*. Berkshire, England: National Foundation for Educational Research.
- Cohen, J. (2001). Social and emotional education: Core concepts and practices. In J. Cohen (Ed.), *Caring classrooms/intelligent schools: The social emotional education of young children* (pp. 3–29). New York: Teachers College Press.
- Crick, B. (1998). *Education for citizenship and teaching of democracy in schools: Final report of the advisory group on citizenship*. London: Qualifications and Curriculum Authority.
- Education Commission of the States. (2005). *National Center for Learning and Citizenship QNA: Resources to assess civic competencies and school climate*. Available on <http://www.ecs.org/qna>
- Education Commission of the States. (2006). State policies for citizenship education database. Available on http://www.ecs.org/ecsmain.asp?page=/html/ProjectsPartners/nclc/nclc_main.htm
- Finnan, C., Schnepel, K., & Anderson, L. (2003). Powerful learning environments: The critical link between school and classroom cultures. *Journal of Education for Students Placed At Risk*, 8(4), 391–418.
- Freiberg, H. J. (1999). *School climate: Measuring, improving, and sustaining healthy learning environments*. London: Routledge.
- Ghaith, G. (2003). The relationship between forms of instruction, achievement and perceptions of classroom climate. *Educational Researcher*, 45(1), 83–93.
- Hahn, C. (1999). Challenges to civic education in the United States. In J. Torney-Purta, J. Schwille, & J.-A. Amadeo (Eds.), *Civic education across countries: Twenty-four national case studies from the IEA civic education project* (pp. 583–607). Amsterdam: The International Association for the Evaluation of Educational Achievement.
- Hess, D. (2004). Discussion in social studies: Is it worth the trouble? *Social Education*, 68(2), 151–155.
- Homana, G., Barber, C., & Torney-Purta, J. (2006). *School citizenship education climate assessment*. Denver, CO: National Center for Learning and Citizenship, Education Commission of the States.
- Homana G., & Greene, J. (2006). *Adolescent participation in student government and voluntary organizations: A comparative study in Australia and the United States*. Manuscript in preparation.
- Husfeldt, V., Barber, C., & Torney-Purta, J. (2005). *Students' social attitudes and expected political participation: New scales in the enhanced database of the IEA Civic Education Study*. College Park, MD: Civic Education Data and Researcher Services.
- Ireland, E., Kerr, D., Lopes, J., Nelson, J., & Cleaver, E. (2006). *Active citizenship and young people: Opportunities, experiences and challenges in and beyond school. Citizenship education longitudinal study: Fourth annual report*. Berkshire, England: National Foundation for Educational Research.
- Kahne, J., Rodriguez, M., Smith, B., & Thiede, K. (2000). Developing citizens for democracy? Assessing opportunities to learn in Chicago's social studies classrooms. *Theory and Research in Social Education*, 28(3), 311–338.
- Kerr, D. (1999). Re-examining citizenship education in England. In J. Torney-Purta, J. Schwille, & J. Amadeo (Eds.), *Civic education across countries: Twenty-four national case studies from the IEA Civic Education Study* (pp. 203–227). Amsterdam: International Association for the Evaluation of Educational Achievement.
- Kerr, D. (2003). Citizenship education in England: The making of a new subject. *Online Journal for Social Science Education*, 2(2). Available at http://www.sowi-onlinejournal.de/2003-2/england_kerr.htm
- Kerr, D. (2005). Citizenship education in England—listening to young people: New insights from the citizenship education longitudinal study. *International Journal of Citizenship and Teacher Education*, 1(1), 74–93.
- Kerr, D., Ireland, E., Lopes, J., Craig, R., & Cleaver, E. (2004). *Citizenship education longitudinal study: Second annual report: First longitudinal study*. Berkshire, England: National Foundation for Educational Research.
- Kerr, D., Lines, A., Blenkinsop, S., & Schagen, I. (2002). *England's results from the IEA International Citizenship Education Study: What citizenship and education mean to 14 year olds* (Research Report 375). Norwich, England: Department for Education and Skills.
- McLaughlin, M. (2004). *Community counts: How youth organizations matter for youth development*. Washington, DC: Public Education Network.

- Morgan, W., & Streb, M. (1999). *How quality service-learning develops civic values*. Bloomington, IN: Indiana University.
- National Council for the Social Studies. (2006). *NCSS mission statement*. Silver Spring: MD. Available at <http://www.socialstudies.org/about/>
- Nelson, J., & Kerr, D. (2005). *International review of curriculum and assessment frameworks, active citizenship: Definitions, goals and practices, background paper*. Berkshire, England: National Foundation for Educational Research.
- Office for Standards in Education. (2005, February). *Citizenship in secondary schools: Evidence from Ofsted inspections 2003/2004*. Available at http://canada.metropolis.net/research-policy/cern-pub/citizenship_education2003_04.pdf#search=%22Citizenship%20in%20secondary%20schools%3A%20evidence%20from%20Ofsted%20inspections%22
- Pigozzzi, M. J. (2006). A UNESCO view of global citizenship education. *Educational Review*, 58(1), 1–4.
- Qualifications and Curriculum Authority. (2006). *National curriculum online*. London, England. Available at <http://www.nc.uk.net/>
- Schulz, W., & Sibberns, H. (2005). *IEA Civic Education Study technical report*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Search Institute. (2000). *40 developmental assets*. Minneapolis, MN. Available at <http://www.search-institute.org/assets/forty.html>
- Torney-Purta, J. (2002). The school's role in developing civic engagement: A study of adolescents in twenty-eight countries. *Applied Developmental Science*, 6(4), 203–212.
- Torney-Purta, J., Barber, C., & Wilkenfeld, B. (2006). Differences in the civic knowledge and attitudes of adolescents in the United States by immigrant status and Hispanic background. *Prospects (A UNESCO journal)*, 36(3), 343–354.
- Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries*. Amsterdam: International Association for the Evaluation of Educational Achievement. Available at <http://www.wam.umd.edu/~iea>
- Torney-Purta, J., & Richardson, W. (2003). Teaching for the meaningful practice of democratic citizenship: Learning from the IEA Civic Education Study in 28 countries. In J. Patrick (Ed.), *Principles and practices of democracy in the education of social studies teachers* (Vol. 2, pp. 25–44). Bloomington, IN: ERIC Clearinghouse for Social Studies/Social Science Education.
- Torney-Purta, J., Schwille, J., & Amadeo, J. A. (1999). (Eds.). *Civic education across countries: Twenty-four case studies from the civic education project*. Amsterdam: Eburon Publishers for the International Association for the Evaluation of Educational Achievement.
- Torney-Purta, J., & Vermeer, S. (2004). *Developing citizenship competencies from kindergarten through grade 12: A background paper for policymakers and educators*. Denver, CO: National Center for Learning and Citizenship, Education Commission of the States.
- Vosniadou, S. (2001). *How children learn* (Educational Practices Series, No. 7). Geneva, Switzerland: IBE/UNESCO.
- Westheimer, J., & Kahne, J. (2003). Teaching democracy: What schools need to do. *Phi Delta Kappan*, 85(1), 34–66.

Home reading environments and children's reading performance: A comparative study of 25 countries

Hyunjoon Park
University of Pennsylvania
Philadelphia, Pennsylvania, USA

Background

Sociologists have long been interested in sources of educational differences between children from poor families and children from affluent families. Cultural capital theory has extended our understanding of the ways in which cultural resources at home, independently from and interactively with financial resources, enhance children's education (Bourdieu, 1973; DiMaggio, 1982). High levels of cultural resources that privileged parents possess and pass on to their child account for a significant portion of educational differences among students from different socioeconomic origins.

However, researchers often disagree on which kinds of cultural resources are more relevant for affecting children's education. The traditional view of cultural reproduction has emphasized parental participation in and knowledge of highbrow culture (Bourdieu, 1984; De Graaf, De Graaf, & Kraaykamp, 2000). The most common way of measuring cultural capital in this line of research is to count the instances of participation in high cultural activities, such as visiting museums and art galleries or attending symphony concerts.

On the other hand, there has been a growing interest in the importance of parental reading habits as another indicator of cultural capital beyond the traditional measure of parental participation in and knowledge of highbrow culture. This perspective argues that what is more important for developing children's language and cognitive skills, and what ultimately leads to educational success, is how much parents stimulate the child's intellectual development by offering favorable reading climates at home (De Graaf et al., 2000; Crook, 1997; Farkas, 1996). This line of research has therefore been interested in parental reading behaviors and other indicators of home reading climates. For instance, De Graaf et al. (2000) found that in the Netherlands parental reading habits had stronger effects on children's educational attainment

than did parental participation in beaux arts.

Although not directly related to cultural capital theory, much research on child development, especially in the United States, highlights the importance of home literacy environments that stimulate the development of the child's cognitive and language skills (Bradley, 1985; Farkas & Beron, 2004). Researchers have found substantial differences in home literacy environments between children from high- and low-socioeconomic families, which, in turn, explain educational differences between the two groups of children (Brooks-Gunn, Klebanov, & Duncan, 1996; Duncan, Brooks-Gunn, & Klebanov, 1994). In short, poor home literacy environments of low socioeconomic families are an important mechanism leading to educational disadvantages of children from low socioeconomic backgrounds.

Despite the significant contribution that previous studies have made to our understanding of how cultural resources matter for children's education, there remains an important gap in the literature. Most studies looking at the effects of cultural resources are so far single-country studies. Also, our current knowledge on the issue is primarily limited to western advanced countries. We know little about to what extent countries vary in the effects of cultural resources and what explains such cross-national variation, if any. For instance, De Graaf et al. (2000) pointed out that the effect of parents' high-culture participation on children's educational outcomes may depend on the extent to which highbrow cultural aspects are institutionalized as a part of the standard curriculum of educational systems. Specifically, they posited that highbrow culture will matter more for children's education success in countries like France and Italy where history and philosophy are more important fields in school than in the Netherlands and the United States. Because they looked only at

the Dutch case, they could not directly assess this intriguing hypothesis. And what about the effects of home literacy environments? Are there systematic associations between the magnitude of the effects and specific country-level factors? Examining how national contexts mediate the relationships between cultural resources and children's educational outcomes will help us better understand the mechanism through which family's cultural resources reproduce educational inequality.

Research questions

In this paper, my aim is to extend the literature on home literacy environments and children's education in two significant ways. First, I compare the roles that home literacy environments play for intergenerational transmission of educational inequality in 25 countries. It remains to be seen to what extent the positive effects of home literacy environments found in the United States and other western countries are generalized to countries that have different social contexts of education.

The second contribution of this paper is to examine various aspects of home literacy environments. Here, I look not only at parental reading attitudes but also at two other aspects: parental engagement with the child in various literacy activities, and the number of books in the home. Although related to each other, these aspects may represent different ways in which parents foster their children's literacy skills. For instance, parents' own reading behavior may not necessarily reflect the extent to which they actually engage with their child in various literacy activities to foster children's literacy skills. What is more relevant for boosting children's reading skills may not be the mere reading habits of parents themselves but the degree to which and how parents actually interact with their children.

In regard to the number of books in the home as a reflection of parents' scholarly culture, Evans, Kelly, Sikora, and Tremain (2005) recently examined the association between number of books in the home during childhood and each respondent's occupational attainment in 31 countries. They found significant relationships in a wide range of societies. They also found that the effect of the number of books on occupational attainment occurred through its effect on the respondent's educational attainment. However, interpreting number of books as reflecting scholarly

culture at home is not as straightforward as it might initially appear. In literature on student achievement, especially literature involving international comparisons, number of books is widely used as a proxy of family socioeconomic conditions when other measures of socioeconomic status are not available from the data. With this caution in mind, I examine how the association between the number of books and children's education varies across countries. I assess the relative strength of the effects among the number of books, parental reading attitudes, and parental engagement with the child in home literacy activities.

Beyond descriptions of the different relationship between home literacy environments and children's education in various contexts, I assess whether there is a systematic association between the effects of home literacy environments and the country's economic development level. Comparative research across many countries can offer an opportunity to develop and empirically test some hypotheses that help explain sources of the cross-national variation in the effects. It is an important limitation for the current study to examine only the economic development level as a mediating factor for the relationships between home literacy environments and children's education. I recognize this limitation and plan to include other country-level variables, especially structural features of educational systems, in the revision.

In sum, I address three questions. First, I examine the extent to which each aspect of home literacy environments is related to children's education in 25 countries and the degree to which differences in home literacy environments account for educational differences among students from different socioeconomic origins. Second, I compare the relative effects among the three aspects of home literacy environments. Finally, I investigate whether the effects of home literacy environments systematically vary according to the country's economic development level.

Data

For this study, I use data on student achievement in reading from 25 countries that participated in the Progress in International Reading Literacy Study (PIRLS) 2001 (Mullis, Martin, Gonzalez, & Kennedy, 2003). PIRLS assessed the reading achievement of Grade 4 students in primary schools and collected

extensive information on students' family and school experiences (student questionnaire), family socioeconomic conditions and parental engagement with the child in various literacy activities (parent questionnaire), and various school characteristics and instructional practices (school questionnaire answered by school administrators). The study was based on a two-stage stratified cluster design. First, schools were selected with probability proportional to size, and one intact classroom for Grade 4 was then selected within each school.¹

Originally, 35 countries participated in PIRLS 2001. But neither home literacy environment measures nor parental education, which are key variables for the present analysis, are available for five countries (Israel, Kuwait, Hong Kong SAR, Morocco, and the United States).² Furthermore, even where these variables are available, significant proportions of the participating students did not provide information on them (more than 30% of total respondents) in five countries (Belize, Cyprus, England, the Netherlands, and Scotland). In the end, I have been able to compare only 25 of the countries that participated in PIRLS 2001.

The major advantage of PIRLS over other international surveys of student achievement (such as the Third International Mathematics and Science Study) is that it collected information on various aspects of home literacy environments. The number of books at home, which is one of three measures of home literacy environments used in the present analysis, was often gathered in previous international surveys of student achievement. However, the other two measures of home literacy environments (early home literacy activities and parental reading attitudes) were usually not available.

Despite the significance of PIRLS for research on cultural resources and children's education, an important limitation should be addressed. PIRLS did not collect information on parental participation in and/or knowledge of highbrow culture. Therefore, it is impossible to compare the relative importance of highbrow cultural activities and home literacy activities within countries, and also cross-national variation in this aspect. Given this limitation of data, I look only at the effects of home literacy environments in this study.

Measures

Home literacy environments

Index of early home literacy activities

A parent or guardian of the child was asked to indicate how often he or she or someone else at home engaged in the following activities with the child before the child entered primary school: read books, told stories, sang songs, played with alphabet toys, played word games, or read aloud signs and labels. Each item was based on a three-point scale: 1—never or almost never, 2—sometimes, and 3—often. The index is an average of the responses on the six items (or non-missing items). Higher values of the index indicate higher levels of parental engagement with the child in literacy activities.

Index of parents' attitudes toward reading

The parent was asked whether he or she agreed with the following statements about reading: "I read only if I have to," "I like talking about books with other people," "I like to spend my spare time reading," "I read only if I need information," and "Reading is an important activity in my home." Each item was based on a four-point scale: 1—disagree a lot, 2—disagree a little, 3—agree a little, and 4—agree a lot. After reversing the scale for negative statements (i.e., "I read only if I have to" and "I read only if I need information"), I computed the average score on the five items. Higher values of the index indicate more positive attitudes toward reading.

Number of books at home

The student was asked to state the number of books at home. The parent answered the same question in the home questionnaire. Because the students' reports were more comprehensive than the parents' reports, I used the number of books at home as reported by students. But if the student's report was not available, I substituted the parent's report for it. For both students' and parents' reports, the response was based on a five-point scale: 0–10 books (1), 11–25 (2), 26–100 (3), 101–200 (4), and more than 200 (5). I used the linear specification that has values from 1 to 5 for the number of books.

1 See Martin, Mullis, and Kennedy (2003) for more detailed information on the sampling of PIRLS 2001.

2 As I describe later in this paper, I use seven items to construct the index of early home literacy activities. In Hong Kong SAR, one of the seven items was not asked. In this version of the paper, I exclude Hong Kong.

The outcome measure

The major outcome of analysis in this study is the student's reading achievement score. The score was scaled to have a mean of 500 points and a standard deviation of 100 points for all students across the countries that participated in PIRLS 2001. Instead of a fixed value for the reading literacy scale, PIRLS provides five plausible values for each student, which should be used simultaneously to obtain estimates of population parameters.

Socioeconomic background and other individual characteristics

To measure a student's socioeconomic background, I used parental education and the location of the school the student attended. Parental education is the educational attainment of whichever parent (mother or father) had the higher level of education. Originally, five levels of educational attainment were distinguished: below lower secondary, completed lower secondary, completed upper secondary, completed post-secondary but not university, and completed university or higher. Because, in several countries, the number of students whose parents had not gone beyond lower secondary education was so small, I combined this category with the category of completed lower secondary. The same consideration led me to combine the category of completed post-secondary but not university with completed university or higher. I therefore compared three levels of education: completed lower secondary or below, completed upper secondary, and completed tertiary. School location was identified through school principals' reports on whether their school was located within an urban or a suburban or rural area.

I also included gender, the number of children, family structure, and language minority status in the models. Comparative studies of student achievement have documented the significant advantage of female students over male in reading achievement in almost every country (OECD, 2001). The negative relationship between number of siblings and a child's education is well established across a variety of societies (Powell, Werum, & Steelman, 2004). PIRLS did not directly ask participants to state the number of siblings but only the number of children in the home. The number of children is included in the model as a continuous variable. For family structure, I distinguished families with one adult and families with two or more adults.

PIRLS did not collect information on whom the student lived with but only the number of adults living together. Therefore, it is not possible to identify whether two adults living together with the child were two parents. With respect to language minority status, I distinguished two groups of students: those students who spoke the language of the test at home always or almost always and those students who did not.

Methods

In this study, I used two different methods. I first conducted OLS regression analyses for each of the 25 countries separately to examine the effects of the three measures of home literacy environments within countries. After assessing the effects separately across countries, I applied a multilevel model technique to the pooled data across the 25 countries (refer Bryk & Raudenbush, 1992). Multilevel models are particularly useful to identify the extent to which the effects of home literacy environments vary across countries and whether the cross-national variation depends on the economic level of countries. Specifically, the two-level models are estimated with student as the first-level unit and country as the second-level unit. In the student-level equation, the reading score for student i in country j is predicted as follows:

$$(\text{Reading literacy})_{ij} = \beta_{0j} + \beta_{1j}(\text{Early Home Literacy Activities})_{ij} + \sum_k \beta_{kj} X_{kij} + r_{ij}$$

Here, the intercept, β_{0j} , represents the mean reading score in country j , adjusted for differences among countries relating to the other student characteristics included in the model (note that all the student-level variables center on grand means). β_{1j} is the slope of the index of early home literacy activities affecting student reading scores in country j , and r_{ij} is the student-specific error. The effects of parental education and student's demographic characteristics are represented through β_{2j} to β_{kj} .

In the two-level models, the coefficients in the first-level equation serve as dependent variables in the second-level equation. Thus, the country-level equations are:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{GDP per capita})_j + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{GDP per capita})_j + \mu_{1j}$$

$$\beta_{kj} = \gamma_{k0} + \mu_{kj}$$

To examine how the within-country effect of early home literacy activities on children's reading performance varied according to the country's economic level, I modeled the slope of early home literacy activities (β_{1j}) to be predicted by the country's GDP per capita and random errors (μ_{1j}).³ In the model, γ_{11} indicates the impact of GDP per capita on the slope of early home literacy activities, while γ_{10} indicates the average slope of early home literacy activities for the countries with GDP per capita corresponding to the average among the 25 countries (i.e., GDP per capita centers on the grand mean). I also modeled each country's mean reading score (β_{0j}) to vary across countries as a function of GDP per capita. γ_{k0} represents the overall effect of the k th control variable at the student level, and u_{kj} indicates a random error. In other words, in

the model, all the effects associated with student-level control variables are assumed to vary randomly. The same specification was applied to the separate model for the index of parental attitudes toward reading and the model for the number of books.

Results

Levels of home literacy environments

Table 1 presents national averages of the three measures of home literacy environments among the 25 countries, along with mean scores for reading achievement and GDP per capita. Countries are sorted in order of the number of books at home. Remember that the number of books was measured on a five-point scale, where 1 represented 0–11 books, 2 related to 11–25 books, 3 represented 26–100, 4 was 101–200, and 5 was more

Table 1: National Averages of Home Literacy Environment Measures

	Number of books	Parental attitudes toward reading	Early home literacy activities	Mean score of reading	GDP per capita (PPP \$)	Number of students
Sweden	3.717 (1.098)	3.346 (0.688)	2.174 (0.373)	561 (96)	25,400	5,483
Norway	3.598 (1.117)	3.371 (0.673)	2.217 (0.363)	499 (83)	31,800	3,104
Iceland	3.536 (1.092)	3.267 (0.620)	2.275 (0.355)	512 (72)	27,100	3,065
Latvia	3.500 (1.137)	3.073 (0.571)	2.347 (0.375)	545 (81)	8,300	2,886
Hungary	3.447 (1.252)	3.377 (0.623)	2.352 (0.364)	543 (65)	13,300	4,445
Czech Republic	3.386 (1.067)	3.214 (0.649)	2.281 (0.342)	537 (70)	15,300	2,666
Canada	3.378 (1.154)	3.261 (0.655)	2.409 (0.387)	544 (67)	29,400	6,863
New Zealand	3.372 (1.210)	3.271 (0.707)	2.440 (0.399)	529 (73)	19,500	2,086
France	3.220 (1.180)	3.077 (0.630)	2.286 (0.392)	525 (66)	25,700	3,173
Slovak Republic	3.207 (1.105)	3.274 (0.640)	2.329 (0.355)	518 (75)	12,200	3,687
Bulgaria	3.139 (1.473)	3.098 (0.788)	2.343 (0.471)	550 (92)	6,600	3,301
Singapore	3.113 (1.175)	2.927 (0.656)	2.161 (0.433)	528 (71)	24,700	6,860
Slovenia	3.099 (1.147)	3.209 (0.618)	2.364 (0.352)	502 (62)	18,000	2,859
Russia	3.084 (1.218)	3.054 (0.678)	2.412 (0.406)	528 (64)	8,800	4,040
Germany	3.069 (1.182)	3.041 (0.780)	2.183 (0.381)	539 (103)	26,600	6,656
Greece	2.996 (1.174)	3.252 (0.722)	2.345 (0.419)	524 (75)	19,000	2,175
Lithuania	2.902 (1.097)	2.940 (0.693)	2.228 (0.386)	543 (93)	8,400	2,471
Italy	2.827 (1.225)	3.089 (0.742)	2.381 (0.392)	541 (81)	25,000	3,394
Macedonia	2.380 (1.151)	3.084 (0.662)	2.394 (0.413)	442 (90)	5,000	2,877
Romania	2.369 (1.262)	2.821 (0.750)	2.278 (0.472)	512 (66)	6,800	3,534
Moldova	2.183 (1.196)	2.782 (0.662)	2.176 (0.430)	492 (92)	3,000	3,466
Turkey	2.060 (1.115)	2.782 (0.698)	1.994 (0.477)	449 (70)	7,000	4,988
Argentina	2.027 (1.157)	2.868 (0.686)	2.245 (0.468)	420 (72)	10,200	2,332
Colombia	1.983 (1.152)	2.975 (0.737)	2.185 (0.470)	442 (66)	6,300	4,637
Iran	1.788 (1.106)	3.000 (0.663)	1.843 (0.493)	414 (86)	7,000	7,142

Note: Countries are sorted in order of the averages of number of books at home. Values in parentheses are standard deviations.

3 Information on GDP per capita for the 25 countries was obtained from the *World Fact Book*, available on the following website: <http://www.umsl.edu/services/govdocs/wofact2003/rankorder/2004rank.html>. The figures are mostly for the year 2002 or 2001.

than 200 books. In the table, Sweden (3.717) shows the highest average number of books, followed by Norway and Iceland. Iran, Colombia, Argentina, and Turkey show the lowest averages.

The next column in the table presents the national averages of the index of parental attitudes toward reading. In the range of 1 to 4, higher values of the index indicate more positive attitudes toward reading. The three countries showing the most positive attitudes are Hungary, Norway, and Sweden. Turkey, Moldova, and Romania show the lowest averages. Note, however, that the averages for these lowest countries are 2.8. In other words, even in these countries, people, on average, agree a little with the statement, "Reading is an important activity in my home" (1—disagree a lot, 2—disagree a little, 3—agree a little, 4—agree a lot).

The third column shows the national averages of the index of early home literacy activities. The index indicates the average extent of parental engagement with the child in six literacy activities before primary school (1—never or almost never, 2—sometimes, 3—often). Parents in New Zealand, Russia, and Canada are more likely to be engaged in literacy activities with their child than are parents in the other countries. Parents in Iran, Turkey, and Singapore show the lowest level of parental engagement. As with the index of parental attitudes toward reading, it is interesting to see that even in those countries with the lowest averages, parents show a fairly high level of engagement with their child in literacy activities. The averages of 1.84 in Iran and 1.99 in Turkey indicate that in the two countries with the lowest levels, the average level of parental engagement is close to "sometimes."

Comparison of the top and bottom countries in relation to each home literacy environment measure reveals a degree of relationship between the average level of home literacy environment and the national economic level. For instance, the top three

countries for number of books are highly developed countries, while the bottom seven countries (except Argentina) are economically poor countries. Table 2, which presents correlations among the three home literacy environment measures, GDP per capita, and mean scores of reading achievement, confirms this observation. Although the correlation between GDP per capita and the index of early home literacy activities is relatively weak, correlations of GDP per capita with the number of books ($r = 0.651$) and with the index of parental attitude toward reading ($r = 0.579$) are fairly strong. In other words, home literacy environments tend to be more favorable in more developed countries. The positive correlation coefficients between mean scores of reading and the three measures of home literacy environments also indicate that countries with more favorable literacy environments tend to show better performance in reading.

Mediating the effects of parental education

To assess the extent to which home literacy environments mediated the effects of socioeconomic background on children's education, I examined how the reading gap between students whose parents had tertiary education and those whose parents did not complete lower secondary education changed after controlling for each measure of home literacy environments. For this purpose, I estimated five different models for each country separately. The first model (Model 1) included parental education and other individual characteristics but no home literacy environment measures. The reading gap between students whose parents had tertiary education and those whose parents had not gone beyond lower secondary education, estimated from Model 1, served as the baseline difference. In Model 2, the index of early home literacy activities was added to Model 1. Comparison of Model 1 and Model 2 thus showed the extent to which the difference in parental engagement

Table 2: Correlation ($N = 25$)

	(1)	(2)	(3)	(4)	(5)
(1): Early Home Literacy Activities	1				
(2): Parental Attitudes toward Reading	0.445 *	1			
(3): Number of Books	0.529 **	0.780 **	1		
(4): Mean Score of Reading	0.523 **	0.459 *	0.827 **	1	
(5): GDP per Capita	0.172	0.579 **	0.651 **	0.448 *	1

Note: * $p < .05$ ** $p < .01$.

in literacy activities with the child accounted for the effect of parental education on student's reading performance. Models 3 and 4 added to Model 1 the index of parental attitudes toward reading and the number of books, respectively. The final model (Model 5) added all three measures of home literacy environments to Model 1.

Figures 1A through 1C present the reading gaps by parental education across the five models in the 25 countries. Countries are separated into three groups, depending on the magnitude of the effect of parental education revealed in Model 1 (high group, middle group, and low group). For instance, among the 25 countries in this study, Hungary shows the largest reading gap (77 points) between students whose parents had tertiary education and their counterparts whose parents did not complete lower secondary education, a finding apparent after I had controlled for demographic and other individual characteristics. The gap is reduced to 75 points, 65 points, and 63 points when the index of early home literacy activities, the index of parental reading attitudes, and the number of books at home, respectively, is added to the model. When all three measures are included, the reading gap is reduced by 25%—from 77 points in Model 1 to 57 points in Model 5.

Although not universal across all the countries, the mediating role of the index of early home literacy activities is rather weak compared to the index of parental reading attitudes and the number of books. Except for a few countries, controlling for the index of early home literacy activities reduces the effect of tertiary education by less than 10% in most countries. Moreover, in many countries, the reduction in the reading gap by parental education seems to be rather modest even after controlling for all three measures of home literacy environment (less than 30% in Model 1), although the reduction is fairly substantial in Moldova (55%), Romania (51%), Russia (51%), and Bulgaria (43%). Except for Russia, the effect of tertiary education remains significant in all the countries after taking into account differences in the three measures of home literacy environments. In short, the results indicate that although differences in home literacy environments between high- and low-educated parents account, to some extent, for the reading gap by parental education, a significant proportion of low-educated parents in many countries show a fair level of engagement with the child in literacy activities, have positive attitudes toward reading, and have a large number of books at home.

Figure 1A: Reading Gaps between Students Whose Parents Had Tertiary versus Below Lower Secondary Education (High Group)

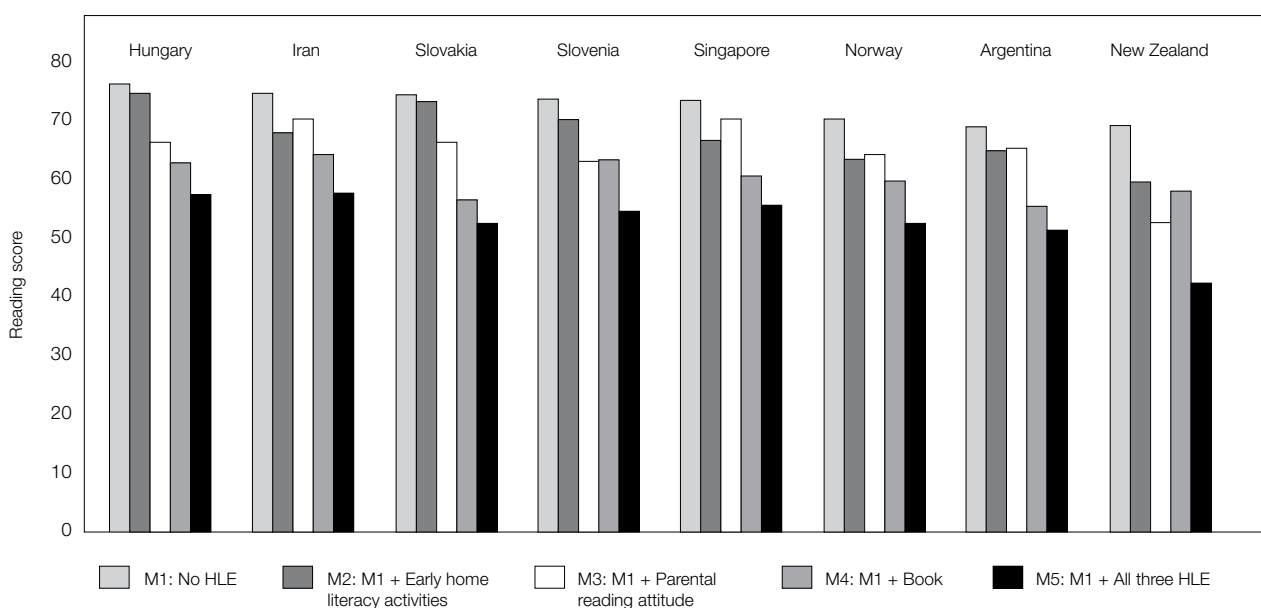


Figure 1B: Reading Gaps between Students Whose Parents Had Tertiary versus Below Lower Secondary Education (Middle Group)

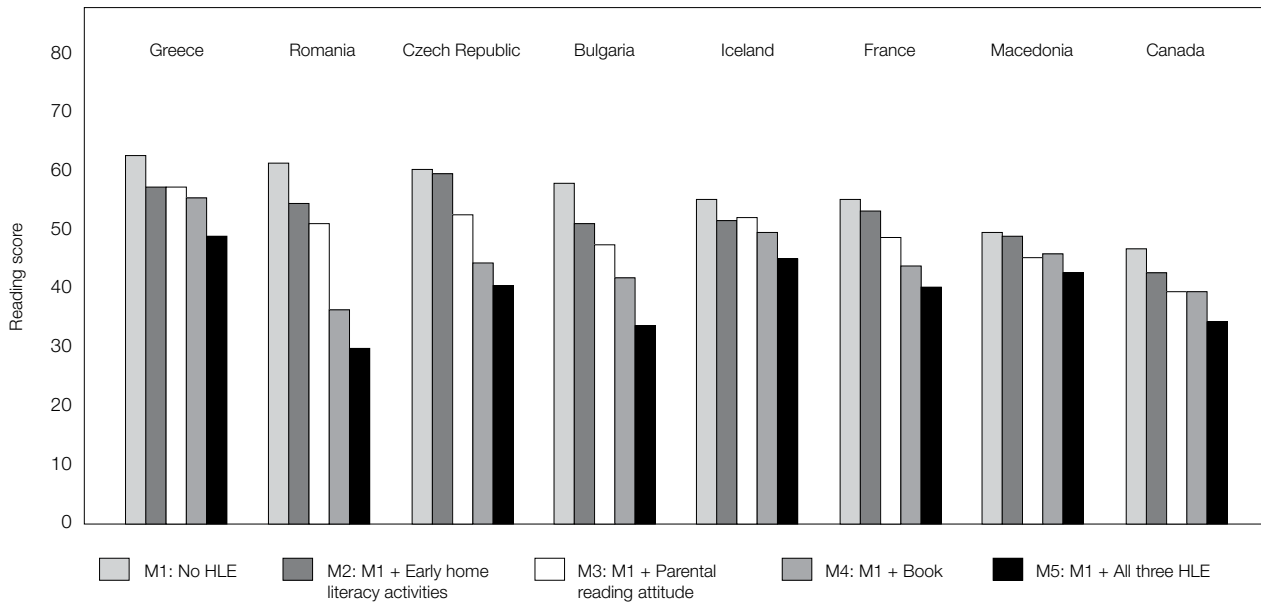
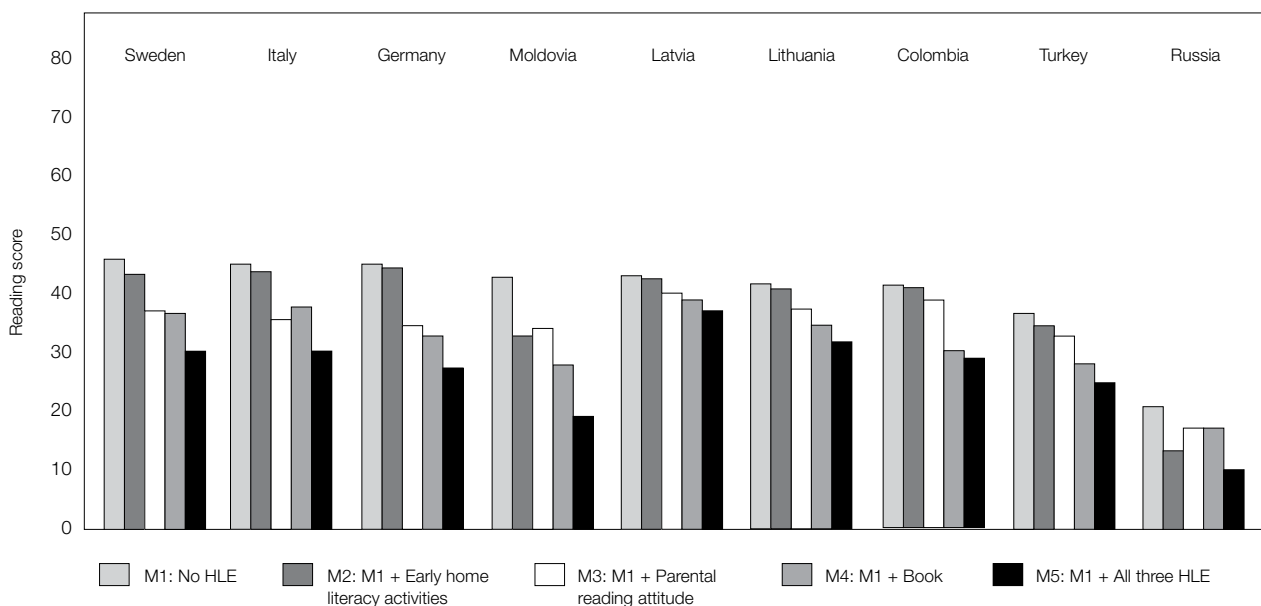


Figure 1C: Reading Gaps between Students Whose Parents Had Tertiary versus Below Lower Secondary Education (Low Group)



Effects of home literacy environments

This section presents my findings on the extent to which home literacy environments related to children's reading performance in each country after I had controlled for parental education and other individual characteristics. Figure 2A shows changes in the reading score per unit change in the index of early home literacy activities. Remember that the index has three values to

indicate the level of parental engagement in literacy activities: 1—never or almost never, 2—sometimes, and 3—often. Therefore, a one-unit increase in the index corresponds to a change from the response of never or almost never to the response of sometimes, or the response of sometimes to the response of often. For example, we can see that in New Zealand, children

whose parents engaged often with them in literacy activities have a 39-point higher average score on reading than those whose parents engaged sometimes. In turn, children whose parents engaged sometimes have a 39-point higher average than those whose parents were never or almost never engaged. The top five countries in the magnitude of the effect are New Zealand, Norway, Greece, Canada, and Iceland. Colombia, Slovakia, Germany, Macedonia, and Lithuania show the weakest effect of the index of early home literacy activities. In fact, I found that early home literacy activities did not significantly contribute to

increasing children's reading performance in Colombia and Slovakia, once I had taken parental education and other individual characteristics into account.

Figure 2B presents the effect of the index of parental attitudes toward reading, measured by changes in the reading score per unit change in the index. Parents indicated their degree of agreement with positive statements on reading behaviors through use of a four-value scale (1—disagree a lot; 2—disagree a little; 3—agree a little; and 4—agree a lot). A one-unit increase in the index indicates the difference between two consecutive categories. In Slovenia, the country with

Figure 2A: Effects of Early Home Literacy Activities

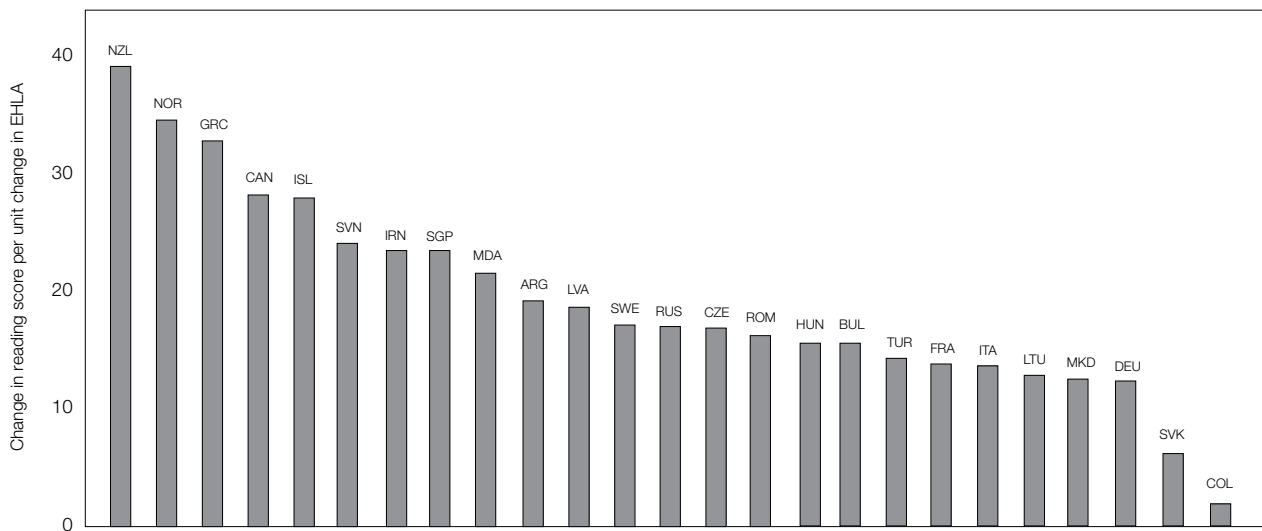
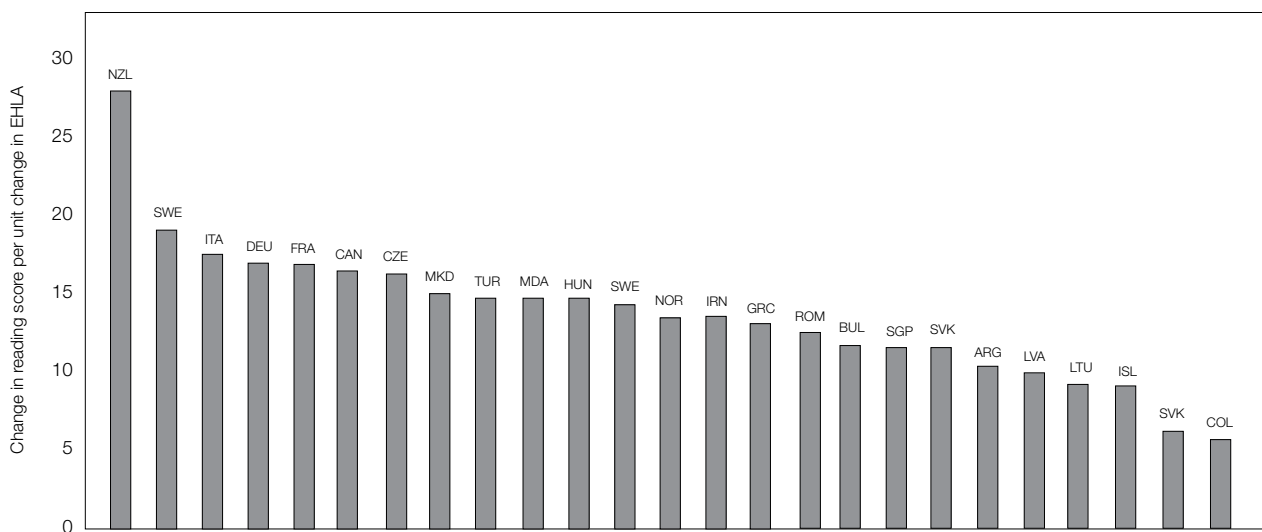


Figure 2B: Effects of Parental Attitudes toward Reading



the second largest effect, the reading gap associated with a one-level increase in parental reading attitudes is 19 points. In Colombia, the country with the weakest effect, the gap is 6 points. It is interesting to see that the index of parental attitudes toward reading is significantly associated with children's reading performance in all 25 countries.

Figure 2C shows the effect of the number of books at home on children's reading performance. The number of books was originally measured on a five-point scale (see under "Levels of Home Literacy Environments" above). Because I used the scale as a continuous variable with values from 1 to 5, interpreting what a one-unit increase in the number of books means is not straightforward. Singapore shows the strongest effect (18 points by one unit increase in the number of books); Russia has the weakest effect (5 points). The effect is significant in all 25 countries.

Because of different units of measurement, it is difficult to assess the relative effects among the three measures of home literacy environments. To facilitate the comparisons, I rescaled each variable so that the mean and the standard deviation of it became 0 and 1, respectively, for all students in the 25 countries. Thus, a one-unit increase in each variable of home literacy environments corresponds to a one standard deviation increase. This means that the effect of each measure is measured by the change in the reading score per one-standard-deviation-increase for each measure of home literacy environments. Furthermore, because

the variables were rescaled to have a mean of 0 and a standard deviation of 1 across all students in the 25 countries, the effects are directly comparable across countries.

Figure 3 depicts the effects of the three home literacy environment measures in each country as changes in the reading score per a standard deviation change. Note that the effects were derived from models that included all three measures at the same time in addition to parental education and other individual variables. Although not universal, the overall pattern revealed from the figure is the relatively stronger effect that number of books has over early home literacy activities and parental attitudes toward reading. This pattern is evident in 20 out of the 25 countries.

Cross-national variation in the effects of home literacy environments

Figure 3 reveals substantial variations across countries in the effects of the three measures of home literacy environments. But which national factors are associated with the country-level differences in the effects of home literacy environments? Many potential candidates of country-level variables may account for cross-national variations in these effects. However, I decided in this study to consider only the impact of the country's economic development level.

To assess the extent to which the effects of the three measures of home literacy environments vary according to the country's economic development

Figure 2C: Effects of Number of Books at Home

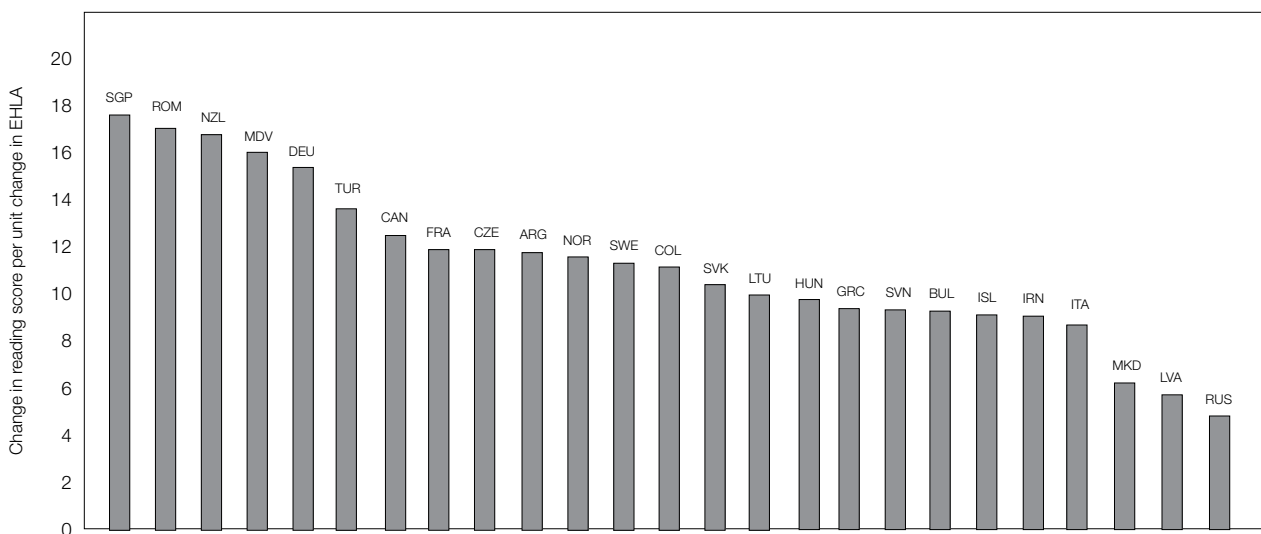
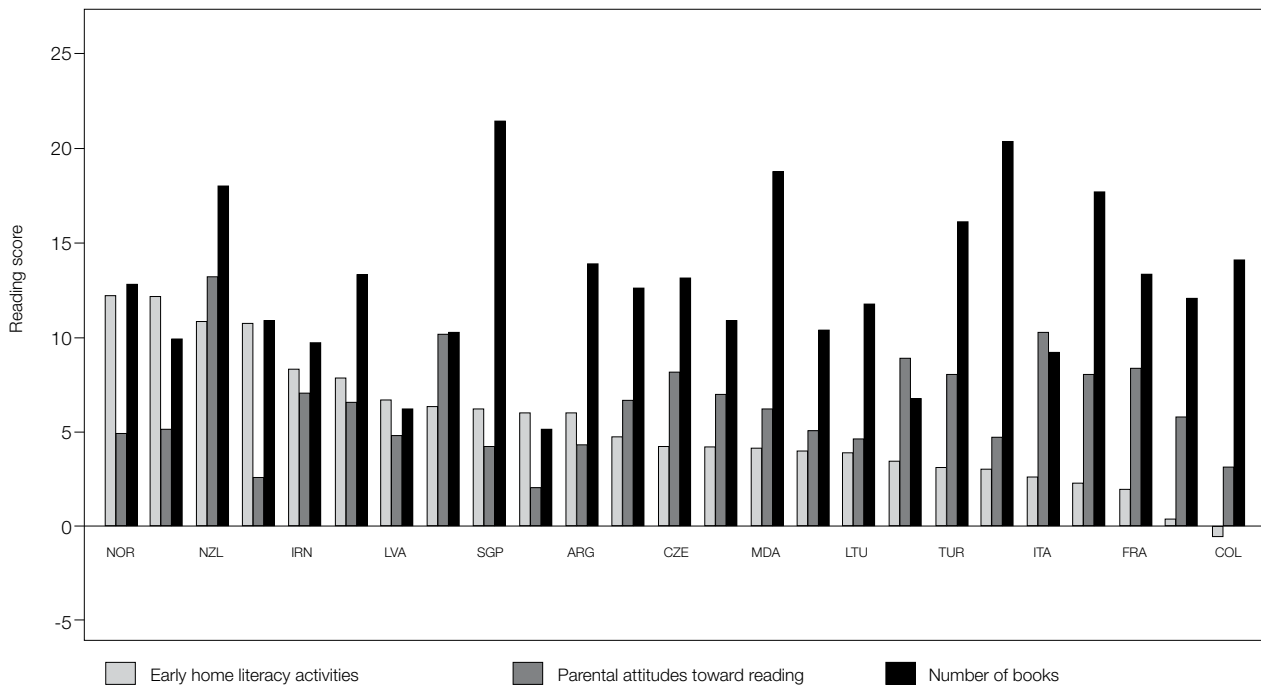


Figure 3: Relative Effects of Home Literacy Environments



level, I used the two-level model technique. Table 3 presents the results of the two-level models. The table shows whether the within-country effect of each home literacy environment measure varied significantly in relation to the country's economic development level as indicated by GDP per capita. To simplify presentation, the effects of other student-level variables are not presented. In the first column, the index of early home literacy activities, the average within-country effect of early home literacy activities is 19.173 for countries with average GDP per capita.⁴ The cross-level interaction term between the effect of early home literacy activities and GDP per capita is significantly positive, indicating that the contribution of early home literacy activities to children's reading performance increases along with increases in the country's economic development level. The table also shows that the total between-country variance of the slope of early home literacy activities is 52.4 and that about 17% of this variance is explained by GDP per capita. Note, however, that the impact of GDP per capita on the slope of early home literacy activities is

not quite as substantial, given that the slope increases only by 0.26 per \$1,000 increase in GDP per capita.⁵

The second column in the table shows some evidence of a relationship between the within-country effect of parental attitudes toward reading and the GDP per capita. Although significant at only the 90% level, the impact of GDP per capita on parental reading attitudes is positive, indicating the greater contribution of parental reading attitudes on children's reading achievement in economically developed countries. GDP per capita accounts for 12.5% of the total between-country variance of the slope for parental reading attitudes. Although evidence indicates the significant cross-national variation in the effect of parental reading attitudes by the country's economic development level, the magnitude of the impact of GDP per capita is quite modest, similar to the case for early home literacy activities: the effect of parental reading attitudes increases only by 0.13 per \$1,000 increase in GDP per capita.

The third column in Table 3 shows that the interaction term between GDP per capita and the slope

⁴ Remember that GDP per capita was centered on the grand mean in the country-level model.

⁵ For the analyses, GDP per capita was scaled as the unit of \$1,000.

Table 3: Cross-National Variation in the Effect of each HLE Measure by GDP per Capita (Two-Level Model)

	EHLA	PATR	BOOK
Index of Early Home Literacy Activities	19.173 *** (1.464)		
X GDP per capita (\$1,000 unit)	0.262 * (0.124)		
Index of Parental Attitudes toward Reading		14.061 *** (0.844)	
X GDP per capita (\$1,000 unit)		0.134 ^ (0.072)	
Number of Books at Home			11.373 *** (0.683)
X GDP per capita (\$1,000 unit)			0.020 (0.063)
Variance of the slope of each HLE measure without interaction with GDP per capita	52.400 ***	15.566 ***	10.357 ***
Variance of the slope of each HLE measure with interaction with GDP per capita	43.404 ***	13.624 ***	10.346 ***
% of variance explained by GDP per capita	17.2%	12.5%	0.1%

Notes: The effects of other student-level variables are not presented for simplicity.

*** $p < .001$ ** $p < .01$ * $p < .05$ ^ $p < .10$.

of the number of books at home is not statistically significant. In other words, although the country variation in the effect of number of books on children's reading performance is considerable (variance = 10.357, $p < .001$), the economic development level explains little of the cross-national variation in the effect of number of books.

Discussion

The results of this study show significant roles of home literacy environments for enhancing children's reading performance. Even after controlling for parental education and other individual characteristics, I found the index of early home literacy activities, the index of parental attitudes toward reading, and the number of books at home to be significantly associated with children's reading performance in almost all 25 countries. Although home literacy environments mediate to some extent the effect of parental education, evidence suggests that in many countries a significant proportion of low-educated parents still engage often with the child in literacy activities, have positive attitudes toward reading, and have a large number of books at home.

There was also evidence of considerable cross-national variations in the effect of each home literacy

measure, and of a systematic association between countries' economic levels and the effect of early home literacy activities: the higher the economic level, the stronger the effect. A similar pattern was observed for the effect of parental attitudes toward reading, but the association seems rather weak. The country's economic level explained little of the cross-national variation in the effect of the number of books at home.

Which countries are more successful in maintaining low degrees of educational inequality associated with family background? Which factors may explain such cross-national variations in the degree of educational inequality? These questions, which can be properly addressed by large-scale cross-national comparative research, are important from both theoretical and practical perspectives. Comparative studies can address the specific ways in which the effects of family background on children's educational outcomes are contingent on national contexts. Understanding how national contexts mediate educational stratification can help educational policymakers and researchers in a society context develop programs and policies that are more effective in reducing educational inequality.

In this regard, future research should pay more attention to identifying important country-level factors beyond the economic level, as these may explain

cross-national variations in the effects of home literacy environments and other cultural resources. One promising candidate would be the structural features of national education systems. Recent studies have found that the institutional characteristics of education systems, such as stratification of school systems and curriculum standardization, shape the influence of

family's socioeconomic status on student reading achievement (Park, 2005). It would be interesting to determine how such institutional arrangements affect the relationship between children's education and cultural resources as well as the family's socioeconomic status.

References

- Bourdieu, P. (1973). Cultural reproduction and social reproduction. In R. Brown (Ed.), *Knowledge, education and cultural change* (pp. 71–112). New York: John Wiley & Sons.
- Bourdieu, P. (1984). *Distinction: A social critique on the judgment of taste*. Cambridge, MA: Harvard University Press.
- Bradley, R. H. (1985). The HOME Inventory: Rationale and research. In J. Lachenmeyer & M. Gibbs (Eds.), *Recent research in development psychology* (pp. 191–201). New York: Gardner.
- Brooks-Gunn, J., Klebanov, P., & Duncan, G. (1996). Ethnic differences in children's intelligence test scores: Role of economic deprivation, home environment, and maternal characteristics. *Child Development*, 67, 396–408.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and data analysis methods*. CA: Sage Publications.
- Crook, C. J. (1997). *Cultural practices and socioeconomic attainment: The Australian experience*. Westport, CT: Greenwood Press.
- De Graaf, N. D., De Graaf, P. M., & Kraaykamp, G. (2000). Parental cultural capital and educational attainment in the Netherlands: A refinement of the cultural capital perspective. *Sociology of Education*, 73, 92–111.
- DiMaggio, P. (1982). Cultural capital and social success: The impact of status culture participation on the grades of high school students. *American Sociological Review*, 47, 189–201.
- Duncan, G., Brooks-Gunn, J., & Klebanov, P. (1994). Economic deprivation and early child development. *Child Development*, 65, 296–318.
- Evans, M. D. R., Kelly, J., Sikora, J., & Tremain, D. J. (2005). *To the scholar go the spoils? The influence of parents' scholarly culture on offspring's occupational attainment in 31 countries*. Paper presented at the annual meeting of the International Sociological Association, Research Committee 28, Los Angeles, August 18–21, 2005.
- Farkas, G. (1996). *Human capital or cultural capital? Ethnicity and poverty groups in an urban school district*. New York: Aldine de Gruyter.
- Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. *Social Science Research*, 33, 464–497.
- Fuwa, M. (2004). Macro-level gender inequality and the division of households in 22 countries. *American Sociological Review*, 69, 751–767.
- Martin, M. O., Mullis, I. V. S., & Kennedy, A. M. (Eds.). (2003). *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report*. Chestnut Hill, MA: Boston College.
- Organization for Economic Co-operation and Development (OECD). (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: Author.
- Park, H. (2005). *Cross-national variation in the effects of family background and school on student achievement: The relevance of institutional and policy contexts*. Unpublished dissertation, University of Wisconsin-Madison.
- Powell, B., Werum, R., & Steelman, L. C. (2004). Linking public policy, family structure, and educational outcomes. In C. Dalton & K. Albright (Eds.), *After the bell: Family background, public policy and educational success* (pp. 111–144). New York: Routledge.

Student government and voluntary organizations: A comparative study of Australia and the United States¹

Gary Homana and Jeff Greene
University of Maryland
College Park, Maryland, USA

Introduction

To maintain a healthy democratic society, active political participation is essential. Yet, while research suggests that support for the democratic process among adults remains strong across many countries (Klingemann, 1999), traditional forms of political participation such as voting, signing petitions, and contacting political representatives appear to have declined, especially in industrialized nations (Dalton, 1999). This decline appears especially pervasive among younger cohorts.

In the United States, for example, voter turnout among youth has declined significantly since 1972, the year when 18- to 21-year-olds were first permitted to vote (US Bureau of the Census, 2002). According to Keeter and colleagues (2002), about 66% of youth and young adults have never engaged in civic activities such as contacting an elected official, participating in a protest or demonstration, or writing a letter to a newspaper. At the same time, in Australia, as part of the 1999 IEA Civic Education Study in that country, Mellor, Kennedy, and Greenwood (2001) found that 86% of the adolescents surveyed expected to vote. This figure is not surprising, however, since voting is compulsory in Australia. Beyond voting, 87% of the students did *not* plan on joining a political party, another 87% did *not* consider being a candidate for a government office, and 76% had *no* plans to write letters to a newspaper about social or political concerns. McAllister's (1998) Australian study further suggests that civic education programs that integrate civics into the school curriculum may be necessary to secure youth political participation beyond simply voting. In fact, recent studies confirm that thoughtful and respectful discussion of political issues in these types of programs not only leads to increased student expectations of voting as adults, but also increased community activism, political interest, and commitment to the

rights and responsibilities of citizenship (Campbell, 2005; Kahne, Rodriquez, Smith, & Thiede, 2000). To gain an understanding of what is happening in the civic engagement and education of adolescents worldwide, the 1999 International Association for the Evaluation of Educational Achievement's (IEA) Civic Education Study (CivEd) measured political knowledge, attitudes, and engagement among almost 90,000 14-year-old students in 28 countries (Torney-Purta & Amadeo, 2004; Torney-Purta, Lehmann, Oswald, & Schulz, 2001). This paper provides a secondary analysis of this database to examine youth civic participation more closely. More specifically, this current cross-national study examines student government and voluntary participation among 14-year-olds in Australia and the United States in order to investigate how these patterns are associated with expected adult political participation.

To understand meanings, clarify findings, and provide insights regarding youth political participation, a theoretical orientation for the study is necessary. According to Pittman (2001), the goal of youth development is to develop a range of competencies that will allow young people to engage in all aspects of life, including civic engagement and participation. This goal suggests that youth participation requires situated opportunities that allow members to identify, share, and develop a context for learning.

As participants within social contexts, people develop a network of relationships that facilitates a sense of belonging and community that can influence adult political participation. Through these engagements and interactions, identity becomes a social process. Wenger (1998) provides a comprehensive model of the social processes leading to involvement, called legitimate participation in communities of practice. The communities of practice model represents a broad

1 The comments and assistance of Judith Torney-Purta are gratefully acknowledged.

conceptual framework that includes how people make meaning out of their lives and the world, the way they interact to sustain mutual agreement, how community is defined and formed by its members, and the way identity is shaped within the context of community. In essence, the communities of practice model reflects who we are and the way we talk, work, plan, and engage in collective action to address common issues. We could expect a more developed notion of communities of practice among youth, particularly one with explicit or implicit political content, to have an impact on development extending into adulthood (see Torney-Purta et al., 2001, ch. 2, for further discussion of this model).

Induction into civic communities of practice seems to be lacking among today's youth, and perhaps this is associated with decreased political engagement and socialization. Putnam (2000), arguing from a somewhat narrower set of constructs, suggests that social capital is essential to creating the network of relationships necessary for civic participation and sustaining democracy. Social capital emphasizes that the greatest resources available to human beings are the social relationships that exist among them. These resources include factors such as trust, goodwill, and fellowship, each of which allows for the development of shared norms and values in society. Accumulation of these resources affects the ability of individuals in a society to interact with one another, serves as a base of understanding, and helps create a reserve of support for social, economic, and political structures within that society. Thus, it would seem that the integration of young people into learning communities where they could develop a political and civic identity, surrounded by peers and mentors, could be a positive step toward encouraging future political and civic involvement.

There is strong theoretical support for exploring the idea that the communities into which youth enter, and the relationships those communities foster, have an impact on youths' preparation as adult members of society, especially, how these experiences influence networks and expectations of adult civic behavior. This cross-national secondary analysis of data from CivEd examines the influence of student government and voluntary participation communities of practice on projected adult political participation and trust among 14-year-olds in Australia and the United States. Our hypotheses are that youth participation in

these communities of practice will be associated with increased levels of expected political participation. If this is the case, this study will give research-grounded support for continued investment of resources and effort into these types of programs among youth. By examining data from both the United States and Australia, we hope to support the robustness of the notion of communities of practice.

Research on youth participation in student government and voluntary organizations

In theory, student participation in extracurricular affairs, and especially student government, should enhance adult political participation, especially to the extent that such participation involves voting or decision-making on issues important to the school community. For example, participation in a student council should influence the likelihood of students' voting as adults, with an indirect impact through the development of civic knowledge. Similarly, while evidence indicates that youth are abandoning traditional forms of political participation (US Bureau of the Census, 2002; The Tarrance Group, Inc., & Lake, Snell, Perry and Associates, 1999), they are engaging in alternative forms of civic participation such as voluntary organizations, community service, and service-learning in record numbers, suggesting the potential of these organizations and activities for promoting political participation (Independent Sector, 2002; Keeter et al., 2002; Torney-Purta et al., 2001).

At the same time, less evidence is available regarding the impact of student involvement in governance and voluntary associations on levels of political trust or types of engagement other than voting. As shown below, the literature reveals modest and sometimes even conflicting research in the areas of student government and voluntary organization participation.

Impact on political participation

One question that precedes the issue of the impact that high school student governments have on future political participation is whether students enter high school with political interests, and if those interests influence the degree and capacity of their involvement in student political organizations. Research suggests that students develop political attitudes and concepts well before high school (Hess & Torney, 1967). In addition, Eyler (1982) found that students' political

attitudes as they entered high school predicted their involvement in high school politics, suggesting that high school students do engage in political ideation.

A number of studies provide evidence of the impact of high school activity on future adult political participation. In their panel study, Beck and Jennings (1982) found a direct link between high school activities and young adult participation. High school activities also had a small but significant indirect impact on young adult participation through the mediating variables of young adult civic orientation and youth civic orientations. Other studies found that high school extracurricular participation had a direct impact on adult participation in voluntary associations, and was a strong predictor of voting behavior independent of other factors such as adult socioeconomic status (Hanks & Eckland, 1978). Otto (1976) found that adolescent social integration was the strongest individual predictor of adult social integration, including political participation. Sigel and Hoskin (1981) showed that student participation in high school extracurricular activity positively influenced both voting and non-voting activities. All of these studies provide some support for the belief that involvement in student government can positively influence later political behavior, although more specific and statistically rigorous studies are needed. In addition, many of these studies were conducted with a previous generation and may not reflect current thinking among today's youth.

More narrowly defined research supports the findings of the broader studies. Hanks' (1981) survey of high school students during their senior year, and again two years later, showed that high school political participation had a direct effect on all measures of adult political participation, including discussion of political issues, campaign participation, and voting. Glanville's (1999) study of extracurricular participation and political activity in early adulthood found, after controlling for personality and adult voluntary organization membership, that extracurricular activities had a direct positive effect on the likelihood of working for a campaign, attending political events, and providing monetary support for political campaigns. Smith (1999) showed that extracurricular participation in Grade 12 significantly increased the likelihood of young adult political participation. Finally, Damico, Damico, and Conway

(1998) investigated why women participated more in extracurricular activities in high school but less in politics as adults. They found that participation in student government differed by curriculum track, with more academically talented students engaging in higher numbers. They also found that high school participation in extracurricular activities increased the likelihood for political involvement later in life.

Verba, Schlozman, and Brady (1995) produced the most direct study of the effects of participation in student government. While the data on high school participation was retrospective, it nonetheless showed that participation in high school government was strongly associated with later political involvement. This was one of the few studies to clearly identify student council involvement as an independent factor and to analyze the impact of this involvement on adult political behavior.

Research regarding youth volunteerism and political participation exists primarily in studies focused on community service and service-learning. It seems logical that involvement in these types of activities leads to increased political participation. However, while studies have found support for increased future volunteering, greater awareness of community needs, and commitment to service (Melchior, 1999; Perry & Katula, 2001; Simon & Wang, 2002), the effects of participation in these types of activities on political participation do not appear substantial.

Several studies have examined the impact of volunteer programs on political, social, and moral attitudes and behaviors, important antecedents of political participation. Finkel (1985), in a longitudinal study of the Survey Research Center's 1972–1974–1975 election study, corroborates other research suggesting that political participation can serve as support for political institutions by encouraging behaviors and attitudes associated with incremental social change. Yates (1999), in her study of political-moral engagement among high school students who volunteered as part of a year-long course in social justice, found that students discussed issues related to moral responsibility and their own capacity to impact change through political participation. Giles and Eyler (1994) reported that among volunteer college students, self-efficacy rose significantly, and that the students aspired to leadership roles so as to have an impact on the political system. Finally, Roker, Player,

and Coleman (1999) argued that youth participation in volunteering and campaigning promotes political awareness, knowledge, and understanding. They suggested that gender, ethnicity, family, friends, locality, and religion influence youth social and political participation.

Prior research makes a case for a connection between membership in student government in high school and adult political participation. Unfortunately, it is important to point out that each study reviewed suffers from at least one limitation, for example, a failure to isolate student government from other activities, a dependence on retrospective data collection, or a failure to differentiate different types of participation. In addition, there is a dearth of studies on involvement in student government other than in the United States.

There is also evidence to support the relationship between volunteer programs and increased cognitive ability, future volunteering, political, moral, and civic identity, and social and civic responsibility. Several of the studies demonstrated that, compared to non-members, volunteers display more active forms of participation, as well as increased civic and democratic attitudes. Few empirical studies, however, support the connection between these programs and political participation such as voting, and writing to elected officials, especially on a cross-national level. Because CivEd is the most comprehensive database on the civic education of adolescents, this study serves as an opportunity to more closely examine the relationships between student government participation, volunteerism, and expected adult political participation.

Impact on political trust

A thorough review of available sources revealed no systematic studies of the influence of participation in student government on levels of political trust. Certainly, some studies looked at democratic values (Damico et al., 1998; Otto, 1976), but this is not the same as generalized trust in the political system or specific trust in certain political actors or organizations. Therefore, this study fills a gap in the research concerning the less tangible impact of student government participation.

Similarly, there is limited evidence regarding the impact of student involvement in voluntary

associations on political trust. Stolle (1998) determined that more diverse and engaged associations and those with weak ties include more trusting people and that these individuals may actually self-select into these organizations. The author suggested that little is known about the process that makes members in voluntary associations more trusting and cooperative, suggesting further research is necessary. In another study, La Due Lake and Huckfeldt (1998) found that social capital, which in part is characterized by trust, is generated through personal networks and that increasing levels of politically relevant social capital increases the likelihood of individual political participation. They acknowledged, however, that individuals who are politically engaged may self-select into these activities.

Brehm and Rahn (1997) analyzed the General Social Surveys from 1972 through 1994 to determine the causes and consequences of civic engagement and interpersonal trust. They determined that there is a positive, reciprocal connection between the two factors and that the relationship is stronger from civic participation to trust, rather than the reverse. They contended that the nature of the relationship was such that it could result in either a “virtuous” cycle, where both factors lead to positive outcomes for civic engagement and trust, or a “vicious” cycle, leading to negative outcomes.

Finally, in an international study of primarily European countries, Muller and Seligson (1994) looked at democratic change across two time periods—1972–1980 and 1981–1990. In general, interpersonal trust seemed to be an effect, not a cause, of democracy. In fact, the authors suggested that a high level of trust is unrelated to a country’s level of democracy, and that low levels of trust may not impede the process.

In conclusion, studies of the impact of volunteer organizations on political trust are limited and somewhat contradictory. As a result, we consider this study provides opportunity to examine the relationship more definitively, specifically on a cross-national level.

Summary

Overall, the literature regarding the impact of adolescent participation, student government, and voluntary organizations on future political behavior is inconsistent and difficult to compare across studies. The

IEA Civic Education Study, however, provides a single, comprehensive source of data on the civic behaviors, attitudes, skills, and knowledge of adolescents in schools worldwide. It provides a means to examine these phenomena cross-culturally, using a common instrument created and validated by a multinational group of scholars. Utilizing this opportunity, this paper examines the relationship between measures of student government and voluntary participation and projected adult political participation and trust among 14-year-olds in Australia and the United States. The particular aim is to investigate if these forms of civic involvement are likely to lead to more politically engaged adults. Our hypotheses are that involvement in both student government and voluntary organizations will be associated with higher levels of expected political participation and trust, and that this effect will be consistent across countries' different cultural contexts.

Method

The IEA study

In 1994, the International Association for the Evaluation of Educational Achievement (Amsterdam) began planning a study of civic education. A case study phase from 1994 to 1998 used qualitative data to craft an instrument that was then used, in 1999, to measure political knowledge, attitudes, and engagement. In all, 28 countries and almost 90,000 students participated in the study of 14-year-olds. Publications and data from the IEA study are available at <http://www.wam.umd.edu/~iea>

Countries and variables chosen for this analysis

The United States and Australia were chosen as data sources for the current investigation for several reasons. Each country varies in its orientation to political issues and involvement. Thus, our interest in the two countries was both for investigation of civic education and possible differences in the process that might be conditioned by culture. In addition, we were interested in examining countries with high levels of student involvement. In Australia, 34% of students participating in CivEd reported taking part in student council or government; 33% of students in the United States did likewise. In terms of voluntary organization participation, 33% of Australian students and 50% of United States students reported involvement. These

rates of involvement were among the highest of all 28 countries.

The independent variables chosen for this study included student government involvement, volunteer organization participation, and gender. Student government or school council involvement and volunteer organization involvement were measured with single items on the IEA survey (Torney-Purta et al., 2001). Item A, within a larger section of items regarding participation in various organizations, asked whether the student had participated in "a student council/student government [class or school parliament]" and item H asked about participation in "a group conducting [voluntary] activities to help the community."² Both questions had no/yes responses.

The dependent variables in the study were political trust and anticipated adult political participation. Trust was measured using the mean of scores to three questions on the IEA survey. The three questions were grouped into a section with a stem that asked: "How much of the time can you trust each of the following institutions?" The three items of interest were: "The national [federal] government [in _____ (the national seat of government)]," "the local council or government of your town or city," and "National Parliament [Congress]." Responses to each included "never," "only some of the time," "most of the time," "always," and "don't know." The scores for each item ranged from 1 for "never" to 4 for "always," with "don't know" coded as 0 (treated as missing). These questions had been part of a scale for trust developed for the original analysis (Torney-Purta et al., 2001).

Adult political participation was measured using two scales: informed voting and conventional participation. Past analyses of the IEA data had shown the validity and exclusivity of these two separate scales for adult participation (Torney-Purta, Barber, & Richardson, 2004). Both scales used responses to items in the "Political Action 2" section of the IEA instrument. This section asked students to answer the question "When you are an adult, what do you expect that you will do?" in reference to various political activities. Item response choices were scored from 1 to 4 and read as: "I will certainly not do this," "I will probably not do this," "I will probably do this," and "I will certainly do this," with "don't know" treated as missing. The expected informed voting scale

2 Bracketed wordings were provided for translators.

was created using the mean of answers to the items "Vote in national elections" and "Get information about candidates before voting in an election." The conventional participation scale was created using the mean sum of answers to the items "Write letters to a newspaper about social or political concerns" and "Be a candidate for a local or city office." Each score could vary from 1 to 4. Students with missing data on any of the above items were removed from the analysis.

Analysis

Analyses of variance (ANOVA) were conducted to explore the hypotheses of interest in this study. Specifically, we hypothesized that student government and voluntary organization participation would be significantly related to more positive answers to political trust and both adult political participation scales. Gender was included, as it had been of interest in other IEA analyses, but we had no hypotheses about its impact.

Results

SPSS 11.0 was used for all analyses. Cronbach's α for the trust, informed voting, and conventional participation scales were .71, .74, and .69, respectively. ANOVA analysis requires a check for the homogeneity of variance among the variables (Pedhazur, 1997). These tests were conducted, and while no variable met this assumption, the results were considered valid due to three factors. First, sample sizes were very high, ranging from 300 to 1,000 per cell. Second, the standard errors of the cells were such that the higher errors were associated with larger sample sizes, thus creating heterogeneity of variance but actually biasing the results against finding statistically significant effects. As long as the larger standard error belongs to the group with the larger sample size, the analysis retains its conservative nature and a p -value of .05 can be used as a reasonable criterion (Lomax, 2001). Finally, as can be seen below, the p -values for our findings are all well below .05, further substantiating confidence in the tests of statistical significance.

Trust analyses

The main effect for student government participation was statistically significant in both Australia and the United States (see Table 1). In both countries, involvement in student government was statistically

significantly associated with higher scores on the trust scale. In addition, in the United States, girls' mean trust was higher than boys'. However, the main effects in the United States must be qualified due to the presence of a statistically significant interaction. Post-hoc analyses showed that boys not involved in student government had a statistically significantly lower mean trust score than did all three other groups. There were no statistically significant differences between the other groups. Parallel results were found with voluntary organization participation and trust. In Australia, the only statistically significant difference was a main effect, with those participating in voluntary organizations having a higher mean trust score than those who did not participate. In the United States, main effects showed higher means for those involved and for girls. The interaction was also significant, however, qualifying these results in the same way as above, with boys who were not involved having a mean trust score statistically significantly lower than the mean trust scores for all other groups.

Informed voting

In both Australia and the United States, main effects demonstrated that those who participated in student government had a statistically significantly higher mean likelihood of voting and getting information about candidates than those who did not; girls had a higher informed voting score than boys (see Table 1). Statistically significant interactions qualified these results in both countries, however. In Australia, boys and girls involved in student government had statistically significantly higher informed voting mean scores than females not involved, who in turn had statistically significantly higher means than boys who were not involved. In the United States, the only statistically significant interaction effect was for boys not involved. Their mean scores were significantly lower than the mean scores for all other groups.

On examining the results by voluntary organization participation, we found that in both Australia and the United States, those participating had statistically significantly higher mean informed voting scores than those who did not participate. In addition, girls had statistically significantly higher mean scores than boys. No significant interactions were found. The total model effect sizes for these models were small (see Table 1).

Table 1: ANOVA Results for Australia and the United States

		Student government			Voluntary organization		
		Trust	Informed voting	Conventional participation	Trust	Informed voting	Conventional participation
AUSTRALIA							
Participation	Yes	2.733	3.310	1.970	2.688	3.254	2.017
	No	2.584	3.058	1.778	2.598	3.084	1.757
	<i>F</i>	23.949**	59.671**	40.696**	8.225**	25.625**	70.159**
Gender	Male	2.670	3.148	1.865	2.662	3.121	1.891
	Female	2.647	3.219	1.883	2.624	3.217	1.884
	<i>F</i>	.583	4.794*	.349	1.483	8.053**	.050
Interaction	Yes/Male	2.743	3.310	1.976	2.667	3.208	1.781
	No/Male	2.552	2.986	1.754	2.581	3.034	1.987
	Yes/Female	2.723	3.309	1.965	2.709	3.301	1.734
	No/Female	2.617	3.130	1.801	2.615	3.133	2.048
	<i>F</i>	.813	4.925*	.919	.021	.011	3.087
df		1,1994	1,2428	1,2374	1,1973	1,2407	1,2352
Effect size ^a		.012	.028	.016	.004	.015	.029
UNITED STATES							
Participation	Yes	2.821	3.376	2.100	2.821	3.324	2.087
	No	2.738	3.219	1.920	2.709	2.988	1.863
	<i>F</i>	8.885**	57.059**	15.583**	17.302**	111.286**	53.495**
Gender	Male	2.745	3.101	2.028	2.728	3.061	1.952
	Female	2.814	3.297	1.991	2.802	3.251	1.998
	<i>F</i>	6.159*	33.765**	1.304	7.574**	35.782**	2.223
Interaction	Yes/Male	2.817	3.278	2.134	2.815	3.208	2.058
	No/Male	2.673	2.925	1.848	2.641	3.034	1.846
	Yes/Female	2.826	3.376	2.065	2.827	3.301	2.117
	No/Female	2.803	3.219	1.991	2.777	3.133	1.879
	<i>F</i>	4.649*	8.481**	10.797**	5.277*	1.922	.176
df		1,2166	1,2408	1,2221	1,2143	1,2389	1,2197
Effect size ^a		.01	.048	.020	.015	.067	.026

Notes: * $p < .05$; ** $p < .01$.^a Adjusted R^2 for entire model.

Conventional participation

In the student government analyses, the only statistically significant difference for Australia between conventional participation scores was that those involved had higher mean scores than those who were not involved (see Table 1). There was no significant interaction. In the United States, main effects showed involvement was associated with higher mean conventional participation scores compared to those for students not involved, and that girls had higher mean scores than boys. A statistically significant interaction in the United States qualified these results, however, showing that uninvolved boys had the lowest scores of

all groups, while uninvolved females had lower scores than males involved in student government.

In terms of voluntary organization participation, the only statistically significant results were main effects for involvement in both countries. Those involved in voluntary organizations had statistically significantly higher conventional participation mean scores than those who were not involved.

Summary

The hypotheses of this study were supported. In each country, student government and voluntary organization involvement seems to be associated with positive civic engagement outcomes such as

increased trust, informed voting, and conventional participation. This is positive news for advocates of school-based civic initiatives. A concerning finding was that in the United States males not involved in student government had the lowest average scores on all three measures. Gender main effects for voluntary organization involvement were found less often, but when significant, they suggested that girls had higher scores than boys. Only in the trust analysis was there a significant interaction, and its results were similar to many of the United States' findings that the trust scores for uninvolved boys were lower than the trust scores for all other groups. In addition, it is evident that the positive effects of student government and voluntary organization participation transcend cultural context, thus providing support for the idea that these communities of practice provide an "element" fundamental to youth.

Implications

Although young people's involvement in traditional forms of political participation has declined, the research presented here suggests that this trend might be ameliorated through increased participation in student government and voluntary organizations. In both Australia and the United States, participation in these activities was positively statistically significantly associated with more favorable civic outcomes, including increased trust and greater expectations to be an informed voter and an active citizen. Also, the character of these activities seemed to be fulfilling some of the aims of those who organize them—student councils solving school problems and volunteer organizations building a positive sense of obligation to the community. Thus, these communities of practice seem to be transcending culture, and adding to the social capital of their larger communities, with the aforementioned benefits related to youth's expected adult political participation.

These results also suggest the importance of voluntary organization participation in promoting civic engagement. As the research review has shown, youth are involved in a wide range of non-traditional volunteer experiences to benefit the community. These experiences can positively influence youths' likelihood of long-term community organization participation. The findings reinforce the fact that youth are seeking alternative forms of civic engagement to address

problems in their communities and suggest that these types of experiences could be redesigned to positively influence traditional civic behaviors, attitudes, skills, and knowledge.

Through its examination of the cross-cultural nature of these issues, this study has raised new areas for research. Should we expect that these student government and voluntary organization structures and organizations have similar expectations and standards across all cultures and countries? If they exist, what are the roles of these programs in different countries? Does the notion of "voluntary" mean different things in different countries? In essence, how different are these experiences across countries? Asking questions and studying these types of areas will provide a better understanding of how to foster civic engagement and hopefully will lead to a firmer grasp of the nature of the civic environment itself. Such an understanding could provide guidance for policymakers and school leaders looking to address the declining political interest of youth and to spur young people toward greater civic engagement and responsibility.

Previous research and this study all support the idea that involvement, be it student council based or within a voluntary organization, has the potential to lead to positive civic engagement, especially for adolescent males who are particularly susceptible to the negative effects associated with a lack of involvement. Given other research showing similar concerns (Barber, 2004; Torney-Purta, 2004), it would seem that communities must provide more opportunities for boys to become involved, and support systems to assist them in that process.

So how and where can youth, particularly boys, become involved in student government and voluntary organizations? We suggest that schools remain the best option to promote political participation among our young people. Torney-Purta (2002) has found that schools that rigorously teach civic content and skills and promote democratic classrooms to encourage participatory action among its members achieve high-quality results in civic engagement outcomes. Schools can provide rich opportunities for students to learn and engage in interactions with other students and adults, to forge a community based on common values, respect, trust, and acceptance of human diversity, and to participate in voluntary associations that encourage civic knowledge, attitudes, and practices. Rather than

reducing opportunities for integration within learning communities of practice with peers and mentors, schools should provide more avenues for young people to become involved with political and civic organizations. Such organizations can help students develop self-efficacy and instrumentality, and build social capital.

The concept of communities of practice (Lave & Wenger, 1991, 2002; Wenger, 1998) suggests that schools can become vibrant learning communities that cultivate civic knowledge or engagement and encourage youth to actively participate in voluntary associations for democratic ideals. This process requires schools to provide youth with opportunities for learning that facilitate robust partnerships within which students are involved in an exchange of shared decisionmaking that enhances their motivation to participate in associations that contribute to deliberative democracy. It also requires creating democratically engaged communities where characteristics and traits are enmeshed into a school's social fabric so that the qualities of citizenship and political participation are embraced and practiced by all its members. As we have seen earlier, one mechanism to help accomplish this goal is through the integration of volunteer experience with school discussion of community problems.

As we have noted, however, not all service-related experiences support political engagement. If service to the community is to become a viable option for citizenship and political participation, it will require design strategies that encourage youth to explore and develop these skills. By linking service to political engagement, these types of activities could become a particularly fruitful approach to developing civic and political practice among youth because it would help them acquire a democratic self through deliberate practice. More specifically, these types of experiences would help students acquire an understanding of democracy and their role in it, and how to participate as a member of a democratic society for the common good of that society.

What are the implications for future research on student government and voluntary association participation of students in schools? First, our

review of the literature indicates a need for studies investigating the influence of voluntary organization participation on political involvement of youth. This need suggests that research examines issues related to citizenship, democracy, and civic engagement, with the aim of ensuring more effective outcomes in these areas. Second is the need for civic education researchers to establish scientific inquiry norms. And, third, researchers should explore promising theories and practices from a range of other fields and consider their application across civic education research.

Furthermore, the IEA study database offers opportunities to examine measures related to voluntary association participation through a variety of analyses. It allows, for example, examination of relationships concerning students' participation in government-related responsibilities, social movement-related citizenship, and attitudes toward one's nation, women's political economic rights, and political activities. In short, CivEd provides tremendous opportunities for practitioners, researchers, and policymakers to explore adolescent civic knowledge, attitudes, and behavior. As they engage in this work, it is our hope that they will delve more deeply into the role of voluntary organizations as a way to understand adolescent civic practice and its impact on future adult political participation (Torney-Purta et al., 2004).

Conclusion

The findings of this study support the idea that youth involvement influences a person's predicted adult behavior. Indeed, as demonstrated by this work, the IEA Civic Education Study serves as a potential wealth of information about civic education endeavors within countries, and as a tool to compare what is universal about that endeavor. By engaging in statistically rigorous, internationally informed research, this IEA study, and hopefully future studies like it, can inform longitudinal work on adolescent through adult political behavior. This information, in turn, should lead to concrete suggestions for ways schools can promote future political and civic engagement through extracurricular involvement in learning communities of practice.

References

- Barber, C. H. (2004). A cross-national analysis of the relation of educational trust to expected educational attainment. In I. Papanastasiou (Ed.), *Proceedings of the IRC-2004 CivEd & SITES*. Nikosia, Cyprus: University of Cyprus Press.
- Beck, P. A., & Jennings, M. K. (1982). Pathways to participation. *The American Political Science Review*, 76, 94–108.
- Brehm, J., & Rahn, W. (1997). Individual-level evidence for the causes and consequences of social capital. *American Journal of Political Science*, 41, 999–1023.
- Campbell, D. E. (2005). *Voice in the classroom: How an open classroom environment facilitates adolescents' civic development*. University of Maryland: Center for Information and Research on Civic Learning and Engagement. Available at <http://www.cived.net>
- Damico, A. J., Damico, S. B., & Conway, M. M. (1998). The democratic education of women: High school and beyond. *Women & Politics*, 19, 1–32.
- Dalton, R. J. (1999). Political support in advanced industrialized democracies. In Pippa Norris (Ed.), *Critical citizens: Global support for democratic governance* (pp. 55–57). Oxford: Oxford University Press.
- Eyler, J. (1982). Test of a model relating political attitudes to participation in high school activities. *Theory and Research in Social Education*, 10, 43–62.
- Finkel, S. E. (1985). Reciprocal effects of participation and political efficacy: A panel analysis. *American Journal of Political Science*, 29, 891–913.
- Giles, D. E., & Eyler, J. (1994). The impact of a College community service laboratory on students' personal, social, and cognitive outcomes. *Journal of Adolescence*, 17, 327–339.
- Glanville, J. L. (1999). Political socialization or selection? Adolescent extracurricular participation and political activity in early adulthood. *Social Science Quarterly*, 80, 270–290.
- Hanks, M. (1981). Youth, voluntary associations and political socialization. *Social Forces*, 60, 211–223.
- Hanks, M., & Eckland, B. K. (1978). Adult voluntary associations and adolescent socialization. *The Sociological Quarterly*, 19, 481–490.
- Hess, R. D., & Torney, J. V. (1967). *The development of political attitudes in children*. Chicago: Aldine Publishing Company.
- Independent Sector. (2002). Giving and volunteering in the United States. Available at <http://www.independentsector.org/programs/research/gv01main.html>
- Kahne, J., Rodriguez, M., Smith, B., & Thiedes, K. (2000). Developing citizens for democracy? Assessing opportunities to learn in Chicago's social studies classrooms. *Theory and Research in Social Education*, 28, 311–338.
- Keeter, S., Zukin, C., Adolina, M., & Jenkins, K. (2002). *The civic and political health of the nation: A generational portrait*. College Park, MD: The Center for Information and Research on Civic Learning and Engagement.
- Klingemann, H-D. (1999). Mapping political support in the 1990s: A global analysis. In P. Norris (Ed.), *Critical citizens: Global support for democratic governance* (pp. 31–56). Oxford: Oxford University Press.
- La Due Lake, R., & Huckfeldt, R. (1998). Social capital, social networks, and political participation. *Political Psychology*, 19, 567–583.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Lave, J., & Wenger, E. (2002). Legitimate peripheral participation in communities of practice. In M. Lea & K. Nicol (Eds.), *Distributed learning*. London: Routledge/Falmer (pp. 56–63).
- Lomax, R. G. (2001). *Statistical concepts: A second course for education and the behavioral sciences* (2nd ed.), Mahwah, NJ: Lawrence Erlbaum Associates.
- McAllister, I. (1998). Civic education and political knowledge in Australia. *Australian Journal of Political Science*, 33, 7–23.
- Melchior, A. (1999). *Summary report: National evaluation of Learn and Serve America*. Waltham, MA: Center for Human Resources, Brandeis University.
- Mellor, S., Kennedy, K., & Greenwood, L. (2001). *Citizenship and democracy: Students' knowledge and beliefs: Australian fourteen year olds and the IEA Civic Education Study*. Canberra, ACT: Department of Education, Training and Youth Affairs.
- Muller, E. N., & Seligson, M. A. (1994). Civic culture and democracy: The question of causal relationships. *The American Political Science Review*, 88, 635–652.
- Otto, L. B. (1976). Social integration and the status-attainment process. *American Journal of Sociology*, 81, 1360–1383.
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). Orlando, FL: Harcourt Brace.
- Perry, J., & Katula, M. C. (2001). Does service affect citizenship? *Administration and Society*, 33, 330–365.
- Pittman, K. J. (2001). *Preventing problems, promoting development, encouraging engagement: Competing priorities or inseparable goals?* Takoma Park, MD: The Forum for Youth Investment.
- Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.
- Roker, D., Player, K., & Coleman, J. (1999). Young people's voluntary and campaigning activities as sources of political education. *Oxford Review of Education*, 2, 185–198.

- Sigel, R. S., & Hoskin, M. B. (1981). *The political involvement of adolescents*. New Brunswick, NJ: Rutgers University Press.
- Simon, C. A., & Wang, C. (2002). The impact of AmeriCorps Service on volunteer participants: Results from a two-year study in four western states. *Administration and Society*, 34, 522–540.
- Smith, E. S. (1999). The effects of investments in the social capital of youth on political and civic behavior in young adulthood: A longitudinal analysis. *Political Psychology*, 20, 553–580.
- Stolle, D. (1998). Bowling together, bowling alone: The development of generalized trust in voluntary associations. *Political Psychology*, 19, 497–525.
- The Tarrance Group, Inc., & Lake, Snell, Perry and Associates. (1999). *New Millennium Project—Phase 1: A nationwide study of 15–24 year-old youth*. Washington, DC: National Association of Secretaries of State.
- Torney-Purta, J. (2002). The school's role in developing civic engagement: A study of adolescents in twenty-eight countries. *Applied Developmental Science*, 6, 203–212.
- Torney-Purta, J. (2004). Adolescents' political socialization in changing contexts. *Political Psychology*, 25, 465–478.
- Torney-Purta, J., & Amadeo, J. (2004). *Strengthening democracy in the Americas through civic education*. Washington, DC: Organization of American States.
- Torney-Purta, J., Barber, C. H., & Richardson, W. K. (2004). Trust in government-related institutions and civic engagement among adolescents in six countries. *Acta Politica*, 39, 360–406.
- Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen*. Amsterdam: International Association for the Evaluation of Educational Achievement. Available at <http://www.wam.umd.edu/-iea>
- US Bureau of the Census. (2002). Voting and registration in the election of November 2000. Washington, DC: Author. Available at <http://www.census.gov/prod/2002pubs/p20-542.pdf>
- Verba, S., Schlozman, K. L., & Brady, H. E. (1995). *Voice and equality*. London: Harvard University Press.
- Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge, MA: Cambridge University Press.
- Yates, M. (1999). Community service and political-moral discussion among adolescents: A study of mandatory school-based programs in the United States. In J. Youniss & M. Yates (Eds.), *Roots of civic identity: International perspectives on community service and activism in youth* (pp. 16–31). Cambridge: Cambridge University Press.

Predicting the political involvement of European adolescents

Rainer H. Lehmann
Humboldt University
Berlin, Germany

Introduction

The second IEA Civic Education Study (CivEd) was designed to assess levels of civic knowledge, civic concepts, and society-related attitudes as well as dimensions of political participation among 14-year-old students across diverse educational systems. During 1999, nearly 93,000 14-year-old students from 28 countries from Europe, North America, South America, Asia, and the Pacific region were administered a test of civic knowledge and skills. At the same time, their concepts of democracy, government, and citizenship were surveyed and their political attitudes and activities as well as their expectations of future civic engagement—electoral participation in particular—were ascertained. The major cross-national findings from the study were published by Torney-Purta, Lehmann, Oswald, and Schulz (2001).

During the following year, over 50,000 upper secondary students from 16 countries, ranging in age from 16 to 19 years, received a similar test of civic knowledge and skills, including some new items that were intended to tap more advanced levels and additional topics of civic knowledge. Some of these new items referred to the domain of “economic literacy.” With respect to the survey of civic concepts, attitudes, and expected engagement, the same instruments were used as in the study of 14-year-olds. The respective report was published by Amadeo, Torney-Purta, Lehmann, Husfeldt, and Nikolova (2002). The response patterns of the upper secondary students were in some respects more differentiated than those of the younger population.

Despite some evidence for the influence of the historical context, that is, the national discourse at a given point in time, the similarities within countries across populations were generally insufficient to explain the outcomes of the survey in such terms. This is not to say that respective analyses should be abandoned altogether. The scope of the present paper, however, is restricted to the 14-year-old students. It attempts to shed more light on the relationship between context

and student characteristics by applying more finely graded analytical techniques than was possible in the preparation of the original reports.

Results from the survey of 14-year-old students indicated that students in most of the 28 participating countries had an understanding of fundamental democratic values, processes, and institutions, although arguably this understanding lacked depth. Also, in most of the participating countries, young people agreed that good citizenship includes the obligation to vote. Thus, it was—and continues to be—one of the crucial research issues embedded in the Civic Education Study, to find out what makes students understand democracy as a political system and how these processes relate to their future participation in democratic elections. It is of particular interest here to find out whether—and if so, to what degree—the aim of having knowledgeable and engaged citizens can be achieved through civic instructions and democratic practices in school. In pursuit of these questions, the international report *Civic Knowledge and Engagement* (Amadeo et al., 2002) contained a series of multiple regression models that focused on differences among students across and within countries and in which these two important aspects of citizenship—civic knowledge as measured by the cognitive test and civic engagement operationalized by the likelihood to vote in future elections—served as the dependent variables (Torney-Purta et al., 2001, p. 145 ff). These OLS regression models were aimed at determining the impact of individual factors, home environment, and school practices on young people’s civic literacy and political participation.

The respective research questions—spelled out more fully below—were intrinsically linked to fundamental tenets of theories of social learning. In this frame of reference, the students were primarily seen as participants in concurrent discourses at different levels, offering modes of perception and thinking, goals of aspiration, and models of present and future activity (Torney-Purta et al., 2001, p. 20 ff).

In their analysis of the young people's intention to vote as adults, Torney-Purta et al. (2001) found that civic knowledge was the most important independent factor in explaining the expectation of voting in the future (p. 150 ff). Thus, even after controlling for other statistically relevant effects, the students found that students with high scores on the cognitive CivEd test were more likely to express their intention to participate in future elections than were those students with weaker results. Students' reports of having learned in school about the importance of voting in national elections turned out to be the second substantial predictor in the estimated regression model. Frequency of watching news on television as well as perceptions of an open classroom climate for discussions also showed considerable positive associations with the expectation to vote. Some of these findings were quite similar to the multivariate analyses of data from 10 countries investigated in the first IEA Civic Education Study 1971 (Torney, Oppenheim, & Farnen, 1975). Torney and her colleagues suggested that schools do have an important role in fostering civic knowledge and skills and in motivating political participation.

Beyond the academic interest in the emergence of civic knowledge and political involvement, widely shared and growing concerns over diminishing participation rates in public elections justify an investigation of potential electoral participation as a criterion in its own right. If the earlier findings as to the conducive roles of an open classroom climate and learning about the importance of voting at school—apart from civic knowledge—stand closer scrutiny, then this information provides strong justification for having corresponding forms of civic and citizenship education as an obligatory component of school education.

In a review of the existing analyses of the IEA Civic Education data, two aspects of achieving progress over existing work stand out. First, there is certain potential in adding context specificity to the existing, rather general, approaches to data analysis. Second, there are opportunities to profit from data analytic techniques that give credit to the nested structure of the data.

Apparently, previous analysts have found it difficult to relate specific findings to the relevant national contexts—or the respective “national discourses” postulated in the multi-faceted reference model underlying the Civic Education Study of 1999/2000.

One would expect to be able to identify in the data influences of the historical and political settings within which they were collected. It should be noted that in the planning stage of data analysis, item-by-country interaction effects were expected to have influenced the test results. However, the respective search rendered little more than instances of chance variation. Any traces of the historical and political context are therefore more likely to be present at the across-construct level than at the individual-item level. It seems plausible, then, to look for substantive as well as structural similarities among countries whose political orientation includes certain regional affinities (as is clearly the case in, for example, the Nordic and the Baltic countries). As far as can be seen, little work has been done so far to detect such similarities.

The second inadequacy of the existing work in this domain is the current shortage of analyses that pay proper respect to the distinction between individual and class-level effects. Such differentiation is essential with respect to the reported claims as to the potential influence of classroom environments. Published results in terms of individual-level or class/school-level effects clearly need to be corroborated by way of appropriate and readily available multilevel techniques of data analysis. In the original reporting on the key criteria of the study, including the intention to vote in future elections, the multilevel structure of the data was not considered. While the present paper revisits some of the issues left for further research in 2001, its primary intention is to fill a part of this gap.

Research questions and hypotheses

This paper explores the effects of students' expressed political interest, exposure to political news, and civic knowledge, as well as school-based civic programs, on students' likelihood to vote. In order to focus on the specific political and historical situation in Europe at the end of the 20th century, five non-European samples in the Civic Education Study of 1999/2000 are disregarded, namely Australia, Chile, Colombia, Hong Kong SAR, and the United States of America. As a consequence, the paper is based on data-sets from the 23 European countries that participated in the second Civic Education Study ($N = 68,000$).

Based on the earlier exploratory analyses, it is hypothesized (1) that students' political interest, political engagement, and civic knowledge as (distal)

indicators of political self-efficacy have a significant effect at the individual level on the intention to vote in future national elections. Furthermore, it is expected (2) that these effects remain significant even after controlling for the students' home literacy resources and their educational aspirations.¹ Similarly, it is anticipated, as a classroom-level effect, to find (3) students' *average* perceptions of the classroom climate (as being conducive to open discussions) positively associated with the anticipation of electoral participation. Also, it is presumed (4) that the *average* perceptions of an open classroom climate and the *common* experience of having learned how to vote in school have independent positive associations with the students' willingness to vote in future elections. The decisive step beyond what has been reported so far lies in the proper separation of individual-level (hypotheses 1 and 2) and class-level effects (hypotheses 3 and 4).

This set of research questions implies the need to consider a number of effects simultaneously. However, priority is given to a twofold research interest. The first is to examine the relationship between the adolescents' political interests and their anticipation of participating in future elections. The second is to investigate the role that school-based practices might have in this context. While the first of these two perspectives seeks to relate the findings from the IEA Civic Education Study to the notion of the adolescents being embedded in a larger—national or even international—political context, the second aspect pays tribute to the fact that young people, as they participate in educational experiences, share a common school-based environment, a point that has to be properly considered in the actual data analysis. On these grounds, differences and/or similarities among Western and Eastern European countries are examined with the aim of arriving at country-profiles that pay respect, both descriptively and structurally, to the variation, at the different levels of analysis, of political interest, civic literacy, and adolescents' reports on having learned at school about voting.

Data

As mentioned, the original data-set referring to the European components of the IEA Civic Education Study (14-year-olds) comprised approximately 68,000 students. For the present purposes, this data-set had to be restricted on two accounts:

1. Exclusion of all cases with one or more missing values on any of the variables included on the multilevel analyses; and
2. Exclusion of all classrooms with fewer than 10 complete student data-vectors.

As a result, 61,582 (unweighted) cases from 3,391 classrooms in the 23 countries considered were available for the analyses.

Method

Two-level regression models were fitted, with expectations of 14-year-old students' own expectations of voting in future national elections as the outcome variable.² This data structure is given by way of the two-stage random sampling procedure in the IEA Civic Education Study. Generally, only one classroom was selected at random from each sampled school (probability proportional to size of school), so it is not possible to disentangle school from classroom effects. Only in the case of Cyprus were two classes per school sampled. This, in principle, would have allowed a three-level model for this country. However, for reasons of equal treatment of countries, this was not done in the preparation of the present paper. Rather, the Cypriot schools were treated as "pseudo-classrooms," a decision that may have to be reconsidered in subsequent work. Eventually, for each of the 23 European data-sets, two separate two-level models were estimated so that individual and school-level effects could properly be distinguished.

Technically, the following procedure was applied in the examination of student differences in terms of their intentions to vote.

- In the first step, Level 1 predictors of the students' intention to vote were introduced into a regression model, following the guiding hypotheses (1) and

¹ These are proxies for the cultural capital and for the academic support a student finds in his or her family.

² Although the item measuring students' intentions to vote was an ordinal four-point scale, this variable was treated as if it were continuous. Since this item was not embedded in a concept with multiple indicators facilitating the construction of a proper interval-level scale, this compromise had to be accepted for pragmatic reasons.

(2) specified above. All predictors were considered as fixed factors, that is, slopes were not assumed to vary across schools within any one country investigated.

- In the second step, the variance attributable to the class level after controlling for all individual-level effects was estimated by computing the intra-class correlation coefficients. This quantity may be interpreted as an upper limit to the accumulation of systematic effects explaining between-class differences.
- In the third step, the Level 2 predictors referred to by hypotheses (3) and (4) were incorporated into the model.

For the analyses presented here, the following predictor variables were included:

- **Student-level predictors** in the explanatory models:

a) Demographic characteristic

- *Gender* (dichotomous, female = 1, male = 0): It has been reported in recent studies that male voters are more willing to participate in elections than are females (see, for example, International Institute for Democracy and Electoral Assistance, 2004).

b) Home environment

- *Home literacy resources* (number of books in the home; five ordered categories): This proxy of cultural capital as a family-based resource has been used successfully in other international studies and has proved to be a very consistent predictor of educational achievement. It correlates highly with the educational level of parents.
- *Exposure to political news on television* (Likert scale, four ordered categories): Students reported how often they watched news broadcasts on television (“never,” “rarely,” “sometimes,” “often”). Regular exposure to news on television can be seen as an indicator of active interest in political issues.
- *Expected years of further education* (seven ordered categories from 0 = “none” to “1 or more years”): This variable is considered to reflect the student’s and his or her parental educational aspirations as well as the school

type or track/stream the student expects to attend in future years.

c) Student characteristics

- *Civic knowledge* (international mean³ = 100; standard deviation = 20): maximum level (ML) estimates of content knowledge and skills in civic-related topics that are acquired particularly, but not exclusively, through course-work at school.
- *Expressed political interest* (four-item Likert scale, overall Cronbach’s alpha reliability coefficient across countries .74): *Example*—“I know more about politics than most people my age.”
- *Expected participation in political activities* (partial-credit IRT scale, international mean² = 10; standard deviation = 2): The *maximum likelihood* (ML) estimates reflect students’ expectations to participate in conventional political activities such as joining a political party, writing letters to a newspaper, or being a candidate for a local office (for details on the scaling of polytomous items in the Civic Education Study, see Schulz, 2004).

d) School-related factors

- *Classroom climate open for discussions* (partial-credit IRT scale, international mean² = 10; standard deviation = 2): The ML estimates indicate the individual students’ perceptions of freedom to express opinions and of their involvement in discussions in class.
- *Students’ reports of having learned about the importance of voting at school* (singleton with four ordered categories): In the international OLS regression model of likelihood to vote, this variable appeared as a significant predictor across all 28 IEA Civic Education countries.

Apart from the dummy-coded variable “gender,” all other independent variables were standardized either at the national (within-country) or at the international level.

In relation to **class-level predictors**, *average number of books at home*, and *average expected years of further education* were considered as classroom-level indicators that might have independent compositional effects on the students’ expectations of participating in future

³ This mean and the international means cited in the following points are based on the 28 countries of the full study, including five non-European countries.

elections. As no consistent findings emerged from this investigation, these aggregates were eventually not included in the analyses. Instead, the following two aggregate variables were found to be relevant in relation to testing the hypotheses listed above:

- *Average classroom climate*: For the purpose of validating the findings from single-level analyses in the original reports, it was deemed important to determine whether the effects of a classroom climate open for discussion are mediated only through the individual perception of students or whether the common perception at the class level (presumably a more valid indicator of the actual process) has an additional effect on the students' willingness to vote in future elections.
- *Class-level (average) reports on having learnt in school about voting*: To distinguish between individual-level effects, possibly limited to more active, more attentive students' assertions as to their intentions to participate in future elections, the possibility of aggregate effects over and above the individual effects was also investigated.

With respect to the two class-level measures included, part of the analytic strategy involved probing possible cross-level interaction effects, that is, influences of class-level aggregates on the slopes relating individual-level predictors to the dependent variable. The special statistical software used for the analyses presented here is the HLM program, version 5 (Bryk & Raudenbush, 1992; Raudenbush, Bryk, & Congdon, 2000).

To test the notion that countries which share certain historical, cultural, political, educational, and geographical characteristics, and thus demonstrate a substantial level of regional cohesion, the 23 countries under investigation were grouped in a preliminary, rather conventional way. The following six groups of countries were identified and considered in some of the analyses:

- *Western Europe*: Belgium (French), England, Germany, and Switzerland;
- *Nordic countries*: Denmark, Finland, Norway, and Sweden;
- *Baltic countries*: Estonia, Latvia, and Lithuania;

- *Southern Europe*: Cyprus, Greece, Italy, and Portugal;
- *East-Central Europe*: Czech Republic, Hungary, Poland, Slovak Republic, and Slovenia;
- *Balkans and Russia*: Bulgaria, Romania, and Russia.

It was expected that the assumed background similarities and presumably corresponding patterns of political discourse would be reflected to some extent in the ways in which the key concepts considered here related to one another.

Results

Table 1 contains country mean scores for five key variables of the set to be considered subsequently, in addition to regional mean scores computed in order to allow for an inspection of among-country similarities within regions. While there are certainly specific national characteristics evident in the table, for example, the exceptionally strong assertions of Norwegian students as to open classroom discussions or the apparent Swedish emphasis on teaching how to vote within the group of Nordic countries, the following specific regional characteristics can also be noted:

- On average, all *Nordic countries* appear to foster open discussions in the classroom.
- Typically, students from *Western Europe* anticipate relatively modest likelihoods of participating in future national elections.
- Young *Southern Europeans* have particularly high expectations with regard to their future voting behavior.
- Students from *all three post-communist regions* allege that in their school experience (which for these students had occurred entirely in post-transition years), there is little encouragement of open classroom discussions (all means are substantially below the international average).
- In addition, adolescents from *East-Central Europe* express high likelihoods of participation in future elections.
- Students from *Eastern Europe*, that is, the *Baltic countries* as well as the *Balkans/Russia* group, attained relatively unfavorable mean scores on the civic knowledge test in addition to reporting limited experience with open classroom discussions.

To validate the initial classification of countries, a cluster analysis⁴ was performed with four additional variables over and above those listed in Table 1: (1) home literacy resources, (2) exposure to TV news, (3) expected years of future education, and (4) current political activities. The table also contains the inferred cluster memberships. The symbols used to identify the clusters are intended to indicate the relatively high degree of reproducing the original taxonomy: **N** “Nordic,” **W** “Western Europe,” **S** “Southern Europe,” **C** “East-Central Europe,” **B** “Baltic countries,” **R**

“Residual group.” The respective entries show the *Baltic* countries all included in the respective cluster, three out of four of the constituents in each of the Nordic and the *Western* clusters, two each of the constituents in the original *Southern* and *East-Central* clusters, and one constituent in the *Residual* cluster. The substantive relevance of these observations is linked to the fact that student perceptions of their educational, social, and political environments are, indeed, related to the conventional classification in terms of political geography.

Table 1: National and Regional Means for Five Selected Variables and Cluster Memberships in a Six-Cluster Solution

Country	Cluster	Political interest	Political knowledge	Open class climate	Learn to vote	Intention to vote
Nordic Countries		2.09	102.92	10.21	2.37	3.15
Denmark	<i>N</i>	2.09	100.4	9.99	2.27	3.28
Finland	<i>N</i>	2.06	109.3	10.01	2.20	3.13
Norway	<i>N</i>	2.13	102.9	10.63	2.42	3.27
Sweden	<i>W</i>	2.07	99.1	10.23	2.60	2.93
Western Europe		2.20	98.07	10.04	2.36	2.79
Belgium	<i>B</i>	2.25	94.7	9.30	2.44	2.80
England	<i>W</i>	2.10	99.4	10.02	2.32	3.05
Germany	<i>W</i>	2.27	99.8	10.36	2.28	2.80
Switzerland	<i>W</i>	2.16	98.3	10.42	2.38	2.49
Southern Europe		2.41	104.41	10.24	2.73	3.35
Cyprus	<i>S</i>	2.63	107.9	10.44	2.96	3.72
Greece	<i>S</i>	2.49	108.1	10.50	2.93	3.39
Italy	<i>C</i>	2.29	105.4	10.36	2.54	3.07
Portugal	<i>R</i>	2.23	96.2	9.65	2.46	3.17
East-Central Europe		2.29	104.16	9.74	2.56	3.23
Czech Rep.	<i>B</i>	2.14	102.6	9.46	2.33	2.70
Hungary	<i>N</i>	2.24	101.6	9.38	2.53	3.41
Poland	<i>C</i>	2.37	110.6	10.38	2.86	3.36
Slovak Rep.	<i>C</i>	2.48	105.4	10.15	2.83	3.56
Slovenia	<i>N</i>	2.24	100.6	9.32	2.21	3.10
Baltic Countries		2.29	93.05	9.71	2.41	2.88
Estonia	<i>B</i>	2.25	94.0	9.74	2.39	2.75
Latvia	<i>B</i>	2.33	91.5	9.60	2.48	2.88
Lithuania	<i>B</i>	2.28	93.6	9.77	2.38	3.02
Balkans and Russia (Residual)		2.44	96.27	9.64	2.73	2.97
Bulgaria	<i>B</i>	2.38	97.6	9.29	2.35	2.54
Romania	<i>R</i>	2.50	91.6	9.47	3.07	3.18
Russia	<i>N</i>	2.44	99.6	10.14	2.73	3.14
Overall Mean		2.28	100.4	9.94	2.52	3.08

Note: Regional statistics are based on equal weights for countries (“senate weights;” cf. Sibberns & Foy, 2004).

⁴ SPSS routine “Quick Cluster,” with number of clusters set to 6 a priori.

In international comparisons of educational achievement, such as the Trends in Mathematics and Science Study (TIMSS) or Programme for International Students Assessment (PISA), where the focus is on core curriculum areas, a major portion of the criterion variance is regularly found to be associated with schools or classrooms. Because the promotion of reading, mathematics, and science skills is, to a large degree, a monopoly of educational institutions, the particular setting in which learning takes place can be expected—and has been demonstrated—to “homogenize” the clientele that it is serving. This consideration is much less clear with respect to the criterion of expecting to participate in future elections. It is therefore of considerable interest to find answers to the question of whether and, if so, to which degree and as a function of which determinants, the expected likelihood to participate in national elections depends on a student’s membership in a particular class. An estimate of the variance components for students and classrooms across Europe with regard to this criterion (on the basis of the whole international data-set comprising all 23 European samples (equally weighted)) found only 18% of the total European variance in the likelihood to vote associated with the classroom level. It follows that the remaining 82% of the international variance must be attributable to the individual student level, despite considerable differences among the countries involved. Such estimates do not pay due respect, however, to the considerable differences in terms of predictive relationships both within and across countries. It seems appropriate, therefore, to enquire into such specific effects by running identical analyses for each country, where likely determinants of voting expectations are also considered (see Table 2).

The *intra-class coefficients*, with only the individual effects within the analyzed countries taken into account, range from 0.11 in Cyprus to 0.33 in Latvia. These results confirm that in most European countries surveyed, the among-school differences in the expressed anticipation of voting in national elections are comparatively small. Some of the exceptions (Bulgaria, 0.25; Estonia, 0.26; Latvia, 0.33; Slovak Republic, 0.24) have rather large ethnic minorities.

In regard to the Level 1 predictors of intention to vote in the multilevel analyses, it is perhaps surprising to find that the independent influence of *home literacy resources* as a proxy for the social and cultural capital

nowhere exceeds $\beta = 0.10$ when the hierarchical data structure is properly taken into account; in more than half of the countries, the effect has even failed to attain statistical significance. The variable *expected years of further education*, which may be taken as another indicator of social aspirations, does show the expected positive effect in most countries. With the exception of Sweden, Germany, and Russia, students who are likely to attain higher levels of education are also more certain as to their future active participation in elections. The reasons behind these exceptions are unclear and may require additional inquiry on the basis of the national data-sets. While *exposure to political news on TV* is, among 14-year-olds, still very much a matter of the home environment, it is also related to their focus of interest. Thus, it was quite plausible to assume that this variable would be “universally” associated with the degree of certainty as to the students’ future participation in elections. Students whose families and who themselves place some value on being informed politically are, as a rule, clearly more likely to exercise their right of voting than are those students and their families who do not believe in the relevance of such information.

Civic knowledge as measured by the respective test is even more consistently related to the anticipation of voting. It is the best predictor in 18 out of the 23 countries surveyed, with the exceptions mostly explainable by collinearities with other relevant variables: “expected political engagement” (and “voting: learned in school”) in Finland and Switzerland; and “political news on TV” in Cyprus. Latvia and Russia share the specific finding that the gender effect in favor of female adolescents is just as pronounced (Russia) or even larger (Latvia) than is the association with knowledge.

Wherever this gender effect has attained statistical significance, it operates in the direction of girls expressing a greater inclination to vote in future elections. Why in three of the four Western European countries surveyed and also in the Czech Republic the inclination to participate in future elections tends to be more pronounced among males, albeit insignificantly, constitutes an idiosyncrasy that perhaps merits further analysis.

The students’ *intentions to get engaged in conventional politics* are also associated with highly uniform effects, if compared across countries. In all three regions

Table 2: Multilevel Analyses Associating 14-year-old Students' Anticipated Participation in National Elections with Individual and Class-level Characteristics

Predictor	Scandinavia					Western Europe					Southern Europe			
	Denmark	Finland	Norway	Sweden	Belgium (Fr.)	England	Germany	Switzerland	Cyprus	Greece	Italy	Portugal		
Student Level														
Gender	0.11	0.04	0.07	0.18	0.16	-0.08	-0.05	-0.08	0.04	0.15	0.00	-0.07		
Civic Knowledge	0.19	0.11	0.25	0.27	0.26	0.20	0.22	0.13	0.13	0.20	0.24	0.23		
Expected Pol. Engagement	0.01	0.17	0.02	0.18	0.04	0.14	0.14	0.19	-0.03	-0.04	0.04	0.16		
Open Classroom Climate	0.09	0.05	0.07	0.10	0.10	0.08	0.09	0.02	0.10	0.16	0.07	0.05		
Home Literacy Resources	0.03	0.04	0.04	0.10	0.03	0.07	0.10	0.09	-0.03	-0.02	0.00	-0.02		
Voting: Learnt in School	0.12	0.23	0.15	0.16	0.14	0.14	0.15	0.14	0.08	0.06	0.16	0.09		
Political News on TV	0.12	0.05	0.15	0.08	0.12	0.15	0.05	0.03	0.14	0.14	0.11	0.08		
Years of Further Education	0.08	0.14	0.07	0.02	0.08	0.09	0.02	0.11	0.07	0.07	0.09	0.05		
Political Interest	0.15	0.06	0.07	0.04	0.02	0.04	0.12	0.06	0.11	0.07	0.08	0.09		
Intra-class correlation (Level 1 only)	0.15	0.24	0.15	0.22	0.22	0.20	0.13	0.26	0.11	0.18	0.19	0.10		
Class Level														
Mean: agg. open climate	-0.02	0.00	0.03	0.04	0.07	-0.00	-0.01	-0.11	-0.10	-0.01	0.01	0.08		
agg. learn vote	0.09	0.09	-0.07	-0.10	-0.03	0.00	0.09	-0.04	0.12	(-0.14)	(-0.14)	-0.06		
Slope voting in school:	-0.00	-0.01	-0.01	0.02	0.10	-0.02	0.03	-0.01	-0.03	-0.02	0.01	0.00		
agg. open climate														
Slope political interest:	0.04	-0.25	0.05	0.04	0.03	0.03	(-0.11)	-0.01	-0.11	-0.09	-0.12	-0.09		
agg. learn vote														
Intra-class correlation (Level 1 and Level 2)	0.14	0.20	0.15	0.21	0.17	0.20	0.14	0.26	0.12	0.18	0.19	0.12		

Note: Standardized regression coefficients. Bold: coefficient significant at a 95% level; bold in parentheses: $0.05 < p(\alpha) < 0.10$.

Table 2 (contd.): *Multilevel Analyses Associating 14-year-old Students' Anticipated Participation in National Elections with Individual and Class-level Characteristics*

Predictor	Eastern Central Europe					Baltic Countries				Balkans and Russia	
	Czech Rep.	Hungary	Poland	Slovak Rep.	Slovenia	Estonia	Latvia	Lithuania	Bulgaria	Romania	Russia
Student Level											
Gender	-0.06	0.05	0.17	0.13	0.12	0.21	0.24	0.07	0.10	0.01	0.16
Civic Knowledge	0.26	0.20	0.22	0.17	0.16	0.24	0.21	0.20	0.17	0.22	0.16
Expected Pol. Engagement	0.16	0.11	0.09	0.05	0.11	0.17	0.17	0.06	0.20	0.04	0.10
Open Classroom Climate	0.01	0.08	0.07	0.04	0.06	0.06	0.08	0.05	0.08	0.02	0.15
Home Literacy Resources	0.04	0.06	-0.01	0.04	0.01	0.03	-0.02	0.05	0.07	0.05	0.02
Voting: Learnt in School	0.19	0.16	0.12	0.14	0.13	0.17	0.11	0.05	0.07	0.18	0.13
Political News on TV	0.06	0.13	0.15	0.12	0.12	0.11	0.17	0.13	0.15	0.10	0.10
Years of Further Education	0.14	0.06	0.13	0.07	0.12	0.09	0.08	0.07	0.08	0.06	0.00
Political Interest	0.08	0.03	(0.04)	0.10	0.05	0.08	0.05	0.10	0.09	0.04	0.09
Intra-class correlation (Level 1 only)	0.25	0.18	0.22	0.24	0.18	0.26	0.33	0.20	0.25	0.21	0.20
Class Level											
Mean: agg. open climate	0.01	-0.03	0.01	-0.03	0.09	-0.03	0.02	0.08	0.06	-0.04	0.02
agg. learn vote	0.08	-0.01	(0.13)	-0.04	0.19	-0.03	0.15	0.06	-0.05	0.06	0.01
Slope voting in school:	-0.00	0.01	0.04	0.03	0.06	0.05	0.01	-0.06	0.03	0.01	0.03
agg. open climate											
Slope political interest:	-0.02	0.02	-0.12	0.06	.002	(-0.11)	-0.04	0.00	(-0.13)	-0.01	0.09
agg. learn vote											
Intra-class correlation (Level 1 and Level 2)	0.25	0.18	0.22	0.25	0.18	0.26	0.32	0.20	0.25	0.21	0.20

Note: Standardized regression coefficients. Bold: coefficient significant at a 95% level; bold in parentheses: $0.05 < p(\alpha) < 0.10$.

comprising post-communist countries and, with the exception of Belgium (French), in Western Europe as well, this relationship is consistently significant. Finland, Sweden, and Portugal provide three more examples. Only Denmark, Norway, and three Southern European countries offer counter-examples. The students' expressed political interest shows a fairly similar pattern—a statistically significant positive relationship with the anticipation to participate in future national elections, and one that is, with very few exceptions, explainable in terms of multicollinearity.

The two variables of primary concern here—*classroom climate open for discussions* and the students' assertions as to having *learned about voting in school*—are positively associated with anticipated voting in almost all countries. The somewhat smaller classroom climate effect has failed to reach statistical significance only in Switzerland, the Czech Republic, and Romania. While this observation is not new, it confirms the existing interpretations given in the first international IEA Civic Education Study report (Torney-Purta et al., 2001) in the light of an adequate handling of concurrent effects at two levels of analysis: students (Level 1) and classrooms (Level 2).

Thus, the findings discussed so far are in many respects similar to the results of the published OLS regression models. In each case, the listed standardized regression coefficient (β) has referred to the *independent* contribution of a Level 1 predictor, now also controlling for Level 2 influences. The fact that the Level 1 predictors were entered into the equations as centered around the grand mean implies that the respective effects are not restricted to relationships within the class (Level 2).

As far as the aggregate-level hypotheses—numbers (3) and (4) in the list above—are concerned, the two-level analyses have produced rather incoherent findings. For the most part, the respective coefficients vary substantially in terms of magnitude and even in terms of sign/direction. As was, perhaps, to be expected, the average tendency in a given class to participate in future elections (class mean) does not appear to be subject to consistent class-level influences of either openness to debate or learning how to vote—with Slovenia as the only, yet quite marked, exception. The same appears to hold for the relation between learning in school how to vote and the inclination to exercise voting rights: only French-speaking Belgium

and, to a much lesser degree, Poland offer examples of the common perception of an open classroom climate strengthening the relationship between “voting instruction” and “intending to vote.”

An effect of particular interest can be found in the way in which the aggregate level of “voting instruction” moderates the effect of individual interest on the inclination to vote. In six countries from all six regions considered here—Finland, Germany, Italy, Poland, Estonia, and Bulgaria—learning how to vote in school appears to have contributed to *disassociating* the political interest of individual students from their intention to participate in future elections. In fact, this moderating effect (or “cross-level interaction”) appears to be present in all four Southern European countries sampled, even if statistical significance is not attained throughout.

One final observation related to the similarities between students from distinct national settings may, perhaps, be added to this rather long list of statistical detail. The cluster-analysis that had, at least temporarily, justified the approach of grouping student samples by regions (and which, in turn, derived from the conventional classifications of political geography) was based on “surface findings” (mere vectors of country means) alone. The relational analyses conducted, but not presented here, confirmed this grouping at a more substantial level. For instance, the close pair-wise similarities in the patterns of regression coefficients between Denmark and Norway, England and Germany, Cyprus and Greece, Poland and the Slovak Republic, Estonia and Latvia gave clear examples of settings in which similar political circumstances were reflected in similar student response patterns.

Conclusion

In summary, it appears justified to arrive at the following conclusions with respect to the students' expectations as to their future participation in national elections:

1. Most of the variation of the 14-year-olds' “inclination to participate in future elections” exists between students and not between classrooms.
2. Schools/classrooms appear to have a consistent, but moderate, influence on their students' inclinations to exercise their voting rights when they are of age.
3. The respective predictors of students' participation

in future elections—in particular, “learning how to vote in school” and “open classroom climate”—proved to be effective, even when student and home background characteristics as well as the hierarchical data structure were controlled for.

It is of particular interest that the individual-level relationship between “political interest” and “inclination to participate in future elections” is in many countries subject to or can be manipulated by classroom-level intervention. The supposedly “natural” tendency for politically interested students to envisage their role in the electorate can be transformed into

a more equitable distribution: it is not unlikely that “voting instruction” has the effect of mobilizing future voters whose basic interests are somewhat distant from politics.

Clearly, these claims, if causally interpreted, are far from being established. The IEA Civic Education Study has been a cross-sectional assessment of civic knowledge and skills, political concepts and activities, affects and orientations related to citizenship and civic education. It may be, however, that some of the relationships encountered in the data lend themselves to a more rigorous testing of the respective hypotheses.

References

- Amadeo, J.-A., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement: An IEA study of upper secondary students in sixteen countries*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Bryk, A. S., & Raudenbush, S.W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage Publications.
- International Institute for Democracy and Electoral Assistance (IDEA). (2004). *Voter turnout in Western Europe since 1945: A regional report*. Stockholm: Author.
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. T. (2000). *HLM 5 hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International.
- Schulz, W. (2004). Mapping student scores to item responses. In W. Schulz & H. Sibbern (Eds.), *IEA Civic Education Study: Technical report* (pp. 127–132). Amsterdam: International Association for the Evaluation of Educational Achievement.
- Sibbern, H., & Foy, P. (2004). Weighting procedures. In W. Schulz & H. Sibbern (Eds.), *IEA Civic Education Study: Technical report* (pp. 61–67). Amsterdam: International Association for the Evaluation of Educational Achievement.
- Torney, J. V., Oppenheim, A. N., & Farnen, R. F. (1975). *Civic education in ten countries: An empirical study*. Stockholm/New York: Almqvist & Wiksell/John Wiley and Sons.
- Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: The civic knowledge and engagement of adolescents*. Amsterdam: International Association for the Evaluation of Educational Achievement.

Reading achievements in urban and rural communities: A comparative analysis of equity in education

Aistė Elijo

*Vilnius University, National Examinations Center
Vilnius, Lithuania*

Abstract

On the basis of PIRLS 2001 data, we consider the issue of equity in education in rural and urban communities in three groups of countries: Eastern-European, Scandinavian, and Anglo-Saxon. Results show that Scandinavian countries have the highest level of equity in education, whereas the Eastern-European and Anglo-Saxon countries show different tendencies of disparity. In Eastern Europe, students from rural communities do not perform as well as those from urban communities, whereas in Anglo-Saxon countries, students from rural communities perform better than their peers from urban communities. Our investigation indicates that, in the case of Anglo-Saxon countries, the difference in achievement is mainly because the social and economic family

backgrounds of students tend to be different in cities and towns compared with villages. In Eastern-European countries, however, the differences in achievement cannot be attributed solely to home background—the locality factor explains variance additional to that predicted by a student's home background situation. This paper documents a further analysis, conducted with the aim of unveiling what is behind the school locality impact in Eastern-European countries. The analysis identified two explanatory factors. The first related to the “average home background situation of students in a school” (or, in other words, social and educational peer-environment), and the second reflected quality of education, which was not the same in rural and urban communities.

Introduction

For years, the question of equity in education and the impact of various social, economic, and educational factors on students' educational achievements have been of great interest not only to researchers in education, economics, and other social sciences, but also to policymakers. We are particularly interested in investigating whether education systems in various countries provide a comparable quality of education in both rural and urban schools, and whether differences that are found in some countries can be attributed solely to the different home background situation. A number of researchers have already addressed some of the issues we are interested in, but their findings do not yet provide complete answers to our questions. A short overview of their main findings follows.

Blau's (1999) analysis of the effect of parental income on children's cognitive, social, and emotional development concluded that the effect of current income (total household income during the calendar year prior to the year of assessment) is small. Blau found that the effect of “permanent” income (income averaged over all periods) is substantially larger, but

relatively small when compared to family background characteristics, such as parental education and household structure.

Jensen and Seltzer (2000) found that individual, family, and neighborhood factors all influence the further education decisions of young Australian students.

Lee and Croninger (1994) modeled the influence of both home and school environment on the literacy development of children. Although home factors seemed to have the stronger impact, the authors focused on school factors, arguing that schools have major responsibilities to equalize the development of their students. Despite this, the authors contended that it is easy and common for schools “to ascribe the learning disadvantages of their less affluent students to deficient home environments” (p. 318).

Thirunarayanan (2004) compared students' achievements in different content areas by school location in the United States, and found that students in central-city schools performed statistically significantly worse in many subject areas than did students in

suburban schools. Thirunarayanan noted that students of urban schools are more likely than their peers in suburban localities to face a number of social and economic disadvantages, including poverty, illiteracy, unemployment, crime, and teacher shortages.

Brown, Anfara, and Roney (2004) paid special attention to schools' organizational health (teacher affiliation, resource support, academic emphasis, and institutional integrity) in their search for possible reasons for the difference in educational achievement between high-performing suburban and low-performing urban schools. They argued that "improving school climate and culture by focusing on the technical, managerial, and institutional levels of organizational health" (p. 452) would decrease this disparity.

However, not all authors claim urban communities are poor in terms of developing students' knowledge and abilities. Israel, Beaulieu, and Hartless (2001) analyzed the influence of family and community social capital on educational achievement. Having noted that, in socioeconomic terms, rural areas lag behind suburban and urban ones, they went on to argue that low-capacity rural communities might develop an attitude and atmosphere that neither supports nor promotes high educational achievements. The authors reaffirmed "the significant role of parents' socioeconomic status in shaping their children's educational performance" (p. 61), and added that both family and community social capital help children excel in school, although the influence of community social capital is usually indirect.

Johansone and Foy (2004) analyzed the similarities and differences between patterns of reading achievements of primary school students in the European Union (EU) countries and (then) candidate countries. They found no correlation between general reading achievements and whether a country was a current EU member or a candidate country; comparisons of achievements between school communities revealed differences within countries. In some candidate countries, children in rural communities seemed to be at a disadvantage, as their average results were statistically significantly lower than the results of their peers in urban or suburban communities.

Elijio (2004), in an earlier analysis of the impact of various home factors on students' reading achievement in a number of countries, found the relationship between home background factors and students' reading achievement was similar across countries. This relationship held for purely home background factors, such as parents' level of education, educational resources at home, and literacy activities. However, the impact of school locality factors across countries was not clear.

In general, the literature indicates the need for further investigation of the impact of school locality on students' educational achievements. Data from different countries and different researchers are consistent in some ways (e.g., noting either small or more noticeable relationships between locality and achievements), but different in others (e.g., finding that students from urban schools sometimes perform worse and sometimes better than their peers from other localities). This paper attempts to investigate the tendencies in different countries and to analyze the interaction of other factors related to students' achievements within one locality.

Data and methods

The data used for this analysis are from the database for the Progress in International Reading Literacy Study (PIRLS) 2001, an international assessment of students' reading achievements, conducted by the International Association for the Evaluation of Educational Achievement (IEA). The survey population was Grade 4 students from 35 countries. Students' reading comprehension achievements were measured using a variety of test booklets, and Item Response Theory (IRT) scaling methodology was used to generate student scores for analysis and reporting. The PIRLS 2001 scale score average across countries was set to 500 and the standard deviation to 100. Data on background factors were collected using student, parent (home), teacher, and school questionnaires.

In the investigation documented in this paper, my colleagues and I mainly employed Hierarchical Linear Modeling (HLM 6.02). For the analysis, we used weights HOUWGT, described in Gonzalez and Kennedy (2003). We grouped countries according to the following classification:

- *Eastern-European countries:* Hungary, Latvia, Lithuania, Macedonia, Moldova, Romania, Russia, Slovakia¹
- *Anglo-Saxon countries:* England, New Zealand, Scotland, the United States
- *Scandinavian countries:* Iceland, Norway, Sweden.

For the HLM analysis, we used the following representatives from each group of countries: Lithuania (LTU) for the Eastern-European countries group, England (ENG) for the Anglo-Saxon countries group, and Sweden (SWE) for the Scandinavian countries group. The three countries had similar average reading achievements, making it easier to compare the models.

Results

Home educational factors

Similar to that shown by Eljio (2004), using ANOVA, home educational factors, such as parents' level of education and home educational resources, had a similar impact on reading achievement in all three groups of countries. Although student achievement differed across the groups, the expected tendencies worked similarly in all of them. Similar results were obtained by using HLM. The first few equations of our analysis therefore included the following variables:

- RA: Reading achievement in Lithuania (LTU), England (ENG), and Sweden (SWE), respectively
- BOOKS: Number of books at home
- FEDU: Father's level of education (aggregated variable, having three categories—lower than secondary (0), secondary (1), higher than secondary (2))
- MEDU: Mother's level of education (aggregated variable, having three categories—lower than secondary (0), secondary (1), higher than secondary (2))
- EHLA: Early home literacy activities (an index made from parents' answers to questions relating to the home literacy activities prior to the child entering primary school, and with three categories—low (0), medium (1), and high (2))

- SATR: Students' attitude toward reading (an index made from the students' answers to questions relating to their attitude toward reading, and with three categories—low (0), medium (1), and high (2)).

The linear regression on reading achievement with BOOKS as a predictor gave the following results:

$$RA (SWE) = 535.29 + 13.51 \cdot BOOKS + u_0 + r,$$

$$RA (ENG) = 536.66 + 19.04 \cdot BOOKS + u_0 + r,$$

$$RA (LTU) = 525.99 + 11.79 \cdot BOOKS + u_0 + r.$$

We can easily see that each higher category of books at home positively influenced average achievements of students by 11.79 to 19.04 scale points. All coefficients were statistically significant ($p < 0.01$). The highest influence of number of books is apparent in England, whereas in Sweden and Lithuania the influence of number of books at home on students' reading achievement is very similar. Although the coefficients differ a little in the three equations, the same tendency is evident.

Similar calculations for linear regression on reading achievement with FEDU and MEDU as predictors produced these results:

$$RA (SWE) = 544.96 + 14.88 \cdot FEDU + u_0 + r,$$

$$RA (ENG) = 522.10 + 21.45 \cdot FEDU + u_0 + r,$$

$$RA (LTU) = 513.74 + 21.42 \cdot FEDU + u_0 + r.$$

In this case, we can see that each higher category of father's education increases reading achievement by about 15 to 21 scale points. All coefficients were statistically significant ($p < 0.01$). The influence of father's education on student reading achievement is very similar in England and Lithuania, but a little lower in Sweden. Nevertheless, the same tendency is evident in all three equations.

We obtained analogous equations in regard to mother's education:

$$RA (SWE) = 542.09 + 17.05 \cdot MEDU + u_0 + r,$$

$$RA (ENG) = 555.33 + 20.77 \cdot MEDU + u_0 + r,$$

$$RA (LTU) = 511.77 + 20.21 \cdot MEDU + u_0 + r.$$

¹ Some other Eastern-European countries that took part in PIRLS 2001 were not included in the analysis because the number of students from rural communities was very small.

It is obvious that the FEDU and MEDU variables are related, and we were not surprised to obtain the following results when we used both of the variables as predictors:

$$RA(SWE) = 531.81 + 9.97 \cdot FEDU + 14.71 \cdot MEDU + u_0 + r,$$

$$RA(ENG) = 549.57 + 14.96 \cdot FEDU + 14.05 \cdot MEDU + u_0 + r,$$

$$RA(LTU) = 499.84 + 17.32 \cdot FEDU + 13.30 \cdot MEDU + u_0 + r.$$

In these equations, the similarity of the results in all three countries is very evident; the differences (except in the case of FEDU in Sweden) are negligible.

We also analyzed the regression on reading achievement, with the EHLA (early home literacy activities) as a predictor. The results for the three countries were as follows:

$$RA(SWE) = 550.85 + 11.03 \cdot EHLA + u_0 + r,$$

$$RA(ENG) = 519.80 + 26.07 \cdot EHLA + u_0 + r,$$

$$RA(LTU) = 528.49 + 9.48 \cdot EHLA + u_0 + r.$$

Again, as in the previous examples, EHLA showed a positive influence on the reading achievement of students in all three countries. All coefficients were statistically significant ($p < 0.01$). In this case, the highest impact can be seen in England, but in Sweden and Lithuania this influence was more than two times lower.

Students' attitude toward reading (SATR) is another important factor related to students' reading achievement. However, it is difficult here to define which variable influences the other: does students' reading achievement rise because of the more positive attitude toward reading, or is a better attitude toward reading simply a result of higher reading literacy? We did not explore this issue in the current investigation because our focus was on analyzing the impact of various factors on reading achievement and comparing the general relation between the factors in the three groups of countries. Therefore, in the regression, we considered SATR as a predictor of students' reading achievement. We obtained the following results:

$$RA(SWE) = 512.86 + 31.39 \cdot SATR + u_0 + r,$$

$$RA(ENG) = 509.15 + 33.83 \cdot SATR + u_0 + r,$$

$$RA(LTU) = 510.16 + 22.00 \cdot SATR + u_0 + r.$$

The regression equations show that SATR is a very strong predictor of students' reading achievement: with each higher category of SATR, reading achievement rises by up to 34 scale points. All coefficients were statistically significant ($p < 0.01$). In this case, we can see not only that students' attitude toward reading had a smaller impact on students' achievement in Lithuania than it did in Sweden and England, but also that, in all countries, the coefficients are high, and the relationship is naturally similarly positive.

The educational background variables that we have been investigating are clearly related. We therefore estimated the regression equations using all variables—BOOKS, FEDU, MEDU, EHLA, and SATR—as predictors of reading achievement. Our results follow.

$$RA(SWE) = 481.91 + 7.28 \cdot BOOKS + 6.06 \cdot FEDU + 9.06 \cdot MEDU + 4.57 \cdot EHLA + 28.56 \cdot SATR + u_0 + r,$$

$$RA(ENG) = 464.87 + 9.65 \cdot BOOKS + 10.84 \cdot FEDU + 9.01 \cdot MEDU + 16.22 \cdot EHLA + 34.76 \cdot SATR + u_0 + r,$$

$$RA(LTU) = 466.94 + 7.32 \cdot BOOKS + 14.22 \cdot FEDU + 8.59 \cdot MEDU + 4.96 \cdot EHLA + 20.69 \cdot SATR + u_0 + r.$$

In these equations, all coefficients (except for 4.96 next to EHLA for Lithuania) were statistically significant ($p < 0.05$). Otherwise, we can see the equations are similar. We can therefore conclude that, in all three countries, home background variables had a similar impact on students' reading achievement.

School locality

We have seen that home background factors worked similarly in all three groups of countries. But could the same be said for the school level factors? We were especially interested in the school locality factor because it can indicate variation in the quality of education provided in different localities throughout the countries. We therefore carried out a regression on students' reading achievements, with the school locality (SLOCALITY) variable as a predictor. This variable has two categories: 0—rural and 1—city/town. We obtained the following regression equations:

$$RA(SWE) = 558.39 + 0.82 \cdot SLOCALITY + u_0 + r,$$

$$RA(ENG) = 570.39 - 22.04 \cdot SLOCALITY + u_0 + r,$$

$$RA(LTU) = 513.87 + 35.23 \cdot SLOCALITY + u_0 + r.$$

The coefficient 0.82 next to SLOCALITY in the regression equation for Sweden is clearly negligible: it has a standard error of 8.01 and $p = 0.919$. Therefore, we can confidently conclude that, in Sweden, school locality does not influence students' reading achievement. In the case of England and Lithuania, the regression coefficient next to SLOCALITY was statistically significant at the $p < 0.01$ level. However, the influence of SLOCALITY in Lithuania is opposite to that for England. On the basis of our results, we could expect students from city/town schools in Lithuania to perform better than their peers from village schools by about 35 scale points. In England, we could expect students from city/town schools to perform worse than their peers from rural schools by about 22 scale points.

In terms of further investigation, it is important to know if our findings regarding locality apply just to the three countries or to all countries within each group of countries. As can be seen from Figures 1 to 3, each based on ANOVA results, we can indeed draw the same conclusions for all countries in each group. In the Scandinavian countries, we did not find a statistically significant difference between the results of students from rural and from urban communities. In the Anglo-Saxon countries, students from rural schools performed better than their peers from urban schools. The differences for all these countries were statistically significant. In the Eastern-European countries under analysis, students from cities and towns performed better than their peers from rural schools.

These results raise a number of questions. What is the difference between the urban and rural communities that makes locality an important factor in the Eastern-European and Anglo-Saxon countries? Is this difference solely due to different social and educational family backgrounds in these localities (an assumption very often made by the educational authorities), or is it because of the differences in the education provided by the schools in the different types of communities? Is the similarity of the results in urban and rural communities in the Scandinavian countries caused by the same quality of education provided throughout the country, or is it the result of the high level of social and economic equality within the countries?

We carried out deeper investigation in an effort to find some answers. The analysis that follows relates

Figure 1: Average Reading Achievement Results of Students from Urban and Rural Communities in Scandinavian Countries

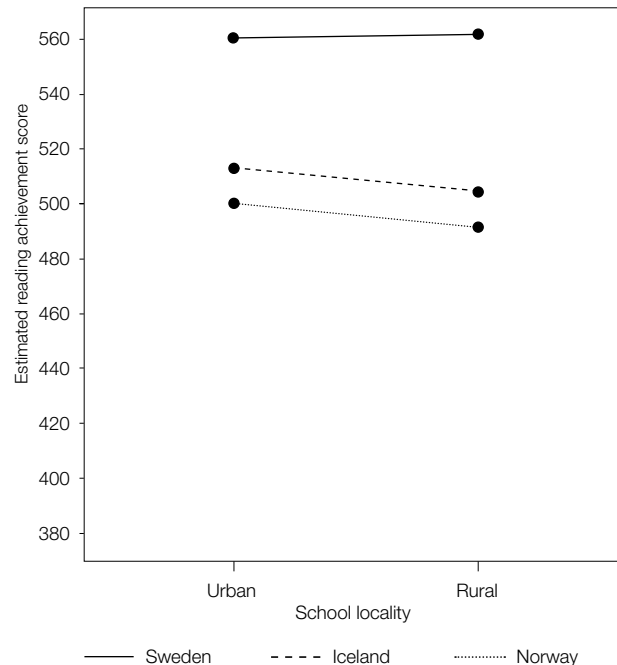


Figure 2: Average Reading Achievement Results of Students from Urban and Rural Communities in Anglo-Saxon Countries

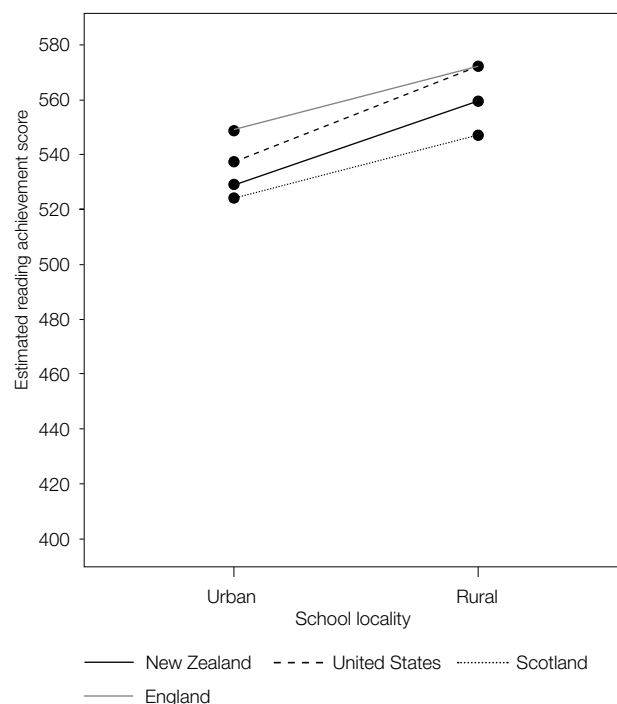
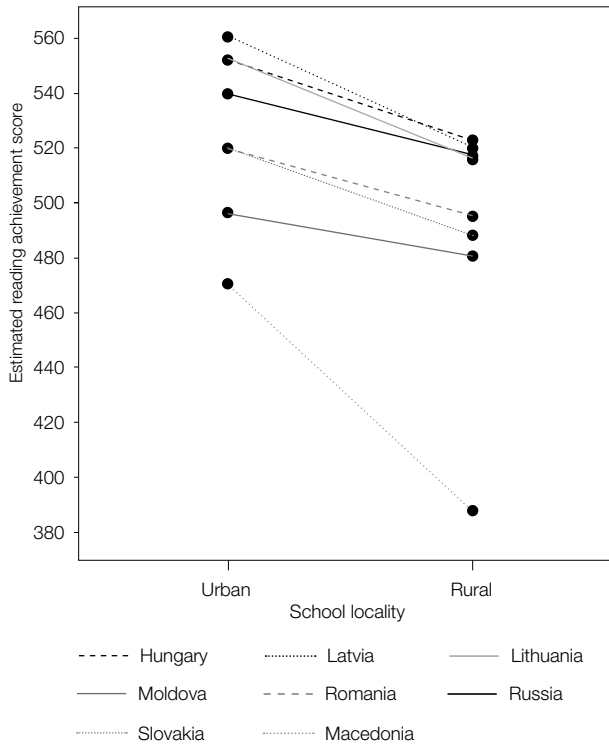


Figure 3: Average Reading Achievement Results of Students from Urban and Rural Communities in Eastern European Countries



to representatives of the groups of countries where differences in achievement are found, namely, England and Lithuania.

School locality versus home background factors

As mentioned previously, one of the hypotheses in relation to the impact of school locality on students' reading achievement is that it is because the home background situation is simply worse or better in a certain type of locality. For example, in the Eastern-European countries, it is only natural to expect that students in the villages come from homes where factors related to reading achievement are less favorable than they are for students in cities and towns. An example of data supporting this view can be seen in Table 1. Similarly, in the case of Anglo-Saxon countries, we

would expect to find that the generally less favorable home backgrounds found in urban localities would explain the lower results of the students from urban schools (see, for example, Thirunarayanan, 2004). With these considerations in mind, we carried out a regression on reading achievement using SLOCALITY and the home background factors that we had already investigated. Our aim was to determine if the hypothesis held true in the case of England and Lithuania.

School locality and home background factors in England

We obtained the following regression equations for England:

$$RA(ENG) = 540.65 - 4.94 \cdot SLOCALITY + 18.92 \cdot BOOKS + u_0 + r.$$

In this case, the coefficient -4.94 next to SLOCALITY was statistically insignificant ($p = 0.507$).

$$RA(ENG) = 560.06 - 10.30 \cdot SLOCALITY + 21.42 \cdot FEDU + u_0 + r.$$

Here, the coefficient -10.30 next to SLOCALITY was statistically insignificant, with $p = 0.222$.

$$RA(ENG) = 528.79 - 11.58 \cdot SLOCALITY + 26.05 \cdot EHLA + u_0 + r.$$

In this case, the coefficient -11.58 next to SLOCALITY was statistically insignificant, with $p = 0.142$.

$$RA(ENG) = 526.60 - 22.62 \cdot SLOCALITY + 33.91 \cdot SATR + u_0 + r.$$

With this equation, the coefficient -22.62 next to SLOCALITY was statistically significant, with $p = 0.002$.

These equations show us that, in the case of England, only students' attitude toward reading (SATR) acted as an independent factor from school locality, whereas all the other factors related to home background annihilated the effect of school locality. This result verifies the hypothesis that the school locality factor in Anglo-Saxon countries is an indicator

Table 1: Percentage of Students Who Indicated Certain Number of Books at Home in Lithuania

School locality	Number of books at home					Total
	0-10	11-25	26-100	101-200	>200	
Urban (percentage of students)	5.5	14.3	39.2	19.3	21.7	100
Rural (percentage of students)	17.8	21.9	38.5	13.3	8.5	100

of the different home background factors of students there. As such, we have no reason to assume there is a difference in the quality of education provided in rural and urban schools.

School locality and home background factors in Lithuania

We obtained the following regression equations for Lithuania:

$$RA(LTU) = 504.13 + 29.73 \cdot SLOCALITY + 11.38 \cdot BOOKS + u_0 + r.$$

All coefficients were statistically significant ($p < 0.01$).

$$RA(LTU) = 492.24 + 29.97 \cdot SLOCALITY + 20.59 \cdot FEDU + u_0 + r.$$

Again, all coefficients were statistically significant ($p < 0.01$).

$$RA(LTU) = 502.08 + 35.70 \cdot SLOCALITY + 9.22 \cdot EHLA + u_0 + r.$$

All coefficients were statistically significant ($p < 0.01$).

$$RA(LTU) = 484.61 + 34.42 \cdot SLOCALITY + 21.84 \cdot SATR + u_0 + r.$$

In this case, all coefficients were statistically significant ($p < 0.01$).

From these equations, it is evident that, for Lithuania, the coefficient next to SLOCALITY is virtually unaffected by the inclusion of the home background factors. Even in the cases of number of books at home and parental level of education, the best indicators of the home educational background situation, the SLOCALITY factor, continues to play a very significant role, with the regression coefficient being equal to about 30. Even after including a number of home background factors together in the regression equation, we obtained a regression equation in which the SLOCALITY factor remained very significant:

$$RA(LTU) = 483.90 + 26.54 \cdot SLOCALITY + 8.72 \cdot BOOKS + 14.76 \cdot FEDU + 5.90 \cdot EHLA + u_0 + r.$$

In this case, all coefficients were statistically significant ($p < 0.05$).

These results did not allow us to conclude that in the Eastern-European countries school locality factor was an indicator of different home background situation, as it was in the Anglo-Saxon countries. We therefore decided to investigate what the school locality factor in the Eastern-European countries represented.

Dissolution of school locality factor for Lithuania

Although, in regard to Lithuania, we could not attribute the effect of the school locality factor solely to the particular home background situation of an individual student, we were still left with the question of what is behind the school locality factor. It is logical to assume that it was not the locality itself that influenced the reading achievement of the Grade 4 students who participated in PIRLS 2001, but rather that certain features that related to the type of locality contributed to the lower or higher reading literacy or general educational achievements of these students. The hypotheses that come to mind are twofold. The first relates to the difference in quality of education between cities/towns and villages. Although this assumption is an unpleasant one for policymakers, it is plausible. Social and economic differences in the circumstances surrounding lives in cities/towns and rural communities can also easily influence the quality of education, especially in regard to school facilities and teachers' qualifications. Teachers with higher qualifications generally prefer to work in a city or a town rather than a village. However, we still cannot comfortably draw the conclusion that the difference in the quality of education between the urban and rural schools is responsible for the difference of about 30 scale points in the reading achievement of the Grade 4 students. To do this, we need to investigate if there are other factors behind the factor of school locality.

This led to our second hypothesis, which posited that, in addition to the home background situation of the individual student, the overall home background situation of the students in the class or school somehow influences the overall motivation and educational aspirations in the class/school, which in turn influences individual reading achievement. If that is the case, and if the overall situation is different in rural and urban schools, then this finding might also explain the effect of this factor.

With this hypothesis in mind, we calculated the overall home background factors for each school (i.e., for the sampled Grade 4 students at each school) and if used them as school-level factors. We took the home background factor that showed the highest impact on students' reading achievement (FEDU) and calculated the aggregated school-level factor SFEDU (father's education), which we included in the further regression

estimations. We obtained the following hierarchical linear regression equations:

$$RA(LTU) = 433.65 + 16.61 \cdot SLOCALITY + 71.10 \cdot SFEDU + u_0 + r,$$

$$RA(LTU) = 434.68 + 16.09 \cdot SLOCALITY + 53.74 \cdot SFEDU + 18.23 \cdot FEDU + u_0 + r,$$

$$RA(LTU) = 422.79 + 16.97 \cdot SLOCALITY + 59.16 \cdot SFEDU + 18.63 \cdot FEDU + u_0 + r,$$

$$RA(LTU) = 440.26 + 15.21 \cdot SLOCALITY + 57.22 \cdot SFEDU + 10.52 \cdot BOOKS + u_0 + r,$$

$$RA(LTU) = 426.03 + 17.93 \cdot SLOCALITY + 67.98 \cdot SFEDU + 8.70 \cdot EHLA + u_0 + r.$$

In all cases, the coefficients were statistically significant, with $p < 0.05$.

These hierarchical linear regressions show that inclusion of the school-level factor that indicated the overall home educational situation of the students in the school sharply lessened the impact of the school locality factor, although it still remained a significant one. We can therefore conclude that, for Lithuania, the school locality factor acts alongside the home background of an individual student, and that it can be divided into two main factors: one related to the average home background situation of students at school, and the other to the fact that the quality of education in rural and urban communities is not the same.

Conclusions

The Scandinavian countries in our analysis showed the highest level of equity in education, whereas the Eastern-European and Anglo-Saxon countries showed different tendencies of inequity. In Eastern Europe, the reading achievement of students from rural communities was lower than the achievement of students from urban communities. In Anglo-Saxon countries, the performance of students from rural communities was higher than that of their peers from urban communities. Our investigation indicates that, in the case of Anglo-Saxon countries, the difference in achievement comes mainly from the different social and economic family backgrounds evident in cities/towns and villages, whereas in the Eastern-European countries, the differences in achievement cannot be attributed solely to different

home background situations. In these countries, the locality factor also acts in addition to a student's home background situation. Further analysis of the school locality impact in Eastern-European countries led us to conclude that the school locality factor can be divided into two main factors. The first relates to the "average home background situation of students in a school" (or, in other words, the social and educational peer-environment), and the second to the finding that the quality of education provided in rural and urban communities is not the same.

References

- Blau, D. M. (1999). The effect of income on child development. *The Review of Economics and Statistics*, 81(2), 261–276.
- Brown, K. M., Anfara Jr., V. A., & Roney, K. (2004). Student achievement in high performing, suburban middle schools and low performing, urban middle schools: Plausible explanations for the differences. *Education and Urban Society*, 36(4), 428–456.
- Elijio, A. (2004). Impact of home factors on students' reading achievements. *Lithuanian Mathematics Journal*, 44(special issue), 441–446.
- Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 user guide for the international database*. Chestnut Hill, MA: Boston College.
- Israel, G. D., Beaulieu, L. J., & Hartless, G. (2001). The influence of family and community social capital on educational achievement. *Rural Sociology*, 66(1), 43–68.
- Jensen B., & Seltzer, A. (2000). Neighborhood and family effects in educational progress. *The Australian Economic Review*, 33(1), 17–31.
- Johansone, I., & Foy, P. (2004). PIRLS 2001 results in the context of the European Union Expansion. In C. Papanastasiou (Ed.), *Proceedings of the IRC-2004* (Vol. 3, pp. 36–45). Cyprus: Cyprus University Press.
- Lee, V. E., & Croninger, R. G. (1994). The relative importance of home and school in the development of literacy skills for middle-grade students. *American Journal of Education*, 102(3), 286–329.
- Thirunarayanan, M. O. (2004). The "significantly worse" phenomenon: A study of student achievement in different content areas by school location. *Education and Urban Society*, 36(4), 467–481.

Secondary analysis of PIRLS 2001 Norwegian data¹

Victor H. P. van Daal, A. Charlotte Begnum, and Ragnar Gees Solheim
National Center for Reading Education and Reading Research
University of Stavanger, Norway

Herman J. Adèr
(Formerly of the) Clinical Epidemiology and Biostatistics Department
VU University Medical Center
Amsterdam, the Netherlands

Abstract

Data from 3,459 students in Grade 4, of whom 98% were born in 1991 in Norway and who took part in the Progress in International Reading Literacy Study (PIRLS) 2001, were analyzed to determine which variables in isolation and which constructed scales had an impact on the Rasch-scaled scores for reading comprehension. Urdu students obtained lower comprehension scores on narrative texts. In general, better results were obtained on both expository and narrative texts by students who did not have a negative attitude toward reading, had more books at home, had developed emergent literacy

skills, and whose mothers had higher levels of education. Students with fathers with higher levels of education did better only on expository texts. The students who were better at understanding narrative text were those who spoke Norwegian with adults at home, had entered school with higher literacy skills, and undertook more reading activities at home. No interactions of the aforementioned factors with gender, age, or SES were found. These results are discussed in a model that links the early steps in language development to learning outcomes at school.

Introduction

Reading comprehension

Reading comprehension may be the most important skill we need in order to profit from education and training. We learn about all sorts of subjects through reading books, and students learn many skills both in and outside school through perusing written instruction. We also need to be able to comprehend text to do our jobs properly and, simply, to live an everyday life. Essentially, we need to be able to understand what written messages are about. Perhaps the area where reading comprehension really has “added value” is when we read books for pleasure. Thus, it is important to understand what reading comprehension is and how we can teach it to children, and perhaps to adults who have not yet had opportunity to acquire it.

The challenge of understanding what reading comprehension entails is daunting. The psychological process seems to be hugely complex. However, the assumption that reading comprehension builds on

listening comprehension has proven a good starting point (Kintsch & Rawson, 2005). According to these authors, comprehension largely depends on automatic processes that help us build up a representation of the text at hand. Automatic processes are processes that do not require conscious effort to execute them, such as listening comprehension (in one’s native language). Another process that needs to be automatic is word recognition. Word identification processes need to be automatic so that we have resources available for understanding what the text is about. For an extensive discussion of automaticity in reading, see van der Leij and van Daal (1999).

A text is represented at several levels that include a linguistic structure, a semantic representation, and a so-called situation model, that is, a mental model of what the text is about. Although this model adequately describes how adult readers may process a text in order to understand it, the processes involved in

¹ The authors thank Alexandra Gottardo (Wilfrid Laurier University, Canada), visiting professor at the National Center for Reading Education and Reading Research, for comments on this manuscript.

learning to comprehend what we read and in teaching children how to understand a text seem even more complicated.

Perfetti, Landi, and Oakhill (2005) suggest that the essential skills children should acquire include the following:

1. The parsing of meaning and forming of sentences into a text representation;
2. Building up a situation model on the basis of the text representation; and (most importantly)
3. Drawing inferences, that is, making the text coherent, because no text is completely explicit.

The skills that underpin these understanding processes include:

1. General knowledge—knowledge about the world, how things work, etc.;
2. Linguistic knowledge—being familiar with the phonology, syntax, and morphology of the language; and
3. Vocabulary—having a lexicon, with fully specified meanings and the phonological, orthographic, and syntactic characteristics of words available.

Finally, the model developed by Perfetti and colleagues assumes that the real “bottleneck” in reading for meaning is decoding skill, that is, quick word recognition (see also Perfetti, 1985).

This paper is not about how these cognitive processes work, however, but about factors that influence the reading comprehension of Grade 4 students in school. Our aim in preparing this paper was to explore which variables, at different levels, affect reading comprehension performance. Because reading comprehension skill builds on other skills, we hypothesized that, in the first place, circumstances at home, along with individual differences between children, make a difference. We also assumed that both the school (at a more general level) and teachers and their teaching of reading make a difference.

In this present paper, we concentrate on a secondary analysis of the Norwegian PIRLS 2001 data. In particular, we endeavor to answer the question of what determines Grade 4 Norwegian children’s performance in reading comprehension. An answer to this question is needed in light of an evaluation of initiatives taken by the Norwegian government to improve the reading

comprehension skills of children in Norway. To this effect, a strategic plan, titled *Gi Rom for Lesing!* (*Make Space for Reading!*), was launched in 2003. The effects of the implementation of this plan can be assessed by comparing the PIRLS 2001 data with the PIRLS 2006 data, which are due for release soon.

The *Gi Rom for Lesing!* plan has three main aims: (i) to improve reading skills and increase reading motivation; (ii) to improve teacher’s skills in the teaching of reading, to provide materials for libraries, and to stimulate the use of libraries; and (iii) to increase people’s awareness that reading is important for learning and working. In this present study, we used the results of PIRLS 2001 as a baseline to assess the efficacy of *Gi Rom for Lesing!* and looked into two factors that are typical of the Norwegian situation. First, and overall, girls perform better than boys in reading comprehension (a relatively large difference compared with this gender difference in other countries). Second, Norwegian has two written languages (one of which might be harder to acquire than the other).

Reading comprehension assessment in PIRLS 2001

The assessment of reading comprehension in PIRLS 2001 involved the following process. Students read a series of texts and then answered both open-ended questions and multiple-choice items. The texts and the questions were designed to ensure they were as free of cultural bias as possible, but that they still had meaning or provided potential student engagement, despite translation. Effort was also made to make the texts equivalent across several factors: topic and theme appropriateness; fairness and sensitivity to racial, gender, ethnic, and religious considerations; the nature and level of linguistic features; and density of information. Furthermore, texts were made equal with respect to readability, using the Fry Index.

Four of the eight texts used in the assessment were expository texts; the other four were narrative texts. After data cleaning, Item Response Theory (IRT) was used to estimate scale scores for understanding narrative texts, for understanding expository texts, and for the combination of the two, because each student was only given two texts of each kind to keep the testing time as short as possible (Martin, Mullis, & Kennedy, 2003).

Method

Sample

The Norwegian PIRLS 2001 sample had 3,459 students in Grade 4. Of these students, the 98% born in 1991 were selected. Twenty-six students were absent during the testing sessions, and another 40 did not report gender on the questionnaires. We tested 1,644 girls and 1,775 boys. The testing was conducted in Bokmål (one of the two Norwegian languages) for 1,261 girls and 1,378 boys. The numbers for Nynorsk (the other language) were 383 girls and 397 boys. Ninety-one percent of the children were born in Norway. Of the children born abroad, three had as their mother tongue Sami (which is spoken in the north of Norway), 56 had Swedish or Danish, 77 English, 11 Urdu, and 102 had another language as their mother tongue.

Variables

Measurements were taken at four levels: student, parent/guardian, teacher, and school. The students answered questions on the following: activities outside school, including literacy activities; reading activities in the school; reading related to homework; the use of computers; their attitudes and self-concepts with respect to reading; what they thought of their schools; and more. The parents/guardians were questioned about how they supported their child with literacy activities at home (both at the time of the research and when their child was young), about their own reading activities, and about their socioeconomic status, education, and so on. Teachers reported on their teaching, assessment, and remediation of reading comprehension; their training; and their experience as a teacher. The principal filled in the school questionnaire, which dealt with school characteristics, including the size of the school, the resources available, and so on. The following describes the variables and the scales that were constructed.

• Student questionnaire

- Month and year of birth.
- Reading activities outside school: This scale consisted of items like “I talk with my family about what I’m reading” and “I listen to someone at home read aloud to me.”
- Amount of reading material: This scale was constructed from items such as “How often do you read comics, magazines, subtitles on TV?” “How often do you borrow books from the library to read for fun?”
- The amount of time students spent watching TV or videos outside school.
- Reading activities within school: This scale was constructed from items that included “My teacher reads aloud to the class” and “I read silently on my own.”
- Activities after reading in class: This included items such as “I write about what I’ve read” and “I talk with other students about what I’ve read.”
- How often students read for homework.
- The amount of time students spent reading for homework.
- The availability of a computer.
- Students’ self-reported attitudes toward reading: This contained items such as “I think reading is boring.”
- Students’ estimation of their reading skill.
- What the students thought about their school.
- Students’ reports of negative events at/aspects of school, such as bullying.
- Language spoken by students when little.
- How often students spoke Norwegian at home.
- How often students spoke Norwegian with adults at home.
- The number of books in students’ homes.
- Wealth indicators, such as TV-sets, faxes, etc.
- The number of members of the family at home.
- The number of children living at home.
- Whether students were born in Norway and, if not, how long they had lived in the country.
- Whether students’ father and/or mother were born in the country.

• Parent/guardian (home questionnaire)

- Who completed the questionnaire.
- Early literacy/language activities at home, such as singing songs.
- Number of years in kindergarten.
- At what age the student entered primary education.
- Early reading skills upon entrance to primary school, such as knowing the letters.
- Reading activities done at home.
- Type of contact with the school and attitude toward school.
- Reading habits of the parent/guardian.
- Reasons parent/guardian had for reading.

- Number of books at home.
- Number of children's books at home.
- Level of education of father and of mother.
- Employment situation of father and of mother.
- Father's job and mother's job.
- How well-off parent/guardian thought they were.
- Annual household income.

• Teachers

- Number of students in total and number at Grade 4 level in their class.
- Description of the reading level of their class.
- Number of students who had problems understanding Norwegian.
- Number of students needing and receiving remedial reading instruction.
- Number of students receiving enrichment reading instruction.
- Language instruction: This included curriculum aspects, intensity, frequency, and assignment of language homework.
- Curriculum aspects of reading instruction.
- Total time for reading instruction.
- Time spent on formal reading instruction.
- Frequency of reading instruction.
- Grouping of students for reading instruction.
- Resources used for reading instruction.
- Variety of texts used.
- Differentiation of instruction to students with different reading levels.
- Reading activities in reading instruction.
- Assignments for students after reading.
- Reading comprehension strategies.
- Media other than books used in reading instruction.
- Use of computers in reading instruction.
- Use of library and reading corner.
- Homework assignments.
- Reading difficulties: This included resources available and the action the teacher took on finding reading problems.
- Assessment of progress in reading.
- Opportunities provided by the school for professional development.
- Discussion of reading curriculum with colleagues.
- Contact with students' parents.

- Expectancy with respect to students' performance at the end of Grade 4.
- Training and experience.

• School

- Enrolment of students.
- Location of the school.
- Facilities such as library available.
- Characteristics of students enrolled, such as language, home background, etc.
- Characteristics of the instruction, such as amount of time, curriculum.
- Characteristics of reading instruction in the school.
- Materials available for reading instruction.
- Emphasis on decoding/reading comprehension.
- Library and other resources.
- Contacts with students' homes.
- School climate.
- Teacher collaboration.
- Tasks for the principal (who filled in the questionnaire).

Analysis

The data were analyzed according to the following plan. First, we looked at whether the items had discriminative power and whether we could reduce the data set by combining items. For both questions, we used the strategy of splitting, at random, the data file into two parts, so that the results of identifying discriminative items and constructing scales in one part of the data could be cross-validated against the results in the other part.

Second, we multiplied imputed missing data, using the SPSS 13 missing values procedure MVA (regression imputation). However, we estimated the missing values on the basis of the whole file, in order to get optimal estimates. We then split the files for the next steps of the analysis. The following analyses were done on one half of the imputed data sets, and later repeated on the other half, to detect whether imputing the missing values had an effect on the modeling process described below.

The next step was to run GLM (multivariate analysis of covariance) to find out which items or constructed scales contributed to variability in the total comprehension score, the score for narrative texts, and the score for expository texts. These scores were defined as the dependent variables, whereas gender, language

(Bokmål, Nynorsk), and being an Urdu student (or not) were defined as independent variables. Age and all other items and constructed scales were considered covariates.

Our final step was a multilevel analysis. We used MLwiN 2.0 (Rasbash, Steele, Browne, & Prosser, 2005) to assess which factors (either isolated items or constructed scales) determined the Rasch-scaled score for overall reading comprehension, comprehension of expository texts, and comprehension of narrative texts. We used the same strategy as above of splitting the files. On finding a satisfactory model, we conducted checks, using the other half of the file and retained factors that could be identified in both halves. Within the multilevel analyses, we first assessed the influence of main factors. We assessed interactions only for factors in which we had a theoretical interest (age, gender, Urdu, SES, and whether or not students were in a combination class), or for factors that we assumed to be main factors but that for some reason did not have an influence on their own. In short, all relevant variables were entered first, and then the model was reduced by taking out variables that did not significantly contribute to explaining variance in the dependent variables.

We expected age would be important, as children within one class can vary widely with respect to age and, therefore, with respect to maturity, which may lead them to differentially profit from both instruction at school and stimulation at home. As previously noted, gender differences in reading performance appeared to be relatively large in Norway in PIRLS 2001, and so we were interested in determining the possible causes of these differences.

The main reason to use multilevel modeling is that the PIRLS design pertains to a home/student level that is nested under the school/teacher level. The home and the student levels need to be combined because they are dependent: one student always belongs to one household. In doing the same for the teacher and school combination, we ignored the fact that there were a few schools with more than one class (and teacher) taking part in PIRLS 2001. Furthermore, with multilevel modeling, it is possible to more precisely examine the interaction patterns of factors at different levels. For an overview of the steps in the data analysis, see Figure 1.

Results

Data exploration: Descriptives

For all variables, we looked at the parameters of the frequency distribution. Items with no or almost no variance were removed. Items that were too skewed were also removed.

Data reduction: Scales

The next step was to compute scales for those variables that were thought to measure the same thing. In Tables 1 to 4, we present the complete list of variables, with Cronbach's α for the scales constructed. For all scales, a similar α was found in the other half of the data, except for the teacher scale 7—frequency of different activities with the computer—for which we found a considerably higher α of .54 in the other half. When no α is given, this means that either no reliable scale could be constructed or the question consisted of only one item.

Missing values analysis

We carried out this analysis with the MVA procedure of SPSS 13.0, using the regression method, which allows adding random error. As recommended by Rubin (2004), imputation was carried out three times (multiple imputation), to assess the influence of missing data on the modeling process.

Multivariate analysis of covariance

No differences in reading comprehension performance between the students who took the test in either of the two written Norwegian languages were found. This variable was therefore dropped from the multilevel analyses.

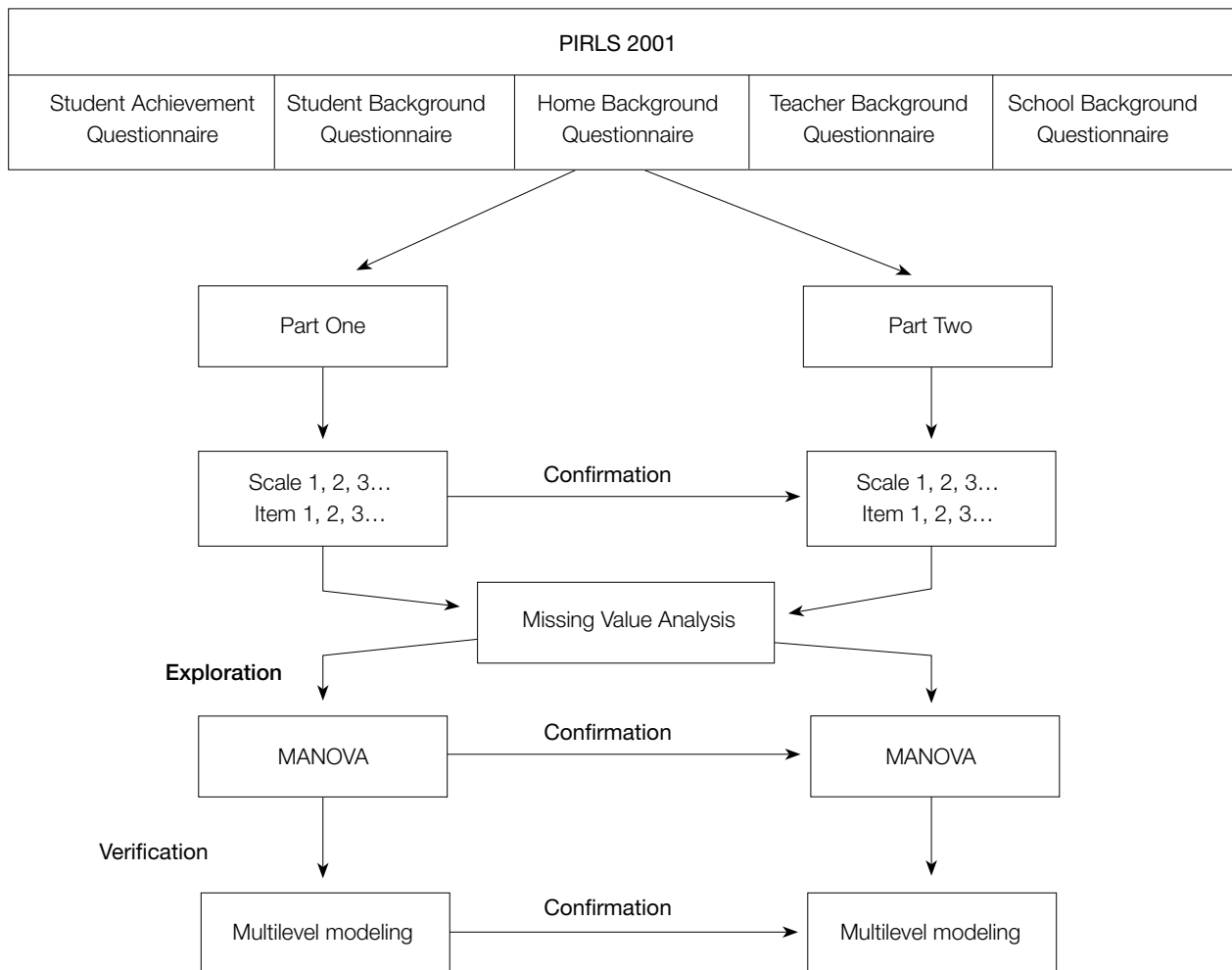
Multilevel modeling

The results are summarized in Table 5. We report only the effects of variables and constructed scales that were consistent across data splits and multiple imputations. Table 5 gives the parameter and standard error estimates for the first imputation.

Home level

Students who had developed good early reading skills at home were better at comprehending both narrative and exploratory texts than were children whose early reading skills were not as proficient. Students with better-educated mothers did better on both types of

Figure 1: Plan for Secondary Analysis of PIRLS 2001 Data for Grade 4 Norwegian Students



texts than did children with mothers who were not as well educated. Students with better-educated fathers did better than did students with less well-educated fathers on expository texts only. Finally, students who undertook more reading activities at home did better on narrative texts than students who undertook fewer reading activities at home.

Student level

Urdu students performed less well than non-Urdu students on narrative texts, as did students in general who did not speak Norwegian at home with adults. Students scored higher on both text types when they had a less negative attitude toward reading. Students also scored higher on the total comprehension score when they reported fewer negative “things” happening at school. However, this was not consistent across types of text. Finally, the more books students reported

being available at home, the better they performed on the comprehension of both expository and narrative texts.

School level and teacher level

Students who had better reading skills when they entered school performed better on comprehending narrative texts.

Discussion

We first discuss the factors that unexpectedly did *not* contribute to the reading comprehension performance of Grade 4 students in Norway. Most importantly, we could not find systematic differences between boys and girls. However, in some parts of the data, we found large groups of boys attaining low total comprehension scores, and likewise for the narrative texts. We did not find this pattern for comprehension

Table 1: Scale Constructed in Relation to Student Questionnaire

Question	Scale number (Item numbers from booklet)	Cronbach's α
Gender	(1)	
Age	(2)	
Reading activities outside school	Scale 1; (3)	.61
Amount of reading materials	Scale 2; (4)	.62
How often student borrowed books from library to read for fun	(5)	
Time spent watching TV or videos outside school	(6)	
Reading activities within school	Scale 5; (7)	.48
Activities after reading in class	Scale 6; (8)	.77
How often students read for homework	(9)	
Time spent reading for homework	(10)	
Availability of a computer	(11, 11a, 11b)	
Attitude toward reading	Scale 10a negative; (12) Scale 10b positive; (12)	.49 .59
Estimation of reading skill	Scale 11; (13)	
Thoughts about school	Scale 12; (14)	.58
Reports on negative aspects of school	Scale 13; (15)	.67
Language spoken when little	(16)	
How often student spoke Norwegian at home	(17)	
How often student spoke Norwegian with adults at home	(18)	
Number of books at home	(19)	
Wealth indicators, such as TVs, faxes, etc	Scale 18 (20)	.62
Size of family at home	(21)	
Number of children living at home	(22)	
Whether student born in Norway and, if not, how long he/she had lived in the country	(23, 23a)	
Whether father and/or mother were born in the country.	(24, 25)	

of expository texts. A similar picture emerged for age. Overall, we were unable to determine whether older students always did better than younger students within a class. Finally, we could find no differences between the two written Norwegian languages, both through the MANOVA and the multilevel modeling. However, *t*-test calculations on the complete data set (and the *t*-test is, in our opinion, an inadequate technique) did produce differences.

At the home level, we found no effects for the number of years the children had attended kindergarten, the age at which they entered primary school, their parents' reading habits and reasons for reading, their parents' employment situation, and how well-off their parents thought their family was. There is ample evidence that early language activities at home indeed support the development of letter knowledge and phonological processing skills (Bowey,

Table 2: Scale Constructed in Relation to Parent/Guardian (Home) Questionnaire

Question	Scale number (Item numbers from booklet)	Cronbach's α
Who completed the questionnaire	(1)	
Early literacy/language activities at home	Scale 1; (2)	.67
Number of years in kindergarten	(3)	
Age at which child entered primary education	(4)	
Early reading skills at time of entering primary school	Scale 2; (5)	.91
Reading activities at home	Scale 3; (6)	.73
Type of contact with the school and attitude toward school	Scale 4; (7) Scale 5; (8)	.65 .81
Reading habits of parent/guardian	(9)	
Reasons why parent/guardian read	Scale 6; (10)	.48
Number of books at home	(12)	
Number of children's books at home	(13)	
Level of education of father and of mother	(14)	
Employment situation of father and of mother	(15)	
Father's job and mother's job	(16)	
How well-off parent/guardian thought they were	(17)	
Annual household income	(18)	

1995). However, these skills pertain to decoding processes that children have already mastered during their first few years in primary school rather than to reading comprehension skills that develop later. This could explain why we did not find effects for parents' reading habits and reasons for reading. From school readiness studies, we know that delayed school entry and, therefore, age and reading skills when entering school, do not have long-term benefits (Phillips & Lonigan, 2005). The parents' employment situation and how well-off parents think they are possibly play a lesser role in Norway, where differences in incomes are generally smaller and where all people tend to have better access to basic education than do people in other countries.

At the student level, a surprising number of scales, even though sufficiently reliable, did not predict reading comprehension at all. These included the scales for amount and variety of reading materials, reading in and after class (at school), students' estimates of their reading skill, and what students thought of their schools. Also included here was the scale on wealth

indicators. A possible explanation for these findings is that differences already exist between children by the time they enter school. The variables that did make a difference, however, were those relating to the homes of the children. The scale on wealth indicators obviously tapped into the same influence as the item in the home questionnaire on how well-off the parents thought the family was.

At the teacher level, we found no effect of scales and variables relating to the "input" characteristics of the students, such as their reading level and whether they needed remedial instruction, and also curriculum aspects of language and reading instruction. Null results were also found for the way the reading instruction was organized with respect to teaching, assessment, differentiation (individualization of instruction), resources and homework assignments, and the training and experience of the teacher. Note that these considerations are not entirely in line with older research, in which teachers' variables such as experience and the extent to which they themselves read books were found to play a role (Postlethwaite &

Table 3: Scale Constructed in Relation to Teacher Questionnaire

Question	Scale number (Item numbers from booklet)	Cronbach's α
Number students in total and number at Grade 4 level in class	(1)	
Description of reading level	(2)	
Number of children having problems understanding Norwegian	(3)	
Number students needing and receiving remedial reading instruction	(4)	
Number students receiving enrichment reading instruction	(5)	
Curriculum aspects of language instruction	(6,7)	
Curriculum aspects of reading instruction	(8)	
Total time for reading instruction	(9a)	
Time spent on formal reading instruction	(9b)	
Frequency of reading instruction	(10)	
Grouping of students for reading instruction	(11)	
Resources used for reading instruction	(12)	
Variety of texts used	Scale 2; (13)	.62
Differentiation of instruction to students with different reading levels	(14)	
Reading activities in reading instruction	Scale 3; (15)	.45
Assignments for students after reading	Scale 4; (16)	.62
Reading comprehension strategies	Scale 5; (17)	.66
Media other than books used in reading instruction	Scale 6; (18)	.68
Use of computers in reading instruction	Scale 7; (19, 19a, 19b, 19c)	.54 (19c; in second half of data set)
Use of library and reading corner	(20)	
Homework assignments	(22, 23)	
Reading difficulties	Scale 8; (25)	.40
Assessment of progress in reading	Scale 9; (26, 27, 28)	(27) .61
Opportunities provided by the school for professional development	(29)	
Discussion of reading curriculum with colleagues	(30)	
Contact with students' parents	(31)	
Expectations of students' reading performance	(32)	
Training and experience	(33–43)	

Ross, 1992). These researchers found, however, effects in relation to the amount of time spent on teaching the class. The null effect of having high expectations of the students in the class is not in line with research by Pressley (1998).

Let us now turn to the factors and variables that were found to have an effect on reading comprehension

at the home and the student levels. The factors at the home level were reading activities done at home, the number of books at home, the mother's and father's education level, and developing early reading skills. The finding that reading activities at home have a positive effect on reading comprehension at school does not come as a surprise. It can be explained as simply more

Table 4: Scale Constructed in Relation to School Questionnaire

Questions	Scale number (Item numbers from booklet)	Cronbach's α
Enrolment of students	(1–3)	
Location of the school	(4, 5)	
Facilities available	(6)	
Characteristics of students enrolled	(7–9)	
Characteristics of the instruction in general	Scale 1; (13) (10–12)	.59
Characteristics of reading instruction in the school	Scale 2; (14) Scale 3; (15) (16, 17, 20)	.77 .52
Materials available for reading instruction	Scale 4; (18)	.51
Emphasis on decoding/reading comprehension	Scale 5 (decoding) Scale 6; (both 19)	.67 .52
Library and other resources	Scale 6; (24)	.92
Contacts with students' homes	(21–23) Scale 7; (27) (25–26)	.56
School climate	Scale 8; (28) Scale 9; (29)	.76 .84
Teacher collaboration	(30, 31)	
Tasks for the principal	(32)	

“time on task.” It is also specific, as it has an effect on narrative texts only. Having many books available will provide opportunities to read “a lot,” and even more so if there is support from the mother, who is usually the parent who engages in reading activities with the children. This support may be even more effective when the mother has a higher education level. The whole process, in turn, may result in students developing a positive attitude toward reading or, at least, in not developing a negative attitude. Taken together, these findings suggest that the home specifically prepares students for understanding narrative texts.

It might be worthwhile to tie the different levels together in one model. Lundberg (2002) has proposed such a model, but presents it with a focus on how reading difficulties come into existence. In short, the model points to two strands of development of reading skill: word recognition and comprehension. Underlying skills for word recognition are phonological recoding (converting letters to sounds) and orthographic

processing (fast recognition of familiar words), which are, in turn, based on phonological development and automaticity. Comprehension is underpinned by vocabulary and by general language skills (syntactic and semantic development), and vocabulary is also supported by phonological development. It is obvious that basic cognitive processes such as memory and perception are also involved. Of relevance for this paper is how the variables and factors at the respective levels of home, student, teacher, and school contribute to the development of reading skill and, more specifically, to the development of reading comprehension skill. It may, however, be realistic to expect relatively more influence of home and student variables than of school and teacher variables, as has been the case in the research on the development of reading skills in much younger children (54 months old), for whom we would expect a relatively large influence of school and teacher variables, especially in relation to the formal instruction of reading (Connor, Son, Hindman,

Table 5: Results of the Multilevel Analysis

Level	Variable/Constructed Scale	Total First Part	Total Second Part	Expository First Part	Expository Second Part	Narrative First Part	Narrative Second Part
Home	Reading activities			0.527 (0.200)	0.690 (0.282)		
	Early reading skills			0.979 (0.172)	1.178 (0.192)		
	Education mother			0.301 (0.100)	0.316 (0.100)	0.297 (0.297)	0.401 (0.101)
	Education father			0.465 (0.095)	0.261 (0.097)		
Student	Books at home			0.698 (0.179)	0.598 (0.180)	0.507 (0.183)	0.382 (0.180)
	Speak Norwegian to adults at home					-0.809 (0.258)	-0.646 (0.264)
	Urdu as first language					-4.827 (2.408)	-9.320 (1.987)
	Negative attitude toward reading			1.319 (0.312)	3.211 (0.485)	1.531 (0.330)	2.603 (0.416)
School and Teacher	Negative things at school	1.608 (0.535)	1.114 (0.556)				
	Skills when started school					1.448 (0.469)	0.873 (0.434)

Note: Estimates and standard errors (in parenthesis) for parameters of the final multilevel model.

& Morrison, 2005). Instead, in this research, the main factors of influence found were home learning environment and family socioeconomic status.

Support for the Lundberg model comes from different sources. Language development starts long before children go to school. Newborns are initially able to discriminate among all the sounds of the world's languages. However, in time, children gradually refine their recognition to encompass the sounds of the language in which they grow up, while their ability to discriminate the sounds of other languages completely disappears (Juszyk, 1997; Vihman, 1996). From studies with children with dyslexia, we know that their phonological representations are less stable than these representations for other children (Elbro, 1996) and that they therefore have trouble acquiring a properly functioning vocabulary (Scarborough, 1990). Hart and Risley (1995) reported pronounced differences in vocabulary among three-year-old children. These differences were completely dependent on the number of words used in the children's home environments. Moreover, as Scarborough and Dobrich (1994) found, about 10% of reading ability can be explained by someone reading aloud to children. Although the model is far more complicated than can be described in this paper, one more piece of evidence is worth mentioning. Children who are read to daily acquire a schema for the structure of stories (Cain, 1996). It is

this schema that plays a role later, when the children need to read not only stories but also expository texts at school.

Finally, we need to mention a limitation of this study. We have not yet been able to verify the model sketched above because many other variables influence reading comprehension. Given the complex relationship between decoding and comprehension, and especially with respect to the developmental trajectory, it seems valuable to subsume word decoding under the model. We have therefore assessed word-decoding skill in the PIRLS 2006 project in Norway. The (verified) model should be seen, however, as a set of interrelated hypotheses. As a consequence, specific parts of the model need to be elaborated. In this respect, we need to mention a few issues we are currently working on. First, what teachers do in the classroom may differ from what they report they do. This consideration is not just a matter of giving socially desirable responses. A lack of training could result in the teachers not knowing exactly what to do in reading comprehension lessons. Second, it may make a difference if teachers are able to appropriately interpret the requirements of the reading curriculum. And last, but not least, we need to assess, through cross-country modeling, which factors are universal and which are country-specific in reading comprehension.

References

- Bowey, J. A. (1995). Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement. *Journal of Educational Psychology*, 87, 476–487.
- Cain, K. (1996). Story knowledge and comprehension skill. In C. Cornoldi & J. V. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 167–192). Mahwah, NJ: LEA.
- Connor, C., Son, S-H., Hindman, A., & Morrison, F. J. (2005). Teacher qualifications, classroom practices, family characteristics, and preschool experience: Complex effects on first graders' vocabulary and early reading outcomes. *Journal of School Psychology*, 43, 343–375.
- Elbro, C. (1996). Early linguistic abilities and reading development: A review and a hypothesis about distinctness of phonological representations. *Reading and Writing*, 8, 453–485.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.
- Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 209–226). Oxford: Blackwell Publishing.
- Lundberg, I. (2002). The child's route into reading and what can go wrong. *Dyslexia*, 8, 1–13.
- Martin, M. O., Mullis, I. V. S., & Kennedy A. M. (Eds.). (2003). *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.
- Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.
- Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford: Blackwell Publishing.
- Phillips, B. M., & Lonigan, C. J. (2005). Social correlates of emergent literacy. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 173–187). Oxford: Blackwell Publishing.
- Postlethwaite, T. N., & Ross, K. N. (1992). *Effective schools in reading: Implications for educational planners*. Hamburg: International Association for the Evaluation of Educational Achievement.
- Pressley, M. (1998). *Reading instruction that works: The case for balanced teaching*. New York: Guilford Press.
- Rasbash, J., Steele, F., Browne, W., & Prosser, B. (2005). *A user's guide to MLWin, Version 2.0*. Bristol: Centre for Multilevel Modelling, University of Bristol.
- Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Scarborough, H. S. (1990). Very early language deficits in dyslexic children. *Child Development*, 61, 1728–1743.
- Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review*, 14, 245–302.
- van der Leij, A., & van Daal, V. H. P. (1999). Automatization aspects of dyslexia: Speed limitations in word identification, sensitivity to increasing task demands, and orthographic compensation. *Journal of Learning Disabilities*, 32, 417–428.
- Vihman, M. M. (1996). *Phonological development*. Oxford: Blackwell Publishing.

The effect of multilingual policies on performance and progression in reading literacy in South African primary schools

Sarah Howie, Elsie Venter, Surette van Staden, and Michelle van Gelder
University of Pretoria
South Africa

Abstract

South Africa's multicultural society has a rich diversity of languages, as indicated by the number of official languages of the country (11 dominant languages have been declared). South African children start their learning at school in their home language, and this practice continues until Grade 3. However, in the majority of schools, the language of instruction changes: from Grade 4, more than 80% of South African students learn in a second language. This paper seeks to ascertain (i) to what extent South African students are proficient readers in their main language (or alternatively the language in which they have received reading instruction for four years), and (ii) what progress

they make after an additional year of instruction (i.e., what value is added to that language) after the language switch. To address these research questions, descriptive and inferential statistics from the Progress in International Reading Literacy Study (PIRLS) 2006 were applied to determine differences in achievement and correlations among home language, language of instruction, and achievement. The raw data from PIRLS 2006 were used, although these are unweighted and not yet scaled. The results, therefore, should be treated as broad patterns and with caution until they are confirmed (or otherwise) with the scaled and weighted data, which are scheduled for release in November 2007.

Introduction

South Africa participated in the International Association for the Evaluation of Educational Achievement's (IEA) Progress in International Reading Literacy Study (PIRLS) for the first time in PIRLS 2006. This was the most complex study conducted on reading literacy in South Africa to date. While most countries participating in the study tested students at only one grade level, South Africa chose to test students in two grades: 4 and 5. The children, from just over 400 schools, were each assessed in one of South Africa's 11 official languages. Thus, for each child, the language used was the language of the reading instruction he or she had received in Grades 1 to 3.

South African children receive tuition at school in their home language until they reach Grade 3. However, in the majority of schools, the language of instruction changes, and in Grade 4 more than 80% of South African students learn in a second language. The South African situation is complex because most white, Indian, and colored children continue to receive their schooling in the same language of instruction from Grades 1 to 12, namely English or Afrikaans, while most children speaking African

languages at home switch at Grade 4 to either English or Afrikaans, despite the current government language policy permitting learning to take place in their home language throughout school.

The aim of this paper is to report on the progress and development of reading skills and knowledge of South African students from Grade 4 to Grade 5. Specifically, the variation in performance is explored to determine if it is possible to detect the contextual factors that are outlined in the South African students' achievement profiles. The paper therefore addresses the following research questions:

1. To what extent are South African students proficient readers in their main language (or, alternatively, the language in which they have received reading instruction for four years)?
2. What progress do they make after receiving an additional year of reading following the language switch?
3. What is the relationship between the students' performance in reading literacy in the language of the test and the South African students' home language?

Literature review

The importance of literacy has become evident in the 50 years since the United Nations declared it to be a basic human right along with the right to adequate food, health care, and housing. Literacy education has indeed become a tool to help address what might be perceived to be more pressing needs for food, health care, and housing. Fuchs and Woessmann (2004) use the definition of reading literacy employed by the Organization for Economic Co-operation and Development (OECD). They describe reading literacy as the capacity to understand, use, and reflect on written texts in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society.

Reading literacy is therefore understood not as a basic skill, but rather as a goal. It is an important means of functioning effectively in education and of developing as an individual, within and outside school, today and in later life, in further education, at work, and in leisure activities (Linnakyla, Malin, & Taube, 2004). Viewing literacy as a social practice means that reading represents a multitude of evolving human activities with language at its centre (Landis, 2003). According to Frost, Madsbjerg, Niedersee, Olofsson, and Sorensen (2005), reading is an activity used for interpersonal communication, but it also depends on intrapersonal sources such as motivation, attention, imagination, memory, comprehension, and language. Most current theories of reading development stress the fundamental importance of phonological skills to learning to read (Nation & Snowling, 2004). Some researchers, such as Wood, Hill, Meyer, and Flowers (2005), note that phonemic awareness, vocabulary, and fluency variables seem essential for effectively predicting reading performance. Others, such as Beech (2005) and Hempenstall (2004), argue that reading literature pays specific attention to how children progress through different phases of reading according to defined stages of development.

The consequences for students who cannot read or who struggle to read in the early grades are well documented. Leslie and Allen (1999) cite Juel (1988), who reported that 88% of children who scored in the lowest quartile for reading comprehension remained below the 50th percentile at the end of Grade 4.

In terms of reading performance, three systems seem to be of major influence: the home, the school,

and the students themselves. Strickland, Ganske, and Monroe (2002) compiled a list of what is known about successful readers and writers. According to these authors, successful readers have normal to above average language skills and have opportunities to identify letters and environmental print. Children who become successful readers have exposure to adults who involve them in purposeful literacy experiences during the early childhood years. As a result, these children have a fair measure of pleasurable, motivating early childhood experiences with books and literacy. Strickland and colleagues stated that successful readers are also likely to be influenced by responsive adults who listen and talk to them and who are likely to engage them in activities such as rhyming and singing, thus creating an awareness of the internal structure of spoken words. Successful readers are likely to attend schools that provide students with frequent and intensive opportunities to read and write, and that build upon early childhood experiences with opportunities for students to learn the nature of the alphabetic system. Successful readers experience overall progress that is steady and sure, despite periodic difficulties, and in line with this have the ability to build on informal experiences with literacy from early years as they encounter more formal and complex tasks.

When it comes to what is known about students at risk of failure, Strickland et al. (2002) state that some factors pertain to the child's personal development and others to the group or situation they reside in. Children who are particularly at risk of encountering reading difficulties typically have a history of preschool language impairment, limited proficiency in English, or come from homes where a non-standard dialect of English is spoken. Students at risk often have parents who had difficulty learning to read, they may come from poor neighborhoods, and they are likely to attend schools in which classroom practices are deemed ineffective. Strickland and colleagues also point out, however, that none of these factors is an automatic barrier to literacy, and that none of them functions in isolation as a single causal factor of reading difficulties.

The reality for most students in South Africa is one of reading in a second language once they progress to Grade 4. Before Grade 4, the premise is that learning to read takes place in the children's mother tongue (or first language). Education policy aims to accommodate students' language needs by means

of code-switching in the classroom, where teachers switch between English and their students' respective mother tongues. Howie (2003) reiterates that the South African education system has the challenge of providing quality education to a multicultural learner population, the members of which among them speak 11 different official languages. English as a first language is spoken by less than 10% of the population. As one of the languages most used by schools (the other being Afrikaans), English is not the most frequently spoken language at home. Thus, second language acquisition, mastery, and learning is a reality for the majority of students in South Africa. Admiraal, Westhoff, and de Bot (2006) refer to such a system as one of immersion, where a language that is not the language of the larger society is used as a medium of instruction.

According to Ely (2005), children master the rudimentary aspects of their native languages during the first years of life. By age three, they should have reached a large and varied lexicon. By age five, their command of a language is relatively sophisticated. This sophistication should increase and progress as the child enters school and learns to read. D'Angiulli, Siegel, and Maggi (2004) cite a growing body of evidence that shows that the development of reading skills for students for whom English is a second language is similar to the development of reading skills in children for whom English is a first language. Gersten and Geva (2003) support the notion by stating that both English students and English second language students seem to take similar paths of development in specific pre-reading skills such as phonological awareness.

By Grade 4 in the South African education system, many children are immersed in a second language curriculum in which English is the language of teaching and learning (as opposed to mother-tongue teaching and learning). According to Verhoeven (1990), second language students face two types of difficulties, namely inter-lingual learning problems caused by mother-tongue interference, and intra-lingual learning problems, caused by the structure of the second language. Verhoeven (1990) states that there is agreement that word recognition is a critical part of reading and of learning to recognize words, with students using three representational systems, namely phonemic mapping, recognition of orthographic patterns, and direction recognition of words already represented in memory. Children acquiring reading

in a second language may experience difficulty with all three recognition processes (Verhoeven, 1990). The present study therefore uses scores of students as obtained in the mother tongue. Our hypothesis here is that learning difficulties are more pronounced in cases where the above-mentioned recognition processes are absent, even for reading performance in mother tongue.

Method

Sample

The South African sample for PIRLS 2006 consisted of 441 schools offering schooling at Grade 4 level. A proportional allocation was done firstly by province and subsequently by language from 15,182 schools. From this sample, Grade 5 students, where available, were also included in the study. The realized sample involved only 434 Grade 4 schools, because in seven cases the information about the schools (and their replacements) was lacking to such an extent that the schools could not be traced. The realized sample for Grade 5 was 402. The study generated achievement data for 16,288 Grade 4 students and 14,711 Grade 5 students, across nine provinces and 11 languages.

Instruments

We used PIRLS achievement instruments to generate percentages from raw scores for every student and to establish the test language. We took great trouble during the preparation stages of the study to determine and confirm the language of instruction of each class, and we use the terms "test language" and "language of instruction" interchangeably in this paper. An English test was developed as a national option in South Africa to gain some insight into students' reading skills in English as a second language, but although information about home language was sourced from this instrument, the national option as such is not the focus of this paper.

Data analysis

In our preliminary analysis, we used South African data from PIRLS 2006 that were collected and captured during the last part of 2005 and early 2006. It is important to keep in mind that this paper considers preliminary data only. Data are currently being cleaned and verified internationally. At the time of writing, no scaling or weighting had been done. We

therefore used descriptive and inferential statistics to address and attempt to answer the research questions. The data under consideration have been constructed from the achievement data where the language of the test and the home language were known. The number of Grade 4 students complying with the above-mentioned requirement was 15,776. In Grade 5, the home language was available for 14,335 students.

To determine the normality of the data, we plotted a histogram, shown in Figure 1. Although the raw scores are, strictly speaking, on the ordinal measurement scale, for the purposes of this preliminary analysis, the percentages can be regarded on the interval scale of measurement. We first considered the mean performance across test languages and then the performance per grade in the 11 official languages. The final step was to look at the correspondence, if any, between home language and test languages.

Results

Normality of the data

The sample distribution of the variable “percent” is positively skewed, but normality is assumed for the purposes of this paper. The national mean for performance in reading over all languages in Grade 4 was 13.18% ($SE = 0.1176$). For Grade 5, the mean was 18.17% ($SE = 0.1516$), as depicted in Table 1. A t -test revealed a statistically significant difference of 4.99 between the two means. As expected, the difference indicates a progression in the reading abilities of students from Grade 4 to Grade 5.

Language of instruction

Table 2 presents the mean percentages of each test language per grade. In all 11 languages tested, some progression was evident from Grade 4 to Grade 5. The largest improvement in reading percentage between Grade 4 and Grade 5 occurred in Afrikaans, with a difference of 9.36%. English followed closely, with an improvement of 7.88%, while isiZulu showed the

Figure 1: Histogram of Reading Performance in Percentages for Grade 4 and Grade 5 South African Students Participating in PIRLS 2006

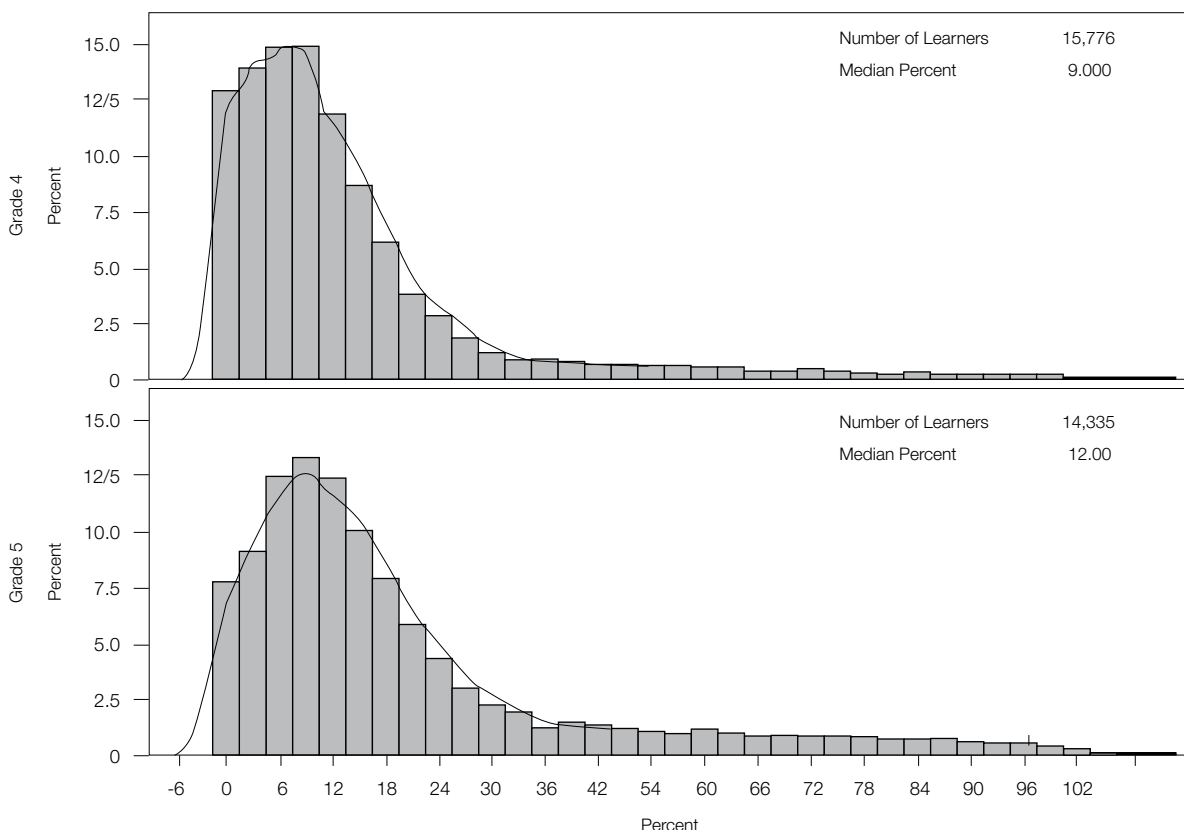


Table 1: Descriptive and Significance Statistics for the National Means for Grades 4 and 5 Students

Grade	N	Mean %	Std. Err.	t-value	p > t
4	5,776	13.18	0.1176	-26.23	< 0.0001*
5	14,335	18.17	0.1516		

Note: * At a 5% level of significance.

lowest progression (2.20%). As a group, the Nguni languages (isiNdebele, isiXhosa, isiZulu, and SiSwati) presented the smallest improvement. This finding was expected, as the indigenous languages are traditionally from an oral and not a written tradition.

Home language versus language of instruction

Another aspect under investigation was the correspondence between a student's home language

and the language of the test. In Grade 4, students who wrote the test in their home language constituted 62.72% of the total number of students. In Grade 5, this percentage was slightly higher at 64.40%. Intuitively, one would think that a test in a student's home language would be more accessible to the student. In other words, one would expect a higher level of reading literacy when a student has access to language instruction in the same language as his or her

Table 2: Descriptive Statistics for Language Used in Test Sat by Grades 4 and 5 Students

Test language	Grade	N	Mean %	Difference
Afrikaans	4	1,674	24.54	9.36
	5	1,667	33.90	
English	4	2,797	20.93	7.88
	5	2,679	28.81	
isiNdebele	4	776	9.16	2.51
	5	810	11.67	
isiXhosa	4	1,567	8.09	2.65
	5	1,513	10.74	
isiZulu	4	2,217	10.83	2.20
	5	1,751	13.03	
Sepedi	4	1,480	8.46	2.45
	5	1,293	10.91	
Sesotho	4	952	10.55	4.94
	5	973	15.49	
Setswana	4	1,147	10.27	4.79
	5	1,050	15.06	
SiSwati	4	1,270	9.04	2.73
	5	1,066	11.77	
Tshivenda	4	886	10.11	3.49
	5	773	13.60	
Xitsonga	4	1,010	9.71	3.74
	5	760	13.45	

Table 3: Descriptive and Significance Statistics for Same and for Different Languages for Grades 4 and 5 Students

Grade	Language	N	Mean %	Std. Err.	t-value	p > t
4	Same	9,894	13.88	0.1565	7.64	< 0.0001*
	Different	5,882	12.02	0.1729		
5	Same	9,232	18.72	0.1968	4.89	< 0.0001*
	Different	5,103	17.17	0.2330		

Note:* At a 5% level of significance.

home language. Table 3 presents the results of a *t*-test to ascertain the difference in performance at each grade between students for whom the home language and the test language corresponded and students for whom the home and test language differed. The difference between the two categories in the case of Grade 4 was 1.86 and for Grade 5 the difference was 1.65. Both differences were statistically significant at the 0.05 level, with *t*-values of 7.64 and 4.89 respectively.

English, Afrikaans, and indigenous languages

Before presenting the results for this section, it is necessary to briefly explain South Africa's language policy as it relates to education. This policy states that mother-tongue education is compulsory in the Foundation phase (Grades 1 to 3). A switch to either English or Afrikaans as language of instruction is prescribed in Grade 4. The implementation of this policy is not ideal in every school for different reasons. In some instances, parents insist that students are taught in English, right from the beginning, in Grade 1. Often, in an urban setting, it is not practical to choose an indigenous language because of the variation among languages in the catchment area and the subsequent mix of students from different language backgrounds at any one school. With this in mind, we compared the performance in the three test languages (English, Afrikaans, and the indigenous languages as a group) in terms of mother-tongue speakers and non-mother-tongue speakers.

Two groups of students were considered. Group 1 consisted of all students who indicated English as their home language and who wrote their answers to the test in English. Group 2 constituted students who had a home language different from English, but who wrote their answers to the test in English. To determine the statistical significance between the average performances of the groups, we performed a

t-test and followed this procedure for Afrikaans and the indigenous group as well. Table 4 presents the results.

In the case of English, a substantial difference was evident between students who were English mother-tongue speakers and students who were English second-language speakers. The difference in Grade 4 for English was 26.78%. In Grade 5, this figure was 29.88%. In Afrikaans, the difference in performance between students who wrote the test in their mother tongue and those who were second-language Afrikaans speakers but wrote the test in Afrikaans was markedly lower at 5.36% and 10.81% for Grades 4 and 5 respectively. Amongst the African indigenous languages, this difference was extremely small and virtually non-existent at three-quarters of a percent in each grade. This small difference was statistically significant, but probably only as a result of a very large sample size and therefore considered not meaningful in this context.

The English mother-tongue speakers in Grade 5 were the only group that attained an average of above 50% (52.44%). This low percentage is a cause for great concern. The message is clear: South African students cannot read at a sufficient level. The majority of students receiving instruction in English when English is not their mother tongue face a serious problem. The mean percentages of 14.83 at Grade 4 level and of 22.56 at Grade 5 for this group (see Table 4) paint a bleak picture indeed of the level of education and achievement in South Africa, even when the increase between the two grade levels is taken into account. The implications for further education and the economic development and contribution of these students to the country are almost too awful to imagine. If it is considered that children in this group have had English as the language of instruction (as per definition of the

Table 4: Descriptive and Significance Statistics for Whether Languages Students Used in Test Were Same as or Different from Home Language

Grade	Language	Home vs test language	N	Mean %	Std. Err.	t-value	$p > t $
4	English	Same	637	41.61	0.9955	32.07	< 0.0001
		Different	2,160	14.83	0.8351		
5		Same	560	52.44	1.0764	29.21	< 0.0001
		Different	2,119	22.56	0.4423		
4	Afrikaans	Same	1,236	25.94	0.6561	4.26	< 0.0001
		Different	438	20.58	1.0266		
5		Same	1,389	35.70	0.6945	6.48	< 0.0001
		Different	278	24.89	1.3602		
4	African	Same	8,021	9.81	0.0912	4.68	< 0.0001
		Different	3,284	9.03	0.1351		
5		Same	7,283	12.89	0.1169	3.28	< 0.0001
		Different	2,706	12.16	0.1837		

administered test language) since beginning school in Grade 1, the gravity of the situation increases.

The rest of the indigenous languages speakers are even worse off. They have to change from mother-tongue instruction to English in Grade 4, and it is clear that they do not have a command of reading even in their own home language.

Conclusions

The project is still at a very early stage, as the data were collected only seven months ago (at time of writing). However, this is the first IEA study where so many languages of instruction have been assessed within one country and where the possibility exists for an extensive exploration of students' performance across such a wide variety of languages.

Given that the analysis is at an early stage, we can only speculate on the reasons for the poor state of reading literacy amongst students in South Africa. The reasons are doubtless varied and will be difficult to

trace. The reading literacy levels could be a function of poor teaching practices, poor or non-existent training in reading practices for the indigenous languages, a grave lack of resources, a lack of a drive or will to alter the situation, the oral tradition of the indigenous languages, and the general role and influence of television. They could also be a function of the times we live in, where instant gratification in so many aspects of life is promoted.

The most recent debates on language of learning in South Africa indicate that the Department of Education is reconsidering its policy on language of learning and may introduce all 11 official languages as languages of learning into the classroom for both primary and secondary schools. PIRLS 2006 is therefore very timely, and the additional national option included by the South African project team is appropriate in terms of providing significant information about the reading literacy levels of children in their first language as well as in English.

References

- Admiraal, W., Westhoff, G., & de Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English. *Educational Research and Evaluation*, 12(1), 75–93.
- Beech, J. R. (2005). Ehri's model of phases of learning to read: A brief critique. *Journal of Research in Reading*, 28(1), 50–58.
- D'Angiulli, A., Siegel, L. S., & Maggi, S. (2004). Literacy instruction, SES and word-reading achievement in English-language learners and children with English as a first language: A longitudinal study. *Learning Disabilities Research and Practice*, 19(4), 202–213.
- Ely, R. (2005). *The development of language*. Boston, MA: Allyn and Bacon.
- Frost, J., Madsbjerg, S., Niedersee, J., Olofsson, A., & Sorensen, P. M. (2005). Semantic and phonological skills in predicting reading development: From 3–16 years of age. *Dyslexia*, 11, 79–92.
- Fuchs, T., & Woessmann, L. (2004). *What accounts for international differences in student performance? A re-examination using PISA data* (CESifo Working Paper No. 1235).
- Gersten, R., & Geva, E. (2003). Teaching reading to early language learners. *Educational Leadership*, 60(7), 44–49.
- Hempenstall, K. (2004). How might a stage model of reading development be helpful in the classroom? *Educational Psychology*, 24(6), 727–751.
- Howie, S. J. (2003). Language and other background factors affecting secondary students' performance in mathematics in South Africa. *African Journal of Research in Mathematics, Science and Technology Education*, 7, 1–20.
- Landis, D. (2003). Reading and writing as social, cultural practices: Implications for literacy education. *Reading and Writing Quarterly*, 19, 281–307.
- Leslie, L., & Allen, L. (1999). Factors that predict success in an early literacy intervention project. *Reading Research Quarterly*, 34(4), 404–424.
- Linnakyla, P., Malin, A., & Taube, K. (2004). Factors behind low reading literacy achievement. *Scandinavian Journal of Educational Research*, 48(3), 231–249.
- Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*, 27(4), 342–356.
- Strickland, D. S., Ganske, K., & Monroe, J. K. (2002). *Supporting struggling readers and writers: Strategies for classroom intervention 3–6*. Portland, OR: Stenhouse Publishers.
- Verhoeven, L.T. (1990). Acquisition of reading in a second language. *Reading Research Quarterly*, 25(2), 90–114.
- Wood, F. B., Hill, D. F., Meyer, M. S., & Flowers, D. L. (2005). Predictive assessment of reading. *Annals of Dyslexia*, 55(2), 193–216.

Equity of achievement: A matter of education structures?¹

Ariane Baye and Christian Monseur

University of Liège, Belgium

Abstract

This paper focuses on the issue of equality in educational outcomes. It investigates whether inequity increases between primary and secondary education, and relates variations in these patterns to the education structures of various industrialized countries. Data from IEA's Reading Literacy Study and TIMSS 2003 offered the advantage of allowing a comparison of equity within education systems at two levels of education, with the outcomes

of such an investigation having the potential to identify equity changes related to institutional changes between primary and secondary education. The study investigated two hypotheses relating to these considerations. However, methodological issues did not allow a satisfactory investigation of the relationship between institutional settings and equity.

Introduction

Industrialized societies present a double challenge for their education systems: they are required to be both effective and equitable. Education systems have to provide a stock of human capital able to meet the demanding challenges of post-industrialized economies (OECD, 2005), but at the same time they must respect the equity demands of democratic societies. Education stakeholders have to face this double constraint; they need to look for pedagogical and institutional means of fostering equity and effectiveness. In relation to this perspective, international surveys can be viewed as unique tools that enable us to analyze how efficient and how equitable countries are and then to consider this information in relation to institutional settings.

Our initial objective in the study outlined in this paper was to focus on the issue of equality in educational outcomes by investigating if inequity increases between primary and secondary education and then relating variations in these patterns to the education structures of various industrialized countries. Education systems are indeed often classified as "selective" or "comprehensive" according to the way they group their students. As Hanushek and Woessmann (2005) note, the central argument against selective systems is that they are *per se* more inequitable, offering a less demanding education environment to less able students, and using tracks (streams) to

group people from socioeconomically disadvantaged backgrounds. Grisay's (2006) analysis, based on Programme for International Students Achievement (PISA) 2003 data, confirms this argument. Grisay notes that, in most comprehensive education systems, the effects of socioeconomic factors on mathematics achievement are partially cancelled out, but that these factors have a strong impact on performance in mathematics in tracked systems. Unlike the data from PISA, the International Association for the Evaluation of Educational Achievement's (IEA) Reading Literacy Study and TIMSS 2003 data offer researchers the advantage of comparing equity within education systems at two levels of education, which could help identify equity changes related to institutional changes between primary and secondary education. The first section of this paper therefore clarifies the concept of equity, and presents the working hypothesis investigated.

Does institutional differentiation increase inequity?

In accordance with the definition proposed by the European Group of Research on Equity of Education Systems (Baye et al., 2005), not all types of inequality should be qualified as inequitable. Education has room for many kinds of inequalities (of access, achievement,

¹ The authors are grateful to their colleagues Aletta Grisay and Dominique Lafontaine (University of Liège) for making comments and suggestions on earlier versions of this paper.

social actualization) (Demeuse & Baye, 2005; Grisay, 1984). Also, education systems may sometimes accept certain kinds of inequalities in order to improve equality at another level. For instance, compensatory programs waive the principle of equality of treatment in the name of more equality of achievement. Some countries may also accept a high level of dispersion of achievement results, provided that the students at the bottom of the distribution have acquired a set of basic skills. These examples illustrate the difficulty of finding a unique definition of equity in an international perspective. As the concept of equity is a matter of justice, each society may answer differently to the central equity question: which are the unfair inequalities?

To overcome the impossibility of finding a common definition of unfair inequalities in various countries, general theories of justice by various researchers and commentators have been investigated. Most of them agree on one point: *academic success in school should not depend on social origin* (Meuret, 2001). For decades, providing universal access to education was considered the best way to ensure this independence. Nowadays, in developed countries where equality of access to compulsory schooling is achieved, this independence refers to equality of achievement.

The studies conducted by the IEA have regularly shown that socioeconomic background is a powerful predictor of students' achievement, at different grades and in different cognitive areas, in all participating countries (see, for example, Elley, 1992; Mullis, Martin, Gonzalez, & Chrostowski, 2004a, 2004b; Mullis, Martin, Gonzalez, & Kennedy, 2003). However, the strength of the relationship may vary from country to country—the lower the correlation, the higher the level of equity. On the other hand, tracked (or streamed) education systems are often pointed out as inequalitarian because how the educational institutions are organized reinforces the socioeconomic inequalities (Hanushek & Woessmann, 2005). For instance, PISA 2003 data suggest that in countries with institutional differentiation (i.e., number of programs offered to 15-year-old students, age of first selection (i.e., entry into a track or stream), and rate of student retention), the relationship between socioeconomic background and achievement is stronger than it is in comprehensive systems (OECD, 2006).

With these considerations in mind, we formulated a two-step hypothesis to analyze this potential link

between education structures and equity:

1. Education systems that group students according to their abilities (or any surrogate factor) tend to have a *higher correlation between the social background variables and the performance*.
2. As education systems present little variability in terms of ability grouping in primary education compared to secondary education, the *correlation between socioeconomic status (SES) and performance should increase between these two levels of education, especially in education systems with ability grouping*.

Data and method

To test these hypotheses, we looked for IEA surveys that met the following criteria:

1. *Provided a survey of one primary education population and one secondary education population:* The IEA Reading Literacy Study 1991 and TIMSS 2003 fulfilled this condition because they surveyed students from Grade 4 (Population 1) and Grade 8 (Population 2).
2. *Included the same measure of the students' family background at both grades:* The IEA Reading Literacy Study 1991 and TIMSS 2003 met this criterion, through "the number of books in the home" variable. TIMSS 2003 also included international items on "educational resources in the home."
3. *Included education systems that differed considerably in how they grouped students within the educational institutions:* IEA Reading Literacy (1991) and TIMSS 2003 again fulfilled this condition.

In this present investigation, we considered only the developed countries (i.e., OECD and European Union members) that participated in these studies.

Although the two IEA studies met the three conditions needed to test the general hypothesis, it must be recognized that both offer a limited set of common variables to capture the social origin of the students, mainly because of the difficulty of obtaining this information from Grade 4 students. The analysis and conclusions on the equity of the education systems are therefore conditioned by the reliability of this measure. We selected "number of books in the home" from among the limited set of variables used in both studies, and for both populations, as a proxy for students' socioeconomic background. We chose this variable because it represents quantifiable

socioeconomic/cultural capital. For an examination of the methodological issues linked to this choice, see the section below headed “Is the SES Measure Reliable?”

We chose the intra-class coefficient as a measure of the homogeneity/heterogeneity of the education systems. It gives quantitative information on the tendency within education systems to have more or less homogeneous classes in terms of students’ performance, a tendency which reflects explicit or implicit grouping policies. We preferred using the intra-class coefficient to other measures of homogeneity/heterogeneity, such as age of the first selection, number of tracks, and intra-class coefficient of the SES, for methodological reasons. Psychometrically, it presents the advantage of being a metric variable. Furthermore, as academic segregation is not independent from SES segregation, controlling for SES grouping should reduce the academic segregation. Kirsch, de Jong, Lafontaine, McQueen, Mendelovits, and Monseur (2002) discuss this issue of the interaction between social and academic segregation. From their analysis of PISA 2000 data, we could assume that both effects usually merge, except for countries with a high number of private schools (because parents have to pay fees to send their children to these schools) and for Korea, where academic segregation appears to be particularly important, and social segregation is low.

Results

Based on IEA Reading Literacy Study data, Figures 1 and 2 present the relationship between the tendency for countries to group students by ability (intra-class coefficient) and their tendency to be inequitable (i.e., to have a relatively high correlation between SES and achievement). These figures show that education systems that group students according to their abilities tend to have a higher correlation between the students’ social background and their performance, as confirmed by the correlation coefficients between both axes of the figures, respectively 0.43 for Population 1, and 0.31 for Population 2 (see also Table 1).

The correlation coefficients between both dimensions (homogeneity/equity) were also computed for mathematics and science, using TIMSS 2003 data (see Table 1). Even though the correlations are not perfect, the recurrence of the findings (i.e., a correlation between the tendency to be inequitable and the tendency to practice segregation) across domains and populations does support the conclusion on the link between institutional settings and equity.

However, Table 1 also shows that the correlation coefficients are higher in reading literacy and in science than in mathematics. Although we need to consider this finding with caution because of the small numbers of countries included in the analysis, it is possible that this difference is inherent to the subject domain of mathematics. Thus, the lesser effect of the socioeconomic environment on student achievement in mathematics and grouping could be because mathematics is more school-based in nature than are the other two domains.

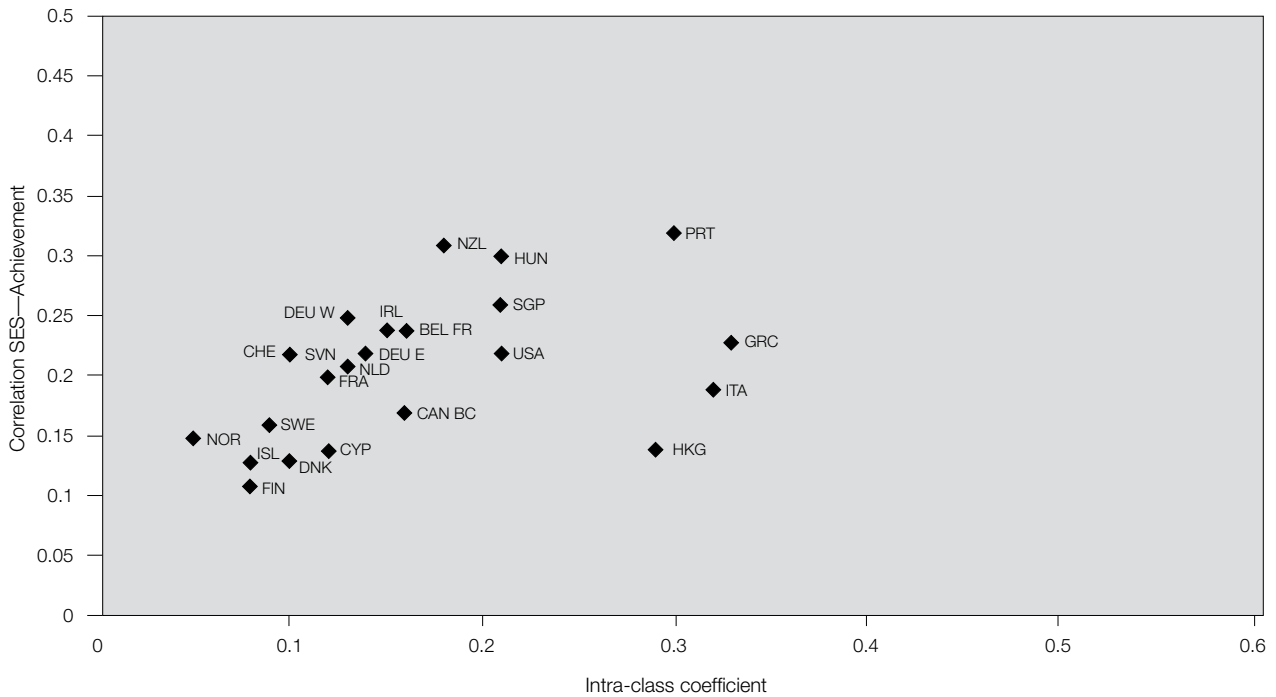
In general, our data confirmed our first hypothesis: homogeneous systems are less equitable than the heterogeneous ones because they tend to group students according to their family background. (For detailed figures, See Appendix Tables A1, A2, and A3.) But does this tendency increase between primary and secondary education? We computed the growth in the intra-class correlation between Populations 1 and 2 and the growth in the correlation coefficients between SES and performance in the two populations. The correlation coefficients between the two growths were respectively 0.04, -0.10, and 0.29 for reading, mathematics, and science.

These results did not confirm the second hypothesis, namely, that education systems with more ability grouping tend to lead to a stronger increase in the relationship between SES and performance between primary and secondary level than do the comprehensive systems. Table 2 shows that the correlation between SES and performance increased

Table 1: Correlation between (1) Intra-class Coefficient and (2) the SES (Books) and Achievement Correlation—IEA Reading Literacy Study 1991 and TIMSS 2003

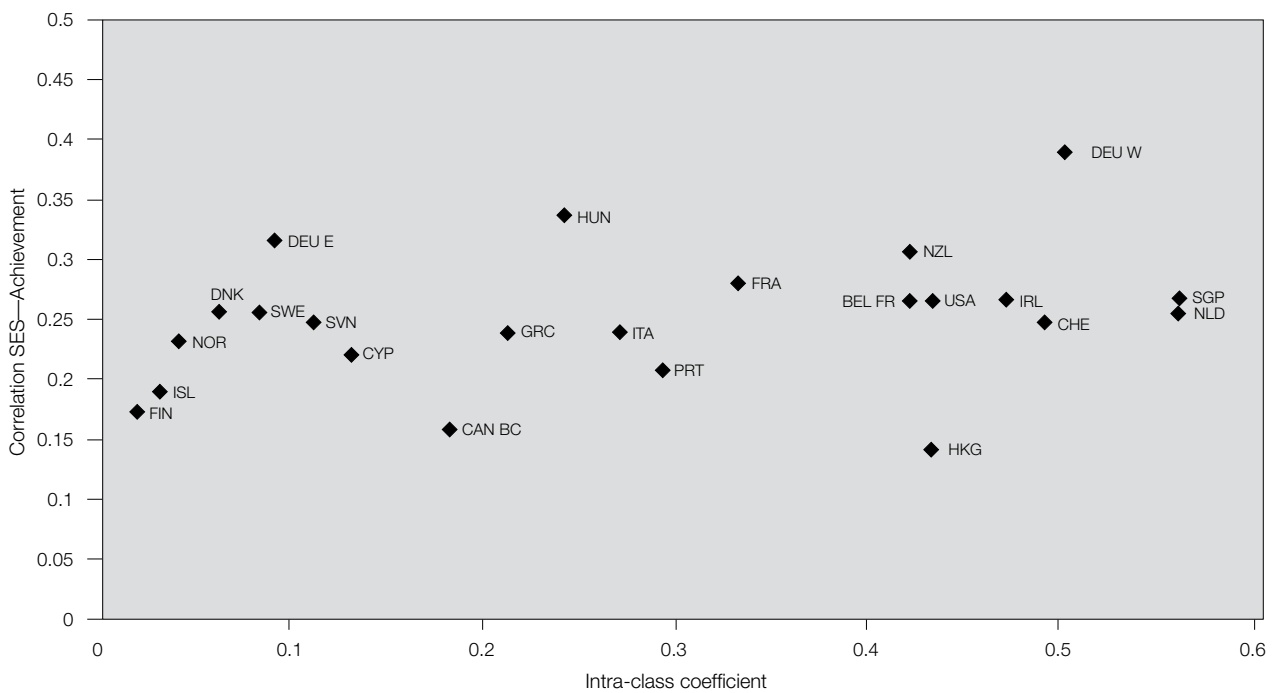
	Population 1	Population 2
Reading Literacy (1991)	0.43	0.31
Mathematics (2003)	0.33	0.15
Science (2003)	0.48	0.45

Figure 1: Intra-class Coefficient (ρ) and Correlation between SES (Books) and Achievement in Reading, Population 1—IEA Reading Literacy Study 1991



Note: Correlation between both axes: 0.43.

Figure 2: Intra-class Coefficient (ρ) and Correlation between SES (Books) and Achievement in Reading, Population 2—IEA Reading Literacy Study 1991



Note: Correlation between both axes: 0.31.

Table 2: Correlation between SES (Books) and Achievement—IEA Reading Literacy Study 1991 and TIMSS 2003

	Population 1	Population 2
Reading Literacy (1991)	0.21	0.30
Mathematics (2003)	0.24	0.31
Science (2003)	0.24	0.33

everywhere, whatever the initial level of correlation observed at the primary level. This finding led us to conclude that, in most education systems, the relationship between socioeconomic background and achievement increases between primary and secondary level, whatever the education structures.

We did not observe the divide we expected between selective and comprehensive systems. Selective systems seem somewhat more inequitable at the primary level. Moreover, their selective structures do not seem to worsen their situation with regard to equity, compared to comprehensive systems, which are more equitable at the primary level but which do not prevent an increase in the correlation between SES and achievement between both levels. We conducted further analyses to improve our understanding of this unexpected result.

Is the SES measure reliable?

A correlation coefficient mainly relies on the latent correlation between both phenomena. Differences in the latent correlation reflect differences in the strength of the relationship. However, the observed correlation differences might also be due to differences in the reliability of the measure from country to country. Variations across industrialized countries in the reliability of the performance scale are usually small. By contrast, the reliability of the SES measure from country to country, and from one population to another, may be questioned. Because “number of books” is a single variable, one cannot directly compute the reliability of the SES measure. However, if number of books at home is a reliable proxy for students’ socioeconomic backgrounds, it should not vary relative to a variable that is presumably uniformly distributed among different social groups, as is the case with gender. We accordingly computed the correlation between SES and performance for boys and girls (Table 3).

From the table we can see that, in some countries, the correlation SES/performance differs considerably between boys and girls. The correlation between books

at home and achievement tends to be higher for girls at the primary level, while the differences tend to be less pronounced, or even reversed, at the secondary level. Note, however, that the increase in the SES/performance correlation between Population 1 and Population 2 correlates at 0.38 with the gender gap in the SES/performance correlation in Population 1, as shown in Figure 3. The correlation coefficient would actually increase to 0.50 if Finland was excluded.

This finding could indicate that, for Grade 4 students, number of books at home is not the best proxy to measure socioeconomic background because it is gender sensitive. For girls, identifying the number of books at home would also be a way to manifest interest in books, which could explain the higher correlation with achievement in reading literacy for girls than for boys. The general evolution of the correlations between books at home and achievement in reading may consequently be affected by this gender gap.

We confirmed the hypothesis of an interaction between an SES proxy and the domain of assessment by contrasting these results with a replication of the analysis in mathematics and science (TIMSS 2003). In mathematics, the gender gap in the correlation between books at home and achievement concerns far fewer countries (cf. Table A6 in the appendix to this paper). This measure seems to be more stable, both at primary and secondary levels. The correlation between gender differences at Grade 4 and the SES/achievement correlation between both grades is 0.05 (compared to 0.38 for reading). In science, the same correlation is also 0.05, but the pattern is different: the correlation between SES and gender tends to be better for boys in some countries, and these gender differences exist at both primary and secondary levels (cf. appendix, Table A7). Here, unlike the situation for reading, the hypothesis persists for Population 2, in that the gender interaction between the SES measure and achievement in science neither correlates with nor sufficiently explains the evolution of the SES/achievement correlation.

Table 3: Correlation between SES (Books) and Achievement, by Gender—IEA Reading Literacy Study, 1991

	Population 1			Population 2		
	Boys	Girls	Difference	Boys	Girls	Difference
Belgium (Fr)	0.26	0.23	-0.03	0.33	0.33	0.00
Canada (BC)	0.17	0.18	0.02	0.18	0.21	0.03
Switzerland	0.21	0.25	0.04	0.31	0.29	-0.02
Cyprus	0.13	0.16	0.03	0.26	0.28	0.02
Germany (E)	0.18	0.27	0.09	0.40	0.37	-0.03
Germany (W)	0.22	0.31	0.09	0.43	0.52	0.09
Denmark	0.11	0.18	0.07	0.30	0.33	0.03
Finland	0.03	0.22	0.20	0.22	0.19	-0.03
France	0.19	0.23	0.03	0.32	0.37	0.05
Greece	0.26	0.20	-0.06	0.28	0.29	0.02
Hong Kong SAR	0.13	0.15	0.02	0.17	0.18	0.00
Hungary	0.31	0.32	0.01	0.41	0.39	-0.02
Iceland	0.11	0.15	0.04	0.26	0.20	-0.06
Ireland	0.25	0.24	-0.01	0.32	0.32	0.00
Italy	0.20	0.19	-0.01	0.28	0.30	0.02
Netherlands	0.22	0.21	-0.01	0.26	0.37	0.11
Norway	0.15	0.17	0.02	0.29	0.31	0.02
New Zealand	0.33	0.30	-0.03	0.38	0.35	-0.03
Portugal	0.32	0.34	0.02	0.25	0.25	0.00
Singapore	0.25	0.28	0.03	0.33	0.31	-0.02
Slovenia	0.19	0.27	0.08	0.31	0.29	-0.02
Sweden	0.13	0.20	0.07	0.28	0.35	0.06
United States	0.22	0.22	0.01	0.35	0.30	-0.05

Because of the possible interaction between SES and gender, we searched for another proxy for SES. We subsequently developed an index of home educational resources that “combined” four items commonly possessed by Grade 4 and Grade 8 students in TIMSS 2003: calculator, computer, study desk, and dictionary. As Table 4 shows, the correlation of this index with achievement in mathematics and science is quite stable for Populations 1 and 2. Table 5 features the relationship between the tendency for a country to group students according to their ability (intra-class coefficient) and the correlation educational resources/achievement. The correlations in this table tell us that

the more selective systems are also the less equitable at the primary level. However, the pattern is reversed at the secondary level. Here, the larger the differences in achievement are between schools, the weaker the relationship is between educational resources and achievement.

Contrary to what we found with the books at home measure, the intra-class correlation growth (increase) between Populations 1 and 2 seems congruent with the growth (decrease) in the correlation coefficients between educational resources and performance between the two populations. (For detailed figures, see Appendix Tables A4 and A5.) The correlation

Table 4: Correlation between SES (Educational Resources) and Achievement—TIMSS 2003

	Population 1	Population 2
Mathematics (2003)	0.27	0.23
Science (2003)	0.24	0.22

Table 5: Correlation between (1) Intra-class Coefficient and (2) the SES (Educational Resources) and Achievement Correlation—TIMSS 2003

	Population 1	Population 2
Mathematics (2003)	0.31	-0.27
Science (2003)	0.57	-0.17

coefficients between the two growths were respectively 0.19 and 0.32 for mathematics and science.

Does this result lead to the conclusion that selective systems are actually more equitable, in the sense that the differentiation intervening at the beginning of secondary education limits, or even reverses, the increasing influence of family resources on achievement? Or are there other factors affecting the SES index? A closer look at the results revealed that the correlation between SES and achievement for Population 2 was particularly low in the more industrialized countries, and tended to remain stable, or even to increase, in countries with low GDP, such as Cyprus, Lithuania, and Latvia. On removing these less industrialized countries from the analysis, we found that the correlation coefficients between the intra-class coefficient increase and the SES impact decreased to 0.53 for mathematics, and 0.62 for sciences.

This instability of the results according to the level of development of the country suggests a “ceiling effect” hypothesis. In the most industrialized countries, it appears that the educational resources index is not discriminating enough because most Grade 8 students do possess the four items at home. Also in industrialized countries, the low level of correlation between education resources and achievement in mathematics and science for Population 2 does not mean there is no link between SES and achievement, but rather that contextual variables prevent a sufficiently robust measure of this dimension and its link with achievement.

Conclusion

Political interest in and demand for equity indicators, and the need to better understand the link between structural choices (variables that are managed politically) and equity and effectiveness reinforce the need for international surveys to gather information in these sensitive domains. In this respect, IEA's studies of achievement across two populations of students provide unique sources of data that allow us to compare equity at different levels of education in various education systems. The two studies from which we drew data for the analysis presented in this paper were the IEA Reading Literacy Study 1991 and TIMSS 2003.

Convergent results on the relationship between education structures and equity led us to investigate the hypothesis that the correlation between SES and student performance in reading, mathematics, and science would be stronger in systems with ability grouping than in those without ability grouping. At this stage, the data do not allow us to give firm support to the hypothesis because of the difficulty of finding a measure of students' socioeconomic background that is equally reliable across different levels of education and different assessment domains.

The analysis performed on data from the two IEA studies indicates that the number of books at home is not always a consistent SES proxy. Its interaction with gender in reading at the primary level suggests that this variable not only measures a quantity of goods at home (probably quite difficult to assess at Grade 4), but may also measure, mainly for girls, the value accorded to books at this age. While more coherent across gender and domains, the measure based on

educational resources at home is also not a relevant SES proxy because the international common items are not discriminating enough for industrialized countries.

Interaction between the SES variable with proficiency (in the case of “books at home”) and ceiling effects (in the case of “educational resources”) affect ability to build equity indicators based on IEA Reading Literacy and TIMSS data, even though these studies offer a unique opportunity to study cross-level effects. Further work on a composite index of socioeconomic background, including information on student status with regard to immigration and national-specific variables on the items possessed at home, is needed in order to build a stronger socioeconomic index,

and to document the issue of the impact of education structures on equity of achievement.

This paper also suggests that a project as ambitious as the one announced here (i.e., an analysis of the equity of education systems in relation to their institutional settings) cannot be achieved through use of one single equity indicator. In addition to consideration of the methodological issues associated with building indicators, this paper emphasizes the need for a system of equity indicators that capture the complexity of education systems and the interaction between phenomena. The variety of the dimensions to be taken into account rests on the importance placed, especially within the political sphere, on the equity issue.

Appendix

Table A1: IEA Reading Literacy 1991

Countries	Correlation Books/ Achievement Pop. I	Intra-class coefficient Pop. I	Correlation Books/ Achievement Pop. II	Intra-class coefficient Pop. II	Difference correlation SES Achievement Pop. 2 – Pop. 1 (col. 3 – col. 1)	Difference intra-class coefficient Pop. 2 – Pop. 1 (col. 4 – col. 2)
	(1)	(2)	(3)	(4)	(5)	(6)
Belgium (Fr)	0.24	0.16	0.32	0.42	0.08	0.26
Canada (BC)	0.17	0.16	0.19	0.18	0.02	0.02
Switzerland	0.22	0.10	0.30	0.49	0.08	0.39
Cyprus	0.14	0.12	0.27	0.13	0.13	0.01
Germany (E)	0.22	0.14	0.38	0.09	0.16	-0.05
Germany (W)	0.25	0.13	0.47	0.50	0.22	0.37
Denmark	0.13	0.10	0.31	0.06	0.18	-0.04
Finland	0.11	0.08	0.21	0.02	0.10	-0.06
France	0.20	0.12	0.34	0.33	0.14	0.21
Greece	0.23	0.33	0.29	0.21	0.06	-0.12
Hong Kong SAR	0.14	0.29	0.17	0.43	0.03	0.14
Hungary	0.30	0.21	0.41	0.24	0.11	0.03
Iceland	0.13	0.08	0.23	0.03	0.10	-0.05
Ireland	0.24	0.15	0.32	0.47	0.08	0.32
Italy	0.19	0.32	0.29	0.27	0.10	-0.05
Netherlands	0.21	0.13	0.31	0.56	0.10	0.43
Norway	0.15	0.05	0.28	0.04	0.13	-0.01
New Zealand	0.31	0.18	0.37	0.42	0.06	0.24
Portugal	0.32	0.30	0.25	0.29	-0.07	-0.01
Singapore	0.26	0.21	0.32	0.56	0.06	0.35
Slovenia	0.22	0.10	0.30	0.11	0.08	0.01
Sweden	0.16	0.09	0.31	0.08	0.15	-0.01
United States	0.22	0.21	0.32	0.43	0.10	0.22
<i>Correlation col. 1 & 2</i>		<i>0.43</i>				
<i>Correlation col. 3 & 4</i>				<i>0.31</i>		
<i>Correlation col. 5 & 6</i>						<i>0.04</i>

Table A2: TIMSS 2003, Mathematics

Countries	Correlation Books/ Achievement Pop. I	Intra-class coefficient Pop. I	Correlation Books/ Achievement Pop. II	Intra-class coefficient Pop. II	Difference correlation SES Achievement Pop. 2 – Pop. 1 (col. 3 – col. 1)	Difference intra-class coefficient Pop. 2 – Pop. 1 (col. 4 – col. 2)
	(1)	(2)	(3)	(4)	(5)	(6)
Australia	0.24	0.25	0.24	0.49	-0.01	0.24
Belgium (Fl)	0.22	0.16	0.25	0.63	0.03	0.48
Canada (O)	0.25	0.19	0.31	0.16	0.06	-0.03
Canada (Q)	0.21	0.15	0.25	0.41	0.04	0.26
Cyprus	0.16	0.10	0.27	0.06	0.11	-0.05
England	0.32	0.24	0.37	0.55	0.05	0.31
Hong Kong SAR	0.15	0.25	0.20	0.58	0.06	0.32
Hungary	0.34	0.24	0.44	0.34	0.10	0.10
Italy	0.10	0.32	0.30	0.28	0.20	-0.04
Japan	0.26	0.06	0.28	0.14	0.02	0.08
Lithuania	0.27	0.23	0.34	0.20	0.07	-0.03
Latvia	0.20	0.23	0.25	0.24	0.05	0.01
Netherlands	0.30	0.16	0.36	0.71	0.06	0.56
Norway	0.23	0.10	0.29	0.11	0.06	0.01
New Zealand	0.32	0.34	0.34	0.45	0.02	0.12
Scotland	0.25	0.18	0.41	0.57	0.16	0.39
Singapore	0.34	0.56	0.34	0.39	0.00	-0.17
Slovenia	0.18	0.12	0.28	0.12	0.10	0.00
United States	0.32	0.29	0.38	0.40	0.07	0.11
Correlation col. 1 & 2		0.33				
Correlation col. 3 & 4				0.15		
Correlation col. 5 & 6						-0.10

Table A3: TIMSS 2003, Science

Countries	Correlation Books/ Achievement Pop. I	Intra-class coefficient Pop. I	Correlation Books/ Achievement Pop. II	Intra-class coefficient Pop. II	Difference correlation SES Achievement Pop. 2 – Pop. 1 (col. 3 – col. 1)	Difference intra-class coefficient Pop. 2 – Pop. 1 (col. 4 – col. 2)
	(1)	(2)	(3)	(4)	(5)	(6)
Australia	0.28	0.25	0.33	0.49	0.06	0.24
Belgium (Fl)	0.22	0.16	0.29	0.63	0.07	0.48
Canada (O)	0.27	0.19	0.32	0.16	0.05	-0.03
Canada (Q)	0.18	0.15	0.28	0.41	0.10	0.26
Cyprus	0.13	0.10	0.27	0.06	0.13	-0.05
England	0.33	0.24	0.41	0.55	0.08	0.31
Hong Kong SAR	0.14	0.25	0.20	0.58	0.06	0.32
Hungary	0.30	0.24	0.40	0.34	0.10	0.10
Italy	0.12	0.32	0.27	0.28	0.15	-0.04
Japan	0.22	0.06	0.28	0.14	0.06	0.08
Lithuania	0.24	0.23	0.28	0.20	0.04	-0.03
Latvia	0.21	0.23	0.23	0.24	0.02	0.01
Netherlands	0.28	0.16	0.40	0.71	0.11	0.56
Norway	0.21	0.10	0.29	0.11	0.08	0.01
New Zealand	0.29	0.34	0.40	0.45	0.12	0.12
Scotland	0.25	0.18	0.47	0.57	0.23	0.39
Singapore	0.36	0.56	0.39	0.39	0.03	-0.17
Slovenia	0.18	0.12	0.28	0.12	0.10	0.00
United States	0.31	0.29	0.42	0.40	0.11	0.11
<i>Correlation col. 1 & 2</i>		0.48				
<i>Correlation col. 3 & 4</i>				0.45		
<i>Correlation col. 5 & 6</i>						0.29

Table A4: TIMSS 2003, Mathematics

Countries	Correlation Educational Resources/ Achievement Pop. I	Intra-class coefficient Pop. I	Correlation Educational Resources/ Achievement Pop. II	Intra-class coefficient Pop. II	Difference correlation SES Achievement Pop. 2 – Pop. 1 (col. 3 – col. 1)	Difference Intra-class coefficient Pop. 2 – Pop. 1 (col. 4 – col. 2)
	(1)	(2)	(3)	(4)	(5)	(6)
Australia	0.28	0.25	0.16	0.49	-0.11	0.24
Belgium (Fl)	0.14	0.16	0.26	0.63	0.12	0.48
Canada (O)	0.27	0.19	0.18	0.16	-0.09	-0.03
Canada (Q)	0.25	0.15	0.13	0.41	-0.12	0.26
Cyprus	0.29	0.10	0.37	0.06	0.08	-0.05
England	0.27	0.24	0.23	0.55	-0.04	0.31
Hong Kong SAR	0.12	0.25	0.14	0.58	0.03	0.32
Hungary	0.36	0.24	0.37	0.34	0.01	0.10
Italy	0.18	0.32	0.21	0.28	0.04	-0.04
Japan	0.22	0.06	0.19	0.14	-0.02	0.08
Lithuania	0.30	0.23	0.30	0.20	0.00	-0.03
Latvia	0.22	0.23	0.21	0.24	0.00	0.01
Netherlands	0.22	0.16	0.16	0.71	-0.05	0.56
Norway	0.30	0.10	0.17	0.11	-0.14	0.01
New Zealand	0.34	0.34	0.22	0.45	-0.12	0.12
Scotland	0.29	0.18	0.26	0.57	-0.03	0.39
Singapore	0.38	0.56	0.25	0.39	-0.12	-0.17
Slovenia	0.34	0.12	0.25	0.12	-0.09	0.00
United States	0.34	0.29	0.23	0.40	-0.11	0.11
Correlation col. 1 & 2		0.31				
Correlation col. 3 & 4				-0.27		
Correlation col. 5 & 6						0.19

Table A5: TIMSS 2003, Science

Countries	Correlation Educational Resources/ Achievement Pop. I	Intra-class coefficient Pop. I	Correlation Educational Resources/ Achievement Pop. II	Intra-class coefficient Pop. II	Difference correlation SES Achievement Pop. 2 – Pop. 1 (col. 3 – col. 1)	Difference Intra-class coefficient Pop. 2 – Pop. 1 (col. 4 – col. 2)
	(1)	(2)	(3)	(4)	(5)	(6)
Australia	0.29	0.25	0.17	0.49	-0.13	0.24
Belgium (Fl)	0.14	0.16	0.26	0.63	0.12	0.48
Canada (O)	0.26	0.19	0.12	0.16	-0.15	-0.03
Canada (Q)	0.22	0.15	0.14	0.41	-0.08	0.26
Cyprus	0.23	0.10	0.31	0.06	0.09	-0.05
England	0.22	0.24	0.23	0.55	0.01	0.31
Hong Kong SAR	0.09	0.25	0.11	0.58	0.02	0.32
Hungary	0.33	0.24	0.32	0.34	-0.01	0.10
Italy	0.19	0.32	0.19	0.28	0.00	-0.04
Japan	0.18	0.06	0.18	0.14	0.01	0.08
Lithuania	0.19	0.23	0.20	0.20	0.02	-0.03
Latvia	0.24	0.23	0.25	0.24	0.01	0.01
Netherlands	0.14	0.16	0.15	0.71	0.01	0.56
Norway	0.22	0.10	0.20	0.11	-0.02	0.01
New Zealand	0.36	0.34	0.26	0.45	-0.10	0.12
Scotland	0.25	0.18	0.25	0.57	0.00	0.39
Singapore	0.40	0.56	0.27	0.39	-0.13	-0.17
Slovenia	0.27	0.12	0.25	0.12	-0.02	0.00
United States	0.32	0.29	0.23	0.40	-0.08	0.11
Correlation col. 1 & 2		0.57				
Correlation col. 3 & 4				-0.17		
Correlation col. 5 & 6						0.32

Table A6: Correlation between SES (Books) and Achievement, by Gender—TIMSS 2003, Mathematics

Countries	Boys	Girls	Difference	Boys	Girls	Difference
Australia	0.27	0.29	0.02	0.19	0.15	-0.03
Belgium (Fl)	0.18	0.11	-0.08	0.25	0.28	0.03
Canada (O)	0.24	0.30	0.05	0.19	0.18	-0.01
Canada (Q)	0.27	0.24	-0.03	0.14	0.12	-0.02
Cyprus	0.29	0.29	0.00	0.37	0.35	-0.02
England	0.30	0.25	-0.05	0.22	0.23	0.01
Hong Kong SAR	0.10	0.13	0.03	0.13	0.15	0.02
Hungary	0.37	0.34	-0.03	0.39	0.35	-0.04
Italy	0.16	0.19	0.03	0.22	0.21	-0.01
Japan	0.19	0.24	0.05	0.19	0.20	0.01
Lithuania	0.31	0.28	-0.02	0.32	0.28	-0.04
Latvia	0.24	0.20	-0.04	0.24	0.20	-0.04
Netherlands	0.23	0.21	-0.02	0.16	0.17	0.01
Norway	0.30	0.31	0.00	0.13	0.19	0.06
New Zealand	0.34	0.35	0.00	0.22	0.22	0.00
Scotland	0.30	0.30	0.01	0.26	0.26	0.00
Singapore	0.37	0.38	0.01	0.22	0.27	0.04
Slovenia	0.36	0.32	-0.04	0.23	0.28	0.05
United States	0.33	0.36	0.03	0.23	0.24	0.01

Table A7: Correlation between SES (Books) and Achievement, by Gender—TIMSS 2003, Science

Countries	Boys	Girls	Difference	Boys	Girls	Difference
Australia	0.31	0.25	-0.06	0.38	0.29	-0.08
Belgium (Fl)	0.20	0.24	0.04	0.32	0.36	0.03
Canada (O)	0.28	0.27	-0.01	0.27	0.29	0.02
Canada (Q)	0.22	0.16	-0.06	0.42	0.40	-0.03
Cyprus	0.14	0.13	-0.01	0.27	0.26	-0.01
England	0.35	0.32	-0.03	0.52	0.45	-0.07
Hong Kong SAR	0.13	0.14	0.01	0.23	0.17	-0.06
Hungary	0.34	0.27	-0.08	0.42	0.42	0.00
Italy	0.13	0.11	-0.02	0.29	0.26	-0.03
Japan	0.20	0.23	0.03	0.34	0.22	-0.12
Lithuania	0.24	0.23	-0.01	0.31	0.24	-0.07
Latvia	0.19	0.22	0.03	0.25	0.21	-0.05
Netherlands	0.30	0.28	-0.02	0.40	0.40	0.00
Norway	0.26	0.17	-0.09	0.29	0.30	0.01
New Zealand	0.32	0.27	-0.05	0.40	0.41	0.01
Scotland	0.23	0.28	0.05	0.31	0.29	-0.02
Singapore	0.37	0.37	0.01	0.41	0.38	-0.03
Slovenia	0.14	0.21	0.08	0.28	0.28	0.00
United States	0.33	0.31	-0.02	0.45	0.41	-0.05

References

- Baye, A., Benadusi, L., Bottani, N., Bove, G., Demeuse, M., García de Cortazar, M., Giancola, O., Gorard, S., Hutmacher, W., Matoul, A., Meuret, D., Morlais, S., Nicaise, J., Ricotta, G., Smith, E., Straeten, M.-H., Tiana-Ferrer, A., & Vandenderghe, V. (2005). Equity in European education systems: A set of indicators. *European Educational Research Journal*, 4(2), 1–151.
- Demeuse, M., & Baye, A. (2005). Pourquoi parler d'équité? [Why do we need to talk about equity?] In M. Demeuse, A. Baye, M.-H. Straeten, J. Nicaise, & A. Matoul (Eds.), *Vers une école juste et efficace: Vingt-six contributions sur les systèmes d'enseignement et de formation [Toward fair and efficient schools. Twenty-six papers on education and training systems]* (pp. 149–170). Brussels: De Boeck.
- Elley, W. B. (Ed.). (1992). *How in the world do students read?* Hamburg: Grindeldruck GMBH.
- Grisay, A. (1984). Quels indicateurs pour quelle réduction de l'inégalité scolaire? [Which indicators reduce which school inequality?] *Direction générale de l'Organisation des Etudes [General Administration for School System Management]*, 19, 1–14.
- Grisay, A. (2006). Que savons-nous de l'«effet établissement»? [What do we know about school effects?] In G. Chapelle & D. Meuret (Eds.), *Améliorer l'école [Improving schools]* (pp. 215–230). Paris: Presses Universitaires de France.
- Hanushek, E. A., & Woessmann, L. (2005). *Does educational tracking affect performance and inequality: Differences-in-differences evidence across countries*. (CESifo Working Paper No. 1415)
- Kirsch, I., de Jong, J., Lafontaine, D., McQueen, J., Mendelovits, J., & Monseur, C. (2002). *Reading for change: Performance and engagement across countries. Results from PISA 2000*. Paris: OECD.
- Meuret, D. (2001). School equity as a matter of justice. In W. Hutmacher, D. Cochrane, & N. Bottani (Eds.), *In pursuit of equity in education: Using international indicators to compare equity policies* (pp. 93–111). Dordrecht: Kluwer Academic Publishers.
- Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Chrostowski, S. J. (2004a). *TIMSS 2003 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Chrostowski, S. J. (2004b). *TIMSS 2003 international science report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Kennedy, A.M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools in 35 countries*. Chestnut Hill, MA: Boston College.
- Organisation for Economic Co-operation and Development (OECD). (2005). *Education at a glance: OECD indicators*. Paris: Author.
- Organisation for Economic Co-operation and Development (OECD). (2006). *Education at a glance: OECD indicators*. Paris: Author.

Characteristics of United States Grade 4 language minority students in an international context: Findings from PIRLS 2001¹

Laurence T. Ogle
*National Center for Education Statistics
Institute of Education Sciences
Washington DC, USA*

David C. Miller
*Education Statistics Services Institute
American Institutes for Research
Washington DC, USA*

Lydia B. Malley
*Education Statistics Services Institute Child Trends
Washington DC, USA*

Introduction

Overview of Grade 8 language minority students

Language minority students are a large and growing segment of most Group of Eight (G8) countries. According to 2001 data from the Progress in International Reading Literacy Study (PIRLS), and as defined in this paper, language minority students made up about 12% of fourth-grade students across the G8 countries that participated in the study.² The percentage of language minority fourth-graders in the participating G8 countries ranged from 4% (Italy) to 16% (Canada) (Table 1). In every participating G8 country except Italy, language minority fourth-graders made up at least 10% of the overall fourth-grade population; in the United States, language minority students made up 15% of the overall fourth-grade population.

This percentage of the school-age population is growing. In the United States, for example, the number of school-age children speaking a language other than English at home increased between 1979 and 2004 from 3.8 to 9.9 million, or from 9 to 19% of the school-age population (United States Department of Education, 2006).

The purpose of this paper is to analyze the educational characteristics and outcomes of language minority fourth-graders in the United States and in the other G8 countries that participated in PIRLS 2001. As defined in this paper, language minority students are those fourth-graders who report sometimes or never speaking the language of the test (PIRLS) at home.³ For example, in the United States, PIRLS was administered only in English. United States language minority students, therefore, are those fourth-graders who spoke English well enough to partake in PIRLS, but who only sometimes or never spoke English at home. Those who always or almost always spoke English at home were considered non-language minority students.

There were also some students who could not read or speak the language of the test and so could not overcome the language barrier of testing. Typically, a student who had received less than one year of instruction in the language of the test was excluded from PIRLS, although this definition was adapted in different countries. Along with students who were unable to take part in the assessment because

1 This paper is intended to promote the exchange of ideas among researchers and policymakers. The views expressed in it are part of ongoing research and analysis and do not necessarily reflect the position of the US Department of Education. Correspondence concerning this article should be addressed to Laurence T. Ogle, National Center for Education Statistics, 1990 K St., NW, Suite 900, Washington, DC 20006. Electronic mail may be sent to larogle@gmail.com. David Miller's co-authors thank him for editing this paper.

2 Japan did not participate in PIRLS 2001.

3 The data from PIRLS indicate that the percentage of G8 fourth-graders who never spoke the language of the test at home was quite small—less than 2% across all participating G8 countries. Such small percentages would have produced unstable estimates for the characteristics analyzed in this paper. As noted, the comparison group for this paper includes those students who sometimes or never spoke the language of the test at home.

Table 1: Group of 8 (G8) Countries, G8 Countries Participating in PIRLS, and Percentage Language Minority Grade 4 Students: 2001

G8 countries	G8 countries in PIRLS	Percentage language minority Grade 4 students
Canada	Canada	16
France	France	13
Germany	Germany	10
Italy	Italy	4
Japan		
Russian Federation	Russian Federation	15
United Kingdom	England (only)	12
	Scotland (only)	11
United States	United States	15

Note: Of the G8 countries, Japan did not participate, and the United Kingdom was represented by England and Scotland, which were reported in PIRLS as separate entities.

Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

of a physical or mental disability, these students contributed to the within-school exclusion rate. Across the participating G8 countries, within-school exclusion rates ranged from 0.3% (France) to 4.7% (United States).

In the United States, data from the National Assessment of Educational Progress (NAEP) document that language minority learners do not achieve at the same level as their native English-speaking peers (Klein, Bugarin, Beltranena, & McArthur, 2004). Yet even though language minority students bring special needs to their schools, research in the United States and other developed countries with this population has been limited. Understanding who these fourth-graders are, the types of schools they attend, and how they perform in reading can help educators and policymakers in the G8 countries address the needs of this increasingly important segment of the student population.

Overview of PIRLS

PIRLS is an assessment of reading literacy conducted by the International Association for the Evaluation of Educational Achievement (IEA). It provides comparative information on children's reading literacy and examines factors that may be associated with the acquisition of reading literacy. Thirty-five countries participated in PIRLS 2001, each drawing nationally representative samples from students in the upper of two adjacent grades with the most nine-year-olds—Grade 4 in most countries, including the United States.

This population was chosen because it represents a key transitional point in children's development as readers. In most countries, children at the end of Grade 4 are generally completing formal reading instruction (Kelly, 2003).

Students' reading literacy was assessed by using a selection of four literary passages (drawn from children's storybooks) and four informational texts. Submitted and reviewed by the PIRLS 2001 participating countries, the literary passages included realistic stories and traditional tales. The informational texts included chronological and non-chronological articles, a biographical article, and an informational leaflet. Approximately half of the questions on the assessment were multiple-choice questions and the other half were constructed-response, which required the students to generate and write their answers. Almost two-thirds of the score points on the assessment came from the constructed-response items (Mullis, Martin, & Kennedy, 2003).

In addition to the reading literacy assessment, several questionnaires were administered as part of the PIRLS 2001 data collection. Students were given a survey that asked about their reading activities in school and outside of school, their feelings about reading and about their school, and information about their family—including the availability and use of educational resources at home. Teachers were asked to complete a survey regarding their training and professional development, instructional reading practices, and assessment methods. School

administrators were asked to complete a survey about the school's enrollment and characteristics, climate, resources, and services, as well as the school's reading curriculum and reading programs. In most countries, parents/caregivers were given a survey that included questions regarding activities used to foster early reading literacy.

Research objectives

This study has several objectives. First, we explore how language minority fourth-graders in the United States compare to United States fourth-graders overall along selected student and school characteristics. Second, we explore how language minority fourth-graders in the participating Group of Eight (G8) countries compare to G8 fourth-graders overall along the same student and school characteristics. Third, we examine if language minority fourth-graders in the United States differ from their G8 peers in terms of the degree to which students with the selected characteristics are under- or over-represented among language minority students. Finally, for the United States and for the average of the G8 countries, we compare student outcomes (i.e., PIRLS assessment scores) of language minority fourth-graders to fourth-graders overall.

For the United States and for the average of the G8 countries, we report PIRLS average scores for fourth-grade language minority students compared to the overall population of fourth-graders. All of these results are shown separately for the selected student and school characteristics.

The G8 countries consist of Canada, France, Germany, Italy, Japan, the Russian Federation, the United Kingdom, and the United States. Of these countries, Japan did not participate in PIRLS 2001; in the United Kingdom, only England and Scotland participated, and they participated as separate countries. As a result, this study is unable to report any Japanese data and reports England and Scotland as separate countries—along with the six other participating G8 countries (see Table 1 above).

We selected the G8 countries for comparison because they are among the most industrialized nations in the world, and are, in economic and development

terms, relatively similar to the United States (Sen, Partelow, & Miller, 2005). Also, language minority individuals are a significant and growing part of the population in these countries. Comparisons to fourth-grade students in these countries permit an analysis of student performance at an early point in their academic careers and an evaluation of the relative similarities and differences between United States fourth-graders and their counterparts in the other participating G8 countries on a variety of demographic and educational variables.

This report is grouped into the following sections based on the type of characteristics examined:

1. *Student demographic and background characteristics*
 - Race/ethnicity—only in the United States
 - Students' reports of the number of books in the home (as a proxy for socioeconomic status)
2. *Student self-perceptions about reading*
 - Students' attitudes toward reading
 - Students' reading self-concept
3. *School characteristics (as reported by the school principal)*
 - Economic composition of school's student population
 - School climate
 - Availability of school resources.

These categories were chosen in an attempt to provide a descriptive profile of language minority students, including their perceptions about reading and information about the schools they attend. We also chose to focus on characteristics that may be associated with low student achievement, such as students coming from homes with 25 or fewer books, students with poor attitudes toward reading, and students from schools with deficient environments. We then examined whether or not students with these characteristics were over-represented among language minority students, compared to fourth-graders overall. Comparisons made in the text of this report were tested for statistical significance at the .05 level. The tests used were dependent *t*-tests.⁴ For further information about the statistical procedures used in this analysis, see the technical notes in the appendix to this paper.

⁴ These descriptive results are presented as a series of univariate comparisons. The individual characteristics examined are very likely to correlate with one another, so it is not possible to rank them according to their size or importance. To do this would require regression analyses, where the characteristics' relative associations with the outcomes are evaluated, taking into account how they may be confounded with one another.

Student demographic and background characteristics

This section presents a brief demographic and background profile of fourth-grade language minority students. The analyses describe whether or not students with selected demographic characteristics are over-represented among language minority students, compared to fourth-graders overall. For example, students from homes with 25 or fewer books are “over-represented” among language minority students if the percentage of language minority students from homes with 25 or fewer books is higher than the percentage of fourth-grade students overall from homes with 25 or fewer books.

Race/ethnicity in the United States

United States data from PIRLS showed that, on average, White fourth-graders outperformed Black and Hispanic Grade 4 students on the combined reading literacy scale. Asian fourth-graders, on average, also performed better than Black and Hispanic Grade 4 students on the combined reading literacy scale (Ogle et al., 2003).⁵

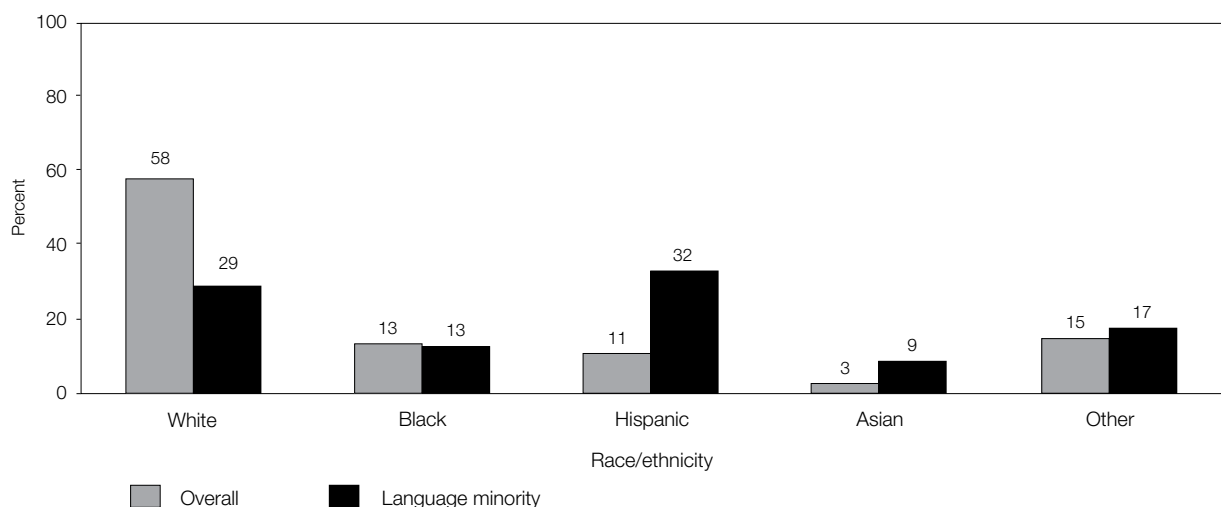
Compared to their proportion among United States fourth-graders overall, White students were under-represented among language minority students. Whereas White students accounted for 58% of US

fourth-graders overall, they made up 29% of the language minority students (Figure 1). In contrast, and as would be expected, Hispanic and Asian fourth-graders were over-represented among language minority students. Hispanic students accounted for 11% of US fourth-graders overall, but they made up 32% of language minority students. As for Asians, they made up approximately 3% of US fourth-graders overall, but 9% of the language minority students. Despite these relative differences, it is important to note that White students accounted for a larger portion of the language minority population than any race/ethnicity group other than Hispanics (where no measurable difference was detected between the percentage of White and the percentage of Hispanic language minority students). Among Whites, Blacks, and Hispanics, language minority students had lower average scores on the PIRLS combined reading literacy scale than did the overall fourth-grade population in that racial/ethnic category (Figure 2).

Number of books in the home

A strong relationship has been found between family socioeconomic status (SES) and a student’s learning outcomes (Coleman et al., 1966; Lemke et al., 2001, 2005; Miller, 2006; Ogle, Miller, & Johnston, 2006; Rothstein, 2004; West, Denton, & Reaney 2000;

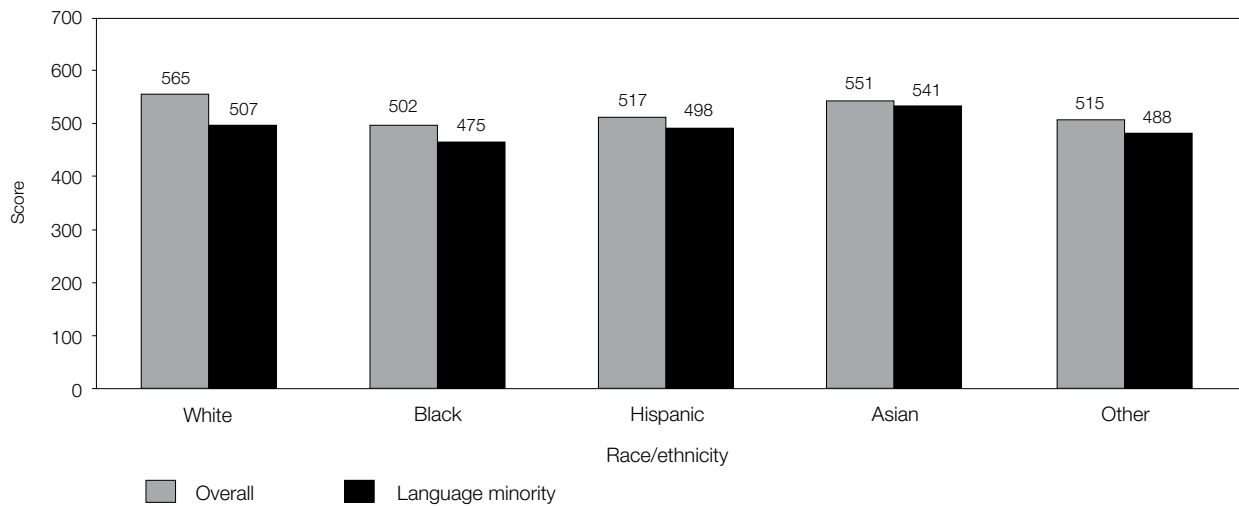
Figure 1: Overall Percentage of United States Grade 4 Students and Percentage of Language Minority United States Grade 4 Students, by Race/Ethnicity: 2001



Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

⁵ Because these racial and ethnic categories are not common across countries, it is not possible to make racial/ethnic comparisons for other countries.

Figure 2: Combined Average Reading Literacy Score for United States Grade 4 Students Overall, and Score for United States Language Minority Grade 4 Students, by Race/Ethnicity: 2001



Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

Williams, Levine, Jocelyn, Butler, & Haynes, 2000). In international studies with children, it is very difficult to develop comparable measures of SES, especially without parent-reported information.⁶ Therefore, since SES measures are limited in PIRLS, this report uses students' reports of the number of books in their homes as a proxy for SES. This variable has been used before in international studies and has been found to be a consistent predictor of students' educational achievement (Beaton et al., 1996; Torney-Purta, Lehmann, Oswald, & Schulz, 2001). Specifically, in this paper we report on the percentages of students who self-report having 25 or fewer books in their homes among fourth-grade language minority students and among fourth-grade students overall.

In the United States, and for the average of the participating G8 countries, fourth-graders who reported 25 or fewer books in their homes were over-represented among language minority students compared to their proportion among fourth-graders overall. That is, while 26% of US fourth-graders overall indicated coming from homes with 25 or fewer books, 38% of language minority students reported coming from homes with the same number of books (Figure 3). Across the G8 countries, on average, 29% of fourth-graders reported coming from homes with 25 or fewer books, but 41% of the G8 language minority students

reported coming from homes with 25 or fewer books.

In the United States and across the average of the G8 countries, language minority students from homes with 25 or fewer books had lower average scores on the PIRLS combined reading literacy scale than did the overall fourth-grade population from homes with 25 or fewer books. In the United States, on average, students from homes with 25 or fewer books overall scored 506 while language minority students from homes with 25 or fewer books scored 473 (Figure 4). Across the average of the G8 countries, students from homes with 25 or fewer books overall scored 509 while language minority students from homes with 25 or fewer books scored 479.

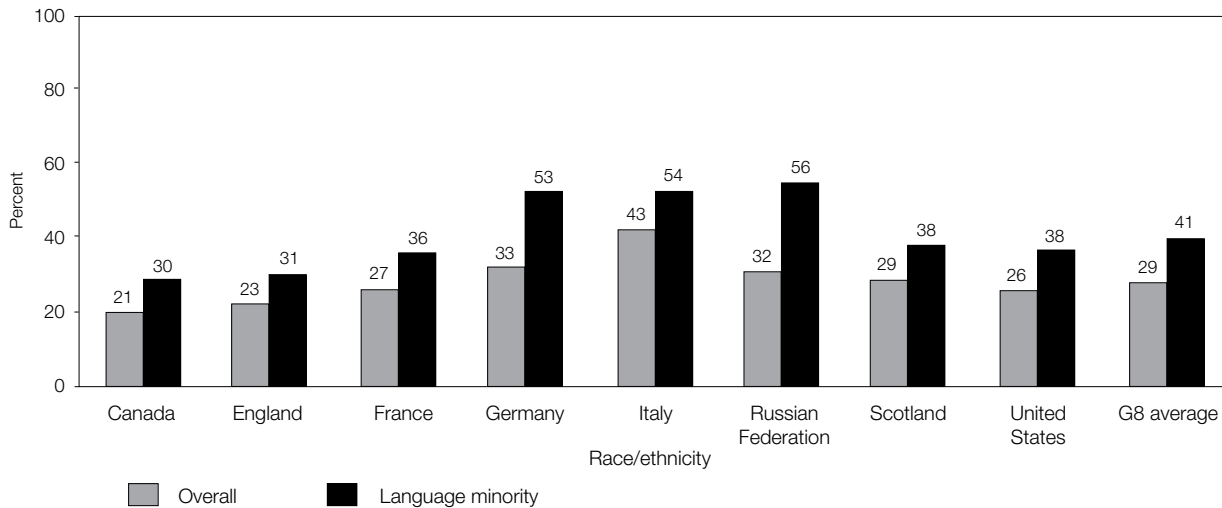
Summary of student demographics and background characteristics

The preceding analyses indicated the following:

- Compared to United States fourth-graders overall, United States language minority students were more likely to be either Hispanic or Asian and more likely to come from homes with 25 or fewer books. Nonetheless, White students accounted for a larger portion of the language minority population than any race/ethnicity group other than Hispanics (where no measurable difference was detected between the percentage of White and Hispanic language minority students).

⁶ Although in many countries a parent questionnaire was administered as part of PIRLS, the United States did not administer a PIRLS parent questionnaire.

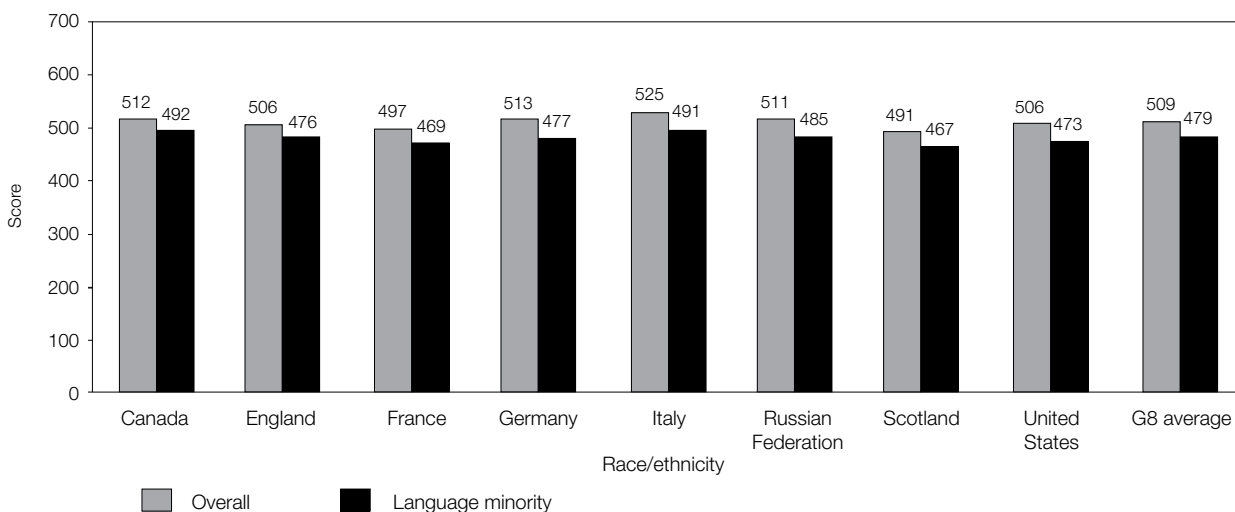
Figure 3: Overall Percentage of Grade 4 Students from Homes with 25 or Fewer Books, and Percentage of Language Minority Grade 4 Students from Homes with 25 or Fewer Books, by Country: 2001



Note: Number of books in the home as reported by the students.

Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

Figure 4: Combined Average Reading Literacy Score for Grade 4 Students from Homes with 25 or Fewer Books, and for Language Minority Grade 4 Students from Homes with 25 or Fewer Books, by Country: 2001



Note: Number of books in the home as reported by the students.

Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

- Across the average of the participating G8 countries, language minority students were also more likely to come from homes with 25 or fewer books compared to G8 fourth-graders overall.
- Across the student demographic and background characteristics, language minority students generally had lower average reading scores than the overall comparative student population.

Student self-perceptions about reading

Student self-perceptions about reading are frequently examined by researchers. PIRLS provides the opportunity to analyze the data cross-nationally through creation of two composite measures: (1) general attitudes toward reading, and (2) reading self-concept. How language minority students generally look on these two measures is examined below.

Students' attitudes toward reading

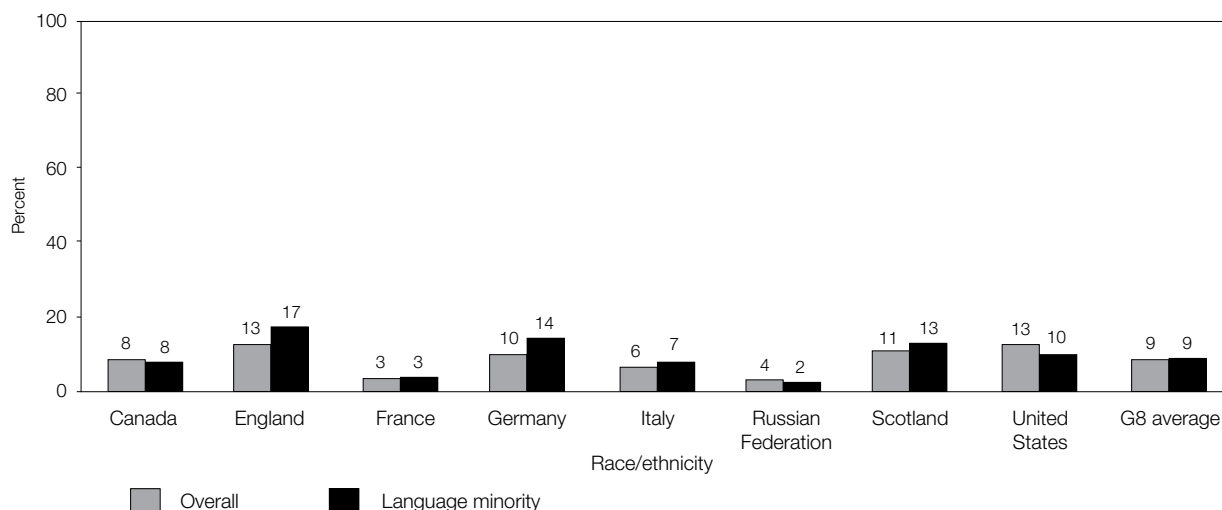
A composite measure of students' attitudes toward reading (SATR) was developed using selected items from the PIRLS student questionnaire (Mullis, Martin, Gonzalez, & Kennedy, 2003). Students were asked about their level of agreement (strongly disagree, disagree, agree, strongly agree) with the following statements:

- I read only if I have to (reverse coded)
- I like talking about books with other people
- I would be happy if someone gave me a book as a present
- I think reading is boring (reverse coded)
- I enjoy reading.⁷

In the United States, fourth-graders who exhibited a poor attitude toward reading were under-represented among language minority students compared to their proportion among fourth-graders overall. In the United States, while 13% of fourth-graders overall indicated a poor attitude toward reading, 10% of language minority students exhibited a poor attitude toward reading (Figure 5). In contrast, across the G8 countries, the same percentage of fourth-graders—both overall and among language minority students—exhibited a poor attitude toward reading (9%).

In the United States and across the average of the G8 countries, language minority students with a low SATR score had lower average scores on the PIRLS combined reading literacy scale than did the overall fourth-grade population with a low SATR score. In the United States, on average, students with a low SATR overall scored 513, while language minority students with a low SATR scored 477 (Figure 6). Across the average of the participating G8 countries, students with a low SATR overall scored 511 while language minority students with a low SATR scored 486.

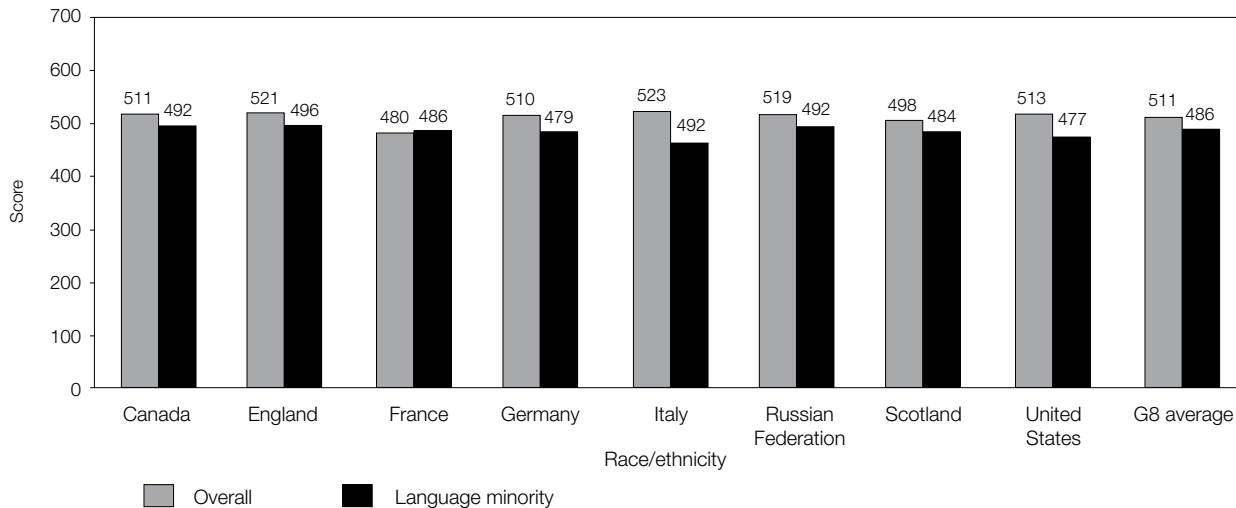
Figure 5: Overall Percentage of Grade 4 Students at the Low Level on the Students' Attitudes Toward Reading (SATR) Index, and Percentage of Language Minority Grade 4 Students at the Low Level on the SATR Index, by Country: 2001



Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

⁷ The SATR average was computed on a four-point scale: disagree a lot = 1, disagree a little = 2, agree a little = 3, agree a lot = 4. High level indicates an average greater than 3 through 4. Medium level indicates an average of 2 through 3. Low level indicates an average of 1 to less than 2. In this paper, a "poor attitude" is equivalent to the low level of SATR.

Figure 6: Combined Average Reading Literacy Score for Grade 4 Students at the Low Level on the Students' Attitudes Toward Reading (SATR) Index, and for Language Minority Grade 4 Students at the Low Level on the SATR Index, by Country: 2001



Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

Students' reading self-concept

A composite measure of students' reading self-concept (SRSC) was developed using selected items from the PIRLS student questionnaire (Mullis, Martin, Gonzalez, & Kennedy, 2003). Students were asked about their level of agreement (strongly disagree, disagree, agree, strongly agree) with the following statements:⁸

- Reading is very easy for me
- I do not read as well as other students in my class (reverse coded)
- Reading aloud is very hard for me (reverse coded).

In the United States, there was no statistically significant difference between the percentage of fourth-graders overall who indicated a low reading self-concept (7%) and the percentage of language minority students who exhibited a low reading self-concept (6%) (Figure 7). However, across the participating G8 countries, whereas 6% of fourth-graders overall indicated a low reading self-concept, 8% of the G8 language minority students exhibited a low reading self-concept. Thus, for the average of the G8 countries, fourth-graders

who exhibited a low reading self-concept were over-represented among language minority students compared to their proportion among fourth-graders overall.

In the United States, there was no statistically significant difference in the scores on the PIRLS combined reading literacy scale between fourth-graders overall with a low SRSC (490) and language minority students with a low SRSC (459) (Figure 8).⁹ However, across the average of participating G8 countries, students with a low SRSC overall scored higher than language minority students with a low SRSC (491 versus 465).

Summary of student self-perceptions about reading and reading performance

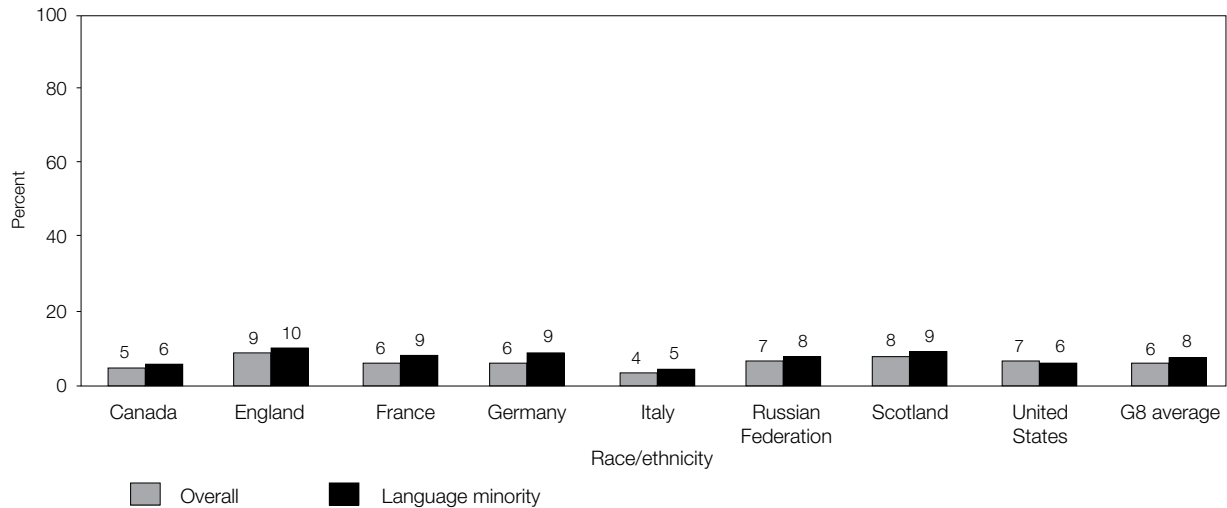
The preceding analyses indicated the following:

- Relatively low percentages of fourth-graders overall in the United States and across the average of the participating G8 countries had poor attitudes toward reading or low reading self-concepts (range from 7% to 13%).

⁸ The SRSC average was computed on a four-point scale: disagree a lot = 1, disagree a little = 2, agree a little = 3, and agree a lot = 4. Responses for negative statements were reverse coded. High indicates an average of greater than 3 through 4. Medium indicates an average of 2 through 3. Low indicates an average of 1 to less than 2.

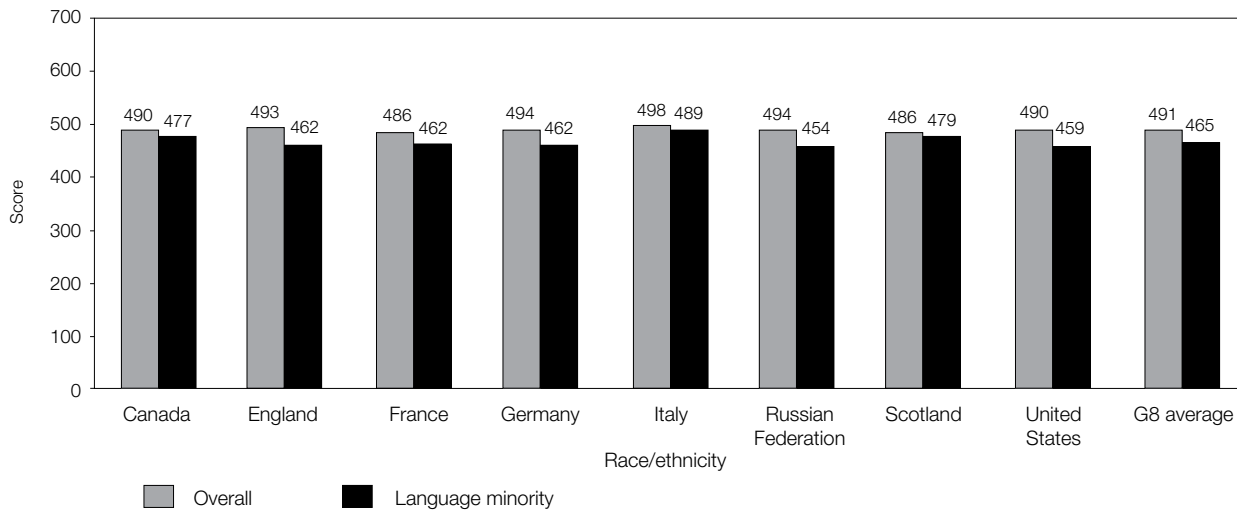
⁹ Some apparent differences may appear large but not be statistically significant. This is due in part to the relatively large standard errors surrounding some of the estimates for particular subgroups (because of a relatively small sample size).

Figure 7: Overall Percentage of Grade 4 Students at the Low Level on the Students' Reading Self-Concept (SRSC) Index, and Percentage of Language Minority Grade 4 Students at the Low Level on the SRSC Index, by Country: 2001



Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

Figure 8: Combined Average Reading Literacy Score for Grade 4 Students at the Low Level on the Students' Reading Self-Concept (SRSC) Index, and for Language Minority Grade 4 Students at the Low Level on the SRSC Index, by Country: 2001



Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

- Among United States fourth-graders, language minority students were less likely to have poor attitudes toward reading than were fourth-graders overall.
- Across the average of the participating G8 countries, language minority students were more likely to hold a low reading self-concept than were fourth-graders overall.
- On average, G8 students overall with a poor attitude toward reading or a low reading self-concept had higher average scores than did the G8 language minority students who had a poor attitude toward reading or a low reading self-concept, respectively. In the United States, the same pattern was found for students with a poor attitude toward reading.

School characteristics

PIRLS used a variety of measures to evaluate school characteristics. For this paper, we chose to focus on several characteristics that may be associated with low student achievement, such as attendance at schools with a relatively large proportion of economically disadvantaged students. By focusing on these schools, we hope to determine if students who attend schools with characteristics that may not foster positive academic outcomes are over-represented among language minority students, compared to students in the overall population.

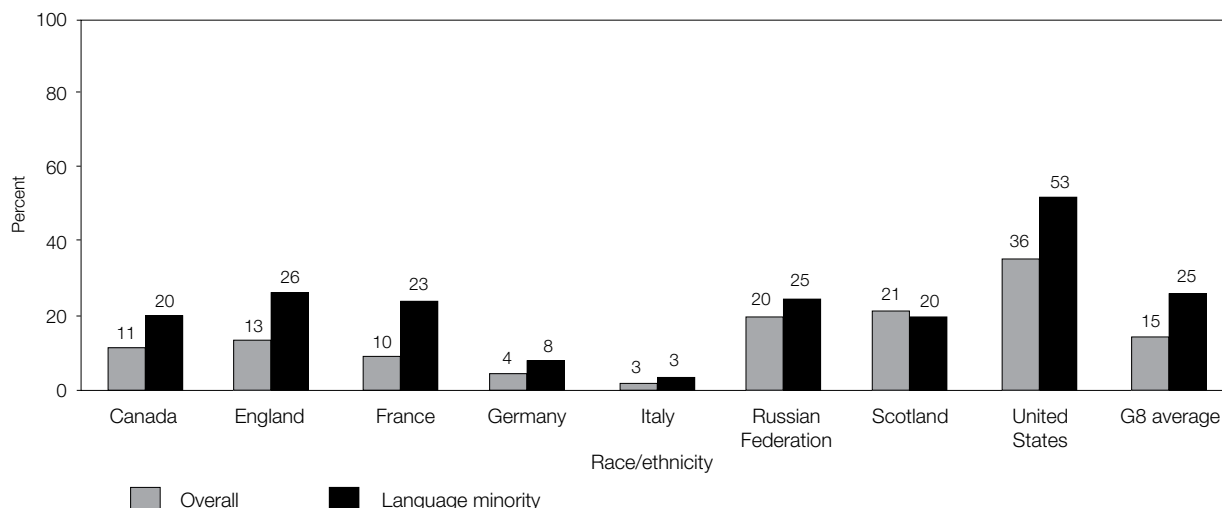
Economic composition of the school's student population

In an attempt to gauge school economic characteristics, PIRLS asked school principals to estimate the percentage of children in their schools who came from economically disadvantaged homes.¹⁰ In this paper, those fourth-graders attending schools where more than 50% of the students reportedly come from economically disadvantaged homes were considered to attend a "disadvantaged school."

In the United States and across the average of the participating G8 countries, fourth-graders who attended disadvantaged schools were over-represented among language minority students compared to their proportion among fourth-graders overall. In the United States, while 36% of fourth-graders overall attended disadvantaged schools, 53% of language minority students attended these schools (Figure 9). On average across the G8 countries, while 15% of fourth-graders overall attended disadvantaged schools, 25% of the G8 language minority students attended these schools.

In the United States and across the average of the G8 countries, language minority students from economically disadvantaged schools had lower average scores on the PIRLS combined reading literacy scale than did the overall fourth-grade population from disadvantaged schools. In the United States, on average, students from disadvantaged schools overall scored 505 while language minority students from disadvantaged schools scored 476 (Figure 10). Across the average of G8 countries, students from disadvantaged schools

Figure 9: Overall Percentage of Grade 4 Students from Disadvantaged Schools, and Percentage of Language Minority Grade 4 Students from Disadvantaged Schools, According to Principals, by Country: 2001

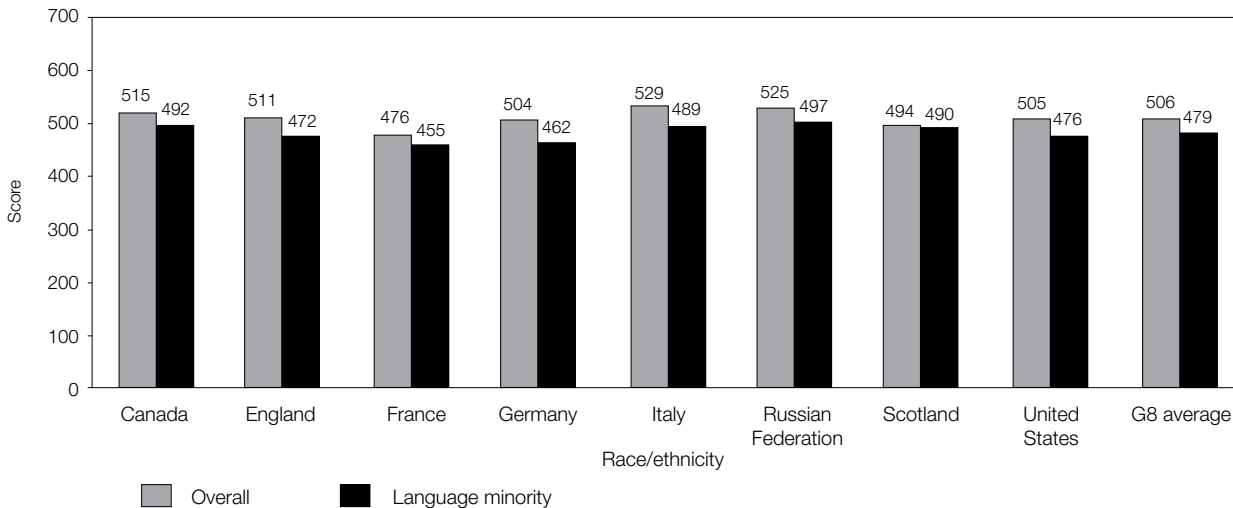


Note: School disadvantage status is based on principals' reports of the percentage of primary-grade students coming from economically disadvantaged homes. Disadvantaged schools were schools where more than 50% of students reportedly came from economically disadvantaged homes.

Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

10 The question in PIRLS was, "Approximately what percentage of students in your school comes from economically disadvantaged homes?" Responses were as follows: (1) 0–10%, (2) 11–25%, (3) 26–50%, and (4) more than 50%. No definition of "disadvantaged" was given to the principals. Rather, PIRLS relied on the judgment of the principals to obtain this measure. Responses differed markedly by country, ranging from 3% of principals in Italy reporting that more than 50% of their students came from economically disadvantaged homes to 36% in the United States reporting the same.

Figure 10: Combined Average Reading Literacy Score for Grade 4 Students from Disadvantaged Schools, and for Language Minority Grade 4 Students from Disadvantaged Schools, by Country: 2001



Note: School disadvantage status is based on principals' reports of the percentage of primary-grade students coming from economically disadvantaged homes. Disadvantaged schools were schools where more than 50% of students reportedly came from economically disadvantaged homes.

Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

overall scored 506 while language minority students from disadvantaged schools scored 479.

School climate

School climate is a key element of the context in which learning takes place. PIRLS measured school climate by asking school principals for their rating of the following:

- Teachers' job satisfaction
- Teachers' expectations for student achievement
- Parental support for student achievement
- Students' regard for school property
- Students' desire to do well in school.¹¹

An index of principals' perceptions of school climate (PPSC) was developed from these items (Mullis, Martin, Gonzalez, & Kennedy, 2003).¹² In most G8 countries, the percentage of students attending schools in which principals reported a low level of school climate was quite small. In fact, this percentage was so small that, on average, PIRLS was not able to estimate an average achievement score for students in any G8 country whose principals reported

a low level of school climate. Consequently, in this analysis, medium and low levels of principal-reported school climate were collapsed into a single category—"moderately to strongly negative."

In the United States, 38% of fourth-graders overall attended schools with moderately to strongly negative principal-reported school climate, and 44% of language minority students attended schools with this level of principal-reported school climate (Figure 11). This difference, however, was not statistically significant. Across all participating G8 countries, on average, 60% of fourth-graders overall attended schools with moderately to strongly negative principal-reported school climate, and 61% of the G8 language minority students attended schools with this level of principal-reported school climate. This difference was also not statistically significant.

Furthermore, these data show the following: a greater percentage of G8 fourth-graders attended schools with moderately to strongly negative principal-reported school climate than did their United States counterparts, both overall (60% versus 38%) and

¹¹ The average is computed on a five-point scale: very high = 1, high = 2, medium = 3, low = 4, and very low = 5. High level indicates an average of 1 to less than 2.33. Medium level indicates an average of 2.33 through 3.67. Low level indicates an average of greater than 3.67 through 5.

¹² Data were not available for England, which did not administer these items (L. Twist, PIRLS National Research Coordinator for England, personal communication, November 16, 2005).

among language minority students (61% versus 44%). As shown in Figure 11, there was generally greater variability across countries with respect to principal-reported school climate than within-country variability between fourth-graders overall and language minority fourth-graders along this characteristic.

In the United States and across the average of the G8 countries, language minority students who attended schools with moderately to strongly negative principal-reported school climate had lower average scores on the PIRLS combined reading literacy scale than did the overall fourth-grade population attending the same type of schools. In the United States, on average, students attending schools with moderately to strongly negative principal-reported school climate overall scored 523 while language minority students attending the same type of schools scored 479 (Figure 12). Across the average of G8 countries, overall students attending schools with moderately to strongly negative principal-reported school climate scored 528 while language minority students attending the same type of schools scored 491.

Availability of school resources

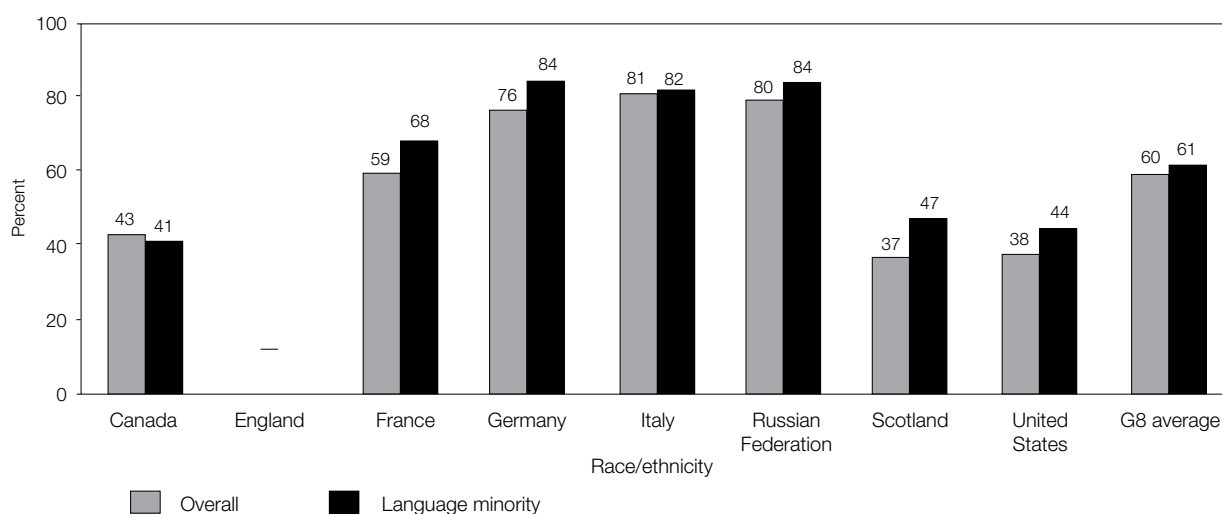
The availability of school resources contributes to the overall instructional context in which learning takes

place. PIRLS created an index of the availability of school resources (ASR) to determine the extent to which a school's capacity to provide instruction was affected by a shortage or inadequacy of staff, basic classroom supplies, instructional materials, or the school's buildings or grounds. Specifically, the index included the following:

- Instructional staff
- Teachers qualified to teach reading
- Instructional materials
- Supplies (e.g., paper, pencils)
- School buildings and grounds
- Heating/cooling and lighting systems
- Instructional space (e.g., classrooms)
- Special equipment for physically disabled students
- Computers for instructional purposes
- Computer software for instructional purposes
- Computer support staff
- Library books
- Audiovisual resources.¹³

A medium or low index score suggested that instruction was affected somewhat or a lot, respectively, by a shortage or inadequacy of resources (Mullis, Martin, Gonzalez, & Kennedy, 2003). In most G8 countries, the percentage of students attending schools

Figure 11: Overall Percentage of Grade 4 Students from Schools with Moderately to Strongly Negative Principals' Perceptions of School Climate (PPSC), and Percentage of Language Minority Grade 4 Students from Schools with Moderately to Strongly Negative PPSC, by Country: 2001

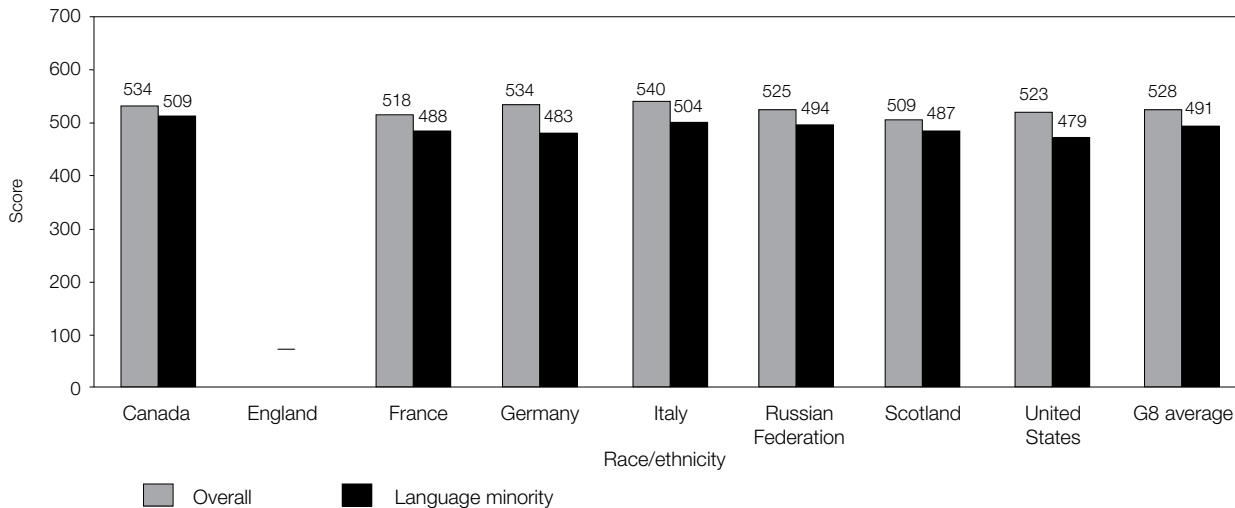


Note: — Not available.

Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

¹³ The average was computed on a four-point scale: not at all = 1, a little = 2, some = 3, and a lot = 4. High level indicates an average of 1 to less than 2. Medium level indicates an average of 2 through 3. Low level indicates an average of greater than 3 through 4.

Figure 12: Combined Average Reading Literacy Score for Grade 4 Students from Schools with Moderately to Strongly Negative Principals' Perceptions of School Climate (PPSC), and for Language Minority Grade 4 Students from Schools with Moderately to Strongly Negative PPSC, by Country: 2001



Note: — Not available.

Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

in which principals indicated a low level of school resources was, on average, quite small. In fact, this percentage was so small that PIRLS was not able to estimate an average achievement score for students in four of the G8 countries whose principals indicated a low level of school resources.¹⁴ Consequently, in this analysis, medium and low levels of school resources were collapsed into a single category—"medium or low."

In the United States, 23% of fourth-graders overall attended schools with a medium or low level of school resources and 31% of language minority students attended schools with this level of school resources (Figure 13). This difference, however, was not statistically significant. Across the average of all participating G8 countries, 40% of fourth-graders overall attended schools with a medium or low level of school resources and 41% of the G8 language minority students attended schools with this level of school resources. This difference was also not statistically significant.

These data also show the following: a greater percentage of G8 fourth-graders attended schools with a medium or low level of school resources than did their United States counterparts (40% versus 23%). As shown in Figure 13, there was generally greater

variability across countries with respect to principals' reports of school resources than within-country variability between fourth-graders overall and language minority fourth-graders along this characteristic.

In the United States and across the average of the participating G8 countries, language minority students who attended schools with a medium or low level of school resources had lower average scores on the PIRLS combined reading literacy scale than did the overall fourth-grade population attending the same type of schools. In the United States, on average, students attending schools with a medium or low level of school resources overall scored 525 while language minority students attending the same type of schools scored 490 (Figure 14). Across the average of participating G8 countries, students attending schools with a medium or low level of school resources overall scored 533, while language minority students attending the same type of schools scored 494.

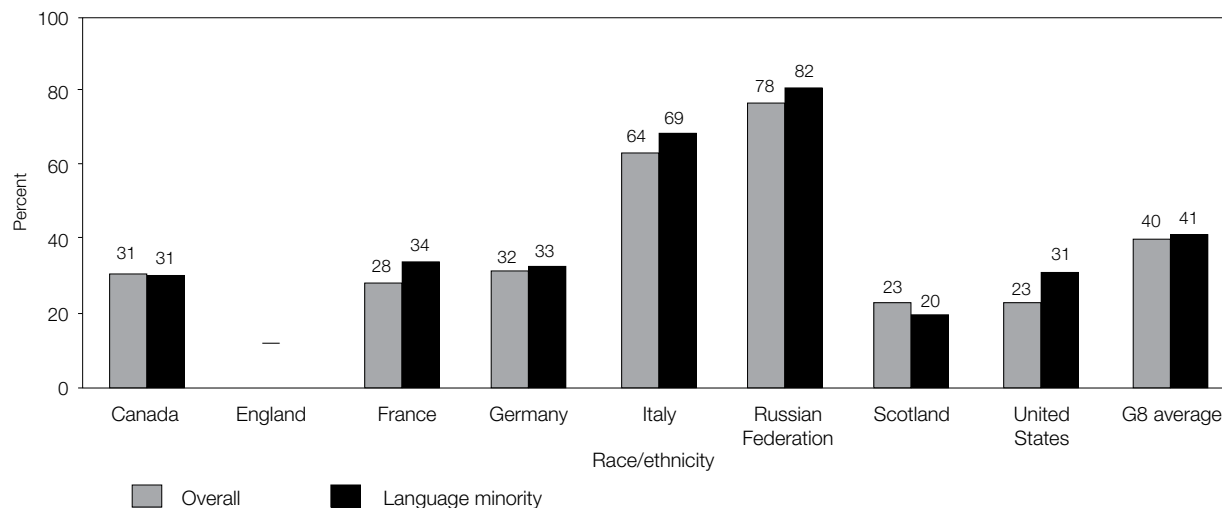
Summary of school characteristics

The preceding analyses indicated the following:

- Greater percentages of language minority students than students from the overall fourth-grade population were attending economically disadvantaged schools. This finding held for both

¹⁴ In addition, data were not available for this index for one country: England.

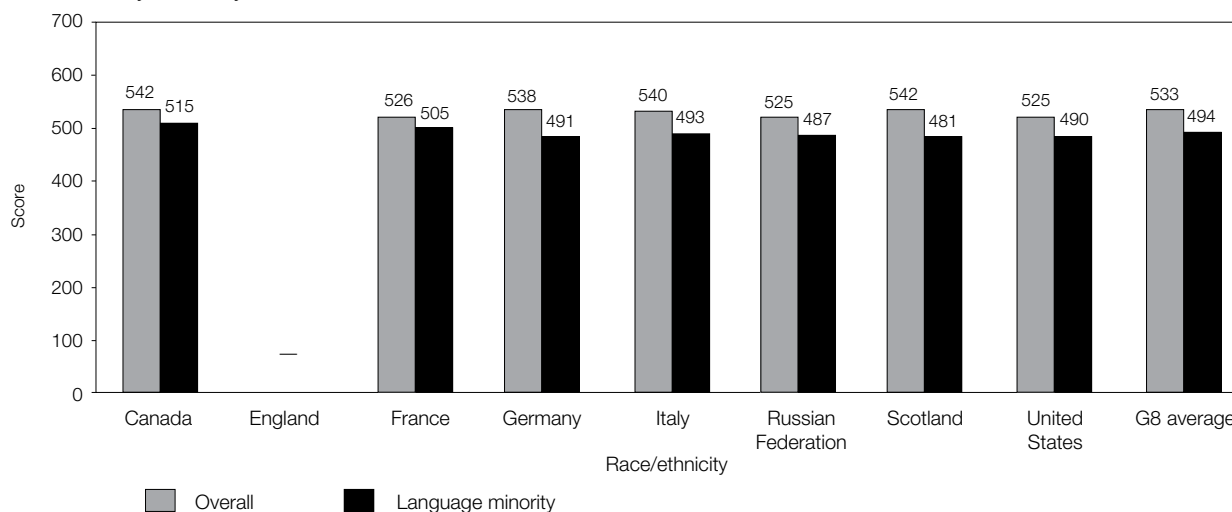
Figure 13: Overall Percentage of Grade 4 Students from Schools with Medium or Low Availability of School Resources (ASR), and Percentage of Language Minority Grade 4 Students from Schools with Medium or Low ASR, by Country: 2001



Note: — Not available.

Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

Figure 14: Combined Average Reading Literacy Score for Grade 4 Students from Schools with Medium or Low Availability of School Resources (ASR), and for Language Minority Grade 4 Students from Schools with Medium or Low ASR, by Country: 2001



Note: — Not available.

Source: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001.

the United States and across the average of the participating G8 countries.

- For the most part, greater percentages of students in the G8 countries, on average, were attending schools with undesirable characteristics (i.e., moderately to strongly negative principal-reported

school climate and medium to low level of school resources) than were United States fourth-graders. A notable exception, however, was the percentage of students attending disadvantaged schools.

- There was generally greater variability across countries pertaining to the school characteristics

compared to the within-country variability observed between fourth-graders overall and language minority fourth-graders along these characteristics.

- In all school types and in both the United States and across the average of the participating G8 countries, the average reading scores of language minority students were lower than the average reading scores of the fourth-graders overall.

Discussion and conclusions

According to the data presented in this paper, United States fourth-grade language minority students relative to the overall United States fourth-grade population were more likely to be Hispanic or Asian, to come from homes with 25 or fewer books, or to attend schools with a relatively large proportion of disadvantaged students. Although fewer language minority United States students than United States students overall had poor attitudes toward reading, their average scores on the PIRLS combined reading literacy scale were lower than those of the overall fourth-grade student population. With few exceptions, United States language minority students scored lower than United States fourth-graders overall across the student and school characteristics examined in this analysis.

Like the United States language minority students, G8 language minority students, on average, relative to the overall fourth-grade population, were also more likely to come from homes with 25 or fewer books and to attend schools with relatively large proportions

of disadvantaged students. G8 language minority students, on average, also consistently scored lower than fourth-graders overall across the student and school characteristics examined in this analysis.

Thus, the general findings from this analysis show that language minority students consistently perform lower than their peers overall. Furthermore, children from homes with 25 or fewer books and children attending disadvantaged schools are over-represented among language minority children, compared with their representation in the overall student population.

Another noteworthy finding is that there was generally greater variability across countries pertaining to the school characteristics compared to the within-country variability observed between fourth-graders overall and the language minority fourth-graders along these characteristics. This finding suggests that principals' responses to these items may carry different meanings in different countries.

However, the general findings about language minority students from this analysis should be interpreted with some caution, as there were some apparent differences along the selected characteristics that were not statistically significant. This was due to large standard errors associated with the estimates for small subgroups. Also, as previously pointed out, some of these findings are based on student self-reports while the school findings are based on principals' perceptions.

Appendix: Technical notes

PIRLS study design and methodology

Ten years prior to PIRLS, the IEA conducted the International Reading Literacy Study of 1991. Although PIRLS was built on the foundation of this 1991 study, it was a new study that was designed using a different assessment framework. The PIRLS framework was developed with the intention of incorporating the latest approaches to measuring reading literacy as advocated by experts in the field (Campbell, Kelly, Mullis, Martin, & Sainsbury, 2001). PIRLS 2001 was the first in a planned five-year cycle of international trend studies. More than 40 countries planned to participate in PIRLS 2006.

PIRLS defines reading literacy as follows:

The ability to understand and use those written language forms required by society and/or valued by the individual. Young readers can construct meaning from a variety of texts. They read to learn, to participate in communities of readers, and for enjoyment. (Campbell et al., 2001, p. 3)

The reading passages and corresponding items in the PIRLS assessment were developed based on two aspects of reading literacy: purposes of reading and processes of comprehension. *Purposes of reading* refers to the two types of reading that account for most of the reading young students do, both in and out of school: (a) reading for literary experience, and (b) reading to acquire and use information. PIRLS 2001 contains about an equal proportion of text assessing each purpose, with narrative fiction used to assess students' ability to read for literary experience, and a variety of informational texts used to assess students' ability to acquire and use information while reading. *Processes of comprehension* refers to ways in which readers construct meaning from the text. These processes include: (a) focusing on and retrieving explicitly stated information, (b) making straightforward inferences, (c) interpreting and integrating ideas and information, and (d) examining and evaluating content, language, and textual elements. Each of these processes is embedded within each purpose of reading (Campbell et al., 2001).

As described in Ogle et al. (2003), a group of distinguished international reading scholars, the

Reading Development Group, was formed to construct the PIRLS 2001 reading framework and to approve the final reading assessment. Each country followed internationally prescribed procedures to ensure valid translations and representative samples of students. Quality Control Monitors were then appointed in each country to monitor the testing sessions at the schools and to ensure that the high standards of the PIRLS 2001 data collection process were met.

The target population set by the IEA for PIRLS was all students enrolled in the upper of the two adjacent grades with the largest proportion of nine-year-olds at the time of test administration. This was referred to as the *international desired population*. National coverage of the international desired population—which determines the *national desired population*—was generally comprehensive. Coverage was 100% in all countries except Canada and Lithuania.¹⁵ Overall exclusions from the national desired population (including both school-level exclusions and within-school exclusions) were generally 5% or under across countries. For the most part, school-level exclusions included very small schools and schools for the disabled. Within-school exclusions generally consisted of students who were unable to take part in the assessment because of a physical or mental disability and students who could not be assessed in the language of the test. The number of schools in the participating countries' *effective target population* ranged from 140 in Iceland to 71,498 in the United States, and the number of students in the participating countries' effective target population ranged from 4,566 in Iceland to almost 3.9 million in the United States (Joncas, 2003).

Data for PIRLS 2001 were collected in the final months of the 2000/01 school year. In the United States, data were collected in the spring of 2001 from both public and private schools. Internationally, more than 150,000 students participated, including almost 3,800 fourth-grade students from the United States. Consistent with the PIRLS sampling requirements, most countries sampled about 150 schools and one intact classroom from each school. School participation

¹⁵ In Canada, national coverage was 60%, including only students in the provinces of Ontario and Quebec. In Lithuania, national coverage was 90%, including only students in Lithuanian-speaking schools.

rates before the use of replacement schools were at least 80% in all but eight countries—including the United States (with a participation rate of 61%). After the use of replacement schools, school participation rates were at least 80% in all but two countries—Scotland (79%) and Morocco (74%). In the United States, the school participation rate after the use of replacement schools was 86%. Student participation rates were at least 90% in all but three countries, with Lithuania the lowest at 85%. The student participation rate in the United States was 96%. The overall participation rate after the use of replacement schools ranged from 69% in Morocco to 98% in Iran (Islamic Republic of), Italy, and Singapore; in the United States, the overall participation rate was 83% (Joncas, 2003).

Statistical procedures

Tests of significance

Comparisons made in the text of this report were tested to establish whether the observed differences were statistically significant. The tests used were dependent *t*-tests, since the comparisons involved comparing a subgroup of a population to the total population. An example is the comparison of the percentage of language minority fourth-graders from homes with 25 or fewer books with the percentage of fourth-graders overall from homes with 25 or fewer books. In such comparisons, the following formula was used to compute the *t* statistic:

$$t = \frac{\text{est}_{\text{sub}} - \text{est}_{\text{total}}}{\text{SQRT}[(\text{se}_{\text{sub}})^2 + (\text{se}_{\text{total}})^2 - 2p(\text{se}_{\text{sub}})^2]}$$

est_{sub} is the subgroup estimate being compared, and $\text{est}_{\text{total}}$ is the total estimate being compared. se_{sub} and se_{total} are the corresponding standard errors of these estimates; *p* is the proportion of the subgroup to the total population. For example, 15% of United States fourth-graders were language minority students, and so *p* for the comparisons for the United States is .15. In all comparisons using *t*-tests, the critical value of *t* is 1.96.

Standard errors

As a result of the complex sampling design in PIRLS 2001, the resulting statistics are more variable than they would have been had they been based on data from a simple random sample of the same size. Several methods are available for calculating precise estimates of sampling errors for complex samples. One such method is the jackknife repeated replication (JRR) technique (Wolter, 1985). The particular application of this technique used with these data files is a paired selection model because it assumes that the sampled population can be partitioned into strata, with the sampling in each stratum consisting of two primary sampling units (PSU), selected independently. A number of statistical software packages can account for complex sampling designs and calculate precise estimates of sampling errors for complex samples. The analyses carried out in this report used the Taylor Series procedure to calculate standard errors. They were done using the AM statistical software package (Cohen, Albright, Cohen, Hollender, Jiang, & Chan, 2003). AM is available for free download at <http://am.air.org>

References

- Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.
- Campbell, J. R., Kelly, D. L., Mullis, I. V. S., Martin, M. O., & Sainsbury, M. (2001). *Framework and specifications for PIRLS assessment 2001* (2nd ed.). Chestnut Hill, MA: Boston College.
- Cohen, J., Albright, L., Cohen, D., Hollender, D., Jiang, T., & Chan, T. (2003). *AM statistical software*. Washington, DC: Jon Cohen and the American Institutes for Research.
- Coleman, J. S., Kelly, D. L., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. United States Department of Health, Education, and Welfare. Washington, DC: United States Government Printing Office.
- Joncas, M. (2003). PIRLS sampling weights and participation rates. In M. O. Martin, I. V. S. Mullis, & A. M. Kennedy (Eds.), *PIRLS 2001 technical report* (pp. 113–133). Chestnut Hill, MA: Boston College.
- Kelly, D. L. (2003). Overview of PIRLS. In M. O. Martin, I. V. S. Mullis, & A. M. Kennedy (Eds.), *PIRLS 2001 technical report* (pp. 1–12). Chestnut Hill, MA: Boston College.
- Klein, S., Bugarin, R., Beltranena, R., & McArthur, E. (2004). *Language minorities and their educational and labor market indicators: Recent trends* (NCES 2004–009). United States Department of Education. Washington, DC: National Center for Education Statistics.
- Lemke, M., Calsyn, C., Lippman, L., Jocelyn, L., Kastberg, D., Liu, Y. Y., Roey, S., Williams, T., Kruger, T., & Bairu, G. (2001). *Outcomes of learning: Results from the 2000 Program for International Student Assessment of 15-year-olds in reading, mathematics, and science literacy* (NCES 2002–115). United States Department of Education, National Center for Education Statistics. Washington, DC: United States Government Printing Office.
- Lemke, M., Sen, A., Johnston, J. S., Pahlke, E., Williams, T., Kastberg, D., & Jocelyn, L. (2005). *Characteristics of United States 15-year-old low achievers in an international context: Findings from PISA 2000* (NCES 2006–010). United States Department of Education, National Center for Education Statistics. Washington, DC: United States Government Printing Office.
- Miller, D. C. (2006, April). *Personal and contextual resources as related to fourth-graders' reading achievement: An international perspective*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., and Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools in 35 countries*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., & Kennedy, A. M. (2003). Item analysis and review. In M. O. Martin, I. V. S. Mullis, & A. M. Kennedy (Eds.), *PIRLS 2001 technical report* (pp. 135–150). Chestnut Hill, MA: Boston College.
- Ogle, L. T., Miller, D. C., & Johnston, J. S. (2006, April). *Characteristics of U.S. fourth-grade low achievers in an international context: Findings from PIRLS 2001*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Ogle, L. T., Sen, A., Pahlke, E., Jocelyn, L., Kastberg, D., Roey, S., & Williams, T. (2003). *International comparisons in fourth-grade reading literacy: Findings from the Progress in International Reading Literacy Study (PIRLS) of 2001* (NCES 2003–073). United States Department of Education, National Center for Education Statistics. Washington, DC: United States Government Printing Office.
- Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the Black–White achievement gap*. Washington, DC: Economic Policy Institute.
- Sen, A., Partelow, L., & Miller, D. C. (2005). *Comparative indicators of education in the United States and other G8 countries* (NCES 2005–021). United States Department of Education, National Center for Education Statistics. Washington, DC: United States Government Printing Office.
- Torney-Purta, J., Lehmann, R., Oswald, H., and Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen*. Amsterdam: The International Association for the Evaluation of Educational Achievement.
- United States Department of Education, National Center for Education Statistics. (2006). *The condition of education 2006* (NCES 2006–071). Washington, DC: United States Government Printing Office.
- West, J., Denton, K., & Reaney, L. (2000). *The kindergarten year* (NCES 2001–023). United States Department of Education, National Center for Education Statistics. Washington, DC: United States Government Printing Office.
- Williams, T., Levine, D., Jocelyn, L., Butler, P., & Haynes, J. (2000). *Mathematics and science in the eighth grade: Findings from the Third International Mathematics and Science Study* (NCES 2000–014). United States Department of Education. Washington, DC: National Center for Education Statistics.
- Wolter, K. M. (1985). *Introduction to variance estimation*. New York: Springer-Verlag.

Teachers' practices in relation to students' civic engagement in three countries¹

Carolyn Barber
Department of Human Development
University of Maryland
College Park, Maryland, USA

Introduction

The development of civic knowledge, skills, and dispositions does not take place in a vacuum; rather, it is influenced extensively by one's proximal and distal surroundings. This belief was central to the development of the International Association for the Evaluation of Educational Achievement's Civic Education Study of 1999 (IEA CivEd: Torney-Purta, Lehmann, Oswald, & Schulz, 2001). In the conceptual framework guiding this study (commonly referred to as the "Octagon Model"), Torney-Purta and her colleagues place the individual at the center of a complex system, most immediately surrounded by teachers, parents, peers, and others who are responsible for translating to the individual the social goals and values of a society. Because CivEd focuses on adolescents, who spend a considerable amount of time participating in school activities both inside and outside of class, consideration of how school acts as a catalyst for development is especially important.

The theoretical frameworks set down by Bronfenbrenner (1988) and Lave and Wenger (1991), which each served as background for the development of the Octagon Model, can help us understand more specifically why schools (and why teachers in particular) are important for adolescents' civic development. In the tradition of Bronfenbrenner, one can think of teachers as acting as a microsystem for students, within which teachers influence students by interacting directly with them. Through these interactions, teachers can pass on their own attitudes to students. Teachers who think that learning about civic education is important can instill in their students an appreciation for the beliefs and behaviors necessary for active democratic citizenship. This transmission is most effective when

teachers are confident in what they teach. In summary, teachers' beliefs—both about themselves and about their subject matter—have the potential to greatly influence students' civic development (see Pajares, 1992; Shulman, 1986).

In the communities of practice framework, the practices of different teachers provide students with different opportunities to rehearse the skills needed for civic engagement. Teachers who regularly set up opportunities to actively engage in learning (e.g., by allowing students to research a topic themselves rather than having them rely on teachers' lectures) give students a chance to exercise the cognitive skills associated with civic engagement. In many cases, teachers may frame such engagement in the context of group work, which additionally requires students to practice working with other students from diverse backgrounds. Further, teachers may shape students' civic attitudes by creating situations in which students practice learning about ideas related to specific aspects of civic development. An appreciation for cultural diversity, for example, may be encouraged by providing students with activities specifically related to this area. Finding inspiration for such new ideas, however, can be difficult. Analyses of textbooks and content standards in the United States reveal that such sources often focus on the most conventional definitions of civic engagement, such as voting and government structure (Gonzales, Reidel, Avery, & Sullivan, 2001; Gonzales, Reidel, Williamson, Avery, Sullivan, & Bos, 2004). As a result, teachers who wish to provide students with such opportunities may need to look beyond the usual standards and curriculum manuals for guidance.

1 The comments and assistance of Amy Hendrickson and Judith Torney-Purta are gratefully acknowledged.

Previous analyses of teachers' effects on students in the IEA CivEd

Several previous analyses of data from IEA's CivEd have begun to link characteristics of teachers to students' civic development. These studies used either single items or simple composites (employing classical test theory techniques) from the teacher survey to create variables, or focused on those items that most closely mirrored the items available from the student survey.

Torney-Purta, Richardson, and Barber (2005) acknowledge in their review of literature on the knowledge and beliefs of teachers of civic-related subjects that few studies explicitly link the characteristics of teachers to student outcomes. Torney-Purta and her colleagues sought to fill the gap by following up a series of descriptive analyses of teachers with simple multilevel models connecting teachers' training and confidence in teaching political topics to students' civic knowledge in several countries. In the United States, for example, students of teachers who reported participating in civic-related professional training had significantly higher civic knowledge than had students of teachers without such training. While such training was not significant in Hungary, Hungarian students with teachers reporting high confidence in teaching political topics had higher civic knowledge than had students with less confident teachers.

The effects of teachers on students' civic development are not limited to civic knowledge. Additional analysis of United States teachers by Torney-Purta and colleagues (Torney-Purta, Barber, & Richardson, 2005) revealed that students whose teachers had participated in civic-related professional activities also had higher expectations of informed voting than students who did not have such teachers. In addition, students with teachers who had a university degree in a civic-related subject also had higher expectations of informed voting than those who had teachers with no such degree. The effect of teachers' degree, however, was smaller than the effect of teachers' participation in professional activities.

More recent analysis has included teacher variables formed from composites of items in more complex models of student outcomes. An analysis by Torney-Purta, Barber, and Wilkenfeld (2007) found that United States students whose teachers considered the use of official standards and curricular guidelines very important to their planning had stronger expectations

of informed voting than had students whose teachers found this use less important. This effect stood after controlling for several characteristics of students and schools, including ethnicity, home background, school socioeconomic status, and openness of classroom climate for discussion.

In the same analysis, Torney-Purta and colleagues found that the gap between Latino and non-Latino students in positive attitudes toward immigrants was larger in schools where teachers more regularly used interactive activities such as role-playing and group projects. In these schools, Latino adolescents' attitudes toward immigrant rights were even more positive, while no difference in attitudes was observed for non-Latino students. Torney-Purta and colleagues interpreted this finding as evidence that the interactive activities, which encouraged students to become more actively engaged in their schoolwork, led to Latino adolescents' increased identification with the immigrant population regardless of whether they themselves were born in the United States.

In their analysis of how a school's citizenship education climate can contribute to civic development, Homana and Barber (this volume) found the use of interactive class activities by teachers in the United States to be a significant predictor of students' civic knowledge. In this analysis, teacher reports of the activities used in their classrooms were included in a model along with many other aspects of school citizenship education climate considered as predictors of knowledge, among them extracurricular activity participation, learning about cooperation, classroom discussion climate, and students' individual reports of their classroom activities.

While the majority of analyses on teachers in CivEd did not employ item response theory (IRT) techniques, a study of teachers' norms of conventional and social-movement citizenship conducted by Husfeldt and Barber (2005) involved the creation of IRT scales using a partial credit model (Masters & Wright, 1997). The scales created in this analysis mirrored the citizenship norms scales created for students (see Torney-Purta et al., 2001). Follow-up multilevel analysis linking teachers' concepts of citizenship to student outcomes found significant effects of teachers' citizenship concepts on students' attitudes toward conventional citizenship norms and their confidence in the value of school participation in Germany, Hungary, and the

United States; and additionally predicted students' social-movement citizenship norms and expectations of participating in political activities in Hungary only.

Summary: The current analysis

In summary, several analyses have begun to consider how teachers' experiences, practices, and beliefs can influence students' civic development along several dimensions. To gain more information, however, we need to use more sophisticated measurement and analytical techniques. One hallmark of CivEd (and of IEA studies more generally) is the use of IRT scales to measure students' knowledge, skills, attitudes, and behaviors. IRT methods are preferable to scales made using classical test theory techniques in a cross-national study like this for two reasons: first, they allow for the more precise estimation of respondent characteristics based on fewer items; second, they can easily be tested to ensure scale comparability across countries or age cohorts (Hambleton, Swaminathan, & Rogers, 1991). The development of IRT scales based on the teacher survey will improve researchers' ability to study civic-related teachers and these teachers' role in civic development.

IRT scales are only practical, however, when scale is thought to capture "individual differences" in a given behavior or attitude (Lubke, 2006). For certain behaviors or attitudes, there may be fewer "individual differences" to capture. For example, while a nationally representative sample of students or adults (or even teachers, more specifically) may differ widely in their view of the importance of civic education, the sample of civic-related teachers of 14-year-olds surveyed as part of CivEd may demonstrate considerably less variation. One could argue that these teachers would not have chosen to teach civic-related subjects if they did not find the subject at least somewhat important. Similarly, there may not be many "individual differences" in how such teachers use planning materials. In either case, it may be more correct to say that certain groups of teachers exist with distinct patterns of importance beliefs or of planning habits and that response to certain items can indicate category membership. When categorical variables (e.g., the Likert-scales used in CivEd survey items) are thought to indicate category membership rather than location on a continuous scale, latent class analysis is

an appropriate technique to use to analyze the data (Dayton, 1998). Just as IRT analysis results in scales that can be used in future analyses, latent class analysis results in categorical group membership variables that can be used in similar ways.

The purpose of this current study is twofold: (i) to create, using IRT and latent class analytic techniques, a series of variables measuring teachers' confidence, classroom activity usage, importance beliefs, and planning habits; (ii) to use, in a multilevel analysis, the resulting variables to predict characteristics of students' civic engagement and development. The analysis focuses on Australia, England, and the United States—three countries with major policy initiatives relating to the formal teaching of civic education around the time of CivEd (e.g., the 1998 Crick Report in England, the Civic Expert Group report of 1994 in Australia, and The National Commission on Civic Renewal of 1998 in the United States).

Data

The IEA, a consortium of educational research institutes from nearly 60 countries, has been conducting comparative education studies for nearly 50 years. In the early 1990s, some member countries asked for a study of civic education that included measures of young people's civic-related attitudes and behaviors. Their aim was to study schools' intentions and practice relating to democracy. The IEA brought to this effort a wide network of research institutes in different countries and a wealth of technical and methodological expertise in cross-national comparative education research. Included in this network were several experts in the areas of Rasch scaling and polytomous IRT models.

CivEd 1999 consisted of two stages. The case studies conducted in the first phase of the study were used as the basis of a consensus process aimed at developing content specifications for a test of students' civic knowledge (with right and wrong answers) and surveys of political attitudes and civic behavior, as well as surveys administered to teachers and school administrators. In the second phase, an international steering committee and national research coordinators developed an instrument (a student test and survey) that would be suitable for younger and older adolescents and would take students about two class periods to complete.

Nationally representative samples of students in the modal grade for 14-year-olds (a total of about 90,000 students from 28 countries) were tested in 1999; upper secondary students ranging in age from 16 to 19 (a total of about 50,000 students from 16 countries) were tested in 2000. Torney-Purta and colleagues (2001) described the scaling and analysis of the 28 countries surveying 14-year-olds, while Amadeo, Torney-Purta, Lehmann, Husfeldt, and Nikolova (2002) described the same work conducted in the 16 countries that assessed upper secondary students.

Teachers in CivEd

A survey of teachers' experiences and attitudes was also developed and administered in Phase 2 of the study. Teachers were administered a survey in 27 of the 28 countries that surveyed 14-year-olds, making data available on teachers in every country surveyed except Colombia. In each school where 14-year-old students were tested, between one and three teachers were identified as teachers of "civic-related topics," as they taught subjects such as government, national history, social studies, and social sciences. While it was preferred that the three sampled teachers were the teachers of the class of students who filled out the survey (i.e., they taught those students), other teachers of civic-related disciplines were surveyed if information from one or more teachers of the tested students could not be obtained. The student samples were nationally representative of 14-year-olds; however, the samples of teachers were not representative of civic-related teachers.

The teacher survey consisted of three sections: work experience and education, views on civic education, and the teaching of civic-related subjects, activities, and lessons. The items in each of these three sections were designed to provide contextual information on students' in-school civic education experiences. Losito and Mintrop (2001) provide additional information on how teachers were selected for surveying, as well as an overview of results from the teachers' survey reported on an item-by-item basis.

Method

This current study focuses on data from the civic-related teachers of 14-year-old students in England, Australia, and the United States. The decision to focus on three countries rather than 27 came from a desire to explore in depth how certain items, and

scales made from these items, functioned differently in different countries (Barber, 2006). Statistical analyses of differential item functioning require comparisons among groups (see, for example, Swaminathan & Rogers, 1990). These become extremely complicated when analyzing 27 groups at a time but are more reasonable when exploring only three simultaneously. The three countries chosen were all from a similar language background. The likelihood of questions relating to how items translated across samples from different countries was lowered because the countries were similar in cultural background, and because civic education is a relevant policy topic in each country. Additional information on the education systems for each of these countries, which further outlines similarities and differences in their civic education systems, can be found in the Phase 1 book of case studies by Torney-Purta, Schwille, and Amadeo (1999).

A total of 589 civic-related teachers who reported that they taught the students in a tested class were surveyed in these three countries (Australia $n = 244$, England $n = 271$, United States $n = 74$; see Husfeldt & Barber, 2005, for a description of the selection process). More civic-related teachers were surveyed in Australia and England than in the United States because in the first two countries more than one teacher was linked to a classroom of students (see Losito & Mintrop, 2001, for details).

Variables

The items of interest for this study come from the section of the teacher survey relating to the teaching of civic-related subjects, activities, and lessons. In particular, the analysis focuses on items relating to teachers' frequency of activity use and their confidence in teaching political topics, and extends upon two previous, exploratory analyses of teachers' items conducted by researchers affiliated with CivEd (Torney-Purta, Barber, & Wilkenfeld, 2007; Torney-Purta, Richardson, & Barber, 2005).

Use of classroom activities

Previous analysis shows that the use of classroom activities is a multidimensional construct (Torney-Purta et al., 2001). In the analysis of data from the student survey of CivEd (Schulz & Sibberns, 2004), a two-factor model of students' perceptions of classroom experiences best fitted the data. One dimension relates to an open classroom climate for discussion; the

other focuses on the perception of a lecture-centered environment. Many students who perceived an open discussion climate also perceived a very lecture-oriented classroom.

Similarly, exploratory analyses of data from teachers suggested the presence of several dimensions of classroom experience. In particular, two dimensions thought to be of greatest theoretical importance to the study of teacher practices in a previous analysis (Torney-Purta, Richardson, & Barber, 2005) were selected for analysis. The more "traditional" classroom activities are thought to be teacher-centered, in the sense that teachers are responsible for administering materials that students must use, assessments that students must take, and questions that students must answer. Five items were thought to relate to the use of such activities. Teachers were asked to rate on a scale of 1 ("never") to 4 ("very often") how often the following events happened in their classrooms:

- The teacher chooses the issues to be discussed in class
- Students study textbooks
- Students work on drill sheets or work sheets
- The teacher asks questions and the students answer
- The teacher lectures and the students take notes.

Items relating to this idea are thought to measure the importance that teachers place on interactive classroom activities of any type. The three items chosen as indicators for this construct (based on the above exploratory analysis and on analysis reported by Torney-Purta et al., 2007) were:

- Students work on projects that involve gathering information outside of school
- Students work in groups on different topics and prepare presentations
- Students participate in role play and simulations.

Confidence in teaching political topics

A teacher's "confidence" can refer to many things, including both level of comfort with the subject matter and efficacy in teaching itself (Shulman, 1986). While very few items in CivEd captured teachers' pedagogical efficacy, many were designed to measure teachers' levels of comfort when teaching various civic-related topics. But even if the focus is solely on the subject matter of civics, teaching confidence is a multidimensional construct. Teachers may feel very adept at teaching about the political process or about

systems of government, but may feel differently about teaching topics less traditionally thought of as "civics," such as women's rights or cultural diversity. Indeed, preliminary exploratory factor analysis of the items addressing teachers' confidence in teaching political topics reported by Torney-Purta, Richardson, & Barber (2005) found at least three such dimensions.

In their exploratory analysis of teachers' confidence, Torney-Purta and colleagues focused on two of the dimensions thought to be most prevalent in civics and other related classes cross-nationally. This current analysis extended their work by further analyzing these same two dimensions. For the first dimension, five items were selected that were thought to best relate to conventional political issues. Experts in the field of civic education considered these five items to be those most important to civic education and those most comparable in meaning across countries (as discussed in Torney-Purta, Richardson, & Barber, 2005). Teachers were asked to report their level of confidence for "dealing with" the following topics, with possible answers ranging from "not at all confident" to "very confident":

- National constitution and state/political institutions
- Citizens' rights and obligations
- Different conceptions of democracy
- Elections and electoral systems
- The judicial system.

The second dimension was hypothesized to represent the confidence that teachers had in teaching topics dealing more with social issues, rather than with political systems. The items thought to indicate confidence in this area asked teachers, "How confident do you feel to deal [sic] with this topic:

- Equal opportunities for women and men
- Cultural differences and minorities
- The media?"

Importance of teaching civic-related topics

In addition to reporting their confidence in teaching various civic-related topics, teachers were asked to indicate the importance they attached to each topic. Teachers responded to each item on a four-point scale: "not important," "less important," "important," "very important." Two sets of items measuring the importance that teachers gave to civic-related topics matched the two sets of items measuring teachers' confidence.

One set measured teachers' *belief in the importance of teaching political topics*; the other measured teachers' *belief in the importance of teaching social topics*.

An important difference between teachers' belief in the importance of activities and their confidence in teaching them is the extent to which there are individual differences among teachers' views of the importance of these subjects. While the items in the teacher survey may have captured individual differences in teachers' confidence levels on a continuum, the items related to importance are better suited for identifying categories of teachers who find certain activities to be more or less important. One reason for this difference has to do with the specific group of teachers being studied. Civic-related teachers are unlikely to find many (if any) civic-related subjects unimportant to teach, although some may find certain subjects more or less important. This consideration is reflected in the fact that only a small percentage of the teachers reported that they considered teaching many subjects "not important" or "a little important." (In fact, there were several items for which no teacher reported the subject unimportant.) Thus, while teacher confidence was analyzed in the present study through use of IRT techniques, which place teachers' response on a scale, teacher importance was also analyzed in the study through use of latent class analysis techniques.

Use of planning materials

Similarly, teachers' responses to items relating to their use of materials when planning civic-related lessons did not appear to fall on a continuous scale. Instead, they seemed to reflect membership into one of several categories, or types, of activity usage. In particular, two categories of activity materials were suggested by the items. On the one hand, several items captured the extent to which teachers use official materials in planning. Teachers were asked to report on a scale of 1 ("not important") to 4 ("very important") each of the following sources:

- Official curricula or curricular guidelines or frameworks
- Official requirements or standards in civic education
- [Approved] textbooks.

On the other hand, the remaining items in this set reflected the extent to which teachers used outside, unofficial sources in their planning. With reference

to the same scale, teachers were asked to report how important each of the following sources was to their planning:

- Your own ideas as to what is important to know in civic education
- Original sources (such as constitutions, human rights declarations)
- Materials published by commercial companies, public institutes, or private foundations
- Self-produced materials
- The media (newspapers, magazines, television).

Measurement of teachers' practices

Item response theory

Items thought to measure teachers' beliefs and practices on a continuum (i.e., their preference for certain activities and their confidence in teaching certain subjects) were analyzed using a series of generalized partial credit models (GPCM: Muraki, 1992). One model was estimated for each dimension outlined above, for a total of four models (conventional activities, interactive activities, political teaching confidence, social teaching confidence). A GPCM analyzes five parameters, or characteristics, of items with four categories: the item's overall difficulty, three category thresholds (or the "relative difficulty" of one item response category as compared to the others), and the item's discrimination (or the extent to which the item responses separate the higher scorers on the overall scale from the lower scorers). Previous IRT analysis of student data from CivEd has relied on partial credit models, which take into account overall difficulty and category thresholds, but not discrimination (which is fixed to a value of 1: see Schulz & Sibberns, 2004). Here, the greater flexibility afforded to the model by estimating discrimination was thought to ensure a better fit for an analysis specifically tailored to civic-related teachers of 14-year-olds.

Item parameters were estimated using PARSCALE 4.1 (duToit, 2003). A "senate weight" was employed when estimating parameters within PARSCALE that took into account the national representation of students to whom the teacher was linked, and allowed each of the three countries to contribute equally to the estimation of item parameters, regardless of sample size. After an initial estimation of item parameters, additional analyses were conducted to ensure item fit and to remove items that did not fit the model.

Scale scores were then created in PARSCALE and set to have a mean of 10 and a *SD* of 2, in line with other attitudinal scales created from CivEd data. Finally, additional analyses with these new scales were conducted to ensure that each item functioned similarly as an indicator of its scale across all three countries.

Latent class analysis

For the analysis of items thought to indicate membership in particular categories (e.g., of importance beliefs or of planning habits), a series of latent class analyses were performed using MPlus 4.1 (Muthén & Muthén, 2006). In a latent class analysis, response patterns are analyzed, as they are thought to indicate membership in any number of categories, or "latent classes," requested by the researcher. The software estimates parameters for the conditional probability of choosing a particular response option given membership in one of the latent classes. In this analysis, there were four conditional probabilities (one for each possible response option) for each item in each class estimated. For example, if a set of two latent classes was measured by three items, then a total of $4 \times 2 \times 3$ or 24 conditional probabilities was reported. In addition, the proportion of respondents expected to fall into each latent class was estimated.

To begin, two latent classes were tested for each of the four models estimated (social importance, political importance, official planning sources, and unofficial planning sources). Models with additional classes were also tested as the data would allow, and then compared to the simpler models to see if there was a significant improvement in model fit. Once a model had been chosen for each construct, MPlus computed the conditional probabilities of each teacher falling into each class, and then assigned the teacher to the category with the greater membership probability. As with the IRT analysis, weights were included in the analysis that took into account national representation of students while allowing each of the three countries to contribute equally to the analysis.

Multilevel modeling

Once scale scores and class memberships had been calculated, they were incorporated into an analysis of students' civic engagement. In particular, a series of multilevel models allowed for student outcomes to be estimated by characteristics of teachers in the students'

schools. This estimate was conducted in a way that recognized that students were nested within schools and linked to teachers in those schools.

To take into account the fact that students were linked to more than one teacher in their school, the two units of analysis used in the multilevel model were schools and students, not teachers and students. Scores for all teachers included in the measurement analysis were considered; if data were available for more than one teacher in a school, then the scores were averaged together. In the case of the dichotomous latent class variables for importance of teaching topics and for the use of planning materials, the modal category membership was used. If teachers were equally split between categories, then the school was randomly assigned to one of the categories. In the United States, only one teacher per school was surveyed, which meant that the teacher's score became the score for the school. In this analysis, teachers can be thought of as "informants" of the teaching practices that occur within the school.

Certain schools did not survey a civic-related teacher who taught the tested class. In other cases, the teacher did teach the tested class but did not answer any questions in a section, making the computation of a scale score impossible. In these cases, data were imputed by taking the average value of the schools nearest to it in the data set as ordered by school identification code. Because school ID codes were ordered by strata (i.e., schools under the same control [public/private] or located in a similar region of the country appear near each other in the data set), this imputation strategy was deemed satisfactory. Less than 10% of schools in England or Australia needed to have data imputed; however, close to 50% of schools in the United States needed to have data imputed. Identical analyses were run in the United States for schools with non-imputed data and schools with the full set of imputed and non-imputed schools. No differences were found between these analyses in terms of significant effects. All together, students in 142 schools in Australia, 128 schools in England, and 124 schools in the United States were included in the sample (student $N = 3,331, 3,043$, and $2,811$, respectively).

Once school scores were created and missing data imputed, multilevel models were tested using hierarchical linear modeling (HLM) software

(Raudenbush, Bryk, Cheong, & Congdon, 2004). Because this present analysis is an initial, exploratory study of how these characteristics of teachers might influence students, only one set of teacher variables was tested as predictors at a time. All together, four sets of analyses were tested: one for teachers' use of activities, one for teachers' confidence, one for teachers' belief in importance of certain topics, and one for teachers' use of planning materials. Each analysis included in its model a series of student variables centered on the group mean, which means that these variables only reduced variability within (not between) groups. The inclusion of these student-level variables was designed to capture individual differences in civic development and differences due to home background, which reduced within-school variance (and, by extension, increased the intra-class correlation, or ICC, of these outcomes). It also made easier the detection of between-school effects due to differences among teachers. The student-level variables included were gender, home literacy (books in the home), discussion of political topics with parents, reading about national topics in the newspaper, trust in the government IRT scale (Torney-Purta et al., 2001), and the internal political efficacy IRT scale (Husfeldt, Barber, & Torney-Purta, 2005)

All student-level IRT scales other than internal political efficacy and trust in government were considered as outcomes. These scales are described in depth in the main summary of the IEA Civic Education Study of 14-year-olds (Torney-Purta et al., 2001) and in a summary of additional analyses conducted through CEDARS (Husfeldt et al., 2005). The outcomes of interest included the following:

- *Scales of civic knowledge* (total scale only)
- *Concepts of citizenship and government*: norms of conventional citizenship, norms of social movement citizenship, economy-related government responsibilities, society-related government responsibilities
- *Political and social attitudes*: attitudes toward ethnic minority, immigrant, and women's rights; attitudes toward anti-democratic groups; positive and protective attitudes toward the nation; trust in the media
- *Expected civic participation*: informed voting, non-voting political activities, community participation, illegal protest

- *Attitudes toward school*: perception of open classroom climate for discussion and confidence in the value of school participation.

Results

Measurement results

Item response theory analysis

Tables 1 to 4 report the results of the IRT analyses for the four sets of items whose responses were thought to fall on a continuous scale (conventional class activities, interactive class activities, confidence in teaching political topics, and confidence in teaching social topics, respectively). Each table reports the items' discrimination, difficulty, and three category parameters. A discrimination score of 0 for an item indicates that it does not differentiate at all among teachers with different confidence levels (or levels of activity usage); most items are expected to have a discrimination parameter around 1. Each category parameter is subtracted from the item's overall difficulty to obtain the difficulty threshold between two categories (e.g., category parameter "c1" is the threshold between the lowest and the second-lowest response options). The average scale difficulty is 0. Additional information about the analyses of model fit and differential item functioning is available in Barber (2006).

Although five items were thought to indicate teachers' use of conventional activities, the item relating to teachers choosing the issues to be covered displayed poor model fit and was removed from the analysis. With the exception of the item relating to students' use of textbooks, the remaining items had somewhat low discrimination, or ability to distinguish between teachers with high or low use of these activities more generally. However, the inclusion of these items was necessary to maintain adequate model fit, so they were retained in the analysis. Among the items least able to discriminate between teachers with different preferences were the frequency with which students worked on worksheets, and the frequency with which teachers lectured and students took notes. These items also varied in their difficulty. It took the least reported use of teacher-centered activities for teachers to report they questioned students regularly, and the most use for teachers to report that they regularly lectured to students while the students took notes.

Table 1: Item Parameters for Use of Conventional Activities Scale

	Discrimination	Difficulty	c1	c2	c3
Students Study Textbooks	1.14	-0.24	1.63	-0.34	-1.29
Students Work on Worksheets	0.54	-0.60	2.71	-0.52	-2.18
Teacher Asks Questions/Students Answer	0.83	-1.70	3.50	-0.95	-2.55
Teacher Lectures/Students Take Notes	0.65	0.51	1.98	-0.80	-1.18

Note: c1–c3 are category parameters.

Table 2: Item Parameters for Use of Interactive Classroom Activities Scale

	Discrimination	Difficulty	c1	c2	c3
Students Work in Groups on Different Topics	1.68	-0.41	1.75	-0.27	-1.48
Students Participate in Role-play and Simulations	1.65	0.08	1.71	-0.36	-1.35

Note: c1–c3 are category parameters.

Although three items were listed as potential indicators of teachers' use of interactive classroom activities, only two were retained after analysis of the items. The third, relating to the frequency with which students worked on projects, had considerably lower discrimination, and the model fit did not change when this less-discriminating item was removed. The items that remained were nearly equal in the extent to which they discriminated among teachers with different levels of use for interactive classroom activities. However, it took less use of such activities for teachers to report that students regularly worked in groups on projects than it did for them to report that students regularly participated in role-plays and simulations, as indicated by the difficulty of these items.

The item relating to teachers' confidence in covering the national constitution and political institutions was the most discriminating, while the item relating to elections was the least discriminating. Overall, the topic that required the least confidence (as determined by item difficulty) was teaching citizens' rights; the topic that required the most underlying confidence was the judicial system.

Items relating to confidence in teaching about equal opportunities for men and for women and about cultural differences were similar in their ability to discriminate among teachers with varying levels of confidence in teaching social topics. Confidence in teaching about the media was less discriminating. As determined by the items' difficulty, teaching about

Table 3: Item Parameters for Confidence in Teaching Political Topics Scale

	Discrimination	Difficulty	c1	c2	c3
Constitution/Institutions	2.36	-0.32	1.13	0.08	-1.21
Citizens' Rights	1.52	-0.76	1.48	0.10	-1.58
Conceptions of Democracy	1.55	-0.44	1.21	0.07	-1.28
Elections/Electoral System	1.36	-0.47	1.27	0.20	-1.48
Judicial System	1.42	-0.28	1.34	0.09	-1.43

Note: c1–c3 are category parameters.

Table 4: Item Parameters for Confidence in Teaching Social Topics Scale

	Discrimination	Difficulty	c1	c2	c3
Equal Opportunities	1.80	-1.00	1.35	0.19	-1.54
Cultural Differences	1.76	-0.77	1.33	0.15	-1.48
The Media	1.08	-1.10	1.31	0.34	-1.65

Note: c1–c3 are category parameters.

media also required the least underlying confidence for a teacher to rate it as a subject he or she was confident in, although expressing confidence in teaching about cultural differences was nearly as easy.

Latent class analysis

Analysis to determine categories of teachers based on the teachers' beliefs of the importance of civic-related topics and their use of various planning materials yielded mixed results. In the analyses of the importance of teaching political topics and the use of official planning materials, the data failed to support even a simple two-class model. This outcome suggests there was not enough variability among teachers in this sample in their views of whether studying political topics is important or in their use of official planning materials to create multiple categories of teachers based on these characteristics.

For the analysis of teachers' beliefs in the importance of teaching about social topics, the data were able to support a two-class model only. The inclusion of additional classes resulted in "complete separation," which meant that teachers who gave a particular response to an item had a 0% probability of falling into one of the classes. This indicates there was insufficient information to fully estimate all parts of the model. The conditional probabilities associated with this two-class model are summarized in Table 5. These suggest that the two classes represent teachers who found teaching about these topics somewhat more important and those who found them somewhat less important. Teachers who responded that they thought these activities were "not important," "somewhat important," or "important" had a higher probability of falling into class 1 than they did of falling into class 2. In contrast, teachers who viewed these activities as "very important" had higher probabilities of falling into class 2. The bolded proportions in Table 5 indicate which class membership was more likely given each individual item response.

A look at teachers' use of unofficial planning materials shows the data similarly supported a two-class model only; the specification of additional classes resulted in complete separation of the classes. The results from the two-class model are summarized in Table 6. Once again, the classes appear to discriminate between teachers who found these activities very important and those who found them relatively less important.

Multilevel models of students' civic engagement

Multilevel analyses revealed many ways in which teachers' practices and attitudes can influence students' civic development. All student-level IRT scales were tested against all sets of predictors (with the exception of internal political efficacy and trust in government, included as within-school controls in all models). However, there were no significant teacher-level predictors for students' attitudes toward anti-democratic groups or their protectionist attitudes toward their nation. Therefore, these outcomes do not appear in any other analyses. Country-by-country results for each of the six teacher predictors (scales or categories) are reported in Table 7. A report of this multilevel analysis follows, organized by teacher predictors.

Teachers' use of classroom activities

The results of multilevel analyses revealed that teachers influenced students the most through their choice of classroom activities. In particular, teachers' use of interactive activities, such as role-playing or group projects, had an extensive effect on students' knowledge, conceptualizations of citizenship and government responsibility, social attitudes, expected participation, and their perception of their school. These results are summarized in Table 8.

The impact of teachers' choice of classroom activities was most noticeable in Australia. Students who attended schools in which teachers regularly used interactive activities had higher civic knowledge than those attending schools in which teachers were less inclined to use these activities. These students also had a stronger attitude toward government responsibilities, both economic-related (ensuring that citizens can make a living) and society-related (e.g., ensuring that citizens have access to education and health care). Students with teachers who regularly engaged them in interactive activities had more positive attitudes toward the rights of women, immigrants, and ethnic minorities, and they were more likely to expect to vote and less likely to protest illegally. Finally, being exposed to opportunities to collaborate with one another in the classroom had a positive effect on students' individual attitudes toward school, increasing their perception of an open classroom climate and their confidence that their participation in school decision-making was valued.

Table 5: Conditional Probabilities for Two-Class Model of Importance of Teaching Social Topics

	Low (#1)	High (#2)
<i>Equal Opportunities</i>		
Not Important	1%	< 1%
Somewhat Important	2%	< 1%
Important	82%	23%
Very Important	15%	76%
<i>Cultural Differences</i>		
Not Important	—	—
Somewhat Important	3%	< 1%
Important	94%	6%
Very Important	3%	93%
<i>Media</i>		
Not Important	<1%	< 1%
Somewhat Important	4%	1%
Important	78%	31%
Very Important	18%	68%

Note: No teacher rated learning about cultural differences as “not important.”

Table 6: Conditional Probabilities for Importance of Unofficial Planning Materials

	Low (#1)	High (#2)
<i>Your Own Ideas</i>		
Not Important	2%	2%
Less Important	8%	20%
Important	54%	74%
Very Important	37%	4%
<i>Original Sources</i>		
Not Important	1%	3%
Less Important	13%	17%
Important	35%	64%
Very Important	51%	16%
<i>Outside Materials</i>		
Not Important	5%	8%
Less Important	36%	49%
Important	44%	42%
Very Important	15%	< 1%
<i>Your Own Materials</i>		
Not Important	< 1%	2%
Less Important	5%	29%
Important	58%	67%
Very Important	35%	3%
<i>Media</i>		
Not Important	< 1%	< 1%
Less Important	< 1%	9%
Important	25%	73%
Very Important	75%	11%

Table 7: Descriptive Statistics (Mean and SD or percentage) by Country for New Teacher Variables

	Australia	England	United States
Use of Conventional Classroom Activities	9.9 (1.9)	9.4 (2.0)	10.3 (2.0)
Use of Interactive Classroom Activities	10.1 (1.8)	9.7 (1.9)	10.3 (2.2)
Confidence in Teaching Political Topics	10.3 (1.8)	9.1 (1.9)	10.8 (1.8)
Confidence in Teaching Social Topics	10.1 (2.0)	9.4 (2.0)	10.5 (1.8)
High Importance Given to Social Topics	49%	53%	58%
High Use of Unofficial Planning Materials	51%	40%	43%

Table 8: Summary of the Effects of Classroom Activities on Civic Engagement in Three Countries

	Australia		England		United States	
	Convention	Interactive	Convention	Interactive	Convention	Interactive
Knowledge	n.s.	1.32* (.55)	n.s.	n.s.	n.s.	n.s.
Social Citizen	n.s.	n.s.	-0.04+ (.04)	n.s.	0.06+ (.04)	n.s.
Econ. Gov.	n.s.	0.06* (.03)	n.s.	n.s.	n.s.	n.s.
Society Gov.	n.s.	0.09* (.04)	n.s.	n.s.	n.s.	n.s.
Imm. Rights	n.s.	0.16* (.05)	n.s.	-0.19* (.07)	n.s.	n.s.
Wom. Rights	n.s.	0.12* (.04)	n.s.	n.s.	n.s.	n.s.
Minor. Rights	n.s.	0.14* (.08)	n.s.	n.s.	n.s.	n.s.
Inf. Voting	n.s.	0.10* (.03)	n.s.	n.s.	n.s.	n.s.
Illegal Protest	n.s.	-0.08* (.04)	n.s.	n.s.	n.s.	n.s.
Class Climate	n.s.	0.14* (.05)	n.s.	n.s.	-0.08+ (.09)	-0.09* (.23)
Confidence in School Part.	n.s.	0.11* (.04)	n.s.	n.s.	n.s.	-0.08* (.03)

Notes: HLM regression coefficients reported.

Standard errors are in parentheses.

Civic knowledge student mean = 100, SD = 20; all other student mean = 10, SD = 2.

+ $p < .10$, * $p < .05$.

The effects of teacher activities on students' civic development were less apparent in England and the United States. However, unlike the situation in Australia, the use of conventional classroom activities, such as working from a textbook and lecturing, did seem to significantly affect students. In England, students with teachers who relied more heavily on conventional activities in the classroom had less strong beliefs about the importance of social-movement citizenship (e.g., the importance of standing up for human rights) and less positive attitudes toward immigrants. In contrast, in the United States, students with teachers who made greater use of conventional activities had stronger beliefs about this type of citizenship. In addition, United States students in schools where teachers relied more upon this conventional, teacher-centered form of instruction perceived their classrooms as less open to discussion. However, United States students with teachers who relied more heavily on interactive classroom activities similarly had less positive perceptions of the openness

of their classrooms to discussion, and also had less confidence in the value of their participation (opposite to the effect in Australia).

Teachers' confidence in teaching

Teachers' confidence in teaching certain civic-related topics had far less of an impact on students' civic development. In fact, there were no significant relationships between confidence and students' civic outcomes in the United States. In Australia and England, confidence in both political and social areas was tested for their respective effects on student outcomes; however, only teachers' social confidence was significant. These results are reported in Table 9.

As with the previous analysis, teachers in Australia had the furthest-ranging effects on their students. Students with teachers who were confident in teaching social topics had stronger beliefs in the importance of both conventional citizenship and social-movement citizenship. At the same time, they also expressed less

Table 9: Summary of the Effects of Confidence in Teaching Social Topics in Australia and England

	Australia	England
Social Citizen	0.03* (.03)	n.s.
Conventional Citizen	0.06* (.03)	n.s.
Trust in the Media	-0.05* (.03)	n.s.
Illegal Protest	n.s.	0.11* (.04)

Notes: HLM regression coefficients reported. Standard errors are in parentheses.

All other students mean, 10, *SD* = 2.

Analysis controls for teachers' political confidence.

**p* < .05.

Table 10: Summary of the Effects of Importance of Teaching Social Topics in Australia and England

	Australia	England
Knowledge	-4.47* (1.64)	n.s.
Social Citizen	0.19* (.08)	0.16* (.08)
Conventional Citizen	0.27* (.09)	0.12+ (.07)
Pos. Attitudes Toward Nation	n.s.	0.29* (.10)
Minority Rights	n.s.	0.27* (.14)
Political Activities	n.s.	0.19* (.11)
Classroom Climate	-0.28* (.12)	n.s.

Notes: HLM regression coefficients reported.

Standard errors are in parentheses.

Civic knowledge student mean = 100, *SD* = 20; all other students mean = 10, *SD* = 2.

+ *p* < .10, **p* < .05.

trust in the news media. In England, students with teachers who had confidence in teaching about social topics were more likely to expect to participate in illegal protest.

Teachers' perception of the importance of teaching social topics

This analysis of teaching about social topics focused on the gap in students' achievement between those with teachers who found teaching about these topics more or less important. Again, no significant results were found in the United States; significant results for Australia and England are summarized in Table 10.

In both Australia and England, students with teachers who believed that teaching about social issues is important had stronger beliefs that both conventional and social-movement activities are important for citizenship. In England, this view of social topics as important had additional positive effects on students' attitudes toward ethnic minorities, on positive attitudes toward their own nation, and on their expectations of participating in conventional political activities not related to voting (e.g., joining a political party

or writing letters). In Australia, in contrast, students with teachers who considered teaching these topics important actually viewed their classes as less open to discussion, and also had significantly lower scores on the CivEd test of civic knowledge.

Teachers' use of unofficial materials

The final analysis, summarized in Table 11, explored whether differences in teachers' preparation led to differences in students' civic development. Australian students with teachers who reported that unofficial materials were important for planning civic-related subjects had more positive social attitudes than had students whose teachers found these materials less important. In particular, the students' attitudes toward political rights of immigrants and ethnic minorities were more positive. These students also had stronger expectations of participating in community-related civic activities in the near future. In contrast, English students whose teachers found these unofficial materials important had stronger beliefs that government should take on various society-building responsibilities.

Table 11: Summary of the Effects of Using Unofficial Planning Materials in Australia and England

	Australia	England
Social Government	n.s.	0.21* (.10)
Immigrant Rights	0.34* (.15)	n.s.
Minority Rights	0.27* (.09)	n.s.
Community Participation	0.21* (.10)	n.s.

Notes: HLM regression coefficients reported.

Standard errors are in parentheses.

All other student mean = 10, SD = 2.

* $p < .05$.

Discussion

The analysis presented in this paper demonstrates how a combination of IRT and latent class analytic techniques can be applied to the teacher survey data collected as part of the IEA Civic Education Study of 1999. Scales of teachers' confidence captured individual differences among teachers as to how capable they felt in teaching various civic-related topics. Scales of teachers' activity usage captured similar differences in the frequency with which the teachers used different learning approaches in the classroom. In contrast, latent class analysis techniques revealed that teachers could be grouped into one of two categories based on the importance they placed on students learning about social topics and the importance that they placed on using unofficial planning materials. Although these categories appeared to discriminate between those teachers who rated these subjects and materials as "important" from those who rated them as "very important," multilevel analyses revealed that the students of teachers in the different groups differed significantly in their knowledge, attitudes, and expected behaviors.

The use of these variables in multilevel analyses also indicates that teachers can have a wide-reaching influence on students' civic development. This influence speaks to the importance of considering multiple dimensions of civic engagement. When discussing teacher effects, one is tempted to focus almost exclusively on how characteristics of teachers influence what students know. While the use of interactive class activities positively related to higher civic knowledge among Australian students, the majority of other effects in all three countries related more to students' concepts of citizenship, their social attitudes, and their opinions about school. School and

teacher effects on students' attitudes can be difficult to explore among students because there is so much variability within schools. By adding within-school controls to account for some of that variability, however, we can start to form a clearer picture of how teachers influence their students. The analyses also demonstrate the importance of considering teachers' attitudes toward social topics, not just those topics most associated with "civics," such as government and politics. In particular, teachers' beliefs as related to social topics had a stronger effect than their beliefs about political topics (in line with findings from cross-national studies by Fouts & Lee, 2005).

The multilevel analyses also illustrate how differently teachers can influence students in Australia, England, and the United States. The majority of significant teacher effects in the analyses were found from looking at the effects of Australian teachers. Previous research using smaller samples supports this finding. Prior (1999) found that Australian social studies teachers valued active citizenship, with a focus on knowledge of current events and community consensus, despite a formal curriculum that advocated teaching about government structure and history. It is not surprising, then, to see that the Australian teachers who had especially high beliefs that social topics are important had students with stronger conceptualizations of both conventional citizenship (including the following of current events) and social-movement citizenship, and that those teachers who engaged their students in more interactive activities increased their students' positive attitudes toward the rights of others in their communities. In addition, the more traditional focus of the formal curriculum may have led many Australian teachers to use other planning materials, which could better help them meet their goals of encouraging active citizenship and community participation.

In England, results from focus groups have found that teachers in this country often center on the role of citizens as skeptics of government, or on the role of government in people's lives (Arnot, Araujo, Deliyanni-Kouimtzi, Rowe, & Tome, 1996). This finding is reflected in the present study in the positive relationship between teachers' beliefs that a subject is important and active citizenship (either conventional or social-movement oriented, and also in active participation in political activities), and in the relationship between the use of additional planning materials and stronger beliefs in society-related government responsibilities. In the most extreme cases of fostering skepticism, teachers who were especially confident in teaching these subjects had students who planned to protest illegally. In contrast, teachers who used the most traditional activities appeared to discourage social-movement citizenship norms. Other results, however, appear to differ from this general finding and are worthy of further study, including the effects of teachers' importance beliefs on increased patriotism, and the *negative* effects of interactive classroom activities on immigrant attitudes.

Fewer significant results were found when observing the United States. Those that we did find suggest that the use of interactive classroom activities takes away opportunities for open classroom discussion or for building confidence in school participation. Indeed, another teacher survey item capturing the frequency of political discussion did not fit on the interactive activities scale, perhaps because in the United States

such structured activities leave less opportunity for open discussion (see Schulz & Sibberns, 2004). Further analysis of teachers' use of discussion, and how it relates to other hands-on activities, is warranted.

The lack of significant effects in the United States might also be attributed to two other factors. First, fewer teachers in the United States than in England and Australia completed this survey, and more information from United States teachers may lead to more powerful analyses in which additional effects can be detected. Second, the United States teachers were more likely than the teachers in England and Australia to report higher confidence in each type of activity, and to report frequent use of both sorts of classroom activities. Teachers in the United States may have been thinking more about social desirability than were the teachers in the other countries, leading to a less accurate portrayal of their actual classroom attitudes and behaviors.

In summary, the teacher variables formed by IRT and latent class analyses give a great deal of information about how teachers can influence students' civic development. Additional analyses should work to follow up these findings in individual countries in greater depth, and consideration should also be given to creating new variables that can be generalized across all countries in the IEA CivEd. By considering teachers in a wider variety of social and policy contexts, we can gain a more complete understanding of how teachers can influence students' development into active, engaged citizens.

References

- Amadeo, J., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement: an IEA study of upper-secondary students in sixteen countries*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Arnot, M., Araujo, H., Deliyanni-Kouimtzi, K., Rowe, G., & Tome, A. (1996). Teachers, gender and the discourses of citizenship. *International Studies in Sociology of Education*, 6, 3–35.
- Barber, C. (2006). *An item response theory analysis of teachers' attitudes toward the civic education of 14-year-olds in three countries*. Unpublished Master's seminar paper, Department of Measurement, Statistics, and Evaluation, University of Maryland, College Park.
- Bronfenbrenner, U. (1988). Interacting systems in human development. In N. Bolger, C. Caspi, G. Downey, & M. Moenhouse (Eds.), *Persons in context: Developmental processes* (pp. 25–50). Cambridge, England: Cambridge University Press.
- Dayton, C. M. (1998). Latent class scaling analysis. *Quantitative applications in the social sciences* (Vol. 126). Thousand Oaks, CA: Sage.
- duToit, M. (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood, IL: Scientific Software International.
- Fouts, J., & Lee, W. (2005). *Education and social citizenship: Perspectives of teachers in USA, Australia, England, Russia, and China*. Hong Kong: Hong Kong University Press.

- Gonzales, M. H., Riedel, E., Avery, P. G., & Sullivan, J. L. (2001). Rights and obligations in civic education: A content analysis of the "National Standards for Civics and Government". *Theory and Research in Social Education*, 29(1), 109–128.
- Gonzales, M. H., Riedel, E., Williamson, I., Avery, P. G., Sullivan, J. L., & Bos, A. (2004). Variations of citizenship education: A content analysis of rights, obligations, and participation concepts in high school civic textbooks. *Theory and Research in Social Education*, 32(3), 301–325.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.
- Husfeldt, V., & Barber, C. (2005, September). *Teachers' concepts of civic education in relation to students' attitudes and engagement*. Paper presented at the Third European Consortium of Political Research Conference, Section on Political Sciences and Citizenship in the Classroom, Budapest, Hungary.
- Husfeldt, V., Barber, C., & Torney-Purta, J. (2005). *Students' social attitudes and expected political participation: New scales in the enhanced database of the IEA Civic Education Study*. College Park, MD: Civic Education Data and Researcher Services.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, England: Cambridge University Press.
- Losito, B., & Mintrop, H. (2001). The teaching of civic education. In J. Torney-Purta, R. Lehmann, H. Oswald, & W. Schulz, *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen* (pp. 158–173). Amsterdam: International Association for the Evaluation of Educational Achievement.
- Lubke, G. (2006). *Discriminating between categorical and continuous latent variables*. Paper presented at the Center for Integrated Latent Variable Research Conference on Mixture Models in Latent Variable Research, College Park, MD.
- Masters, G. N., & Wright, B. D. (1997). The partial credit model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 101–122). New York/Berlin/Heidelberg: Springer.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159–176.
- Muthén, L. K., & Muthén, B. O. (2006). *MPlus user's guide* (4th ed.). Los Angeles, CA: Muthén & Muthén.
- Pajares, F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62, 307–332.
- Prior, W. (1999). What it means to be a "good citizen" in Australia: Perceptions of teachers, students, and parents. *Theory and Research in Social Education*, 27, 215–248.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.
- Schulz, W., & Sibberns, H. (Eds.). (2004). *IEA Civic Education Study technical report*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15, 4–14.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361–370.
- Torney-Purta, J., Barber, C., & Wilkenfeld, B. (2007). Latino adolescents' civic development in the United States: Research results from the IEA Civic Education Study. *Journal of Youth and Adolescence*, 36, 111–125.
- Torney-Purta, J., Barber, C., & Richardson, W. (2005). *How teachers' preparation relates to students' civic knowledge and engagement in the United States: Analysis from the IEA Civic Education Study* (fact sheet). College Park, MD: Center for Information and Research on Civic Learning and Engagement, University of Maryland, College Park.
- Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Torney-Purta, J., Richardson, W., & Barber, C. (2005). Teachers' experience and sense of confidence in relation to students' civic understanding and engagement across countries. *International Journal of Citizenship and Teacher Education*, 1(1), 32–57.
- Torney-Purta, J., Schwille, J., & Amadeo, J. (1999). *Civic education across countries: Twenty-four national case studies from the IEA Civic Education project*. Amsterdam: International Association for the Evaluation of Educational Achievement.

Relationships between parental factors and children's reading behaviors and attitudes: Results from the PIRLS 2005 field test in Taiwan

Chia-Hui Chiu

*Department of Education,
National Chengchi University, Taiwan*

Hwa-Wei Ko

*Graduate Institute of Learning and Instruction
National Central University, Taiwan*

Abstract

The purposes of this study were to investigate the relationship between children's reading behaviors and attitudes and the relationship between parental factors and children's reading behaviors and attitudes. Data for this study came from the PIRLS 2005 field test. The participants were 1,184 Grade 4 students and their parents in Taiwan. The PIRLS student questionnaire and Learning to Read Survey were used for the study. Results revealed a significant relationship between children's positive reading behaviors and attitudes, and between

negative reading attitudes and negative self-efficacy. A rich reading environment, including number of books in the home, and positive parental reading attitudes predict children's positive reading behaviors and attitudes, and vice versa. However, different reading focuses may differentiate children with positive or with negative attitudes toward reading. Parents' awareness of difficulty in fostering reading predicts children's negative attitudes and self-efficacy. This consideration is discussed in the paper.

Introduction

Numerous studies on reading focus on examining the cognitive and social factors related to children's reading abilities, such as phonological awareness (Adams, 1990; MacDonald & Cornwall, 1995), and to children's socioeconomic backgrounds (Bradley, Corwyn, Pipes McAdoo, & García Coll, 2001; Noble, Farah, & McCandliss, 2006). The contribution of affective factors to reading achievement has also received emphasis (Barnett & Irwin, 1994; Matthewson, 1994; McKenna, Stratton, Grindler, & Jenkins, 1995; Paris, Wasik, & Turner, 1991). Affective factors are commonly identified as reading attitudes (Kush, Watkins, & Brookhart, 2005), and attitudes are regarded as the continuum of positive to negative feelings toward reading (McKenna, Kear, & Ellsworth, 1995). Reading attitudes are important components in the reading curriculum (Quinn & Jadav, 1987), and children's development of these attitudes aligns with children's engagement in sustained reading throughout their lives (Cullinan, 1987).

Among the socioeconomic background factors influencing reading are those related to the children's parents, who play a critical role in the literacy

development of their children (Baker, 2003). Parents influence not only children's reading achievement (Fitton & Gredler, 1996; Kush et al., 2005; Lee & Barro, 1998; Olofson & Niedersoe, 1999; Sonnenschein, Brody, & Munsterman, 1996; Weinberger, 1996), but also children's reading motivation (Baker, Scher, & Mackler, 1997; Bratern, Lie, Andreassen, & Olaussen, 1999; DeBaryshe, 1995) and attitudes (Baker & Scher, 2002; Greaney & Hegarty, 1987; Rowe, 1991). In recognition of the importance of parents on children's reading ability, the present study explored the relationship between parents' and children's reading behaviors and attitudes by analyzing data from the field test conducted in Taiwan in 2005 for IEA's Progress in International Reading Literacy Study (PIRLS) 2006.

In most studies exploring the relationship between parents and their children's reading attitudes, the parental factors include reading materials at home, parental reading behaviors, the frequency of reading to the child, and parental beliefs. All these factors contribute to the development of positive reading attitudes among children (Greaney & Hegarty, 1987). Baker and Scher (2002) examined children's

motivation for reading in relation to parental beliefs and home literacy experiences. Their sample included 56 six-year-olds. The results revealed that the beginning readers in general had positive views about reading, and that parental identification of pleasure as a reason for reading predicted children's motivation for reading. Parents' beliefs about reading are also associated with differences in children's home reading activities, motivation, and achievement (Baker, Serpell, & Sonnenschein, 1995).

Among all family background variables, level of parental education is a definite factor in determining children's reading abilities and achievement (Baker et al., 1995; Lee & Barro, 1998). Baker et al. (1995) found that parents with higher levels of schooling placed greater value on education and thus provided more materials and school-related activities for their children than did parents with lower levels of schooling. Weigel, Martin, and Bennett (2006) found that maternal education level significantly predicted literacy belief, and that facilitative mothers were more likely to have higher education levels. However, Psacharopoulos and Woodhall (1985) argued that while the education level of both parents was important, the father's was more so. To sum up, both father's and mother's education influence children's reading.

The literature additionally reveals that reading attitudes are associated with many other variables, including reading achievement, reading behaviors, and gender. Successful readers normally possess more positive reading attitudes than poor readers (Wigfield & Asher, 1984). Kush et al. (2005) found that the relationship between reading attitudes and achievement becomes more closely linked over time. There is also a significant positive relationship between reading behaviors and attitudes (Ley, Schaer, & Dismukes, 1994). Gender plays a role as well. A number of studies establish that girls possess more positive attitudes toward reading than do boys (Abdulrahim, 1997; Diamond & Onwuegbuzie, 2001; Kazelskis, Thames, & Reeves, 2004; Rea, Romine, McKenna, & Griffin, 1997). Yet, regardless of age, gender, and family socioeconomic status (SES), reading activity at home has significant positive influences on students' reading achievement, attitudes toward reading, and attentiveness in the classroom (Rowe, 1991).

Although there are extensive studies on reading, there is a problem with the definitions of items

researched. For instance, the items on motivation are similar to the items on attitudes. Sainsbury and Schagen (2004) argue that attitude toward reading is part of motivation to read. In this study, we adopted exploratory factor analysis to regroup questionnaire items and rename factors in order to provide a more robust definition of the variables investigated. In sum, the purposes of this study were to investigate:

1. The relationship between children's reading behaviors and attitudes; and
2. The relationship between all the parental factors and children's reading behaviors and attitudes.

Method

Participants

Data for this study came from the PIRLS 2006 field test, conducted in 2005. All together, 1,590 participants (comprising students paired with their parents) were sampled in Taiwan. Exclusion of participants (either students or parents) who were absent and did not complete the questionnaires resulted in 1,418 pairs of Grade 4 students (719 girls and 699 boys) and their parents from 25 elementary schools in Taiwan.

Measures

Children's reading behaviors and attitudes

The PIRLS student questionnaire was designed to measure children's classroom experiences and reading for homework, self-perception of their reading ability, their attitudes toward reading, and their out-of-school reading habits, computer use, and home literacy resources.

One of the purposes of the study was to investigate the correlation between parental factors and children's reading behaviors and attitudes. The items about the classroom experience and reading for homework and computer use in the survey were therefore omitted. To gain a clearer definition of variables, we performed an exploratory factor analysis, using a principle components extraction with an oblique rotation.

The student questionnaire contained 23 items about children's reading behaviors and attitudes. The item numbers were 3a–3g, 5a, 13a–13k, and 14a–14d. The estimation yielded four factors with eigenvalues larger than 1.00; these accounted for 48.22% of the total variance (see Table 1).

As shown in Table 1, Factor 1 was loaded mainly with items designed to tap "reading interaction," that is,

interactive reading activities with family members and friends. Factor 2, “negative reading attitudes,” mainly tapped children who disliked school-related reading and reading assignments. Factor 3 related to “negative self-efficacy,” which is children’s self-evaluation of their reading ability. According to Wigfield and Guthrie (1997), self-efficacy is one category of motivation. However, in this study, we took the name *self-efficacy* rather than *motivation* in order to demonstrate a finer aspect of reading. Factor 4 appeared to focus primarily on “positive reading attitudes,” that is, active engagement with and expressions of pleasure about reading.

Four indicators constituted reading interaction. Items were rated on a four-point Likert-type scale, with answers ranging from 1 (never or almost never) to 4 (every day or almost every day). The remaining three factors with items were all rated on a four-point Likert-type scale, with answers ranging from 1 (disagree a lot) to 4 (agree a lot).

Parental factors: PIRLS questionnaire for parents

The PIRLS Learning to Read Survey was adapted and used for this study. The survey investigated child–parent literacy interactions, home literacy resources, parents’ reading habits and attitudes, and home–school connections. The survey also included items that were cultural adaptations of the international questions about parental practices and beliefs in fostering children’s reading ability.

Because items relating to parental factors were not clearly identified in the survey, an exploratory factor analysis was performed. Here, a principle components extraction with an oblique rotation was used. Nineteen items relevant to parental factors emerged. The item numbers were 12, 13a–13e, 14a–14e, 15, 16, 17 (two columns), and 20a–20d. The estimation yielded six factors with eigenvalues larger than 1.00; these accounted for 60.83% of the total variance (see Table 2).

As shown in Table 2, Factor 1 mainly loaded with items designed to tap the “environment factor,”

Table 1: Oblique Rotated Four-Factor Solution of the Structure of Children’s Reading Behaviors and Attitudes, and the Items of Each Factor

Item Description	Factor			
	1	2	3	4
<i>Reading interaction</i>	.55			
3a. I read aloud to someone at home	.53			
3b. I listen to someone at home read aloud to me	.54			
3c. I talk with my friends about what I am reading	.59			
3d. I talk with my family about what I am reading				
<i>Positive reading attitudes</i>				
13e. I need to read well for my future				.50
13f. I enjoy reading				.50
13h. I like to read a book that is chosen by myself				.56
<i>Negative reading attitudes</i>				
13g. The textbooks are boring		.69		
13i. I hate to write the reading reports		.62		
13k. My classmates do not like to read		.57		
<i>Negative self-efficacy</i>				
14b. I do not read as well as other students in my class			.73	
14d. I read slower than students in my class			.74	
Eigenvalue	5.39	1.98	1.65	1.59
% of variance	24.51	9.00	7.49	7.22

Note: Factor loadings of a magnitude less than .50 are omitted.

Table 2: Oblique Rotated Three-Factor Solution of the Structure of Parental Reading Behaviors and Attitudes, and the Items of Each Factor

Item	Factor					
	1	2	3	4	5	6
<i>Purposeful reading behaviors</i>						
13a. When you are at home, how often do you read for work?		.58				
13b. When you are at home, how often do you read for enjoyment?		.72				
13c. When you are at home, how often do you read to get news?		.71				
13d. When you are at home, how often do you read for your education/school?		.61				
<i>Evaluation of reading</i>						
14a. I read only if I have to			.81			
14d. I read only if I need information			.78			
20c. How much do you agree with the following statements about developing child's reading abilities:						
a. It is easier to develop reading abilities than math abilities						
b. It is less difficult to develop reading abilities than math abilities						
c. It is less expensive to develop reading abilities than math abilities			.51			
d. It is more important to develop math abilities than reading abilities						
<i>Parents' reading attitudes</i>						
14b. I like talking about books with other people				.83		
14c. I like to spend my spare time reading				.79		
14e. Reading is an important activity in my home				.73		
<i>Environment factor</i>						
15. The number of books in your home	.59					
16. The number of children's books in your home	.56					
17. The level of education of the child's father	.84					
17. The level of education of the child's mother	.87					
<i>Ways of fostering reading ability</i>						
20a. In addition to reading, what kind of activities does your child spend the most time doing after school:						
a. going to talent cram school						
b. going to academic cram school					.72	
c. playing video or computer games						
d. watching TV						
20b. What do you do to enrich your child's reading abilities:						
a. buy story tapes or CDs						
b. buy extracurricular books						
c. buy language practice books					.66	
d. buy learning software						
e. tell stories						

Table 2 (contd.): Oblique Rotated Three-Factor Solution of the Structure of Parental Reading Behaviors and Attitudes, and the Items of Each Factor

Item	Factor					
	1	2	3	4	5	6
<i>The awareness of difficulty in fostering reading ability</i>						
20d. Do you experience the following problems when trying to enhance your child's reading abilities:						
a. The family lacks extracurricular reading materials						
b. Nobody knows how to teach the child to read in the family						.74
c. The family do not have time to teach the child to read						
d. My child has too much homework to do; she/he does not have time to read extracurricular books						
e. My child does not like reading						
Eigenvalue	4.30	2.15	1.82	1.20	1.09	1.00
% of variance	22.64	11.30	9.57	6.32	5.74	5.27

Note: Factor loadings of a magnitude less than .50 are omitted.

which included the number of books at home and the education levels of parents. Factor 2, named "purposeful reading behaviors," concerned the frequency of purposeful reading activities. Factor 3, which covered "evaluation of reading," included parents who did not value reading as highly as they did mathematics and who saw reading as a tool to get information only. Factor 4 tapped "parents' attitudes toward reading." It shows the extent to which parents enjoyed reading. Factor 5, "ways of fostering reading ability," reported the actions taken by parents to enrich their children's reading abilities. Factor 6 appeared to focus primarily on "the awareness of difficulty in fostering reading abilities." It concerns the extent to which parents know what resources they can use to guide their children's reading at home.

All indicators constituted to the six factor items were rated on a four-point Likert-type scale in which the answers ranged from 1 (never or almost never, disagree a lot) to 4 (every day or almost every day, agree a lot). There were also five-point and six-point Likert-type scales ranging from 1 (the least) to 5 or 6 (the most).

Results

The children's reading behaviors and attitudes

Table 3 provides descriptive statistics of the reading behaviors and attitudes of the children who took part in the study. We examined a one-way ANOVA to ascertain the gender effect on the children's reading

behaviors and attitudes. It revealed a significant gender difference on reading interaction, $F(1, 1182) = 18.00$, $MSe = .54$, $p < .001$. Girls showed more reading interaction than did boys. Boys had more negative reading attitudes than did girls, $F(1, 1182) = 8.98$, $MSe = .38$, $p < .01$, and girls had more positive reading attitudes than did boys, $F(1, 1182) = 55.86$, $MSe = .42$, $p < .001$. However, there was no significant difference on negative self-efficacy. As expected, there was, overall, a significant gender effect in favor of girls.

To understand the relationship between children's reading behaviors and attitudes, we conducted correlation analyses among the variables. The results showed a significant correlation between children's reading behaviors and children's positive reading attitudes ($r = .31$, $p < .001$), a significant correlation between children's self-efficacy and children's positive reading attitudes ($r = -.12$, $p < .001$), and between children's self-efficacy and children's negative reading attitudes ($r = .13$, $p < .001$). There was a small but statistically significant relationship between children's negative reading attitudes and children's reading interaction ($r = .07$, $p < .05$). There was no correlation between children's negative reading attitudes and positive reading attitudes and between the children's reading interaction and self-efficacy (see Table 4). In short, those children who had positive reading attitudes enjoyed interactive reading activities, such as reading to others. They also showed less negative self-evaluation of reading abilities. However, children

Table 3: Descriptive Statistics for the Reading Behaviors and Attitudes of Girls and Boys

Variables	Girls (N = 619)		Boys (N = 565)	
	M	SD	M	SD
1. Reading Interaction	2.20	.71	2.03	.75
2. Negative Reading Attitudes	2.31	.55	2.41	.67
3. Negative Self-Efficacy	2.06	.80	2.13	.88
4. Positive Reading Attitudes	3.57	.57	3.29	.74

Table 4: Correlations between Children's Reading Behaviors and Attitudes

Items	Reading interaction	Negative reading attitudes	Negative self-efficacy	Positive attitudes
1. Reading Interaction	–	.07*	-.05	.31***
2. Negative Reading Attitudes		–	.13***	.00048
3. Negative Self-Efficacy			–	-.12***
4. Positive Reading Attitudes				–

Note: * $p < .05$. ** $p < .01$. *** $p < .001$.

who disliked school-related reading work, such as reading textbooks and doing assignments, showed negative self-efficacy. Interestingly, the children who had negative reading attitudes nonetheless somewhat enjoyed interactive reading.

All parental factors and children's reading

Children's reading interaction

We next carried out a multiple regression so that we could determine predictors of children's reading interaction. The factors considered were the environment factor, purposeful reading behaviors, evaluation of reading, parental reading attitudes, ways of fostering reading ability, and parents' awareness of difficulty they had fostering their children's reading abilities. Table 5 provides the beta (β) coefficients and t values of the predictor variables. As shown in Analysis 1 in the table, the regression equation was significant. The significant positive predictors were the environment factor and parental reading attitudes. There was no significant negative predictor.

Children's negative reading attitudes

We then conducted a multiple regression to determine predictors of children's negative reading attitudes. The factors considered were the same as the children's reading interaction. As shown in Table 5, Analysis 2, the regression equation was significant. The significant

predictors were evaluation of reading and the awareness of difficulty in fostering reading abilities.

Children's negative self-efficacy

The next multiple regression we conducted was to determine predictors of children's negative self-efficacy. The factors considered were the same as the children's reading interaction. As shown in Table 5, Analysis 3, the regression equation was significant. The significant predictors were awareness of difficulty in fostering reading abilities and evaluation of reading. A significant negative predictor was the environment factor.

Children's positive reading attitudes

We conducted this multiple regression to determine predictors of children's positive reading attitudes. The factors considered were the same as the children's reading interaction. As shown in Table 5, Analysis 4, the regression equation was significant. The significant positive predictors were purposeful reading behaviors, the environment factor, and parental reading attitudes. The significant negative predictor was parents' awareness of difficulty fostering their children's reading abilities.

In sum, a home environment rich in reading-related resources predicts children's positive reading behaviors and attitudes. Parents who find it difficult to guide their children's reading contribute to their children's

Table 5: Regression Testing of Specific Predictors of Children's Reading Behaviors and Attitudes

Analysis	β	SE	R^2	t	F
Analysis 1			.036		7.40***
<i>Children's reading interaction</i>					
Environment factor	.08	.03		2.66**	
Purposeful reading behaviors	.06	.03		1.85	
Evaluation of reading	-.03	.03		-0.99	
Parents' reading attitudes	.10	.03		3.01**	
Ways of fostering reading ability	-.03	.03		-0.89	
Awareness of difficulty fostering reading abilities	-.02	.03		-0.71	
Analysis 2			.020		3.96***
<i>Children's negative reading attitudes</i>					
Environment factor	.08	.03		0.90	
Purposeful reading behaviors	.03	.03		0.95	
Evaluation of reading	.09	.03		2.58**	
Parents' reading attitudes	.08	.03		0.90	
Ways of fostering reading ability	-.06	.03		-1.81	
Awareness of difficulty fostering reading abilities	.09	.03		2.88**	
Analysis 3			.036		7.30***
<i>Negative self-efficacy</i>					
Environment factor	-.11	.03		-3.38***	
Purposeful reading behaviors	-.03	.03		-0.95	
Evaluation of reading	.07	.03		2.33*	
Parents' reading attitudes	-.004	.03		-0.14	
Ways of fostering reading ability	.01	.03		0.41	
Awareness of difficulty fostering reading abilities	.10	.03		3.42***	
Analysis 4			.051		6.06***
<i>Children's positive reading attitudes</i>					
Environment factor	.10	.03		3.33***	
Purposeful reading behaviors	.08	.03		2.69**	
Evaluation of reading	-.05	.03		-1.52	
Parents' reading attitudes	.07	.03		2.12*	
Ways of fostering reading ability	-.02	.03		-0.52	
Awareness of difficulty fostering reading abilities	-.07	.03		-2.29*	

Note: * $p < .05$. ** $p < .01$. *** $p < .001$.

negative reading attitudes and negative self-efficacy. However, a home environment with few reading-related resources also predicts children's negative self-efficacy. It seems that parents who are less educated and provide fewer books at home and parents who are aware of difficulties in handling children's reading have children who perceive themselves not well skilled in reading and who report a low level of enjoyment of reading.

The home environment factor

As mentioned above, the environment factor plays an important role in children's reading. This factor

includes parental education and the number of books at home (both children's books and other books). However, in many studies, parental education and the number of books at home predict children's reading achievement, attitudes, and motivation respectively. In this study, parents' education and books at home were grouped together as an environment factor. The correlations between parental education and books at home were significant (see Table 6).

We also carried out an additional analysis of the influence of parental education on children's reading. After pairing the mother's and father's education levels, we were left with four conditions:

Table 6: Correlations for the Environment Factor

Items	Father's education	Mother's education	Number of books at home	Number of children's books at home
1. Father's Education	–	.49***	.27***	.23***
2. Mother's Education		–	.26***	.26***
3. Number of books at home			–	.59***
4. Number of Children's Books at Home				–

Note: * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 7: Descriptive Statistics for Father's (F) Level and Mother's (M) Level of Education

Variables	F/M above high school		F above M below high school		F below M above high school		F/M below high school	
	N = 847		N = 88		N = 94		N = 128	
	M	SD	M	SD	M	SD	M	SD
Children's Reading Interaction	2.14	.74	2.18	.76	2.17	.74	1.90	.71
Children's Negative Reading Attitudes	2.36	.61	2.30	.62	2.49	.63	2.32	.63
Children's Negative Self-Efficacy	2.05	.84	2.30	.79	2.18	.81	2.13	.86
Children's Positive Reading Attitudes	3.46	.65	3.40	.70	3.35	.74	3.33	.72

Total N = 1,184*

Notes: "Above high school" includes high school.

To keep the quality of the results, 234 data items with missing values were excluded. Checking found no systematic bias on these excluded data. A total of 1184 data items were used in the results.

1. Father and mother both with higher level of education (at or above high school)¹
2. Father with a higher level and mother with a lower level of education (below high school)
3. Father with a lower level and mother with a higher level of education
4. Father and mother both with a lower level of education (see Table 7).

We then conducted a one-way ANOVA, the results of which revealed that parents' level of education influenced children's reading interaction, $F(3, 1180) = 4.32$, $MSe = .54$, $p < .01$. *Scheffé* post-hoc analysis showed that if both the father's and the mother's level of education was below high school level, their children had significantly fewer interactive reading behaviors than was the case for children whose father and mother both had a higher level of education ($p < .05$). There was also a significant parental education

effect on children's negative self-efficacy, $F(3, 1180) = 2.79$, $MSe = .70$, $p < .05$. When the father's and the mother's level of education was above high school level, their children were marginally significantly less likely to have negative self-efficacy than were children whose fathers had a higher level and whose mothers had a lower level of education (*Scheffé* post-hoc, $p = .08$).

To clarify the relationship between number of books and parental education, we performed a one-way ANCOVA, controlling for the number of books and children's books. The result revealed no parental education effects on all children's factors.

In brief, we found the number of books at home and parental education levels to be significantly correlated. However, when we partialled out the books, parental education was no longer an influential factor in children's reading.

1 High school taken as indicator in line with Weigel, Martin, and Bennett's (2006) study.

Discussion

This study investigated (i) the relationship between children's reading behaviors and attitudes, and (ii) the relationship between all parental factors and children's reading behaviors and attitudes. A component factor analysis produced four child-related reading factors for six parent-related reading factors. The results showed a high correlation between children's positive reading attitudes and positive reading behaviors, and a negative relationship between children's positive reading attitudes and negative self-efficacy. In line with many other studies, gender had a significant effect on the children's reading behaviors and attitudes, with girls having more positive attitudes than boys.

What is particularly worth mentioning in relation to this present study is the near zero correlation coefficient between positive and negative reading attitudes. The positive reading attitudes were about enjoyment of reading in the broader sense and the negative attitudes were about lack of interest in textbooks and school work. The question is why the positive and negative factors did not overlap? The correlation of four children's factors probably could provide some clues to the answer. Both positive and negative reading attitudes were significantly correlated with negative self-efficacy. This finding implies that when children enjoy reading, they see a positive self, and when they have trouble with school reading, they see a negative self. For the second group of children, reading attitudes and skills appear to be limited to and so constrained by school-related reading materials and experiences.

As for parental reading factors, positive parental attitudes and a home environment rich in reading resources positively predicted children's positive reading behaviors and attitudes. Negative parental factors such as having difficulties fostering children's reading abilities, providing few children's books at home, and seeing reading as a functional tool predicted the children's negative reading attitudes and negative self-efficacy. These findings clarify the problem. Parents who provide few books at home and have difficulty fostering their children's reading abilities can affect their children's school skills and contribute to their children's negative feelings about their reading efficacy.

In regard to parents' level of education, we found that when we controlled for the quantity of books at

home, parental education alone did not show its effect on children's reading. This result did not meet our expectation and so warrants investigation. In short, parental education and number of books at home seem bound together. Parents with a high-school and above level of education tend to provide their children with more books and do not have difficulty guiding their children's reading. Their children also tend to have positive reading behaviors and attitudes. Parents whose level of education is below the level of high school tend to provide fewer books at home and have difficulty fostering their children's reading ability. This situation, in turn, appears to affect children's school skills and make them feel even less able in reading.

In conclusion, what we observed in Taiwan in regard to parental factors and children's reading related to both parents' and their children's focus on reading. Those parents and children who saw reading in a broader sense, enjoyed reading. Those who saw reading as school work felt less capable of reading. The findings also confirmed that children's reading behaviors and attitudes were related to home environment.

One technical concern concerning the study relates to the consequence of performing exploratory factor analysis designed to group items into categories. Words with similar expressions were found grouped together, for instance, the words "like" and "enjoy;" the words "hate," "boring," and "don't like." These words, in Chinese, are such that the Chinese-language students participating in PIRLS probably saw any sentences beginning with or using them as asking the same questions. If this assumption is valid, then cross-cultural translations of texts used in tests of reading achievement must be conducted carefully, as must any cross-cultural examination of data from such tests.

Finally, the items concerning the awareness of difficulty in fostering children's reading ability are good indicators of parental practice, and they relate to children's negative reading attitudes and motivation. These items, however, were not in PIRLS, but were cultural adaptations for use in the present study. If researchers want to know what parents do to enhance their children's reading, we suggest they conduct research examining the relationship between the parental factor "awareness of difficulty in fostering children's reading ability" and children's reading behaviors and attitudes.

References

- Abdulrahim, D. N. (1997). Comparison of gender differences in reading attitudes among students in an "at risk program". *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 57(12-A), 5046.
- Adams, M. J. (1990). *Learning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Baker, L. (2003). The role of parents in motivating struggling readers. *Reading and Writing Quarterly*, 19, 87–106.
- Baker, L., & Scher, D. (2002). Beginning readers' motivation for reading in relation to parental beliefs and home reading experiences. *Reading Psychology*, 23, 239–269.
- Baker, L., Scher, D., & Mackler, K. (1997). Home and family influences on motivations for literacy. *Educational Psychologist*, 32, 69–82.
- Baker, L., Serpell, R., & Sonnenschein, S. (1995). Opportunities for literacy learning in the homes of urban preschoolers. In L. M. Morrow (Ed.), *Family literacy: Connections in schools and communities* (pp. 236–252). Newark, DE: International Reading Association.
- Barnett, J. E., & Irwin, L. (1994). The effects of classroom activities on elementary students' reading attitudes. *Reading Improvement*, 31, 113–121.
- Bradley, R. H., Corwyn, R. F., Pipes McAdoo, H., & García Coll, C. (2001). The home environments of children in the United States. Part I: Variations by age, ethnicity, and poverty status. *Child Development*, 72(6), 1844–1867.
- Bratern, I., Lie, A., Andreassen, R., & Olaussen, B. S. (1999). Leisure time reading and orthographic processes in word recognition among Norwegian third- and fourth-grade students. *Reading and Writing*, 11, 65–88.
- Cullinan, B. W. (1987). *Children's literature in the reading program*. Newark, DE: International Reading Association.
- DeBaryshe, B. D. (1995). Maternal belief systems: Linchpin in the home reading process. *Journal of Applied Developmental Psychology*, 16, 1–20.
- Diamond, P. J., & Onwuegbuzie, A. J. (2001). Factors associated with reading achievement and attitudes among elementary school-aged students. *Research in the School*, 8, 1–11.
- Fitton, L., & Gredler, G. (1996). Parental involvement in reading remediation with young children. *Psychology in the Schools*, 33, 325–332.
- Greaney, V., & Hegarty, M. (1987). Correlates of leisure-time reading. *Journal of Research in Reading*, 10, 3–27.
- Kazelskis, R., Thames, D., & Reeves, C. (2004). The elementary reading attitude survey: Factors in variance across gender and race. *Reading Psychology*, 25, 111–120.
- Kush, J. C., Watkins, M. W., & Brookhart, S. M. (2005). The temporal-interactive influence of reading achievement and reading attitude. *Educational Research and Evaluation*, 11(1), 29–44.
- Lee, J. W., & Barro, R. J. (1998). *Schooling quality in a cross section of countries* (Development Discussion Paper No. 659). Harvard, MA: Harvard Institute for International Development, Harvard University.
- Ley, T. C., Schaer, B. B., & Dismukes, B. W. (1994). Longitudinal study of the reading attitudes and behaviors of middle school students. *Reading Psychology*, 15(1), 11–38.
- MacDonald, G. W., & Cornwall, A. (1995). The relationship between phonological awareness and reading and spelling achievement eleven years later. *Journal of Learning Disabilities*, 28(8), 523–527.
- Matthewson, G. C. (1994). Model of attitude influence upon reading and learning to read. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 1131–1161). Newark, DE: International Reading Association.
- McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1995). Children's attitudes toward reading: A national survey. *Reading Research Quarterly*, 30, 934–956.
- McKenna, M. C., Stratton, B. D., Grindler, M. C., & Jenkins, S. J. (1995). Differential effects of whole language and traditional instruction on reading attitudes. *Journal of Reading Behavior*, 27, 19–44.
- Noble, K. G., Farah, M. J., & McCandliss, B. D. (2006). Socioeconomic background modulates cognition-achievement relationships in reading. *Cognitive Development*, 21, 349–368.
- Olofson, A., & Niedersoe, J. (1999). Early language development and kindergarten phonological awareness as predictors of reading problems: From 3 to 11 years of age. *Journal of Learning Disabilities*, 32, 464–472.
- Paris, S. G., Wasik, B. A., & Turner, J. C. (1991). In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research: Volume II* (pp. 609–640). New York: Longman.
- Psacharopoulos, G., & Woodhall, M. (1985). *Education for development: An analysis of investment choices*. Oxford: Oxford University Press for the World Bank.
- Quinn, B., & Jadav, A. D. (1987). Causal relationship between attitude and achievement for elementary grade mathematics and reading. *Journal of Educational Research*, 80, 366–372.
- Rea, D., Romine, B. C., McKenna, M., & Griffin, B. (1997). *A cross-cultural analysis of reading attitude development*. Paper presented at the American Educational Research Association, Chicago, IL.

- Rowe, K. J. (1991). The influence of reading activity at home on students' attitudes towards reading, classroom attentiveness and reading achievement: An application of structural equation modeling. *British Journal of Educational Psychology*, 61, 19–35.
- Sainsbury, M., & Schagen, I. (2004). Attitudes to reading at ages nine and eleven. *Journal of Research in Reading*, 27(4), 373–386.
- Sonnenschein, S., Brody, G., & Munsterman, K. (1996). The influence of family beliefs and practices on children's early reading development. In L. Baker, P. Afflerbach, & D. Reinking (Eds.), *Developing engaged readers in school and home communities* (pp. 3–20). Mahwah, NJ: Lawrence Erlbaum.
- Weigel, D. J., Martin, S. S., & Bennett, K. K. (2006). Mothers' literacy beliefs: Connections with the home literacy environment and pre-school children's literacy development. *Journal of Early Childhood Literacy*, 6(2), 191–211.
- Weinberger, J. (1996). A longitudinal study of children's early literacy experiences at home and later literacy development at home and school. *Journal of Research in Reading*, 19, 14–24.
- Wigfield, A., & Asher, S. R. (1984). Social and motivation influences on reading. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 423–452). New York: Longman.
- Wigfield, A. & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology*, 89(3), 420–432.

Online data collection in SITES 2006: Paper survey versus web survey—do they provide comparable results?

Barbara Nęza Brečko
Educational Research Institute
Ljubljana, Slovenia

Ralph Carstens
IEA Data Processing and Research Center
Hamburg, Germany

Introduction

The role of computers in collecting data is increasing, and many studies have examined aspects of adult populations answering self-administered computer surveys (e.g., Dillman 1998; Lozar Manfreda & Vehovar, 2002). Some studies show that paper-and-pencil methods have certain drawbacks. High costs are one of them and another is that only relatively simple questionnaires can be used (Dillman, 1998). Previous research shows that self-administered computer methods can be a solution for problems regarding costs. There is no extra data entering, no questionnaire printing, and easy distribution of questionnaires. Computer-assisted, self-administered questionnaires (CASAQ) can also overcome the problems of complex paper-and-pencil self-administered questionnaires, because computer programs can handle the flow and logic of the questionnaire.

One of the most obvious positive effects of the computerization of a survey relates to the amount of missing data that occurs in paper-and-pencil mode because of routing and skipping errors. Several studies have demonstrated that computer-assisted data collection reduces the rate of missing data (Tourangeau, Rips, & Rasinski, 2004). However, several studies also show that item non-response is higher with computer-assisted data collection than with traditional paper-and-pencil methods (Lozar Manfreda & Vehovar, 2002).

Most of the authors who have compared different modes of data collection found little difference between the responses obtained by paper-and-pencil mode and the responses obtained by computer-assisted mode, although differences did arise when questions touched on personal aspects, such as drug use and school bullying (see, for example, De Leeuw, Hox, &

Kef, 2003). Respondents seem more willing to report sensitive behavior on computer questionnaires than on paper-and-pencil questionnaires.

Most of the research relating to mode of questionnaire delivery concentrates on response rates, completion rates, and the representativeness of the samples. Very few studies address the question of whether respondents provide different information depending on the mode of the questionnaire delivery (Denscombe, 2006). The issues addressed in this paper therefore concern the quality and reliability of data obtained from computer-based questionnaires or, more specifically, questionnaires available through a website.

Data sources and methods

In 2005, the field trial of the IEA Second Information Technology in Education Study (SITES) 2006 was conducted. SITES 2006 was the first international study in the field of education where participating countries could elect to collect data online—via the web—in addition to the traditional pen-and-paper mode. One of the purposes of the field trial study was to examine whether the two modes provided comparable results and whether the online data collection could be implemented internationally in the main study.

Eighteen countries participated in the field trial, each country with a sample of 25 primary schools. For each country that conducted the trial in mixed mode (used both modes of data collection), standard error was computed for participation rate, and countries where the difference in participation was statistically significant were excluded from further analyses. Ultimately, data from the nine countries that used the mixed mode were used. These countries were

Catalonia, Denmark, Finland, Hong Kong SAR, Italy, Singapore, Slovakia, Slovenia, and Taiwan. To allow comparison between the modes, a split sample design was implemented for the SITES 2006 field test. This process systematically assigned half the selected schools to the online mode and the other half to paper-and-pencil mode (Table 1).

Because the sample of schools was relatively small (25 randomly selected primary schools per country), only teacher questionnaires were used for the in-depth reliability analyses. On average, four teachers, teaching mathematics and science in Grade 8, per school participated in the study. (The number of teachers sampled depended on the frequency with which each teacher used a computer for educational purposes.) The database used for the analyses consisted

Table 1: SITES 2006 Questionnaire Response Mode for Each Country

Country*		Mode		Total
		Paper	Online	
Cntr1	<i>n</i>	57	62	119
	%	47.9%	52.1%	100%
Cntr2	<i>n</i>	33	30	63
	%	52.4%	47.6%	100%
Cntr3	<i>n</i>	46	50	96
	%	47.9%	52.1%	100%
Cntr4	<i>n</i>	73	64	137
	%	53.3%	46.7%	100%
Cntr5	<i>n</i>	38	52	90
	%	42.2%	57.8%	100%
Cntr6	<i>n</i>	74	59	133
	%	55.6%	44.4%	100%
Cntr7	<i>n</i>	46	56	102
	%	45.1%	54.9%	100%
Cntr8	<i>n</i>	54	54	108
	%	50.0%	50.0%	100%
Cntr9	<i>n</i>	43	43	86
	%	50.0%	50.0%	100%
Total	<i>n</i>	464	470	934
	%	49.7%	50.3%	100%

Note: *Because the analyses were conducted on unpublished data, the names of the countries are not disclosed.

of 934 mathematics and science teachers from nine countries, 49.7% of whom completed the paper-based questionnaire and 50.3% of whom completed the web-based questionnaire.

The teachers included in the database complied with the set of rules governing random selection of schools, that is, random assignment of mode to the schools, and response-rate requirements that allowed calculation of unbiased statistics. Because the data were drawn from the field trial, the generalizability of the findings to the main SITES survey needs to be treated with caution. The field trial database consists of several stratified country samples, so the field trial samples were not really probability samples, although they were proportional to explicit stratum distribution.

The paper-based questionnaire and the web-based questionnaire had identical questions, identical wording, and an identical sequence to the questions. The graphical design of the web-based questionnaire was very simple; there was one question on one HTML page. The fact that the two questionnaires were nearly identical meant the measurement error due to the instrument was controlled.

There were no statistical differences between the two samples in terms of three variables: highest level of teacher education ($\chi^2 = 4.373$; $df = 2$; $p = 0.112$), bachelor's degree in science or mathematics ($\chi^2 = 0.060$, $df = 3$ $p = 0.996$), and years of experience teaching ($\chi^2 = 1.597$, $df = 4$, $p = 0.815$).

Results

Chi square tests

In the present study, survey items, not respondents, were of interest. Several statistical analyses, including chi square, were used to examine mode effect. The chi square analyses were conducted on all the items, both at national and international levels. However, because the questionnaire consisted of 333 items, it is beyond the scope of this article to present the analyses for all of them. We therefore present the results for a selection of items only (see Table 2).

As is evident from Table 2, which shows ordinal data, the difference between the two modes in terms of the distribution of responses was negligible. In the presented selection of the 13 variables, two items produced data indicative of some variation between the two modes of questionnaire. We therefore conducted additional tests for these two variables,

Table 2: Distribution of Responses by Data Collection Mode in Relation to the Question: "In your teaching of the target class in this school year, how important is it for you to achieve the following goals?"

	Paper		Online				
	<i>n</i>	%	<i>n</i>	%	χ^2	<i>df</i>	<i>p</i>
<i>T08a1_currgoals/goal importance/world of work</i>							
Not important at all		14.0	3.1	19	4.1		
A little important	66	14.4	72	15.7			
Somewhat important	180	39.4	172	37.5			
Very important	197	43.1	196	42.7	1.198	3	.753
<i>T08b1_currgoals/goal importance/upper edu</i>							
Not important at all	2	0.4	2	0.4			
A little important	14	3.1	11	2.4			
Somewhat important	106	23.2	116	25.5			
Very important	335	73.3	326	71.6	.929	3	.819
<i>T08c1_currgoals/goal importance/assessments</i>							
Not important at all	7	1.5	6	1.3			
A little important	35	7.6	30	6.6			
Somewhat important	187	40.7	217	47.7			
Very important	230	50.1	202	44.4	4.487	3	.213
<i>T08d1_currgoals/goal importance/subj content</i>							
Not important at all	10	2.2	2	0.4			
A little important	19	4.1	25	5.5			
Somewhat important	102	22.2	136	29.9			
Very important	329	71.5	292	64.2	13.186	3	.004
<i>T08e1_currgoals/goal importance/individualize</i>							
Not important at all	2	0.4	2	0.4			
A little important	36	7.9	43	9.5			
Somewhat important	213	46.7	224	49.7			
Very important	205	45.0	182	40.4	2.237	3	.525
<i>T08f1_currgoals/goal importance/inc motivation</i>							
Not important at all	1	0.2	0	0			
A little important	10	2.2	12	2.6			
Somewhat important	101	22.0	140	30.6			
Very important	347	75.6	306	66.8	10.006	3	.018
<i>T08g1_currgoals/goal importance/satisfy parent</i>							
Not important at all	7	1.5	15	3.3			
A little important	106	23.3	104	22.7			
Somewhat important	242	53.2	231	50.4			
Very important	100	22	108	23.6	3.482	3	.323
<i>T08h1_currgoals/goal importance/stud readiness</i>							
Not important at all	5	1.1	4	0.9			
A little important	47	10.2	58	12.7			
Somewhat important	203	44.1	223	48.9			
Very important	205	44.6	171	37.5	5.260	3	.154

Table 2 (contd.): Distribution of Responses by Data Collection Mode in Relation to the Question: "In your teaching of the target class in this school year, how important is it for you to achieve the following goals?"

	Paper		Online				
	<i>n</i>	%	<i>n</i>	%	χ^2	<i>df</i>	<i>p</i>
<i>T08i1_currgoals/goal importance/org skills</i>							
Not important at all	7	1.5	4	9	.882	3	.830
A little important	55	12.0	56	12.3			
Somewhat important	208	45.3	211	46.3			
Very important	189	41.2	185	40.6			
<i>T08j1_currgoals/goal importance/real world ex</i>							
Not important at all	3	0.7	2	0.4	.995	3	.802
A little important	36	7.9	34	7.5			
Somewhat important	164	35.9	177	38.8			
Very important	254	55.6	243	53.3			
<i>T08k1_currgoals/goal importance/learn frm exp</i>							
Not important at all	41	9.0	37	8.1	3.450	3	.327
A little important	149	32.6	126	27.6			
Somewhat important	181	39.6	201	44.0			
Very important	86	18.8	93	20.4			
<i>T08l1_currgoals/goal importance/comm skills</i>							
Not important at all	39	8.5	31	6.8	1.706	3	.636
A little important	125	27.4	126	27.7			
Somewhat important	190	41.6	183	40.2			
Very important	103	22.5	115	25.3			
<i>T08m1_currgoals/goal importance/responsible</i>							
Not important at all	34	7.4	39	8.6	2.782	3	.426
A little important	76	16.6	64	14.1			
Somewhat important	125	27.4	142	31.2			
Very important	222	48.6	210	46.2			

with $p < 0.05$. We used ordinal directional measures (i.e., Kendall's tau) to measure strength and direction between variable T08d1 and mode and variable T08f1 and mode (see Tables 3 and 4). The significance value in both cases was below 0.05, but the value of the test statistics was very close to zero, indicating a weak relationship between the two modes.

Bivariate analyses of all 13 variables showed no systematic differences in terms of administration mode. Some minor significant differences were observed, but no clear pattern could be discerned.

Reliability analyses

Other statistical analyses were done to compare the equity of results obtained by the two different questionnaire modes. One of these was a reliability analysis.

In many areas of research, the precise measurement of hypothesized processes or variables (constructs) poses a challenge. For example, in psychology, the precise measurement of personality variables or attitudes is usually a necessary first step before any theories of personality or attitudes can be considered. In general, in all social sciences, unreliable measurements of people's beliefs or intentions hamper efforts to predict their behavior. The issue of precision of measurement also comes up in applied research, whenever variables are difficult to observe. Reliability and item analysis may be used to construct reliable measurement scales, to improve existing scales, and to evaluate the reliability of scales already in use. The method enables computation of statistics that allow us to build and evaluate scales following the classical testing theory model. The assessment of scale reliability is based

Table 3: Kendall's Tau Calculation for Variable T08d1_currgoals/goal importance/subj content

		Value	Asymp. SE ^a	Approx. T ^b	Approx. sig.
Ordinal by ordinal	Kendall's tau-b	-.069	.032	-2.140	.032
	Kendall's tau-c	-.067	.031	-2.140	.032
N of valid cases		915			

Notes: ^a Not assuming the null hypothesis.

^b Using the asymptotic standard error assuming the null hypothesis.

Table 4: Kendall's Tau Calculation for Variable T08f1_currgoals/goal importance/inc motivation

		Value	Asymp. SE ^a	Approx. T ^b	Approx. sig.
Ordinal by ordinal	Kendall's tau-b	-.094	.032	-2.883	.004
	Kendall's tau-c	-.086	.030	-2.883	.004
N of valid cases		917			

Notes: ^a Not assuming the null hypothesis.

^b Using the asymptotic standard error, assuming the null hypothesis.

on the correlations between the individual items or measurements that make up the scale, relative to the variances of the items.

Internal consistency is probably the method that is the easiest to conduct when we want to estimate the reliability of measuring. It is based on the computation of covariances or correlation coefficients between all the variables measuring the same latent variable. Internal consistency reliability is concerned with the homogeneity of the items comprising a scale (Ferligoj, Leskošek & Kogovšek, 1995). If the items of a scale have a strong relationship to their latent variable, they will have a strong relationship to one another. A scale is internally consistent to the extent that its items are highly inter-correlated. High inter-item correlations suggest that all items are measuring the same latent variable. The purpose of our analysis was to check if the reliability of the scales measuring the latent variables was the same for both data-collection modes.

Internal consistency is equated with Cronbach's coefficient alpha (α). Alpha is defined as the proportion of a scale's total variance that is attributable to a common source, presumably the true score of a latent variable underlying the items (DeVellis, 1991). The range of possible values for coefficient alpha (α) is between 0.0 and 1.0. Proposed interpretation of its

values by Ferligoj et al. (1995) is as follows:

- $\alpha \geq 0.80$: the reliability of the measured variables is excellent
- $0.70 \leq \alpha < 0.80$: the reliability of the measured variables is very good
- $0.60 \leq \alpha < 0.70$: the reliability of the measured variables is sufficient¹
- $\alpha < 0.60$: the reliability of the measured variables is barely acceptable.

In presenting the results, we use the same items as previously. The analyses were conducted on items measuring latent variables derived from previous analyses that used factor analyses and structural equation modeling. Because our focus is on comparison of two modes, we do not discuss the reliability of the scales.

The internal consistency analysis (Table 5) showed the reliability of the scales was quite low for the first two constructs, but that the results were very similar across the two modes. Although there were some differences between the groups according to the mode of data collection, the differences were not statistically significant. In short, the reliability of the scales was not influenced by the mode of data collection.

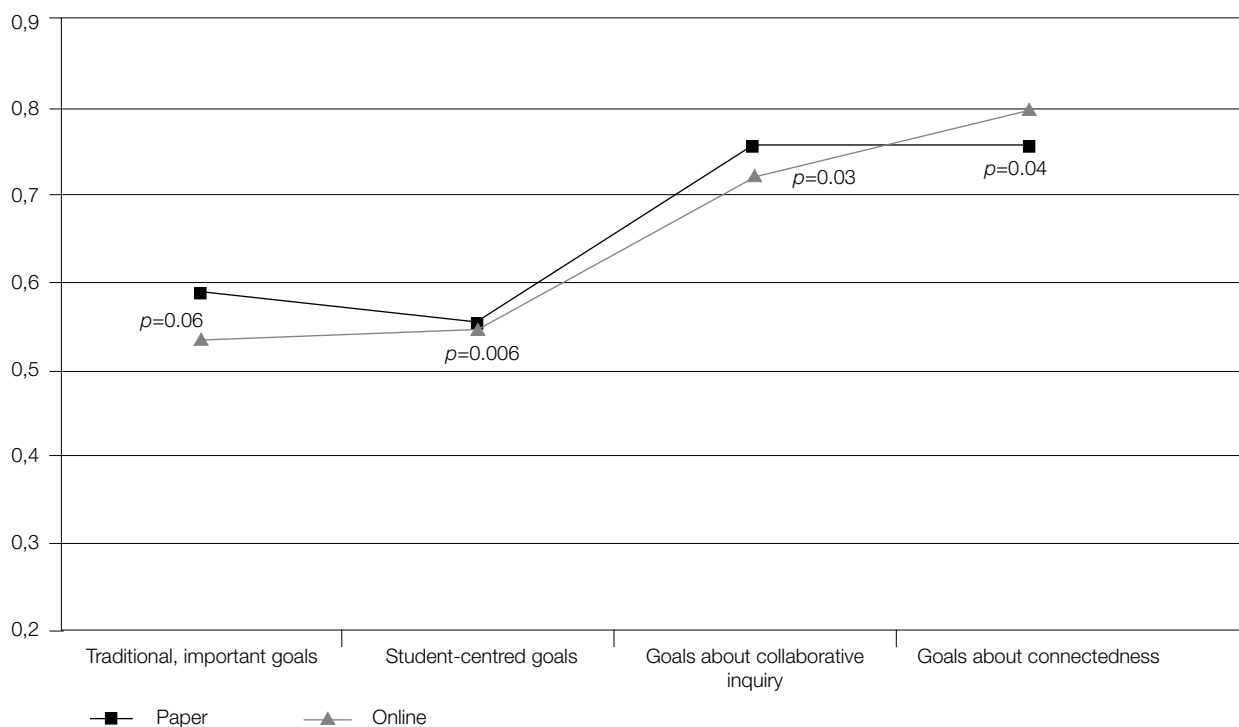
Table 6 and Figure 1 also show that the reliability of the scales was not affected by the mode of the

¹ Note that, in comparison, the *Technical Standards for IEA Studies* (Martin, Rust, & Adams, 1999, p. 73) recommend that reliabilities below 0.70 should be treated and reported with caution.

Table 5: Reliability Analysis in Relation to the Question: “In your teaching of the target class in this school year, how important is it for you to achieve the following goals?”

International database (both modes)		
Traditional important goals	(4 items)	$\alpha = 0.575$
Student-centered goals	(3 items)	$\alpha = 0.552$
Goals about collaborative inquiry	(3 items)	$\alpha = 0.739$
Goals about connectedness	(3 items)	$\alpha = 0.778$

Figure 1: Internal Consistency Difference between the Two Modes (q8)



questionnaire. The same procedures were conducted for other latent variables. The only statistically significant difference found for the latent variables derived from Question 15: “When teaching the target class during this school year, during a typical school week, how often would you engage in the following teaching-related activities?”

Table 7 shows the differences in reliability were significant between the two modes for three constructs (variables), with the reliability being higher for the online data collection mode.

Missing value analysis

During the SITES data entry and later processing, four types of missing value were entered, assigned, or calculated. The designations were “omitted” (including invalid), “not administered,” “logically not applicable,” and “not reached.” While the first three variables relate directly to the respondent’s behavior (such as individuals not responding to a question or countries leaving out an entire question), the last variable (not reached) particularly interested us, as it allowed us to analyze where, in the questionnaire, respondents stopped answering (i.e., dropped out).

Modewise, there was a significant difference in terms of questionnaire completion. About 8.5% of the teachers using the online mode dropped out

Table 6: Internal Consistency Analysis for Paper-Based and Web-based Questionnaires (q8)

	Paper	ODC	<i>p</i> value differences (g1- g2)
	α	α	
Traditional, important goals	0.588	0.559	$p = 0.533$
Student-centered goals	0.554	0.548	$p = 0.908$
Goals about collaborative inquiry	0.755	0.723	$p = 0.287$
Goals about connectedness	0.756	0.799	$p = 0.093$

Table 7: Reliability Analysis in Relation to the Question: “When teaching the target class during this school year, during a typical school week, how often would you engage in the following teaching-related activities?”

	Paper	ODC	<i>p</i> value differences (g1- g2)
	α	α	
Traditional, important instruction	0.632	0.670	$p = 0.319$
Student-centered pedagogy	0.632	0.717	$p = 0.024$
Support collaborative inquiry	0.676	0.757	$p = 0.013$
Support liaison for connectedness in learning	0.692	0.776	$p = 0.006$

before the end of the questionnaire; the drop-out rate for the paper mode was almost negligible. This finding is consistent with other studies of non-response on questionnaires containing many items (Lozar Manfreda & Vehovar, 2002; Vehovar, Lozar Manfreda, & Batagelj, 2001). In all these cases, no devices, such as reminders, were used to limit item non-response. In the case of the SITES 2006 field study, drop-out by individuals clearly occurred at the beginning of either very long matrix questions or at the beginning of the demographics section in the middle of the questionnaire (respondents may have assumed this was the last part in the questionnaire). Figure 2 shows the number of “drop-outs” for each variable for both modes. These findings suggest that shortening the questionnaire and putting the demographic questions at the end of the questionnaire would substantially reduce of the number of missed items.

Discussion

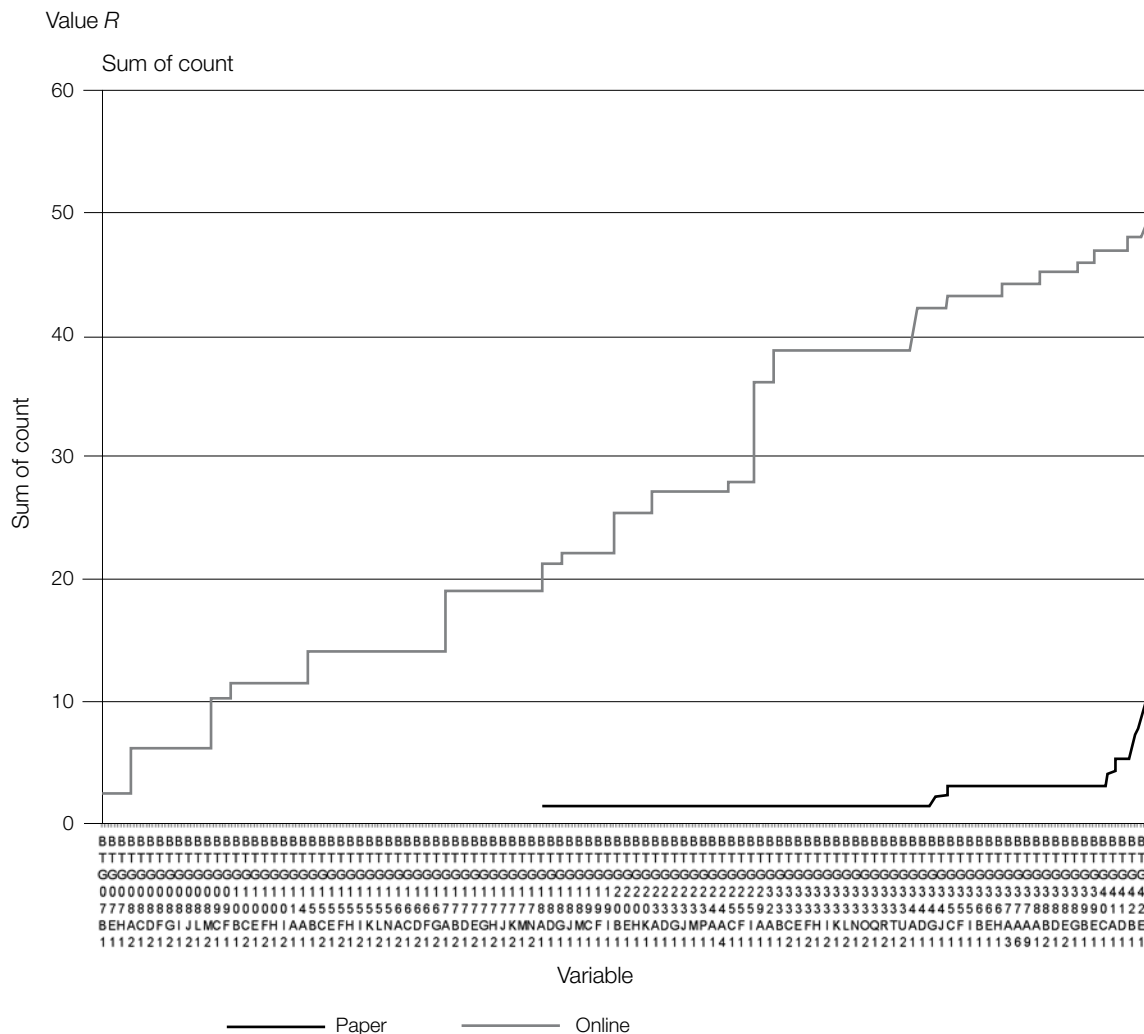
Web-based research has the potential to make the collection of survey data cheaper, quicker, and more user-friendly than conventional methods of survey research involving postal questionnaires, telephone surveys, and face-to-face interviews, a point that is widely recognized by social and market researchers. Weighed against these significant advantages are

the problems of sampling and response rates noted by methodologists. While a large body of research evidence and commentary has emerged in relation to concerns surrounding the collection of data over the internet, and considerable attention has been given to overcoming these limitations, relatively little effort has been put into comparing web-based data with that supplied by conventional data collection techniques. More specifically, do respondents provide the same answers when using the web-based questionnaires as they do when using conventional pencil-and-paper questionnaires?

In the present study, we tested if web-based and paper-and-pencil-based surveys would give comparable results for comparable samples across the two survey modes. The findings presented in this paper are based on data for a specific group of respondents—mathematics and science teachers teaching Grade 8 students from nine countries.

Although the completion rate was higher for paper-based questionnaires, other analyses of the quantitative data showed no or very little mode effect on the quality of the data. Missing data is one of the biggest concerns of the online data methodology, although the literature shows that item non-response is two directional: some studies report a lower incidence of missing data in web-based questionnaires than in traditional pencil-

Figure 2: Number of “Not Reached” Items on Paper-based and Web-based Questionnaires



Note: The vertical axis shows the total number of cases coded as “not reached” versus the sequence of the questionnaire (variables) on the horizontal axis.

and-paper ones; others report a higher incidence of missing data in web-based questionnaires. However, it is not possible to provide controls for item non-response, as this would jeopardize the similarity of the two questionnaires.

The indications from our chi square tests and reliability analyses are that the benefits of web-based questionnaires do not compromise consistency. Furthermore, the findings support previous research that web-based questionnaires generate better-quality data than do pencil-and-paper tests (Denscombe, 2006). Of the 333 questionnaire items, there were only a few for which mode had a statistically significant influence on the information gathered. However, we need to stress that cultural effect was not controlled.

Our reliability analyses showed that the data gathered by the web-based questionnaire tended to be more reliable than that gathered by the traditional questionnaire. In some cases, the web-based questionnaire gave data that were of a statistically significant better quality than the data collected by the pencil-and-paper questionnaire.

Our study differs from other studies examining the mode effect because the data allowed cross-national comparison, and also gave us the advantage of being able to compare data from survey respondents for the two modes from nine different cultures. Most previous research has examined mode effect at the national level. Also, because data collection with the computer is no longer an issue in social science research, we expect

that it will become an option for data collection also at the international level.

Finally, we wish to caution that the findings presented in this paper are based on quantitative data and that there are features and aspects of our larger study that we have not covered here. Also, our analyses were based on small sample sizes at the national level (and did not consider between and across country comparisons). Accordingly, our findings should be seen, at this point, as offering information and conclusions meriting further analysis and validation.

Conclusion

In this paper we examined if paper-and-pencil questionnaires and web-based questionnaires provided comparable information from data collected at the international level for SITES 2006. The analyses showed that although the results obtained from the two modes of data collection were comparable, some differences and issues need consideration. One of the most important issues is the level of drop-out in web-based questionnaires. However, despite the higher rate of missing items for the web-based questionnaire, this method of data collection appears to provide reliable data when compared to that provided by the equivalent paper-based questionnaire.

References

- De Leeuw, E. D., Hox, J. J., & Kef, S. (2003). Computer-assisted self-interviewing tailored for special populations and topics. *Field Methods*, 15(3), 223–251.
- Denscombe, M. (2006). Web-based questionnaires and the mode effect: An evaluation based on completion rates and data contents of near-identical questionnaires delivered in different modes. *Social Science Computer Review*, 24(2), 246–254.
- DeVellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.
- Dillman, D. (1998). *Mail and other self-administered surveys in the 21st century: The beginning of a new era*. Retrieved April 5, 2003 from <http://survey.sesrc.wsu.edu/dillman/papers/svys21st.pdf>
- Ferligoj, A., Leskošek, K., & Kogovšek, T. (1995). *Zanesljivost in veljavnost merjenja [Reliability and validity of measurement]*. Ljubljana: FDV.
- Lozar Manfreda, K., & Vehovar, V. (2002). Do mail and web surveys provide the same results? In A. Ferligoj & A. Mrvar (Eds.), *Development in social science methodology* (pp. 149–169). Ljubljana: FDV.
- Martin, M. O., Rust, K., & Adams, R. J. (1999). *Technical standards for IEA studies*. Delft: Eburon.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2004). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Vehovar, V., Lozar Manfreda, K., & Batagelj, Z. (2001). Sensitivity of E-commerce measurement to survey instrument. *International Journal of Electronic Commerce*, 6, 31–52.

Schools that exceed expectations: A cross-country comparison

Marjeta Doupona Horvat and Alja Krevh
Educational Research Institute
Ljubljana, Slovenia

Abstract

Family income and parents' educational background are strong predictors of a student's reading achievement. Schools with children from wealthier and better-educated families have an easier job to do in terms of teaching children how to read and write than do schools with children from families who are less wealthy and educated. The former group of schools is not always successful, though, and there are schools where students come from

economically disadvantaged homes but have reading achievement above the national levels or above the level that might be expected because of environmental factors. This study looked at schools where students exceeded the expectations associated with family income and parents' educational background. Various variables were tested to determine which were relevant for the effective schools across selected countries.

Introduction

Data from the Progress in International Reading Literacy Study (PIRLS) 2001, conducted by the International Association for the Evaluation of Educational Achievement (IEA), showed a strong connection between educational background of the parents and reading achievement of the students. Family income was also a strong predictive factor. The PIRLS data for Slovenia showed that for schools in which the average family income of students was at a certain level, the average reading achievement of the students in those schools did not fall below a certain margin. Slovenian students achieved an average of 502 score points on the PIRLS test of reading literacy, although in some schools students' reading literacy achievement was as low as 430 points. However, those schools in which the average family income of students was higher than Group 3 (1 being the lowest and 6 the highest group according to family income) had an average reading achievement level of 490 points plus. Thus, those Slovenian schools with students from wealthier families had a "guaranteed" level of reading achievement.

There are countries where differences in reading achievement among schools are small or at least not statistically significant. However, in Slovenia, these differences are quite large. Slovenia had not expected this outcome before its participation in PIRLS because of the country's uniform, centralized school system. In

Slovenia, the state sets the curriculum for the whole country, and all schools but one are public. Slovenia has only 450 elementary schools, and the number of students in any one school grade across the country has not exceeded 20,000 in recent years. However, the differences between the schools in terms of achievement are now highly evident and need addressing.

Although Slovenia is a member of the European Union and although transition from a socialist to a democratic state was peaceful, the main present-day concern relates to the fact that Slovenian society has yet to accept that knowledge is a highly valuable good or to acknowledge that issues associated with poverty need taking in hand. Slovenian society is oriented more toward "good" schools, which are characterized in terms of absolute numbers: high proportions of students who proceed to the gymnasium; high scores on national assessments (despite the fact that results are not announced publicly); and so on. Children from such schools come from well-situated families and have well-educated parents. These people are also generally those with the most societal power (although not necessarily political power).

Observers of a society such as ours at present could gain the impression that people "get what they deserve:" educated and/or wealthy people deserve educated children. While there is nothing wrong per se with this expectation, concern resides at the opposite

end of the societal spectrum—with people from the lower socioeconomic brackets. More specifically, concern centers on two facts: first, authorities in Slovenia have difficulty admitting poverty exists; and second, level of educational achievement relates to family socioeconomic status (SES). A community can act collectively to financially ameliorate the effects of poverty within its schools. However, because people in wealthier communities have the biggest salaries, they can contribute more to their schools, and so the gap between schools in poorer and wealthier communities remains.

According to Postlethwaite and Kenneth (1992), who analyzed the IEA Reading Literacy 1991 data, various factors contribute to effective schools. These factors vary from country to country. Of course, various criteria can be applied to determine which school is effective and which is not. We applied one of them (parents' highest level of educational attainment). In this paper, we present a comparative analysis of differences between effective and non-effective schools in seven countries that participated in PIRLS 2001. Our aim in undertaking this analysis was to identify characteristics of effective schools across countries.

Data

Our analysis focused on PIRLS 2001 data. Because our particular interest was the relationship between home and school factors, we used data from all the PIRLS questionnaires—student, home, school, and teacher. Some of the countries that participated in PIRLS did not use the home questionnaires, so we excluded these countries from our analysis.

Having completed a study about the performance of Slovenian schools in PIRLS (Doupona Horvat, Brečko, Kočič, Krevh, & Rožman, 2006), we were interested in looking at data from countries that we could interpret in relation to our findings and conclusions regarding Slovenia. We eventually selected six countries in addition to Slovenia for our comparison. The countries we selected and the reasons why are as follows:

- *Lithuania*, a former socialist country (as is Slovenia), and a country that became a member of the European Union at the same time as Slovenia did.
- *The Netherlands*, which has been used as a reference country in quite a few secondary analyses studies conducted in Slovenia.

- *Norway*, where changes in the education system in recent years are similar to those that have taken place in Slovenia, and which, like Slovenia, sat at the international average on the PIRLS 2001 reading achievement scale.
- *New Zealand*, for providing a model of a country that has experienced substantial improvements in educational achievement in recent years.
- Germany, another country that has been used as a reference point in many secondary analyses studies in Slovenia,
- *Sweden*, because of the rankings achieved by its Grade 3 and Grade 4 students on the PIRLS international scale of reading achievement. Grade 4 Swedish students were the highest performing in PIRLS 2001. Sweden's Grade 3 students, with an average age of 9.8 years, were more akin in age to the Slovene students who participated in PIRLS 2001. However, the Swedish third-grade students achieved an average achievement score on the reading scale (520 points) that was substantially better than the score achieved by the Slovenian students.

Relationships among some basic variables

Our main aim was to determine effective and non-effective schools across the countries and to compare them according to various school factors. In six of the selected countries (Lithuania did not have data on family income), family income related to students' reading achievement in PIRLS (see Figure 1 and Table 1). The differences between the groups were statistically significant.

In Figure 1, Category 1 is the lowest income category and Category 6 is the highest income category. We have not labeled income in Figure 1 in terms of absolute numbers because not all countries divided income categories according to average income. Some countries retained the PIRLS original categories, which had the lowest-income group as that containing families earning less than \$US20,000 per year, and the maximum category as that containing families earning over \$US60,000 per year. Also, categories within the country were not distributed by normal curve. Nevertheless, what we can discern from the information presented in the figure and the table is that children across the countries were more likely to achieve high reading results if their parents had a high income.

Table 1: Correlations between Family Income and Reading Achievement

Country ID	<i>n</i>	SS_Reg	SS_Total	Mult_RSQ	b00	b00.se	b01	b01.se	<i>t</i> -test
Germany	5,617	2634061	24452633	.11	501.28	3.34	13.66	.82	16.63
Netherlands	1,409	403320	4741619	.07	532.10	5.78	9.17	1.36	6.75
New Zealand	1,890	1302791	15263068	.09	472.31	8.67	15.46	1.62	9.56
Norway	2,994	7813762	19663059	.04	460.97	6.20	10.38	1.35	7.67
Slovenia	2,613	9791832	12920433	.08	470.56	3.07	13.01	1.03	12.59
Sweden (Gd.3)	5,325	1843792	30712496	.06	478.04	5.87	12.02	1.22	9.84

Figure 1: Cross-Country Comparison of the Relationship between Students' Reading Achievement in PIRLS 2001 and the Income of Their Families

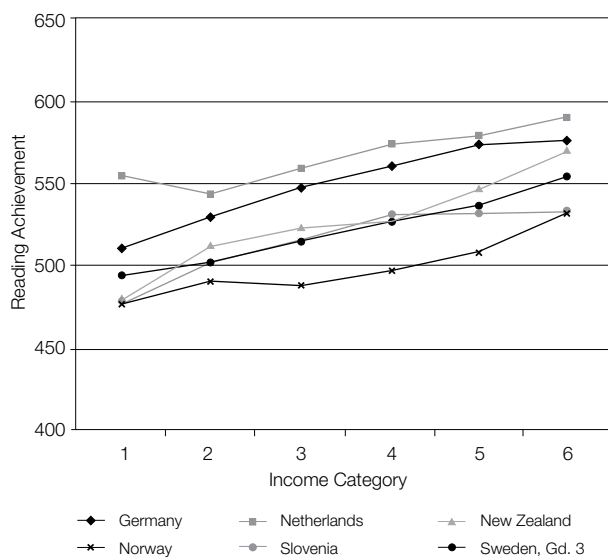


Figure 2: Relationship between Students' Reading Achievement and Parents' Highest Level of Education

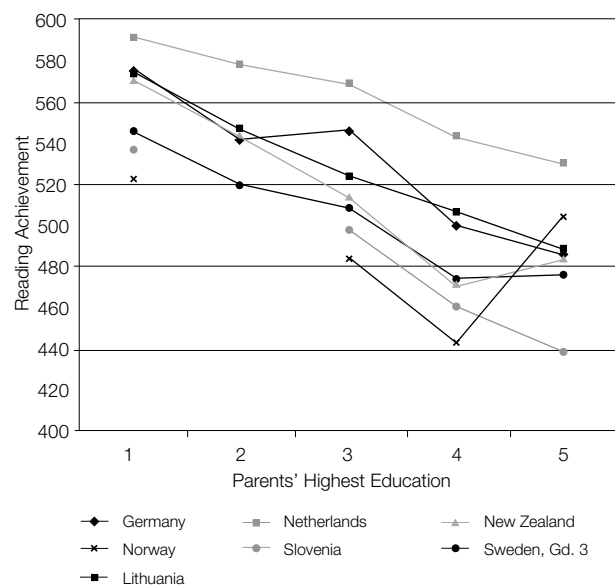


Table 2: Correlations between Parents' Highest Level of Education and Reading Achievement

Country ID	<i>n</i>	SS_Reg	SS_Total	Mult_RSQ	b00	b00.se	b01	b01.se	<i>t</i> -test
Germany	4,891	2224407	21715744	.10	591.14	3.18	-20.12	1.35	-14.92
Lithuania	2,460	1032800	10154767	.10	595.75	5.27	-24.03	2.40	-10.02
Netherlands	2,582	9378542	8317587	.11	612.61	4.28	-16.76	1.29	-13.04
New Zealand	2,037	1667447	16530016	.10	597.39	4.93	-27.57	1.93	-14.28
Norway	3,098	1350317	20199190	.07	541.99	5.62	-19.82	2.27	-8.73
Slovenia	2,796	1366327	13952342	.10	562.37	3.88	-22.38	1.43	-15.66
Sweden(Gd.3)	5,464	2123893	31422788	.07	565.24	3.82	-20.20	1.73	-11.71

We expected to see a similar relationship between reading achievement and parents' highest level of education attainment. In Figure 2, we can see that categories of parental educational attainment are in reverse order from that for the family income variable. Category 1 means at least a university degree while Category 5 means the parents did not complete

elementary education. Two countries have only four categories (Norway and Slovenia did not have the category "finished post-secondary but not university"). Four countries follow the rule that the higher the parents' level of education is, the higher the reading achievement of their children will be. But even in the two countries that are the exception to this rule, we

can assume the same pattern, because Category 5 (“did not finish lower secondary”) is so small that it can be excluded. In Norway, after weighting, there were only six students in Category 5 (i.e., only six students with parents who did not finish elementary school). In New Zealand, 51 students were in Group 5.

Because there is a relationship between the variable “family income” and the variable “parents’ highest education,” we can take only one of them for further comparison. Table 3 depicts the relationship between the two variables. Lithuania did not set a question about family income in the home questionnaire. The correlations between family income and parents’ highest level of education for all other countries included in this study were statistically significant.

Reading achievement by school

Once we had determined the general relationships between basic variables, we looked at those variables

Figure 3: Reading Achievement by School

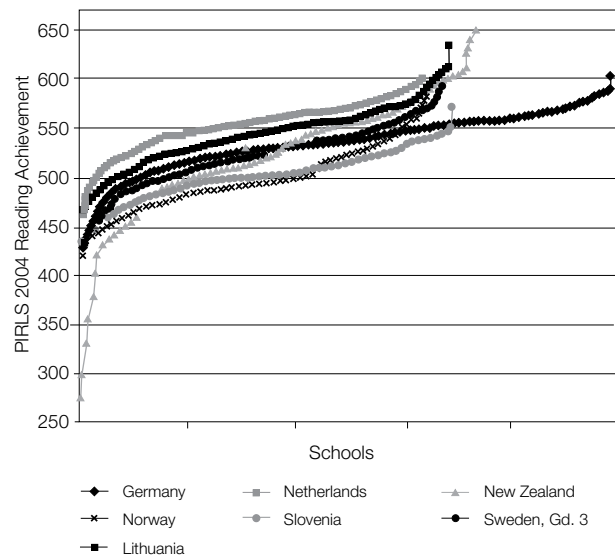


Table 3: Correlations between Family Income and Parents’ Highest Level of Education

		Family income	Parents’ highest education
Germany	Pearson correlation	1	-.384(**)
	Sig. (two-tailed)		.000
	N	5629	4250
Netherlands	Pearson correlation	1	-.472(**)
	Sig. (two-tailed)		.000
	N	1422	1393
New Zealand	Pearson correlation	1	-.368(**)
	Sig. (two-tailed)		.000
	N	1869	1816
Norway	Pearson correlation	1	-.400(**)
	Sig. (two-tailed)		.000
	N	2999	2968
Slovenia	Pearson correlation	1	-.551(**)
	Sig. (two-tailed)		.000
	N	2623	2562
Sweden (Grade 3)	Pearson correlation	1	-.352(**)
	Sig. (two-tailed)		.000
	N	5332	5289

Note: ** Correlation is significant at the 0.01 level (two-tailed).

at the school level. A glance at Table 4 and Figure 3 indicates large differences between schools in all countries. However, these differences were statistically significant only in Slovenia. If we consider the data for PIRLS overall, we find that statistically significant differences between schools were apparent only in the developing countries.

Effective and non-effective schools

For each selected country, we calculated reading achievement for the schools and looked at how it related to parents' highest level of education. Analyses showed that in all seven selected countries, differences between schools, after introduction of the parents' highest education variable, were statistically significant. We built our comparison between effective and non-effective schools on two premises:

1. Higher reading achievement correlates highly with parents' highest education at the level of the individual school.
2. Individual schools where the educational background of the students' parents is, on average, the same (that is, when we compare the schools between themselves) can still have (very) different reading achievements.

These premises led us to assume that there are factors at the school level that make schools effective or non-effective in terms of reading achievement. We accordingly set a regression line between the variable of school reading achievement and the variable of parents' average highest education per school as a criterion for defining effective and non-effective schools. Although we based our decision on the above assumptions about correlations between the variables, it is nonetheless arbitrary. We defined schools ("dots") above the regression line in the following figures as

effective and schools ("dots") below the regression line in the figures as non-effective in terms of reading. Because this is not an ideal representation, residuals vary quite considerably.

In Figures 4, 5, and 6, for Slovenia, Germany, and Sweden, respectively, the horizontal line in each figure represents average parents' highest education per school; the vertical line represents reading achievement. Again, we caution that the regression line is an arbitrary border between effective and non-effective

Figure 4: Reading Achievement According to Parents' Highest Level of Education: Slovenia

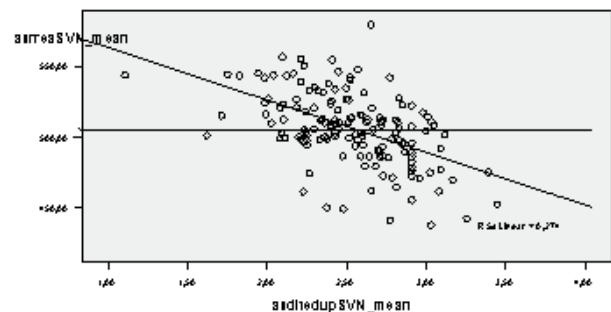


Figure 5: Reading Achievement According to Parents' Highest Level of Education: Germany

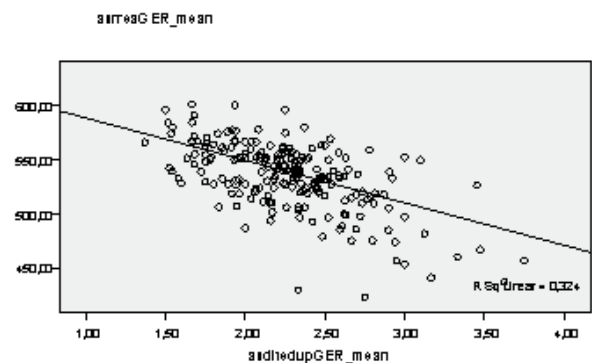
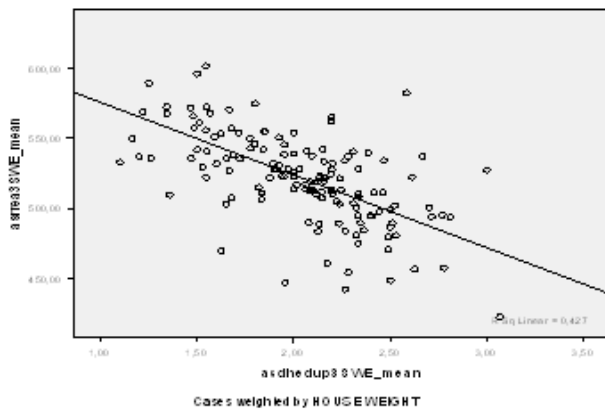


Table 4: Minimum, Maximum, and Average Reading Achievement by Schools: Descriptive Statistics

Country		N	Minimum	Maximum	Mean	Std. Deviation
Germany	mnpv	211	428.09	603.88	534.64	31.200
Lithuania	mnpv	146	443.27	634.35	542.51	32.858
Netherlands	mnpv	135	441.26	600.30	553.56	26.579
New Zealand	mnpv	156	274.32	649.55	522.28	62.671
Norway	mnpv	136	422.59	579.03	496.66	31.299
Slovenia	mnpv	148	431.29	569.91	500.98	25.748
Sweden (Grade 3)	mnpv	144	437.19	600.31	522.11	31.717

Source: PIRLS 2001 database (IDB Analyzer).

Figure 6: Reading Achievement According to Parents' Highest Level of Education: Sweden (Grade 3)



schools. The horizontal line at 502 points represents the average reading achievement for Slovenia.

The same calculations were performed for Lithuania, the Netherlands, New Zealand, and Norway. Here, we split schools into the two categories—effective and non-effective—to allow for further comparisons. However, the two groups were not of the same size; the proportion for each relates to the criterion (regression line). The distribution of the schools in each country is set out in Table 5. In the table, the variable “below_” above” has a value of 1 if the school is effective and a value of 0 if it is not. The proportion of effective schools appears in bold type in the table. The table shows

us that the highest degree of disproportion between effective and non-effective schools—as defined in this paper—is Germany, where the ratio of effective to non-effective schools is almost 2:1. The proportions between the two groups of schools for the remaining six selected countries are more balanced.

Factors contributing to effective and non-effective schools

Location of the school

School principals were asked to identify if their school was located in an urban, a suburban, or a rural area. The relationship between school effectiveness and school location was statistically significant for all seven countries. Figure 7 depicts the relationship between school effectiveness and school location for Germany, Lithuania, the Netherlands, and New Zealand, the countries where the differences were most pronounced. Although Norway, Sweden, and Slovenia had statistically significant differences in terms of the designated relationship, the differences were much smaller than those for the other four countries.

In Germany, 33.7% of the students were attending urban schools. Within this environment, 53% of the students were attending effective schools. In Lithuania, 70% of the children were attending schools in an urban area. Of this (urban) group of students, 45% were in non-effective schools and 55% were in

Table 5: Proportion of Effective and Non-effective Schools per Country

Country ID	below_	above	n	houwgt	mnpv	mnpv_se	pct	pct_se
Germany	.00		3284	2578	513.36	2.84	34.03	3.59
	1.00		4312	4998	552.28	1.87	65.97	3.59
Lithuania	.00		1250	1260	526.97	3.69	49.39	4.06
	1.00		1299	1291	559.29	3.52	50.61	4.06
Netherlands	.00		1958	1940	537.32	3.54	47.18	4.35
	1.00		2154	2172	569.30	1.99	52.82	4.35
New Zealand	.00		990	1021	495.34	4.70	40.96	4.18
	1.00		1510	1472	552.31	5.21	59.04	4.18
Norway	.00		1695	1574	479.53	3.19	45.52	4.98
	1.00		1764	1885	515.60	3.79	54.48	4.98
Slovenia	.00		1490	1471	485.34	2.55	49.84	4.10
	1.00		1462	1481	517.60	2.08	50.16	4.10
Sweden (Gd.3)	.00		2827	2787	502.36	3.91	45.74	5.02
	1.00		3347	3307	534.95	2.96	54.26	5.02

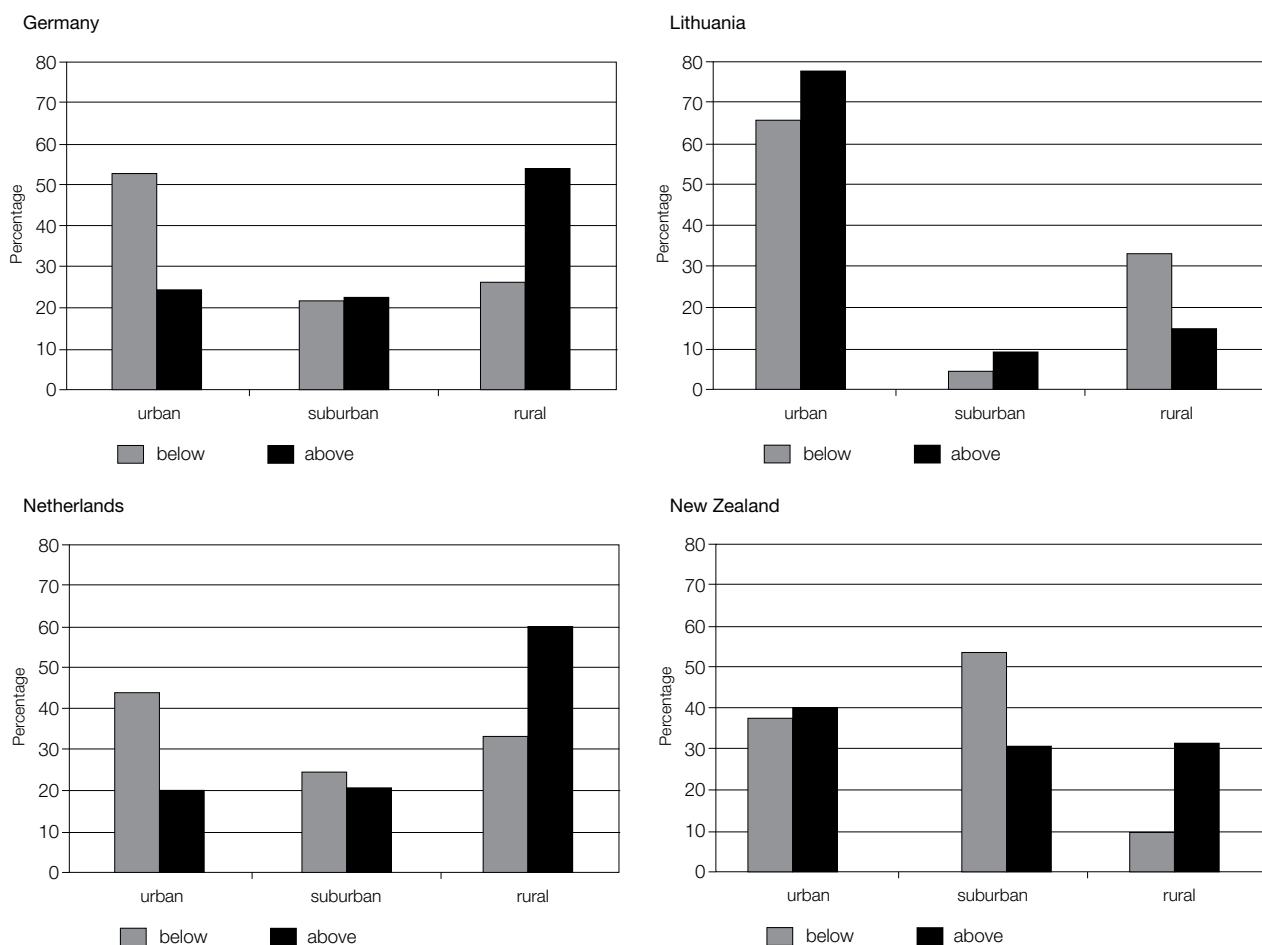
effective schools. In the Netherlands, the situation was almost the reverse of that for Lithuania. Only 31% of the children in the Netherlands were attending schools in urban areas. In these areas, 42% of the children were attending effective schools and 20% of the children were attending non-effective schools. New Zealand had about the same proportion of effective and non-effective schools in urban areas, but there was a difference between the two types in rural areas. Here, 30% of the children were attending effective rural schools and only 10% of children were attending non-effective rural schools.

Teachers' educational attainment

The PIRLS 2001 data showed that level of teacher education has an important influence on the reading achievement of students. In considering this factor in relation to our seven countries, we found only a

few differences because in most countries almost all teachers had a university degree. In Germany, 95% of the students at effective schools had teachers with the highest level of education (a university degree); 87% of students at non-effective schools had students with this level of qualification. In Lithuania, there were no differences between effective and non-effective schools regarding teachers' education. In the Netherlands, 95% of all children had teachers with a university degree; in Norway, the figure was 98.6%. In Sweden, 88% of the students attending effective schools had teachers with a university qualification, and 92% of students attending non-effective schools had teachers with a university degree. Also of interest in relation to Sweden is that teachers of 95% of the students had a teaching certificate, as did teachers of 92% of the students in non-effective schools.

Figure 7: Effective and Non-effective Schools by Location in Four of the Seven Selected Countries



Time spent in seminars on reading

Figure 8 presents the overall time teachers spent at reading seminars in the selected seven countries. In Germany, teachers in effective schools spent less time at seminars than did teachers in non-effective schools. Teachers of 75% of the children at effective schools spent fewer than six hours at the reading seminars; teachers of 67% of the children attending non-effective schools spent the same amount of time. In Lithuania, the situation was similar. In the Netherlands, teachers of 31% of students in effective schools had not attended a seminar on reading; the same was true for teachers of 23% of students at non-effective schools. In New Zealand, teachers of 44% of the students at effective schools had spent fewer than six hours at reading seminars, while teachers of 36% of students at non-effective schools had spent six hours or under at these seminars. In Norway, teachers of 70% of the students at effective schools had attended fewer than six hours of reading seminars, but so too had teachers of 65% of the students at non-effective schools.

The situation in Slovenia was a little different. Teachers of 20% of the students at effective schools had not attended reading seminars; the same was the case for teachers of 11% of the students at non-effective schools. If we look at the two categories “none” and “less [sic] than six hours,” then we can see that both groups of teachers spent roughly the same amount of time at reading seminars. In Sweden, differences were particularly evident in relation to the category 16 to 35

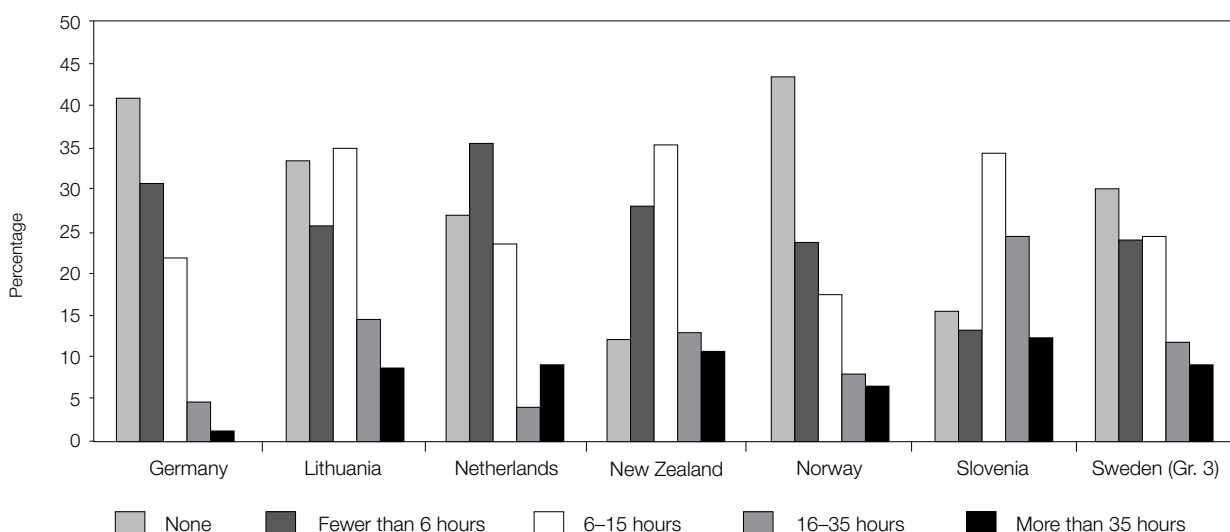
hours. Teachers of 15% of the students at the effective schools spent 16 to 35 hours at seminars while teachers of 8% of the students in non-effective schools spent this amount of time. At the highest category (more than 35 hours), teachers of 12% of the students at non-effective schools in Sweden spent this amount of time at the reading seminars, as did teachers of 7% of students at the effective schools.

Job satisfaction

It was evident from the PIRLS data that teachers of the effective schools gained greater satisfaction from their work than did their colleagues at non-effective schools, perhaps because teachers at effective schools put considerably more effort into the visible effect of their work, that is, student achievement on external assessments and the like. That said, clearer answers are evident when we look at the data relating to one of the questions on the school questionnaire. Here, school principals were asked to rate (“second-hand”) their employees’ level of job satisfaction.

In Germany, the principals reported that teachers of 68% of students in the effective schools were highly satisfied with their job; the same was the case for teachers of 62% of the students at non-effective schools. In Lithuania, the respective figures were 66% and 60%, but the reverse was evident in the Netherlands, where teachers in the less effective schools were more satisfied with their job than were the teachers in the effective schools. In the Netherlands, not one

Figure 8: Percentage of Time Teachers in the Seven Selected Countries Spent at Reading Seminars



principal reported a teacher dissatisfied with his or her job. In New Zealand, teachers of just over 98% of all students were reported to be satisfied with their job. In Norway, teachers of students in effective schools were more content with their work. In Slovenia, the majority of “very satisfied” teachers were in effective schools, but teachers who were only “content” with their work were in non-effective schools. In Sweden, teachers of 25% of the students at effective schools were “very satisfied,” while teachers of 16% of students at non-effective schools had this degree of satisfaction in their work.

Other statistically significant factors from the PIRLS database

Among the many other variables potentially demarcating effective and non-effective schools, the following were statistically significant across the seven countries:

- Amount of time that teachers expected to spend on language homework
- Student progress assessed according to teacher’s own professional opinion
- Age of teacher
- Reading specialist available
- No access to specialists
- School reports that teaching is affected by a shortage of library books
- Cooperation among teachers in the school
- Index of availability of school resources.

Variables found to have statistically significant relationships in some but not all seven countries were:

- Teachers using multiple methods
- Student attitudes toward reading
- School-based programs for teaching
- Teachers’ expectations of students
- The amount of time students reported spending on reading homework.

Conclusion

Effective and non-effective schools are not schools with high or low achievement. Rather, effectiveness is predicated on achievement being higher than one might expect given the circumstances of the students, such as their socioeconomic background. Low achievement appears to be reserved for students from low-income families and neighborhoods, whereas students from

high-income families seem to have a guaranteed higher level of achievement. This “guaranteed achievement” is not theoretical but can be observed in almost every PIRLS 2001 country. Its presence on the reading achievement scale is mediated in terms of the average achievement of and socioeconomic factors evident for students in each country. But other factors also appear to be involved. Schools with students whose parents generally have a high level of educational attainment will still show a spread of reading achievement, but the important point is that the average reading achievement of students in these schools is never as low as that of schools whose students have less well educated parents.

In some countries, this “guaranteed achievement” is set around the national average, which is why ineffective schools can be “overlooked.” However, schools at this level of achievement can be seen as effective if the parents of the children are wealthy and/or well educated. The nature of public opinion is such that these schools tend to be judged according to how the students’ parents and their economic wellbeing are perceived rather than in terms of the true effectiveness of the schools. Essentially, determining effective and non-effective schools should not be made according to society’s ideals and prejudices but rather in terms of which schools need assistance, and what form that assistance should take.

Any effort to identify effective schools immediately poses another concern: is it more likely to expect a school from a low SES environment to exceed the expectations, or to expect that a school that draws its students from a high SES environment will trigger higher-than-expected achievement? From international comparisons, we see that what is deemed high achievement in one country can be considered low in another country. As such, practically any result, no matter how good, can be higher still, even for the best-performing schools within the top-performing countries.

It thus seems that some variables used to determine effectiveness represent a lifestyle in a certain country rather than school effectiveness per se. One such variable is school location. School location does not have a causal link with school effectiveness, but it does indicate where we are most likely to find such schools.

In Germany, the effective schools were as follows: schools in rural areas; schools where teachers were highly satisfied with their jobs; schools where teachers had university degrees; and schools where teachers did not attend more than six hours of reading seminars per year.

In Lithuania, the effective schools were more likely to be in urban areas, to have teachers who spent fewer than six hours per year at reading seminars, and to be those whose teachers were highly satisfied with their jobs. No association was found between effectiveness and teachers' level of education.

In the Netherlands, the most effective schools were those in rural areas and those whose teachers did not attend a seminar on reading in a year. They were also those schools where teachers reported they were less satisfied with their jobs, despite their principals indicating otherwise. Practically all teachers (of 95% of the students) had university degrees, so this did not account for differences between effective and non-effective schools.

For Norway, the (relatively) more effective schools could be found in suburban areas and also included those whose teachers spent fewer than six hours per year attending seminars and who tended to be content with their jobs. In Norway, like the Netherlands, almost all teachers (of 98% of the students) had a university degree.

In New Zealand, effective schools were most frequently found in rural areas. Effective schools were also those where teachers spent fewer than six hours a year at reading seminars. Ninety-eight percent of New Zealand students had teachers who said they were satisfied with their jobs.

In Sweden, the more effective schools were those where teachers spent 16 to 35 hours per year at reading seminars, where teachers were more, rather than less, satisfied with their jobs, and where teachers had teaching certificates.

Finally, in Slovenia, the effective schools were those where teachers spent fewer than six or more than 16 hours per year at reading seminars and whose teachers reported being very satisfied with their jobs.

References

- Doupona Horvat, M. B., Brecko, K., Kocis, A., Krevh, M., & Rozman (2006). *School factors which affect reading achievement* (research report). Ljubljana: Educational Research Institute.
- Postlethwaite, T. N., & Kenneth, N. R. (1992). *Effective schools in reading—implications for educational planners: An exploratory study*. The Hague: International Association for the Evaluation of Educational Achievement.

Changes in reading variations and their relationship with socioeconomic status at school and individual levels in trend countries

Kajsa Yang-Hansen

*Department of Education and Educational Research
Gothenburg University, Gothenburg, Sweden*

Introduction

To explore the changes in students' reading literacy since 1991, the Progress in International Reading Literacy Study 2001 (PIRLS 2001, Mullis, Martin, Gonzalez, & Kennedy, 2003) offered an opportunity for those countries that participated in the International Association for the Evaluation of Educational Achievement's (IEA) Reading Literacy Study (RLS 1991, Elley, 1994) to take part in the so-called 10-Year Trend Study (Martin, Mullis, Gonzalez, & Kennedy, 2003). The Trend Study 2001 (RLS 2001) replicated the reading test and questionnaires administered in RLS 1991. Nine countries took part in the trends study, namely Greece, Hungary, Iceland, Italy, New Zealand, Singapore, Slovenia, Sweden, and the United States.

Changes between the two samples (i.e., between RLS 1991 and RLS 2001) have been observed in the country means of the reading achievement score and of some important background variables relevant to reading and its development. (For summary findings, see, for example, Martin, Mullis, & Kennedy, 2003, pp. 3–5.) However, the changes in country means are not the only feature of interest; the changes between RLS 1991 and RLS 2001 in the reading achievement variation, students' background characteristics, and the relationship between the two are also of great importance.

Previous analyses of the RLS 1991 data showed that the reading achievement test comprised multiple dimensions (Gustafsson, 1998; Yang, 1998). The changes in reading achievement variation between 1991 and 2001 may differ in different reading dimensions as well as at different levels of observation. The same may be true for students' background characteristics, such as socioeconomic status (SES) and the effects of these characteristics on reading dimensions. These changes may reflect changes in educational policies and practices, as well as in social and economic factors in society.

Decentralization and free school choice have become a part of global trends in educational reform (see, for example, Chapman, 1996; Gonnée van Amelsvoort & Scheerens, 1997). Countries may take different actions or a different focus in their educational reforms. Information relating to cross-country differences in changes in reading achievement variation and to the effect of SES on reading may thus also be interesting and valuable knowledge in the international educational context.

The aim of the current study therefore was to examine cross-country differences in reading variance, in SES, and in the effect of SES on reading achievement between RLS 1991 and 2001. Three issues are examined: cross-country differences in the variance changes in reading dimensions at individual and school levels between RLS 1991 and 2001; measurement properties of SES and factor score estimation of SES for each country; and the SES effect on reading properties in all countries.

Hypotheses

In the early 1980s, Burstein and his colleagues noted the impact of educational decentralization on educational equality. In their cross-country comparison study of SES effects on academic achievement (Burstein, Fischer, & Miller, 1980), they showed that the relationship between school SES and achievement was much higher in countries whose educational systems were decentralized and selective, as in, for example, the United States, than in countries with centralized comprehensive systems, such as Sweden. The authors argued that school SES reflects the characteristics of the student intake, and that it mirrors the demographic characteristics of the community. Within the auspices of local control and finance of schools, community characteristics "affect the school's ability to seek better qualified teachers, higher quality learning materials,

and better physical facilities” (Burstein et al., 1980, p. 216). These aspects, in turn, have a strong impact on educational quality and school academic achievement. Thus, a higher relationship between community characteristics, school SES, and academic achievement may be expected in decentralized education systems.

During the 1990s, a trend toward decentralization and deregulation of schooling was observed internationally, especially in terms of the implementation of decentralized school finance and voucher systems with school choice (Beare & Boyd, 1993; Hanson, 1997; Whitty, Power, & Halpin, 1998). Evaluation studies were conducted to examine the consequences of such reforms and policies on educational quality, equality, and efficiency. These studies indicate that the recent school reforms and policy changes may have increased between-school segregation, with respect to school achievement, and school SES and ethnicity composition.

Gewirtz and her colleagues (1995) studied the process of parental school choice, and concluded that free school choice is a key factor in maintaining and reinforcing social class inequalities in the school sector. As Coffey (2001) has pointed out, “... school choice does involve the appropriation of social/cultural capital and the capacity and opportunity to become informed in order to make real choice—and that these are differentiated along class lines” (p. 28; see also, Ball, 1994; Woods, Bagley, & Glatter, 1998).

Studying the effects of school choice in 54 Scottish communities, Willms (1997) found that well-educated parents with higher SES were more likely to practice school choice than were working-class parents during the first few years of reform. However, when the policy was established, more working-class parents were involved in choosing public schools, while the middle class parents simply avoided the public school system and sent their children to private schools. Willms thus concluded that parental choice has a stratifying effect that contributes to the increase in between-school SES and ethnic segregation, which, in turn, is likely to result in greater inequalities in attainment (Willms, 1997). Similar findings of the segregation effect of school choice are also found in studies by, for example, Blair (1994), Conway (1997), Hatcher (1998), Goldring (1995), Levin (1998, 2002), McEwan (2000), Waslander and Thrupp (1997), and Willms and Echols (1992).

In Sweden, to give an example from one of the countries under consideration in this paper, empirical evidence shows that parents’ SES had an increasing effect on school achievement, even though the impact of immigration status remained constant (Skolverket, 2006). Gustafsson (2006) examined between-school segregation in Sweden with respect to parents’ educational level, ethnic background, and the students’ school grades over the previous decade. He found that educational segregation increased during the 1990s in terms of an increased variation in the aforementioned aspects among schools (see also, Skolverket, 1996, 2003).

We can thus assume that school reforms may affect school achievement and SES effects on school achievement through a set of causal links. First, families from different SES backgrounds may exercise the free school choice differently (see, for example, Willms, 1997). Those families that are better off socioeconomically are more likely to take advantage of the school market system. Because families with an immigration background are more likely to be in the lower SES strata, considerable social-class and ethnic variations in school choice may be apparent. Empirical evidence also shows that schools with a greater percentage of students from high-SES family backgrounds provide more effective learning environments, a situation that in turn leads to a higher school average achievement (see, among others, Brooks-Gunn, Duncan, Klebanov, & Sealander, 1993; Leventhal & Brooks-Gunn, 2000; Rutter, Maughan, Mortimore, & Ouston, 1979; Thrupp, Lauder, & Robinson, 2002; Willms, 1992). Thus, family background differences cause a differentiated school choice, which leads to achievement inequalities and social segregation between schools (Gustafsson, 2006; Hoxby, 2001; Skolverket, 2003). Second, communities have varying ability to raise resources for their schools and other related social services. Schools in affluent communities may attract experienced teachers because of their better school resources and more motivated students (Metz, 1990; Orfield & Eaton, 1996). Thus, inequalities in these school contextual factors lead to increased differences in performance between schools.

We can therefore hypothesize that the variation of school-achievement and SES effects on academic achievement increased between year 1991 and 2001. These disparities may be captured by, for example,

differences in the measurement properties in both reading achievement and SES, and the relationship between the two. Moreover, cross-country differences in the variation of school achievement and SES effects may reflect differences in policies and practices between education systems. The findings of the research undertaken by Burstein et al. (1980) support this hypothesis, as does work by Yang (1998) and Yang and Gustafsson (2003).

As we can see from Table 1, countries such as Hungary, Iceland, Italy, New Zealand, Sweden, and the United States have been through a decentralized transformation of their school systems, while the education systems in Greece, Singapore, and Slovenia are more centralized and homogeneous. We can expect that for the countries listed in the former group above, we would see an increase in the between-school

differences in variances of reading achievement in RLS 2001 and a strengthening of the school SES effect on reading achievement. As for Greece, Singapore and Slovenia, we could expect to see no significant change between RLS 1991 and RLS 2001. Note, however, that even though the countries are classified as decentralized or centralized, they differ substantially from one another in the degree of (de)centralization, educational practices, and social and cultural circumstances. Therefore, we can also expect to see variations in the changes in reading variance and SES effects among countries within the same category.

Accordingly, the main hypothesis in the current study is that variation in school reading achievement and school SES effects on reading achievement would have increased for countries that underwent decentralization deregulation reforms and that

Table 1: A Brief Summary of Characteristics and Changes in Education in the Reading Literacy Study Trend Countries

Country	Characteristics
Greece	Highly centralized in all aspects of education. At the local and school level, involvement is particularly limited.
Hungary	In 1989, the educational monopoly of the Hungarian state was abolished and the autonomy of the local and the institutional levels increased. This created a greater variety of schools. Highly decentralized after educational reforms.
Iceland	In 1995, the structure of the system was changed when the new Junior Schools Act was passed. The main feature of the new structure was decentralization, entailing financial and professional independence. Funding of all resources other than instructional materials for compulsory school is the responsibility of the local authorities of the individual schools.
Italy	In 1997, a reform of the Public Administration initiated a process of decentralization aimed at simplifying administrative procedures. This shifted power from the central to the peripheral authorities—the regions and the municipalities—and to the schools. However, school finance is still controlled by central government.
New Zealand	Highly decentralized system already stabilized in the early 1990s. Student enrolment, school type, the socioeconomic demographics of the surrounding community, and the school's property profile are all factors that impact on school funding.
Singapore	The education system in Singapore is characterized by a high degree of homogeneity and cohesion in curriculum coverage across schools. Educational decision making about national policy matters generally is centralized.
Slovenia	The state is responsible for virtually all aspects of education, and education is financed primarily from the state budget, with a small share from municipal budgets.
Sweden	Decentralization and deregulation started in the late 1980s. According to OECD educational indicators, Sweden has become one of the most decentralized education systems among the OECD countries.
United States	The states have delegated the responsibility of operating schools to local governments, which in turn have assigned the role to distribute educational aid, establish policy for graduation requirements and teacher certification requirements, provide curriculum guidance, conduct student assessments, and ensure that efficient and effective school opportunities are made available to every eligible child in each state.

Source: Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Flaherty, C. L. (Eds.). (2002). *PIRLS 2001 encyclopedia. A reference guide to reading education in the countries participating in IEA's Progress in International Reading Literacy Study (PIRLS)*. Chestnut Hill, MA: Boston College.

established free choice of schools during the decade 1991 to 2001. We can also hypothesize that variation in these changes would be evident across countries, depending on the degree of disparity among their respective education systems.

Data

The data analyzed in the current study are from the IEA RLS 1991 and RLS 2001 studies. The target population in both studies was nine-year-olds, most of whom were Grades 3 or 4 students. Table 2 shows the grade tested and the number of students and schools involved in the analysis. For most of the countries, the grade tested is the same in the 1991 and 2001 samples. However, New Zealand and Iceland adjusted their sample grade level in order to obtain comparability of the two studies. The sample sizes in RLS 1991 are much larger compared to those of RLS 2001 in all the countries except for Hungary and Sweden. The relatively small school sample sizes in RLS 2001 may affect the analytical power when testing the current hypothesis with a two-level modeling technique. The number of students in parentheses is the number of students who completed data on all the reading test variables.

In RLS 1991, every student in the sample was given a set of reading tests. All together, there were 15 reading passages comprising three text types in the test, namely, narrative, expository, and documents. The 15 reading passages were divided into two booklets (i.e., Booklet 1 and Booklet 2) and administered to the participants on two separate test occasions (for

detailed information, see Elley, 1994). The same test instrument also was administered to the RLS 2001 sample (Martin, Mullis, & Kennedy, 2003). In the current study, the reading passage scores were used to investigate the dimensionality of the reading achievement test. Each passage score was obtained by summing up the multiple-choice item scores of each passage. To reduce the number of variables in the analysis and to avoid non-convergence of models caused by the high correlation among the variables in the covariance matrix, the document passage scores were further summed into three variable parcels: D1ILMA, D2BTBS, and D2CTTP.

The indicators of socioeconomic status (SES) were selected from the student questionnaire (StQ), in which students were asked to answer some questions on the ownership of a set of household items and their own possessions. The same possession items were administered in RLS 1991 and RLS 2001 in all the countries except Sweden. National options were added in order to capture the economic and social changes in Sweden in the decade between 1991 and 2001. These material possessions comprised two categories. One category represented the family economic situation, such as dishwasher, summer-house, sailing boat, cars, etc. The other signified cultural preferences, with piano, encyclopedia, and books being but three examples. According to Bourdieu (1984, 1997), these possession items are objectified states of economic capital and cultural capital that individuals may be able to use to position themselves into certain social strata.

Table 2: Sample Size for Participating Students and Schools in the Nine Countries in RLS 1991 and RLS 2001

Country	RLS 1991			RLS 2001		
	Grade	Students	Schools	Grade	Students	Schools
Greece	4	3,516 (3,479)	175	4	1,109 (1,107)	68
Hungary	3	3,009 (2,910)	144	3	4,707 (4,696)	216
Iceland	3	3,961 (3,921)	179	4	1,797 (1,767)	65
Italy	4	2,221 (2,206)	154	4	1,590 (1,589)	92
New Zealand	5	3,016 (2,941)	176	4	1,188 (1,166)	73
Singapore	3	7,326 (7,320)	206	3	3,601 (3,597)	98
Slovenia	3	3,297 (3,283)	140	3	1,502 (1,493)	75
Sweden	3	4,301 (4,183)	123	3	5,361 (4,905)	148
United States	4	6,433 (6,338)	164	4	1,826 (1,816)	85

Two variable parcels, HOMEPR and STUDPR, were created by summing up the home possession items and student-owned possession items respectively. Home possession items, representing cultural preferences, such as piano and encyclopedia, were excluded from the parceling. They were instead used at item level to represent the cultural aspect of family SES.

Sampling

Two-stage stratified cluster sample design was used in both RLS 1991 and RLS 2001. In general, schools that could be stratified were sampled at the first stage. At the second stage, one or more classes of the target grade in each sampled school were selected intact. Ultimately, all students in the selected classes were included in the sample (Martin et al., 2003). Data selected in such a way, however, are interdependent and hierarchical in nature because of students in the same classroom being similar and classrooms in the same school being similar. Moreover, stratification in the primary sampling stage may cause an unequal inclusion probability among schools in different strata. To take the effects of unequal inclusion probabilities into account, weights were imposed when analyzing the data.

Analytical techniques

Two-group two-level structural equation modeling (SEM) with latent variables was used to measure reading dimensionality and the effect of SES on reading achievement. An analytical method of this kind has several advantages. First, with two-level SEM, the overall reading achievement variation can be decomposed into variation at the student level and variation at the school level. The reading variance at each level can then be further broken down into variance due to each of the reading dimensions and measurement error.

This decomposition process is done by simultaneously estimating a pooled-with-covariance matrix and a pooled-between-covariance matrix (for technical descriptions, see, for example, Muthén, 1989, 1991, 1994; Muthén & Muthén, 2004). Two-group modeling allows statistical tests of the differences between the two groups (i.e., RLS 1991 and RLS 2001) in both the manifest variables and the parameter estimates. By comparing the chi-square test statistics and the degrees of freedom between more

and less restricted models, one can test whether or not certain parameter(s) significantly differ between two groups (Gustafsson & Stahl, 2004).

Latent variables, being error-free in nature, may help to detect true relationships between theoretical concepts measured by some manifest variables, and to detect which relationships may otherwise be attenuated by measurement and systematic errors. Moreover, factor scores estimated from a measurement model can be used in standard statistical analyses in the same way as an ordinary observed variable. When the set of manifest variables has low reliability and a large amount of missing data, estimation of factor scores may be a better scaling approach to use with a set of manifest variables, compared to, for example, a computed index (i.e., an average score of a set of observed variables; see Yang-Hansen, Rosén, & Gustafsson, 2006). In the current study, individual factor scores were estimated for SES and were used later in the estimation of the effect of SES on reading achievement.

Process for the analyses

The analyses in the current study were carried out in three steps using Mplus in the STREAMS modeling environment (Gustafsson & Stahl, 2004). In the first step, a measurement model of SES was estimated and the factor score estimates were saved. In the second step, the dimensionality of reading achievement was examined by fitting a two-group two-level structural equation model. The distribution of total reading variance into different reading components and levels was achieved as one of the results. Finally, the reading dimensions were regressed on the factor score estimate at both student and school levels in two-level models.

Results

The measurement model of SES

To maintain consistency and comparability of the concept SES, the two data sets, RLS 1991 and RLS 2001, were merged together. The factor score representing the SES latent variable was estimated by fitting a single-factor measurement model to all individual data from the merged data set. Thus, the SES factor score variables in each sample were derived with the same scale of variances and covariance.

Table 3 shows the goodness-of-fit of the SES measurement model in each country. The model fit

was acceptable for all countries, even though the single factor model had a better fit to the data from some countries than from others.

The model fit was considerably better for Hungary, New Zealand, and Singapore than it was for the rest of the countries. The United States and Iceland had a rather poor fit. The poor fit might imply that one single factor is inadequate for explaining the variance in SES indicators. Adding an extra parameter, another factor, or a covariance between two residuals may improve model fit. However, the factor scores estimated by a single-factor unmodified poor-fit model and a modified better-fit model differed only marginally. The correlation between the two-factor score estimates were close to unity, which suggests that the model fit was not important to the usefulness of the factor score estimates (Yang-Hansen et al., 2006). Therefore no further improvement was done for these SES measurement models.

Table 4 presents the factor loadings of the manifest variables measuring SES. Note the similar pattern in

almost all the countries: the indicators of the economic aspect of SES loaded higher than those of the cultural aspect of SES, with HOMEPI being the highest. Exceptions can also be observed. In Sweden, the factor loadings were fairly even among all the SES indicators. BOOKS held the highest factor loading. In the United States, however, over 90% of the SES variation was explained by students' own possessions.

The precision of the factor score estimates is measured by factor score determinacy, which is related to the standard error of an individual's estimated score on a factor. The values for this coefficient range from 0 to 1, with larger values indicating better measurement of the factor by the observed indicators. A missing data modeling technique was used to estimate the measurement model of SES. Mplus detected missing data patterns and estimated a model for each of these patterns, and calculated a factor score determinacy coefficient respectively. For the missing patterns that had lower data coverage, the factor score determinacy coefficients were also rather low. Table 5 shows the

Table 3: Goodness of Fit for the Measurement Models of SES in the Nine Trend Countries

Country	χ^2	df	RMSEA
Greece	14.27	2	.037
Hungary	4.44	2	.013
Iceland	53.71	2	.067
Italy	17.70	2	.045
New Zealand	5.22	2	.020
Singapore	12.61	5	.012
Slovenia	34.42	2	.058
Sweden	228.13	9	.051
United States	79.94	2	.069

Note: $p < .00$.

Table 4: Standardized Factor Loadings in the One-factor Measurement Model of Capital in the Nine Trend Countries

Indicators	Standardized factor loading								
	GRC	HUN	ICE	ITA	NZL	SGP	SVN	SWE	USA
HOMEPI	.75	.72	.57	.69	.67	.68	.69	.42	.57
STUDP	.69	.57	.52	.50	.52	.47	.57	.42	.96
BOOKS	.44	.40	.15	.31	.40	.44	.33	.44	.34
NEWSP	.20	-.12	.16	.23	.16	.25	.26	.29	.12
PIANO						.49		.36	
ENCYC								.39	

Note: GRC = Greece, HUN = Hungary, ICE = Iceland, ITA = Italy, NZL = New Zealand, SGP = Singapore, SVN = Slovenia, SWE = Sweden, USA = the United States.

Table 5: Average Factor Score Determinacy for the General SES Factor

Country	Average factor score determinacy	Cronbach's alpha
Greece	.69 (13 missing patterns and 24 missing cases)	.57 (219 missing cases)
Hungary	.70 (8 missing patterns and 61 missing cases)	.46 (222 missing cases)
Iceland	.54 (10 missing patterns and 44 missing cases)	.35 (262 missing cases)
Italy	.59 (8 missing patterns and 15 missing cases)	.44 (105 missing cases)
New Zealand	.60 (10 missing patterns and 60 missing cases)	.48 (159 missing cases)
Singapore	.67 (18 missing patterns and 9 missing cases)	.53 (314 missing cases)
Slovenia	.68 (10 missing patterns and 11 missing cases)	.50 (341 missing cases)
Sweden	.64 (18 missing patterns and 353 missing cases)	.43 (983 missing cases)
United States	.71 (10 missing patterns and 58 missing cases)	.58 (220 missing cases)

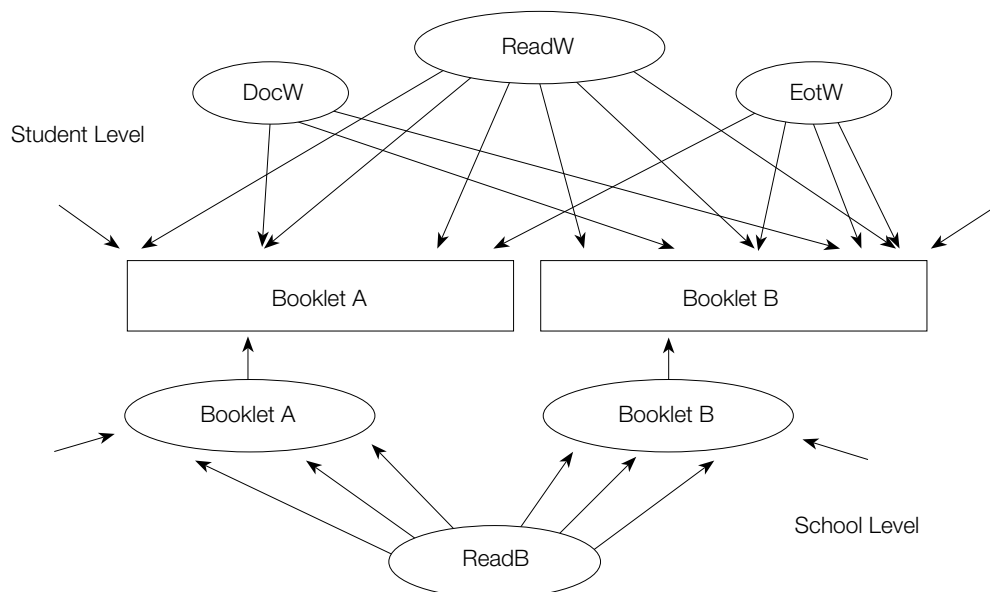
average factor score determinacy and the Cronbach alpha in each country.

In general, the average factor score determinacy was higher than the Cronbach's alpha for each country. As Yang-Hansen et al. (2006) have shown, one of the advantages of using a factor score approach is that factor scores can be estimated for individuals who have not responded to all the items. This approach allows fuller use of the available data, and reduces the risk of bias caused by list-wise deletion of cases with missing data.

In summary, the results showed that the measurement property of the SES factor was consistent across the nine countries, and that the factor score approach ensured the comparability of SES over the two populations.

Changes in variances of reading components between RLS 1991 and RLS 2001

The measurement characteristics of the reading achievement tasks in RLS 1991 have been investigated in previous research (see, among others, Gustafsson, 1997; Yang, 1998; Yang & Gustafsson, 2003). The basic model structure of the reading achievement test is sketched in Figure 1. Here we can see that the two-level multidimensional model fits the data. At the student level, a general reading ability factor (ReadW) is identified along with two narrow residual factors—a document reading factor (DocW) and a speed factor (EotW). At the school level, only one general reading factor is specified to account for the school-level reading variation. This model was tested in the current study for both RLS 1991 and RLS 2001 simultaneously in a two-group model.

Figure 1: Structure of the Reading Achievement Measurement Model

Before conducting the analysis, the design effect was calculated for the reading passage scores of the RLS 1991 and RLS 2001 samples (see Table 6). The design effect (DEFF) expresses the increase in sample size that must be made to achieve the same power as would occur if independent units were sampled. DEFF is often calculated as

$$\text{DEFF} = 1 + \text{ICC} (n - 1), \text{ where}$$

DEFF is the design effect, ICC is the intra-class correlation (ICC), and n is the average cluster size. When DEFF is greater than 2, it means that the cluster effect in the data needs to be taken into account during the estimation because otherwise the standard errors and the chi-square test statistic will be biased (Muthén & Satorra, 1995). For most countries in the present analysis, the design effect of reading passage scores was greater than 2. Exceptions were Iceland in both RLS 1991 and 2001, and Slovenia in RLS 2001, where the design effects were around 2. However, in the current

study, the two-level structure of reading achievement may be interesting even for these two countries within the context of the cross-country comparison.

In general, the ICCs were similar in RLS 1991 and RLS 2001 in Hungary, Iceland, Italy, New Zealand, and Sweden. In Greece, Slovenia, and the United States, the ICCs were larger in the RLS 1991 sample than in the RLS 2001 sample. The ICCs in the Singapore RLS 2001 sample were much higher than those in the Singapore RLS 1991 sample. These differences in ICCs in each country between the two samples were reflected in the changes in the variance components of reading achievement at cluster level.

The next step was to fit the two-group two-level measurement models of reading achievement, as sketched previously, to the RLS 1991 and 2001 data in all countries. They fitted the data very well when no constraints of equality were imposed over the RLS 1991 and RLS 2001 populations. We then carried out chi-square difference tests to test the hypothesis

Table 6: Design Effect of Reading Achievement Scores and SES Indicators in Trend Countries

		GRC	HUN	ICE	ITA	NZL	SGP	SVN	SWE	USA
RLS 2001	Cluster size	16	22	27	17	16	37	20	33	21
	N1BIRD	2.95	3.31	2.04	2.44	3.55	11.44	2.52	2.28	3.40
	N1DGS	2.95	4.57	2.04	3.08	3.55	13.24	2.14	3.56	4.80
	E1WLR	3.70	4.99	1.78	2.44	2.65	10.00	2.33	3.88	3.00
	D1ILMA	2.05	4.57	1.78	3.08	3.10	9.64	1.95	2.60	3.60
	E2SND	2.35	3.94	2.04	2.60	3.55	8.92	1.57	2.28	3.80
	N2SHK	1.60	4.15	2.30	2.76	3.40	11.80	2.52	2.60	4.20
	E2MRM	2.35	3.31	1.52	2.76	2.80	8.20	1.76	2.60	3.20
	D2BTBS	2.65	5.62	2.04	4.36	3.25	11.80	1.95	3.56	4.80
	D2CTTP	2.65	5.20	2.30	3.24	3.40	9.64	1.95	3.24	5.00
RLS 1991	E2TRE	3.55	3.94	2.04	4.04	2.50	8.92	2.33	2.60	4.00
	N2GRP	2.80	4.15	2.30	3.24	3.40	13.24	2.71	2.92	4.60
		GRC	HUN	ICE	ITA	NZL	SGP	SVN	SWE	USA
	Cluster size	20	20	22	14	17	36	23	34	37
	N1BIRD	4.80	2.52	2.05	3.34	3.08	5.20	2.54	2.65	4.60
	N1DGS	4.61	4.04	2.05	3.21	3.08	6.60	2.32	2.65	5.32
	E1WLR	3.47	4.23	2.05	2.56	2.92	4.50	2.54	2.32	3.52
	D1ILMA	4.23	3.28	1.63	3.73	3.40	5.20	2.54	3.31	4.96
	E2SND	4.04	2.90	1.42	2.30	2.44	4.15	2.10	2.65	3.88
	N2SHK	4.99	2.52	1.84	2.30	2.76	4.50	2.32	2.65	4.24
	E2MRM	3.66	2.52	1.84	3.21	2.28	4.50	2.32	2.98	4.24
	D2BTBS	5.75	3.85	2.26	4.51	3.72	4.85	2.54	3.97	5.68
	D2CTTP	4.80	3.66	2.05	3.47	2.92	4.15	2.32	2.98	4.24
	E2TRE	5.37	4.04	2.26	4.25	3.08	4.15	3.42	2.98	5.68
	N2GRP	3.66	3.47	2.26	2.95	2.76	4.85	3.20	2.65	5.68

that model parameters estimated by the measurement model would be equal in the two populations of RLS 1991 and RLS 2001. Comparison was made between a model with complete constraints over the two populations and a model without. As is shown in Table 7, the chi-square difference test was significant for all the countries, except New Zealand.

Table 8 presents the estimations of variance components of each reading dimension at both individual and school level. Iceland had almost no variance in reading achievement at school level in the RLS 2001 population, and the same can be said for Slovenia in relation to RLS 2001. Substantial

Table 7: Chi-square Difference Test between the Complete Constraint and Complete Non-constraint Model

Country	$\Delta\chi^2$	Δdf
Greece	144.51	53
Hungary	248.94	53
Iceland	361.93	53
Italy	71.36	53
New Zealand	57.56	51
Singapore	133.15	53
Slovenia	128.90	53
Sweden	121.53	53
United States	751.09	50

differences between RLS 1991 and RLS 2001 in the school reading variances were found in Greece and Singapore.

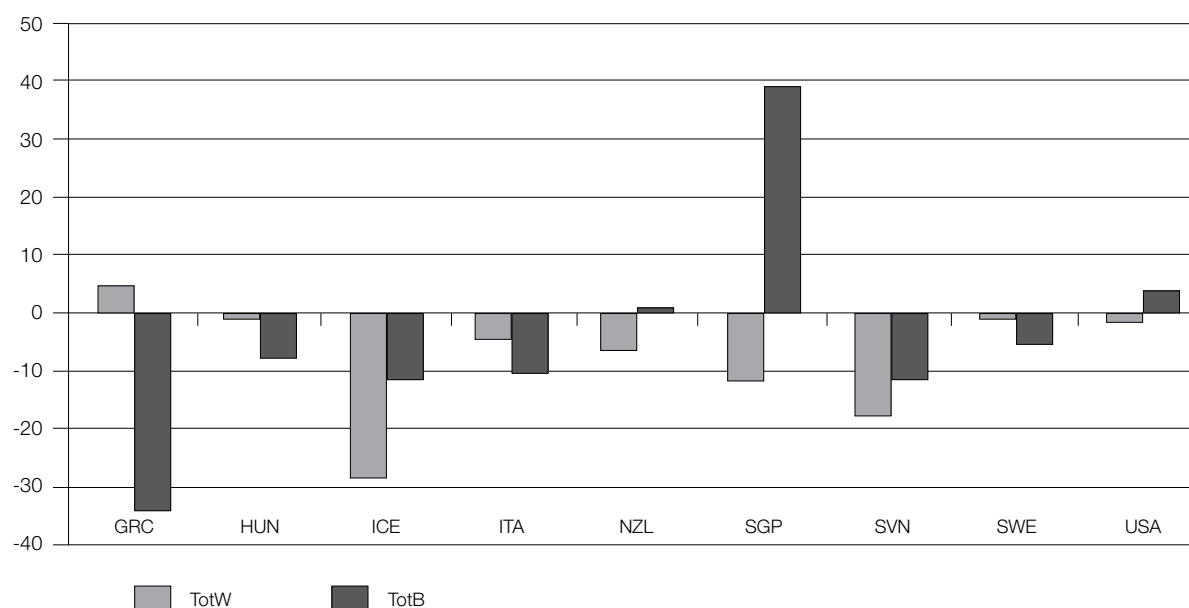
As shown in Figure 2, the reading variance decreased in most countries at both student level and school level across the 10-year period. The largest drop in between-school differences in reading achievement occurred in Greece, followed by Slovenia, Iceland, and Italy. Hungary and Sweden also reduced the school-level reading variation somewhat, while the student-level reading achievement differences were relatively stable over the 10-year period. Singapore had a great increase in school-level reading variance in RLS 2001, but the reading achievement differences between students dropped. A similar pattern of variance changes was evident in New Zealand and the United States. The only country with increased variation between individual students was Greece.

In summary, the results of the two-group two-level measurement model showed that the estimated reading achievement properties differed significantly between RLS 1991 and RLS 2001 populations in all countries except New Zealand. For most of the countries, both between-student reading variance and between-school reading variance decreased. A dramatic increase was found at school level in Singapore.

Table 8: Variance Components of the Reading Components at Individual and School Levels in the RLS 1991 and 2001 Trend Countries

Country	Individual Level					School Level		
	ReadW	DocW	EotW	ErrorW	TotW	ReadB	ErrorB	TotB
GRC01	74.1	2.1	7.4	11.6	95.3	7.4	.8	8.2
GRC91	68.7	2.2	6.6	13.2	90.8	41.2	1.0	42.3
HUN01	72.2	2.6	22.1	12.7	109.5	31.5	1.1	32.6
HUN91	74.1	2.3	20.7	13.5	110.5	39.8	.7	40.5
ICE01	91.84	2.7	10.6	11.7	116.9	.7	.2	.9
ICE91	114.7	3.2	14.5	12.7	145.0	12.2	.2	12.3
ITA01	68.0	2.5	7.9	11.3	89.8	24.6	.8	25.4
ITA91	72.2	2.5	7.3	12.3	94.2	34.2	1.4	35.6
NZL01	101.0	2.7	3.0	11.3	117.9	35.3	.3	35.5
NZL91	105.6	2.5	4.9	11.2	124.3	33.9	.2	34.2
SGP01	66.4	2.4	.5	10.7	79.9	69.6	.3	70.0
SGP91	77.3	1.9	1.2	11.0	91.3	30.6	.1	30.7
SVN01	85.1	1.8	12.2	12.4	111.5	3.4	.4	3.9
SVN91	100.7	2.4	12.9	13.3	129.3	14.8	.6	15.4
SWE01	113.5	4.3	21.8	12.9	152.6	11.2	.3	11.5
SWE91	119.9	3.2	18.3	12.0	153.3	16.4	.3	16.6
USA01	59.8	2.3	8.1	11.2	81.4	32.3	.5	32.8
USA91	66.5	2.1	3.2	10.9	82.7	28.6	.2	28.7

Figure 2: Differences in Reading Variance at Student and School Levels between RLS 1991 and 2001



Note: TotW = total reading variance at within-level (i.e., student level); TotB = total reading variance at between-level (i.e., school level).

Exploring the effect of SES on reading achievement at school level

The reading dimensions in the two-group two-level models were regressed on the SES factor score. The measurement model of reading achievement was completely constrained over RLS 1991 and RLS 2001, such that the consistency of the measurement properties in reading achievement were ensured for the two populations, so allowing the relationship between SES and reading achievement dimensions to be compared on the same ground. The chi-square test statistics and the RMSEA values presented in Table 9

Table 9: Evaluation of Two-Group Two-level Model of SES Effect on Reading Achievement in RLS 1991 and 2001 Trend Countries

Country	χ^2	Df	RMSEA
Greece	881.01	261	.032
Hungary	1313.69	261	.033
Iceland	1369.46	261	.039
Italy	603.90	261	.026
New Zealand	580.60	261	.024
Singapore	999.55	261	.023
Slovenia	826.43	261	.030
Sweden	1206.68	261	.028
USA	1477.53	261	.034

indicate that the model was acceptable in all countries, despite it being rather restricted.

We next conducted chi-square difference tests were in order to examine whether the beta coefficients between SES and the reading dimensions were significantly different between RLS 1991 and RLS 2001. Because, in the current study, one of the hypotheses was that the school SES effect on reading achievement would increase during the 10-year period due to policy changes and reforms in the different education systems, only the school-level beta coefficient (i.e., the correlation between school SES and ReadB) is of interest. Table 10 presents the results of this chi-square difference test.

Significant differences emerged for only four of the nine countries—Iceland, Italy, Singapore, and the United States. This finding implies either that, for the rest of the countries, no significant change occurred in the SES effect on reading achievement at school level between 1991 and 2001, or that the countries' data lacked the statistical power necessary to prove that changes in the school SES effect were significant.

For the countries where the relationship between school SES and school reading achievement differed significantly between RLS 1991 and 2001, the estimates supported the current hypothesis. As seen in Table 11, the beta coefficients between school SES and

Table 10: Chi-Square Difference Test for RLS 2001 and RLS 1991 Populations of the Relationship between SES and Reading Achievement Factor at School Level

Country	$\Delta\chi^2$	Δdf
Greece	3.02	1
Hungary	1.29	1
Iceland	7.46	1
Italy	5.83	1
New Zealand	1.94	1
Singapore	33.08	1
Slovenia	.91	1
Sweden	.47	1
United States	16.13	1

Table 11: Effects of Socioeconomic Status on Reading Dimensions at Individual and School Levels in the RLS 1991 and 2001 Trend Countries

Country	2SES \rightarrow ReadB RLS2001	2SES \rightarrow ReadB RLS1991
Greece	.25	.74
Hungary	.86	.87
Iceland	.23	.51
Italy	.53	.18
New Zealand	.84	.77
Singapore	.89	.66
Slovenia	.54	.71
Sweden	.63	.67
United States	.93	.84

Note: beta coefficients in italics are statistically non-significant.

ReadB were much higher in RLS 2001 than in RLS 1991.

In the United States and Singapore, the school SES explained 80% or more of the variation in school reading achievement in RLS 2001. Iceland and Italy also had moderate correlations between school SES and school reading achievement in RLS 2001, but these relationships were not significant in RLS 1991. Note that the SES factor score is not the latent variable itself but rather an estimate of the individual quantity of the factor. The error involved in the factor score may attenuate the size of the beta coefficient; however, at the school level, the reliability should be better. There is thus reason to believe that the true beta coefficient was higher.

Conclusion and discussion

The factor score estimation of SES and the estimates of the correlation between SES and reading dimensions indicate that the factor score approach has several advantages. First, it can efficiently use all the data at hand to estimate a factor score for those cases with full and partial information on the factor indicators. In the case of SES, there is a large amount of missing data and a low reliability of the indicators. Second, the use of factor score estimates can efficiently reduce the complexity of a model, and overcome the problems of convergence very often involved in large and complex models.

The measurement model of reading achievement fitted each trend country well, and the total reading variation in the achievement test can thus be decomposed accordingly. The reading variation among students is attributed to differences in each student's general reading ability, document reading skills, reading speed, and measurement error. The reading variation among schools can be broken down into the differences involved in school average reading ability and measurement error. The reading achievement properties represented in the measurement model differed significantly between the two populations of RLS 1991 and RLS 2001, with only one exception—New Zealand. This may be because, by the early 1990s, New Zealand already had in place policies allowing parents to choose the school to which to send their children and had given responsibility for financial expenditure to local school boards of trustees. These policies changed little over the following decade. These same reasons may explain the marginal changes in reading variation in New Zealand.

In general, differences in reading achievement between students and between schools diminished, to a larger or a smaller degree, in most of the countries during the decade between 1991 and 2001. This situation suggests that students and schools became more homogeneous in terms of their reading ability in many countries, notably Hungary, Iceland, Italy, Slovenia, and Sweden. In Greece, reading achievement across schools became more homogeneous, whereas their students' reading ability became more heterogeneous. As indicated in Table 1, this pattern of change aligns with Greece's centralized educational policies. Slovenia provides a similar example.

One of the most interesting cases in this study is Singapore. The between-school achievement differences in reading in Singapore increased dramatically over the 10-year period, even though educational provision in Singapore during this time was centralized and homogeneous. One possible explanation might be the different practices between different types of schools and among teachers. As noted in the *PIRLS Encyclopedia* (Mullis, Martin, Kennedy, & Flaherty, 2002), the three different types of schools in Singapore, namely, government, government-aided, and independent, receive varying amounts of funding from the government, while government and government-aided schools “are given greater autonomy and more funds to provide a wider range of innovative enrichment programs for their students” (p. 254). Moreover, teachers are encouraged to use different instruction methods and strategies to suit the learning abilities of their students. However, more information about Singapore’s education system and practices is needed to fully understand the finding.

The pattern of variance changes between 1991 and 2001 in reading achievement in Hungary, Italy, and Sweden is also interesting. The education systems of all three countries have gone through a decentralization process. Such reforms may increase achievement differences between schools, due to the fact that different schools may vary in their ability to get resources and to attract better teachers and children from well-educated parents—factors that relate highly to school achievement level. We may therefore find empirical evidence to support the current hypotheses in the above three countries, if the

school-related factors are brought into the model. It is thus necessary to investigate this issue further in future studies. However, the significant increase of the school SES effect on reading achievement in Italy and the findings in the United States do support the current hypothesis.

One of the important reasons for the minimal support for the hypothesis may be that we did not use restricted sub-samples. Since free school choice is a big city phenomenon, it may be reasonable to use the urban sample in each country to test the hypotheses. Another important reason is lack of statistical power in the data. The hypothesis in this study focused on changes at school level, namely the changes in the school achievement variation and the school SES effect on school achievement. In the RLS 2001 samples for most of the trend countries, the school-level sample was not large enough to allow detection of the true variance and correlation among variables of interest (refer to Table 1 above). Moreover, for most countries, the questions regarding possessions were the same in the student questionnaires for RLS 2001 and RLS 1991. It should be noted that after a 10-year interval, some of these possessions may be too commonly owned to capture socioeconomic differences between families. Lack of variance in SES indicators makes it impossible to discover any significant school SES effect on reading achievement in RLS. However, the current study did find some support for the hypothesis, suggesting that the provision of sufficient and quality data would allow us to obtain a more realistic picture of the changes in reading properties and SES effects between 1991 and 2001 for the trend study participants.

References

- Ball, S. J. (1994). *Education reform: A critical and post-structural approach*. Buckingham: Open University Press.
- Beare, H., & Boyd, W. L. (Eds.). (1993). *Restructuring schools: An international perspective on the movement to transform the control and performance of schools*. London: Falmer.
- Blair, M. (1994). Black teachers, black students and education markets. *Cambridge Journal of Education*, 24(2), 277–291.
- Bourdieu, P. (1984). *Distinction: A social critique of the judgment of taste*. Cambridge, MA: Harvard University Press.
- Bourdieu, P. (1997). The forms of capital. In A. H. Halsey, H. Lauder, P. Brown, & A. S. Wells (Eds.), *Education: Culture, economy, and society* (pp. 46–58). Oxford: Oxford University Press.
- Brooks-Gunn, J., Duncan, G. J., Klebanov, P. K., & Sealand, N. (1993). Do neighborhoods influence child and adolescent development? *American Journal of Sociology*, 99(2), 353–395.
- Burstein, L., Fischer, K. B., & Miller, M. D. (1980). The multilevel effects of background on science achievement: A cross-national comparison. *Sociology of Education*, 53(4), 215–225.
- Chapman, J. D. (1996). *The reconstruction of education: Quality, equality and control*. London: Cassell.
- Coffey, A. (Ed.). (2001). *Education and social change*. Buckingham: Open University Press.
- Conway, S. (1997). The reproduction of exclusion and disadvantage: Symbolic violence and social class inequalities in parental choice of secondary education, *Sociological Research Online* (Vol. 2).
- Elley, W. B. (1994). *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems*. Oxford: Pergamon Press.
- Gewirtz, S., Ball, S. J., & Bowe, R. (Eds.). (1995). *Markets, choice and equity in education*. Buckingham: Open University Press.
- Goldring, E. (1995). *Communal opportunities to learn, parental involvement and school choice: Israel and the United States*. Paper presented at the ESRC/CEPAM Invitation Seminar, Milton Keynes.
- Gonnie van Amelsvoort, H. W. C., & Scheerens, J. (1997). Policy issues surrounding processes of centralization and decentralization in European educational systems. *Educational Research and Evaluation*, 3(4), 340–363.
- Gustafsson, J.-E. (1997). Measurement characteristics of the IEA reading literacy scales for 9–10 year-olds at country and individual levels. *Journal of Educational Measurement*, 34, 233–251.
- Gustafsson, J.-E. (1998). Social background and teaching factor determinants of reading achievement at class and individual levels. *Journal of Nordic Educational Research*, 18, 241–250.
- Gustafsson, J.-E. (2006). *Barns utbildningssituation [Children's educational situation]*. Stockholm: Rädda Barnen [Save the Children, Sweden].
- Gustafsson, J. E., & Stahl, P.-A. (2004). *STREAMS User's guide, Version 3 for Windows 95/98/NT*. Unpublished manuscript. Mölndal, Sweden.
- Hanson, E. M. (1997). *Educational decentralization: Issues and challenges*. Washington, DC: The World Bank.
- Hatcher, R. (1998). Class differentiation in education: Rational choice? *British Journal of Sociology of Education*, 19(1), 5–24.
- Hoxby, C. M. (2001). If families matter most, where do schools come in? In T. Moe (Ed.), *A Primer on American schools* (pp. 285–316). Stanford, CA: Hoover Institution Press.
- Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin*, 126(2), 309–337.
- Levin, H. (1998). Educational vouchers: Effectiveness, choice and costs. *Journal of Policy Analysis and Management*, 17(3), 373–392.
- Levin, H. (2002). A comprehensive framework for evaluating educational vouchers. *Educational Evaluation and Policy Analysis*, 24(3), 159–174.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS trends in children's reading literacy achievement 1991–2001: IEA's study of trends in reading literacy achievement in primary school in nine countries*. Chestnut Hill, MA: Boston College.
- Martin, M. O., Mullis, I. V. S., & Kennedy, A. M. (2003). *PIRLS 2001 technical report*. Chestnut Hill, MA: International Association for the Evaluation of Educational Achievement.
- McEwan, P. J. (2000). The potential impact of large-scale voucher programs. *Review of Educational Research*, 70(2), 103–149.
- Metz, T. M. (1990). How social class differences shape teachers' work. In M. W. McLaughlin, J. E. Talbert, & N. Bascia (Eds.), *The context of teaching in secondary schools*. New York: Teachers College Press.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools*. Chestnut Hill, MA: International Association for the Evaluation of Educational Achievement.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Flaherty, C. L. (Eds.). (2002). *PIRLS 2001 encyclopedia. A reference guide to reading education in the countries participating in IEA's Progress in International Reading Literacy Study (PIRLS)*. Chestnut Hill, MA: Boston College.
- Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54, 557–585.

- Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, 28, 338–354.
- Muthén, B. (1994). Multilevel covariance structure analysis. In J. Hox & I. Kreft (Eds.), *Multilevel modeling: A special issue of Sociological Methods & Research*, 22, 376–398.
- Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. Marsden (Ed.), *Sociological Methodology 1995* (pp. 216–316). Oxford: Blackwell.
- Muthén, L. K., & Muthén, B. O. (2004). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Orfield, G., & Eaton, S. E. (1996). *Dismantling desegregation* (the Harvard Project on School Desegregation). New York: New Press.
- Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. Cambridge, MA: Harvard University Press.
- Skolverket. (1996). *Att välja skola: Effekter av valmöjligheter i grundskolan [Choosing a school: Effects of free choice possibilities in elementary schools]* (No. 109). Stockholm: Skolverket [Ministry of Education].
- Skolverket. (2003). *Valfrihet och dess effekter inom skolområdet. En belysning av begreppet valfrihet och valfrihetens effekter inom skolan [Free school choice and its effects within the school world]* (No. 230). Stockholm: Skolverket [Ministry of Education].
- Skolverket. (2006). *Vad händer i likvärdigheten i svensk skola? En kvantitativ analys av variation och likvärdighet över tid. [What happened with equality in Swedish schools? A quantitative study on variation and equality over time]*. Stockholm: Skolverket [Ministry of Education].
- Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research*, 37(5), 483–504.
- Waslander, S., & Thrupp, M. (1997). Choice, competition, and segregation: An empirical analysis of a New Zealand secondary school market, 1990–93. In A. H. Halsey, H. Lauder, P. Brown, & A. S. Wells (Eds.), *Education: Culture, economy, and society* (pp. 439–459). Oxford: Oxford University Press.
- Whitty, G., Power, S., & Halpin, D. (1998). *Devolution and choice in education: The school, the state and the market*. Buckingham: Open University Press.
- Willms, J. D. (1992). *Monitoring school performance: A guide for educators*. London: Falmer Press.
- Willms, J. D. (1997). *Parental choice and education policy* (No. 12, CES briefing). Edinburgh: Centre for Educational Sociology, University of Edinburgh.
- Willms, J. D., & Echols, F. H. (1992). Alert and inert clients: The Scottish experience of parental choice of schools. *Economics of Education Review*, 11(4), 339–350.
- Woods, P. A., Bagley, C., & Glatter, R. (1998). *School choice and competition: Markets in the public interest?* London: Routledge.
- Yang, Y. (1998). Variance differences in reading achievement at classroom and individual level in 24 countries. *Journal of Nordic Educational Research*, 18(4), 165–175.
- Yang, Y., & Gustafsson, J. -E. (2003). Measuring socioeconomic status at individual and collective levels. *Educational Research and Evaluation*, 10(3), 259–288.
- Yang-Hansen, K., Rosén, M., & Gustafsson, J. -E. (2006). Measures of self-reported reading resources, attitude and activities based on latent variable modeling. *International Journal of Research and Method in Education*, 29(2), 221–237.

Analyzing trends in levels of reading literacy between 1970 and 2001 in Sweden

Monica Rosén¹

Department of Education

Gothenburg University, Gothenburg, Sweden

Introduction

The importance and usefulness of trend measures of achievement as input variables for educational policy development as well as for research in educational sociology is well recognized, but trend measures are hard to obtain. In Sweden, this need has perhaps become more urgent since publication of the results from the first IEA trend study of reading literacy (Gustafsson & Rosén, 2005; Martin, Mullis, Gonzales, & Kennedy, 2003). The results indicated a rather dramatic and unexpected decrease in reading literacy in Sweden among 9- to 10-year-olds. The same pattern has been observed in mathematics and the natural sciences (Mullis, Martin, Gonzales, & Chrostowski, 2004a, 2004b).

Numerous hypotheses have been suggested as explanations for the decrease, most of which concentrate on possible factors within the education system, which in Sweden has undergone considerable change over this period of time. Few of the suggested explanations are, however, based on empirical analyses, so most of the explanatory work remains. Considerably less attention has been paid to potential factors outside the education system, such as societal changes in children's upbringing conditions both inside and outside of the home environment, changes in habits, values, and attitudes, and changes in the world of texts and systems for communication.

One of the key questions in the search for explanations relates to when this decline started. Although Sweden was one of the highest performing countries in the IEA Reading Literacy Study 1991, the decrease may have started before 1991. The IEA trend studies cover the period of the last decade of the 20th century, but the decrease may have started even earlier than this.

This article accordingly summarizes work undertaken so far to analyze the change in reading literacy in Sweden between 1970 and 2001. The overall aim is to further analyze the reported changes in reading achievement in Sweden by presenting a trend analysis that compares the Swedish reading performance in 1970 with the performance in 1991 and 2001, along with possible interpretations. Measures on a common scale for this period may not only contribute valuable information on when this negative trend in reading literacy really started but also information on the size of the decrease. But before we can draw more definite conclusions, we must acknowledge that the construct and the measure of reading achievement have undergone change between the three studies. This change may have implications for our understanding of the Swedish performance in the 1991 as well as in the 2001 study. The changes to the reading achievement construct may also have implications for interpretation of the 30-year trend measure, which may need further adjustments.

Utilizing the data collected in the IEA Reading Comprehension Study in 1970 (Thorndike, 1973), in the IEA Reading Literacy Study in 1991 (Elley, 1994), in the IEA trend study in 2001 (Martin, Mullis, Gonzalez, & Kennedy, 2003), and in PIRLS 2001 (Mullis, Martin, Gonzales, & Kennedy, 2003) offers opportunity to shed some light on this question. Links were ensured in 1991 and 2001 when some reading test instruments from the 1970 study were included in extended measurement designs used in Sweden. Links between the different target populations in these studies also were secured in the Swedish design of PIRLS 2001, where student samples from both Grade 3 and Grade 4 participated in the study.

¹ The work presented in this chapter is part of a collaborative research work between Professor Jan-Eric Gustafsson, Anders Skarlind, and Monica Rosén within a larger research project, SALS, financed by the Swedish Research Council. (SALS is the Swedish acronym for Studies of Reading Literacy in Sweden.) Most of the statistics in this paper were computed by Anders Skarlind, statistician at the Stockholm Institute of Education.

In combination with the methodological advances of matrix sampling and advanced Item Response Theory (IRT) techniques, these data sets enabled us to construct a common scale for reading literacy that covers a period of time beginning 20 years earlier than the IEA 10-year trend study of reading literacy mentioned above. Almost all items in all three studies were included in the IRT analysis, giving a total of 104 items. Common over all three tests were 21 items that originated from the Reading Comprehension Study 1970 (RC 1970). Another 59 items were common to the two latest studies, and the remaining 24 items were specific to RC 1970. The scale was normed to a mean of 500 and a standard deviation of 100, and five plausible values were generated for each student (Skarlind, 2007). To improve the comparability across time, the scale was, in a later analysis, described for different sub-groups of the populations (Skarlind, Gustafsson, & Rosén, 2007), and influences from, for example, differences in age and school start, were taken into account.

Method

This section describes the data sources used, the methodology employed for obtaining comparable measures of reading performance across time, and how the comparable groups were identified.

Subjects

The Swedish data used in this study were drawn from four IEA studies of reading literacy: the RC Study in 1970 (Hansson, 1975; Thorndike, 1973), the Reading Literacy (RL) Study in 1991 (Elley, 1992, 1994; Taube, 1995), the repeat of RL in 2001, and PIRLS 2001 (Mullis, Martin, & Kennedy, 2003; Rosén, Myrberg, & Gustafsson, 2005). The samples were selected from Population 1 (10-year-olds in RC 1970), Population A (Grade 3 students in RL 1991), Grade 3 students from RL 2001, and both a Grade 3 and a Grade 4 sample

from PIRLS 2001. Table 1 presents an overview of the samples and grades.

The target population in 1970 was 10-year-olds, which implied that, in Sweden, students were either in Grade 3 or Grade 4. Therefore, about half of the RC 1970 sample originated from Grade 3 and the other half from Grade 4. In the RL 1991 study, the target population was the upper of the two grades holding the most nine-year-olds at the time of testing (Elley, 1992), which in Sweden was Grade 3. In PIRLS 2001, the target population was defined similarly to the RL 1991 definition ("the upper of the two grades holding the most nine-year-olds") but with an addition that explicitly stated that this definition implied Grade 4 in most countries. Thus, most countries chose Grade 4 as their target grade. Sweden, however, chose to select samples from both Grade 3 and Grade 4 to secure links to the previous studies.

As can be noted in Table 1, there were age and/or grade links between the studies. In 1991, there were no Grade 4 students included in the sample. However, there was some age overlap regardless of grades, and the fact that the 2001 studies included samples from both grades made it possible to give some estimate of reading achievement across time for both grades.

Instruments

The reading scale was based on reading items from all three studies: 43 items from the reading comprehension test in 1970, 66 items from the RL 1991/2001 study, and 98 items from PIRLS 2001. In RC 1970, the text passages were all rather short. Five of them can be characterized as expository prose, and the other three as narratives. In RL 1991, the test covered three text domains: narrative prose, expository prose, and so-called documents. The latter are structured presentations of information, set out in the form of charts, graphs, maps, lists, and sets of instructions.

The length of the passages varied between very short

Table 1: Samples and Grades in IEA Assessments of Reading Comprehension/Literacy

Year	1970	1991	2001	
	Reading Comprehension (RC 1970)	Reading Literacy (RL 1991)	Reading Literacy Repeat (RL 2001)	Progress in Reading Literacy (PIRLS 2001)
Grade 3	901	4, 236	5,361	5,271
Grade 4	979			6,044
Total	1, 880	4, 236	5,361	11,315

to a few pages. In PIRLS, all the passages were quite lengthy (two to five pages) and the text type was labeled as either literary or informational, with the former overlapping completely with narrative prose, and the latter with either expository prose or documents or both. A matrix sampling design was adopted for the passages in PIRLS so that all the passages were rotated over 10 booklets, each holding two passages. However, each student took one booklet only. In RC 1970 and in RL 1991/2001, all students were administered the same test booklets.

In RC 1970, all test items were multiple-choice whereas in RL 1991 and RL Repeat 2001 most items

were multiple-choice and a few were short-response. In PIRLS, about half of the items were multiple-choice, another 25% required shorter written responses, and about 25% demanded extended written responses.

Links between the studies in Sweden were enabled through extended national designs in 1991 and 2001 with regard to both the tests and the target samples. Table 2 presents the reading tests used and their overlaps and links across one another and across time.

Four of the passages (20 items) from RC 1970 were repeated in RL 1991 and also in the RL Repeat in 2001. In 1991, another 14 passages containing 66 items divided into two booklets were administered to

Table 2: Reading Passages and Number of Test Items in IEA Assessments of Reading Comprehension/Literacy

Reading Passage and Items	1970 RC 1970	1991 RL 1991	2001 RL Repeat 2001	PIRLS 2001	
Seal (Exp, 5 items)	Yes				
Robert (Nar, 6 items)	Yes				
Simonides (Nar, 7 items)	Yes				
Musk (Exp, 5 items)	Yes				
Sundew (Exp, 4 items)	Yes	Yes	Yes		
Tailor bird (Exp, 5 items)	Yes	Yes	Yes		
Ernenek (Nar, 7 items)	Yes	Yes	Yes		
Marmot (Exp, 4 item)	Yes	Yes	Yes		Yes
Quicksand (Exp, 3 items)					Yes
Shark (Nar, 4 items) ¹					Yes
Bottles (Doc, 4 items)					Yes
Buses (Doc, 4 items)					Yes
Content (Doc, 3 items)					Yes
Temperature (Doc, 5 items)					Yes
Grandpa (Nar, 6 items) ²					Yes
Tree (Exp, 6 items)					Yes
Postcard (Exp, 2 items)		Yes	Yes	Yes	
Bird and elephant (Nar, 5 items)		Yes	Yes	Yes	
Island (Doc, 4 items)		Yes	Yes	Yes	
No dog (Nar, 6 items)		Yes	Yes	Yes	
Marias timetable (Exp, 2 items) ³		Yes	Yes	Yes	
Walrus (Exp, 6 items) ⁴		Yes	Yes	Yes	

Note: 1 Four items originally

2 Seven items originall

3 Three items originally

4 Seven items originally.

the students. These two booklets were administered in the same manner in the repeat study in 2001. In addition, every student in the PIRLS Grade 3 and Grade 4 samples was administered one of the two booklets from the RL study. We can therefore conclude that links exist between the reading tests in the three studies. The remaining question relates, of course, to the degree to which the underlying reading literacy construct is common for the different reading tests.

Scaling

Test equating is facilitated by the inclusion of bridge items—test items that occur in two or more surveys and that can serve as a bridge between tests that are, to a large extent, different, as long as the tests cover the same content areas. There are different ways to equate tests. The present study used a combination of IRT methodology and percentile equating. The development of IRT methodology makes it possible to equate tests with a limited number of bridge items, and to improve precision by using other items not included in all the equated tests. We can therefore regard the test data from the three studies included here as a large matrix sampling design with test items from the same content area.

Using Bilog-MG (du Toit, 2003), Skarlind (2007) computed an IRT scale based on the test items in RC 1970, RL 1991, and RL Repeat 2001. This allowed comparisons of reading literacy at three points in time. Because it was desirable that all three samples had equal influence in the IRT analyses, the sampling weights were modified for the scaling procedure so that an equal sum of the weights was obtained for the three surveys.

In their analysis of the dimensionality of the RL 1991 test and the reading literacy test used in PIRLS 2001—an analysis that used structural equation modeling—Gustafsson and Rosén (2006) found an almost perfect correlation between the general reading literacy dimensions of both tests. In other words, general reading literacy is the factor that influenced the performance on all the test items. Thus, from a content point of view, the high correlation between the general reading literacy dimensions makes it reasonable to equate the PIRLS test with the IRT scale. Here, the scales were merged by equating percentile by percentile between the IRT scale from RC 1970–RL 1991–RL 2001 and the PIRLS 2001 reading literacy scale.

Previous studies of the dimensionality of the RL 1991 scale show that it is not possible to differentiate between narrative and expository prose. But over and above the general reading literacy dimension are two additional residual factors, one document factor, and one end-of-test factor (Gustafsson, 1997; Gustafsson & Rosén, 2006; Rosén, 1997). The document factors related to most of the document items, and the end-of-test factor related to the last items in each booklet and was interpreted as an effect of reading speed and diminishing motivation. In the PIRLS test, only one residual factor was identified over and above the general dimension. This factor was interpreted as a constructive response factor because it related to items requiring the students to construct written responses.

We can infer from the analyses of the dimensionality in the RL test that the document items that were included in RL 1991 represent a reading literacy dimension that is different from that of the narrative and expository prose. A sub-scale that included narrative and expository items only was therefore computed (Skarlind et al., 2007), and the scales are compared in this paper.

Five plausible values for each scale, the RC-RL Scale and the Nar-Exp Scale, was computed, and each scale was transformed to have a mean of 500 and a standard deviation of 100, in the same manner as that for the IEA reading achievement scales. Skarlind (2006) provides detailed information about the scaling procedures used here.

Obtaining comparable groups

Despite the fact that ages and grades overlapped across these studies, it was not possible to compare the samples in any simple and straightforward manner because of the sampling design adopted for each study. The sample in RC 1970 was an age-based one of 10-year-olds, and students were thus sampled from both Grade 3 (old Grade 3 students) and Grade 4 (young Grade 4 students). During the fall of 1970, when the RC test was administered, the 10-year-olds in Grade 3 represented the older Grade 3 students while the 10-year-olds in Grade 4 represented the younger Grade 4 students. What this means is that we cannot obtain a representative sample from either of the two grades.

The samples in RL 1991 and RL Repeat 2001 were both grade-based samples that included all students in Grade 3, regardless of age. However, RL 1991 included

those students who, at the time of testing (the eighth month of the school year), were between 9.0 years and 10.3 (10 years and three months) of age (Elley, 1992). In RL Repeat 2001, the age in the Grade 3 sample varied between 8.2 to 11.2 years of age. Thus, from these samples, we cannot obtain a representative sample from either the nine-year-olds or the 10-year-olds.

One solution to this problem of comparing students from somewhat differently defined samples is to compare sub-groups of equal age and/or grade across the three studies. Three such groups can be identified:

- *Old Grade 3 students:* This group consists of students who were between 10.0 and 10.2 years of age at the time of testing, and who had started school at the normal time.
- *Grade 3 students:* This group consists of students who were between 9.4 and 10.2 years at the time of testing, and who had started school at the normal time.
- *Young Grade 4 students:* This group contains students who were between 10.5 and 10.11 years at the time of testing, and who had started school at the normal time
- *10-year-old students:* This group contains students in Grades 3 or 4 and who were therefore between 10.0 and 10.11 years old at the time of testing, and who had an early, normal, or late school start.

Table 3 shows the studies for which these groups were available for comparison.

One factor that may influence the results of the comparable age-grade sub-groups is the different amount of schooling at the time of testing between the three samples as a result of the different months of testing. The results of means adjusted for time of schooling are therefore presented together with the unadjusted means on the two scales.

Results

As shown in the methodology section, and despite the fact that the sample definition was quite similar across the surveys included in this study of trends, the survey still differed so much with respect to age and grade that comparisons of complete grades or complete age-cohorts were not possible. Instead, the results in this section are presented for the four comparable and partly overlapping, previously identified sub-groups. In all figures given in this section, the visual distance between 1970 and 1991 is the same as between 1991 and 2001, although the former represents 20 years of time and the latter only 10 years of time. The slopes of the lines in these graphs are therefore somewhat misleading. However, all the trend lines presented are accompanied with a graph in which the differences in means across the years are visualized via bars of score-point differences and *p*-values.

Trends in reading literacy across 30 years of time

Figure 1 presents trend lines of the average performance between 1970 and 2001 on the RC-RL Scale. Included in the RC-RL Scale are almost all the items from the 1970 Reading Comprehension Study and from the Reading Literacy and Reading Literacy Repeat studies of 1991/2001. The figure also presents a mean adjusted for differences in the amount of schooling the students had received at the time of testing.

The overall impression gained from Figure 1 is that of a decrease in average reading literacy skills, which began sometime before 1991. Additional information is presented in Figure 2, which shows the change in score points for each group and period of time along with *p*-values for the estimates. The only group with observed measures at all three time points is the "Old Grade 3 group." At first sight, the drop between 1970 and 1991 for this group appears both large and significant. However, one factor behind the

Table 3: Comparable Sub-Groups of Equal Age and Grade, 1970, 1991, and 2001

Group Label	Age at the time of testing (years and months)	1970 <i>n</i>	1991 <i>n</i>	2001 <i>n</i>
Old Grade 3 students	10.0–10.2	484	1,094	2,601
Grade 3 students	9.4–10.2	3,850	8,855	
Young Grade 4 students	10.5–10.11	922		3,37
10-year-old students	10.0–10.11	1,884		7,424

Figure 1: Trends of Means and Adjusted Means on the RC-RL Scale 1970, 1991, and 2001

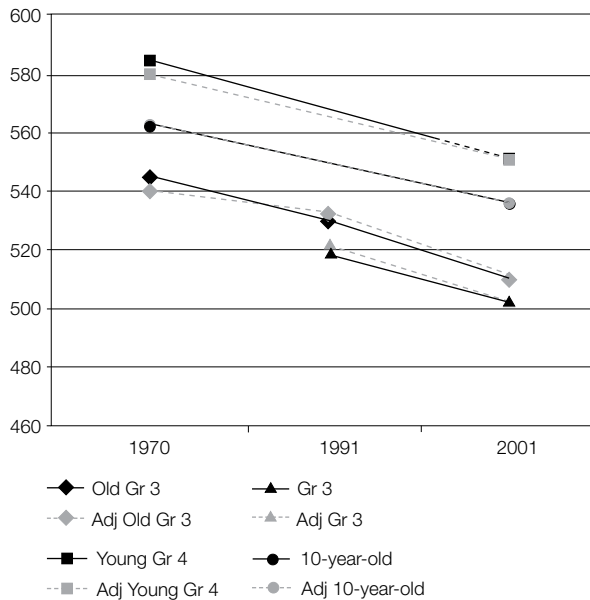
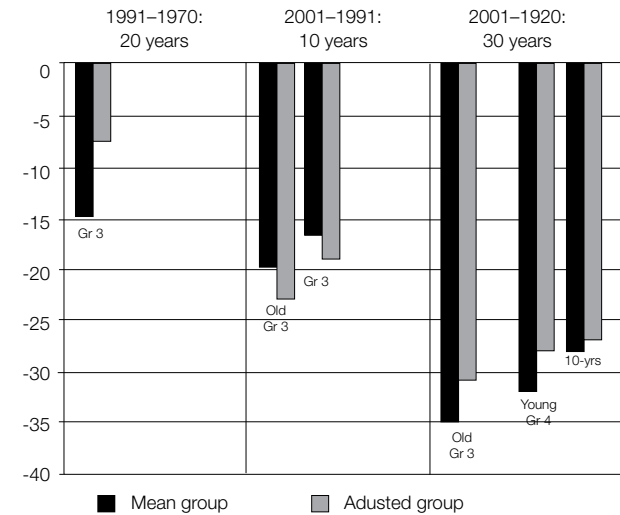


Figure 2: Mean and Adjusted Mean Differences on the RC-RL Scale between 1970, 1991, and 2001



Note: * $p < .05$.

high means in 1970 turned out to be the amount of schooling the students had received at the time of testing. The Reading Comprehension tests in 1970 were administered in late May and early June, the RL test in 1991 was administered in mid March, and the tests in 2001 were administered in mid April. The students in RC 1970 thus had about 2.5 months more of schooling at the time of testing compared to the students in RL 1991, and about 1.0 month more of schooling compared to the students in PIRLS 2001. The adjustments for differences in time of schooling at the time of testing resulted in somewhat different estimates of change, which indicates that this factor influenced the results. When the means were adjusted for these differences, the difference between 1991 and 2001 for the Old Grade 3 group changed from -15 to -8, and was no longer statistically significant.

A comparison of the means in 1970 with the means in 2001 showed substantial drops in achievement for all comparable groups. The total drop seemed to be somewhere between 27 and 32 score points, which is equivalent to about 0.3 of a standard deviation. The mean drops presented in Figure 2 also reveal that a major portion of the total drop occurred for both comparable groups in Grade 3 after 1991. The similar pattern of the trend lines may indicate that this drop occurred for the two remaining age-grade groups as well.

As I discussed in the method section of this paper, part of the results in the above measures of trends may be due to the content of the RC-RL Scale. The reading test in 1970 did not include texts or items from the document dimension of reading literacy, so the inclusion of such items may have affected the estimated measures of performance. Figure 3 shows trend lines that are similar to those above, but these apply only to the scale based on the narrative and expository items, that is, the Nar-Exp Scale.

Both the trend lines in Figure 3 and the bar graph in Figure 4 for the Nar-Exp Scale are quite similar to those for the RC-RL Scale. Figure 5 compares the results from the two scales. The most notable difference here is that the total drop between 1970 and 2001 varies between .30 and .36, which is larger than the drop on the RC-RL Scale. The trend lines indicate that even if the decreasing trend may have started sometime before 1991, the larger portion of it took place after this time. The only group with comparable measures at both 1970 and 1991 shows a significant difference of -16 score points between 1970 and 1991 (before the means were adjusted for differences in schooling); the -10 score points after adjustment did not achieve statistical significance.

The larger decrease over time on the Nar-Exp Scale than on the RC-RL Scale makes good sense because previous analyses of trends in Swedish reading literacy

Figure 3: Trends of Means and Adjusted Means on the Nar-Exp Scale, 1970, 1991, and 2001

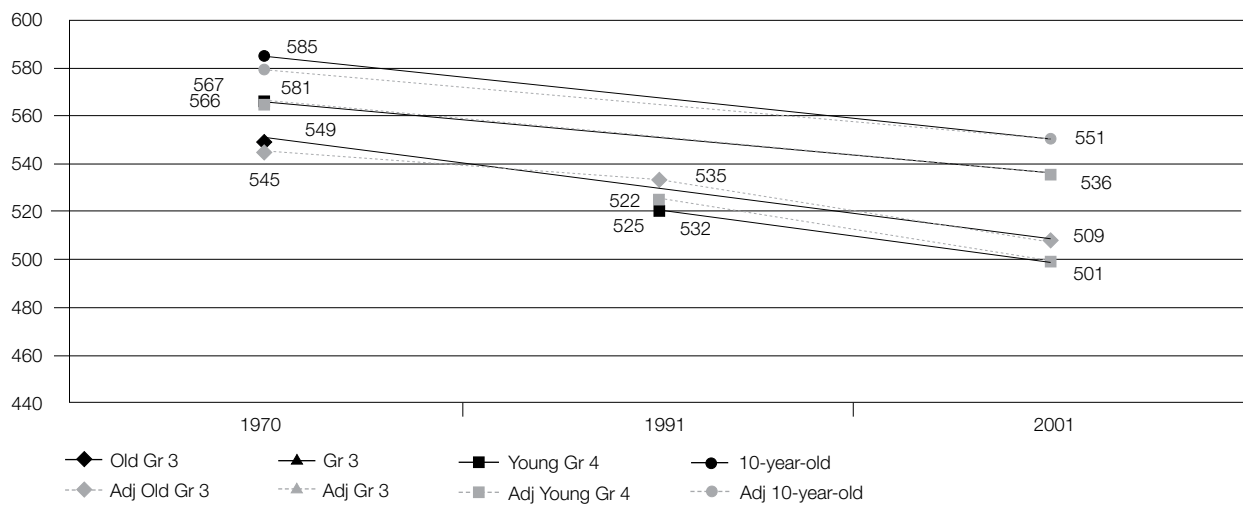


Figure 4: Mean and Adjusted Mean Differences on the Nar-Exp Scale between 1970, 1991, and 2001

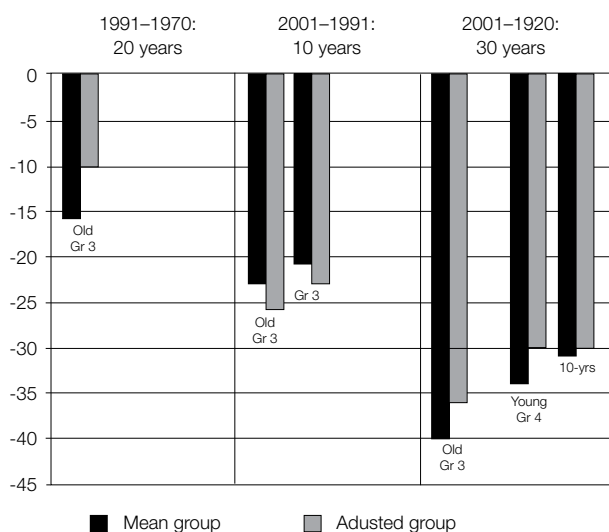
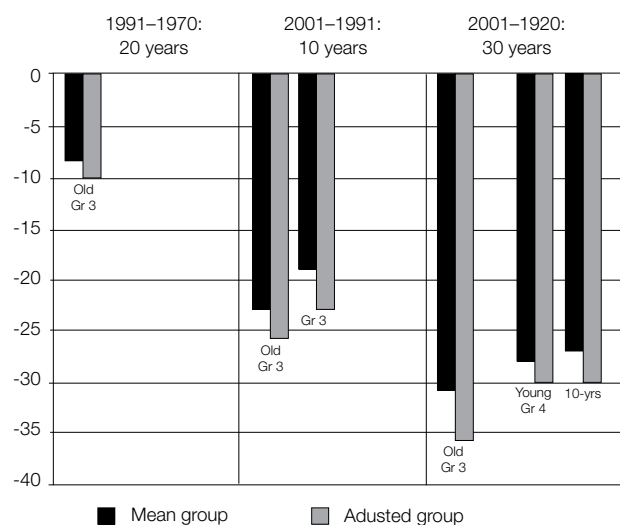
Note: * $p < .05$.

Figure 5: Comparison of Adjusted Mean Differences on the RC-RL and Nar-Exp Scales

Note: * $p < .05$.

between 1991 and 2001 in Grade 3 show a rather large decrease in reading comprehension and in the end-of test/reading speed factor, but no differences in level of document reading performance during this period.

Trends for girls and boys

The trend lines for girls and boys are presented below; all means have been adjusted for differences in the amount of schooling at the time of testing. Figure 6 presents the trend line for boys and girls on the RC-RL Scale. The negative trends of reading performance observed on this scale are much more pronounced for

boys than for girls. Also, the negative trend for boys between 1991 and 2001 appears more pronounced for the two Grade 3 groups than for the two older groups.

Figure 7 shows that the total loss for boys between 1970 and 2001 varies between -34 and -38 score points on this scale, of which about one third seems to have occurred before 1991. For girls, the total drop since 1970 is substantially less; it varies between -.16 and -.20 on the RC-RL Scale. For Young Grade 4 boys, the total decrease is more than twice as large as for the equivalent group of girls.

Figure 6: Trend Lines for Girls and Boys on the RC-RL Scale, 1970, 1991, and 2001

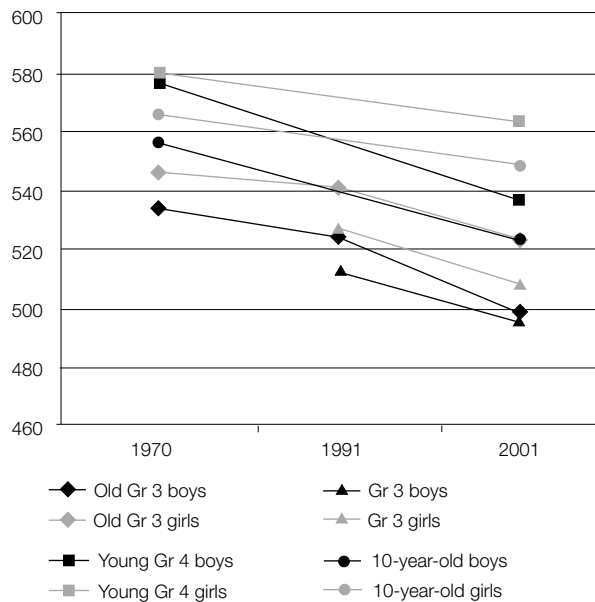
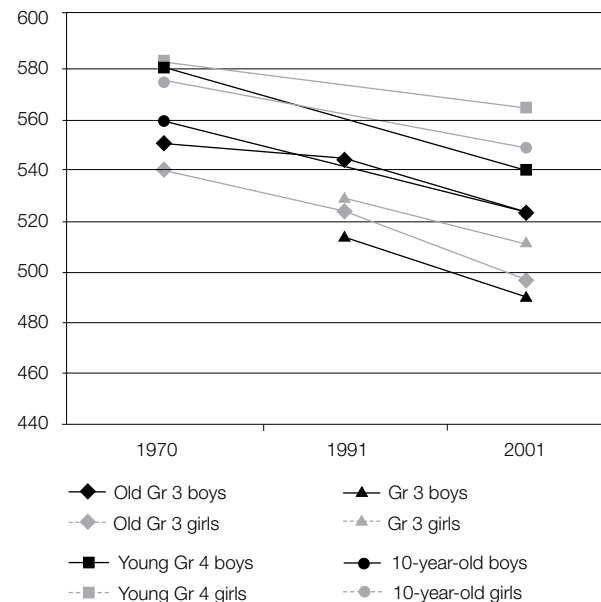


Figure 8: Trend Line for Girls and Boys on the Nar-Exp Scale, 1970, 1991, and 2001

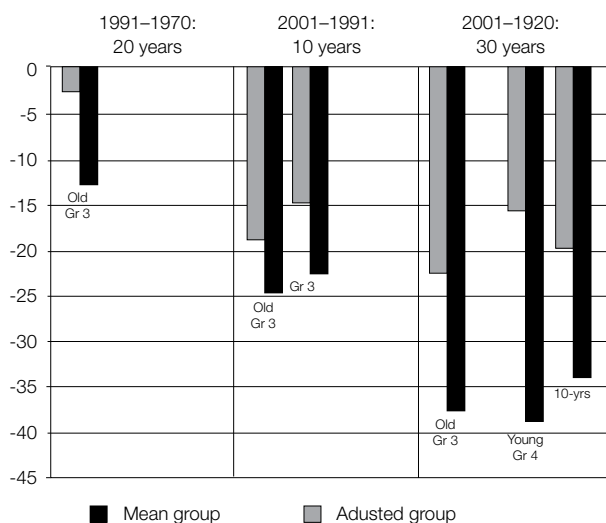


A similar analysis was performed for girls and boys on the Nar-Exp Scale. On this scale, the trend lines for the different age-grade groups, presented in Figure 8, appear different for girls and for boys, although they are all negative. While only a small decrease can be observed for girls between 1970 and 1991, the decrease among the boys is clearly pronounced in all comparable age-grade groups. For girls, the line is

more or less horizontal between 1970 and 1991, while for boys it is clearly negative. Between 1991 and 2001, the line becomes steeper in the negative direction for both girls and boys.

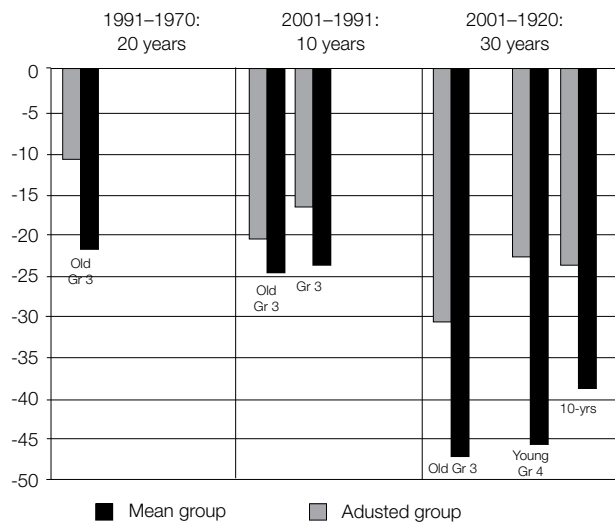
As can be seen in Figure 9, the total drop in performance among boys on the Nar-Exp Scale is even more striking than the one on the RC-RL Scale. It varies between $-.39$ and $-.47$, which is nearly half a

Figure 7: Adjusted Mean Differences for Girls and Boys on the RC-RL Scale between 1970, 1991, and 2001



Note: * $p < .05$.

Figure 9: Adjusted Mean Differences for Girls and Boys on the Nar-Exp Scale between 1970, 1991, and 2001



Note: * $p < .05$.

standard deviation. For girls, the drop is also high, but compared to the drop for boys, it is not strikingly so; it varies between $-.23$ and $-.31$ across comparable age-grade groups

Discussion and conclusions

In this paper, trend measures covering 30 years of time and reflecting reading performance at three time points—1970, 1991, and 2001—were put together in one overall reading scale, the Reading Comprehension-Reading Literacy Scale, and one subscale, the Narrative-Expository Scale. The scales are discussed first, and then the results.

The computation of two IRT scales of reading literacy was enabled by an extended test design in the Reading Literacy (RL) Study 1991 and its repeat in 2001, which, in Sweden, included a test booklet from Reading Comprehension (RC) 1970. The link to the Progress in Reading Literacy Study (PIRLS) 2001 was also enabled through an extended test design, where the students in the PIRLS sample were administered one or the other of the two booklets from RL 1991 in addition to the PIRLS test. The links in themselves are necessary for IRT scaling but do not justify a common scale unless the content can be assumed to be equivalent. One of the main issues relates, of course, to whether the contents of the tests are indicators of the same underlying construct.

The RC-RL Scale was based on the items from all reading passages in RC 1970 and all the items from RL 1991 and RL Repeat 2001. The tests used in these studies appear on the surface to be somewhat different, which is due to the fact that the IEA reading literacy construct has evolved since 1970, a change that is reflected in the tests. However, the tests share so many similarities that a common scale across the test may be justified.

The text passages from 1970 were all either narrative or expository prose and were rather short, and the test items were of multiple-choice type. RL 1991 included one additional text type, the so-called document type, defined, as noted previously, as structured information displays presented in the form of charts, tables, graphs, lists, and/or sets of instruction. The text passages varied in length and the items were basically multiple-choice. In PIRLS 2001, the document dimension was removed from the assessment framework, while the narrative and the expository prose were included, but described

in other terms, namely purposes of reading. (Students read narrative text mainly for literary purposes and the expository texts for informational purposes.) The text passages were longer, and the test items were of both multiple-choice and constructed-response type, but the general dimension in the test almost perfectly correlated with the general dimension in the RL tests. The PIRLS scale was therefore merged to the IRT scale through equating the scores percentile by percentile.

Previous research relating to the RL 1991 and PIRLS 2001 tests shows the correlation between narrative and expository test scores to be too high for these to be regarded as two different reading ability dimensions. The scores from document passages, however, are influenced by an additional factor over and above the general reading dimension, suggesting that it is reasonable to create a subscale based on the narrative and expository test items only, here called the Nar-Exp Scale.

The results showed a rather large drop in mean performance on both of these measures during the period under consideration, although it was even greater on the Nar-Exp Scale than on the RC-RL Scale. Four comparable groups were identified in terms of age and grade that more or less covered Grades 3 and 4 and ages 9 and 10. However, a closer examination of the groups indicated that they were tested at different months of the school year, and that this difference had an unexpectedly large effect on the results.

Adjusting for these differences resulted in a less dramatic mean change between 1970 and 1991 and a more pronounced difference between 1991 and 2001. After these adjustments, the results from the analysis of trends on the RC-RL Scale showed a large decrease in reading achievement, amounting to about 0.3 of a standard deviation for both Grade 3 and Grade 4 between 1970 and 2001. The trend on the Exp-Nar Scale was similar, although the total mean drop was slightly larger than 0.3 of a standard deviation. However, the decline between 1970 and 1991 was not as large as the decline between 1991 and 2001. About one third of the total drop occurred before 1991; the remainder occurred during the 10-year period following 1991. Comparable age and grade groups showed an average decrease of about 10 score points between 1970 and 1991, while the average decrease between 1991 and 2001 amounted to about 23 score points.

One of the main conclusions we can draw is that the decline in reading achievement over the 30-year period was substantial and seems to have started some time before 1991, but that the downward trend escalated during the last decade of the 20th century. Another conclusion is that the amount of schooling students had received at the time their achievement in reading was tested may have been crucially important for results pertaining to comparable age-grade groups in studies of trends in reading achievement. This conclusion may also be valid for studies of trends in other school subjects.

When the trend lines for comparable groups were estimated for girls and boys separately, rather striking differences emerged. The decrease for boys was much larger than for girls on both scales, but particularly so on the Nar-Exp Scale. The drop amounted to almost 0.5 of a standard deviation on this scale while the drop for girls amounted to only 0.25 of a standard deviation. Girls did not show any negative trend on the Nar-Exp Scale between 1970 and 1991, while boys did. After 1991, the lines became more negative for both groups, but the decline was even larger for boys.

The gender pattern was similar on the RC-RL scale, although the differences between boys and girls were not

so high. Boys lost almost 0.35 of a standard deviation unit in mean performance during the 30-year period while the average mean drop for girls was just below 0.20 of a standard deviation. The different trend lines for girls and boys on the two scales indicate that the presence of document items makes a difference. Boys do better on document tasks than they do on narrative and expository tasks. This finding may be indicative of gender differences in reading habits and particularly in spare-time reading habits. Another conclusion from this pattern is that the inclusion of document items in the scale affects the content and thus the interpretation of the scale. The Nar-Exp Scale seems to reflect reading comprehension of continuous texts, while the RC-RL Scale is also influenced by other skills.

A reasonable explanation for the negative trend is, of course, changes in reading habits inside as well as outside of school. The if, how, and why of changes in reading habits remains to be investigated in further research. Numerous factors, directly acting and interacting, have most likely contributed to the results presented here. The focus in this paper has been on establishing measures of change in reading literacy for a 30-year period; the search for explanations is a task for continued research.

References

- du Toit, M. (Ed.). (2003). *IRT from SSI: Bilog-MG Multilog Parscale Testfact*. Lincolnwood, IL: Scientific Software International.
- Elley, W. B. (1992). *How in the world do students read?* The Hague: International Association for the Evaluation of Educational Achievement.
- Elley, W. B. (Ed.). (1994). *The IEA Study of Reading Literacy: Achievement and instruction in thirty-two school systems*. Oxford: Pergamon Press.
- Gustafsson, J.-E. (1997). Measurement characteristics of the IEA reading literacy scales for 9- to 10-year-olds at country and individual levels. *Journal of Educational Measurement*, 34(3), 233–251.
- Gustafsson, J.-E., & Rosén, M. (2005). *Förändringar i läskompetens 1991–2001. En jämförelse över tid och länder* [Changes in reading literacy 1991–2001: A comparison across time and countries] (IPD-rapporter, nr 2005:15). Gothenberg: Göteborgs Universitet, Institutionen för pedagogik och didaktik.
- Gustafsson, J.-E., & Rosén, M. (2006). The dimensional structure of reading assessment tasks in the IEA Reading Literacy Study 1991 and the Progress in International Reading Literacy Study 2001. *Educational Research and Evaluation*, 12(5), 445–468.
- Hansson G. (1975). *Svensk skola i internationell belysning II: Läsnings och litteratur* [Swedish schools in an international light: Reading and literature]. Stockholm: Almqvist & Wiksell.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Kennedy, A. M. (2003). *Trends in children's reading literacy achievement 1991–2001: IEA repeat in nine countries of the 1991 Reading Literacy Study*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Chrostowski, S. J. (2004a). *TIMSS 2003 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Chrostowski, S. J. (2004b). *TIMSS 2003 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College.

- Mullis, I. V. S., Martin, M. O., & Kennedy, A. M. (Eds.). (2003). *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College
- Rosén, M. (1997). *Country differences in reading performance: A reanalysis of the IEA Reading Literacy Study*. Paper presented at the Annual Meeting of AERA, Chicago, March 24–28, 1997.
- Rosén, M., Myrberg, E., & Gustafsson, J.-E. (2005). *Läskompetens i skolor 3 och 4: Nationell rapport från PIRLS 2001 i Sverige* [Reading achievement in grades 3 and 4: National report from PIRLS in Sweden] (The IEA Progress in International Reading Literacy Study: Gothenburg Studies in Educational Sciences 236). Gothenburg: Acta Universitatis Gothoburgensis.
- Skarlind A. (2007). *Test equating of reading literacy tests in Swedish IEA data from 1970, 1991 and 2001*. Unpublished manuscript, Department of Education, Gothenburg University.
- Skarlind, A., Gustafsson, J.-E., & Rosén, M. (2007). *Trends in reading ability in Grades 3 and 4 in Sweden 1970–2001*. Unpublished manuscript, Department of Education, Gothenburg University.
- Taube K. (1995). *Hur i all världen läser svenska elever?* [How in the world do Swedish students read?] (Report Number 78.) Stockholm: Skolverket (Ministry of Education).
- Thorndike, R. L. (1973). *Reading comprehension education in fifteen countries: An empirical study*. Stockholm: Almqvist & Wiksell.

A cross-country comparison of direct and indirect effects of parents' level of education on students' reading achievement

Eva Myrberg and Monica Rosén

*Department of Education, Gothenburg University
Gothenburg, Sweden*

Abstract

During the last few decades, researchers in many countries have established cultural capital in families (and, more specifically, the educational level of parents) as the most important dimension of socioeconomic influence on school performance. Less is known, however, about the factors that actually mediate this influence. The aim of this paper was to investigate the relative importance of various mediating factors, as well as the total effects of parents' level of educational attainment, on the reading achievement levels of children. Data in this comparative analysis come from the International Association for the Evaluation of Educational Achievement's Progress in International Reading Literacy Study (PIRLS) 2001 and include representative samples from seven countries: Bulgaria, France, Hong Kong SAR, Hungary, Italy, Norway, and Sweden. Variables indicating possession of cultural capital in the families were collected from the home questionnaire used in PIRLS. A total achievement

score was used as the outcome variable, and the effects of parental education on reading achievement were estimated with structural equation modeling.

As a starting point, a baseline model was fitted and estimated with a Swedish data set. The results for some 6,000 Swedish fourth-graders were robust and stable. The total effect of parental education was substantial, and almost half of this effect was mediated through other variables, namely the number of books at home, early literacy activities, and emergent literacy abilities at the time at which children started school. The model was then tested using data from the six other participating countries. The model fitted well in almost all the countries. However, the analyses showed considerable variation across the countries. The total effect of parental education varied, as did the pathways for indirect effects mediated through other variables.

Introduction

It has long been known that the home background of students affects their scholastic achievement, such as in reading. The influence of students' home background on school achievement has interested researchers for several decades (Adams, 1990; Coleman et al., 1966; Hanushek, 1986, 1989; Härnqvist, 1992; Snow, Burns, & Griffin, 1998; Statens Offentliga Utredningar/SOU, 1993). There is also evidence of substantial social group differences in first-graders' decoding and reading comprehension skills when general intelligence is controlled for (Hecht, Burgess, Torgesen, Wagner, & Rashotte, 2000). The influence on school achievement of a less favorable background in, for example, the socioeconomic respect, is likely, however, to vary across school systems (Organisation for Economic Co-operation and Development/OECD, 2001, 2004; SOU, 1993; Yang, 2003). Socioeconomic status (SES) is a complex and multi-dimensional concept.

Parents' level of education

During the last few decades in many countries, cultural capital in families (or, more specifically, parents' level of educational attainment) has been proved as the most important dimension of the influence that SES has on school performance (Yang, 2003). How the transmission of cultural capital over generations affects achievement has yet to be fully explained and understood. Bourdieu (1984, 1990, 2002) describes the transmission of cultural capital as a long-lasting process that starts in early childhood. The notion of cultural capital and its transmission over generations is central in the present analysis. By relating academic success to the distribution of cultural capital, this notion offers a theoretical hypothesis that makes it possible to explain the unequal scholastic achievement of children originating from different social classes.

The number of books at home

The number of books at home is a variable that has a well-known association with students' achievement, and it has been used in research as an indicator of children's availability of literacy resources (see, for example, Elley, 1992). The mere availability of print and frequent exposure to texts of different kinds may have benefits for young children's literacy and may promote the development of specific skills and knowledge that enables more efficient subsequent reading. Furthermore, environments that are rich in literature offer many opportunities for joint activities between parents and their children (Paris & Cunningham, 1996).

Early reading activities

Moschovaki (1999) described literary experiences as embedded, to a considerable extent, in family social events rather than as pure reading and writing per se. Within the context of the family, education occurs at every moment, whether it is deliberate and systematic or is realized in an informal way through everyday activities. The set of attitudes, expectations, and information transmitted to the child constitutes informal instruction. Sonnenschein and Munsterman (2002) have also stressed the importance of informal reading activities. Sénéchal and LeFevre (2002), however, found that informal shared reading of storybooks during preschool years seems unrelated to parents' teaching about reading—that is, formal literacy activities. They also found that different types of activities are associated with different outcomes. They furthermore found the link between parents' reports of teaching their children about reading and reading storybooks to their children to be indirect and mediated through children's emergent literacy skills. Reading storybooks at home predicted children's receptive language skills both concurrently and longitudinally, and this relationship held after controlling for exogenous factors such as parents' level of education and endogenous factors such as phonological awareness.

Early reading abilities

Existing evidence suggests a continuous influence of individual differences in early reading skills from kindergarten onwards (Hecht et al., 2000; Sénéchal & LeFevre, 2002; Stanovich, 1986, 2000), although schooling may have equalizing effects (Gustafsson

& Rosén, 2007). Lundberg (1985) found children's reading abilities in Grade 1 to be good predictors of reading achievement in Grade 3. Paris and Cunningham (1996) also concluded from a review of the literature that reading skills in kindergarten and Grade 1 predict reading skills in Grade 2, and that children's developmental continuity in school success or failure is based on their achievement in the primary grades. Although there is vast evidence of the influence of cultural capital on students' achievement, the relative effects of various mediating factors have yet to be fully investigated.

Structural modeling of the effects of parents' level of education

In a previous study, Myrberg and Rosén (2007) investigated direct and indirect effects of parental education with a Swedish sample of third-graders. They found that although the direct effect of parental education was modest (.17) for this sample, the total effect was substantial (.32). This estimate agrees well with what has been found in previous research. The reason why the total effect was stronger than the direct effect is that other variables mediated some of the indirect effects of parental education.

Our aim in this present study was to investigate the total effects of parents' level of education, as well as the relative importance of various mediating factors, on the reading achievement levels of children in different countries. Previous research suggests that the influence of cultural capital varies across countries, and that the influence of parental education is mediated through other variables and with varying strengths in different countries.

Method

Data and variables

The data used in the analysis presented here come from representative samples of 9- to 10-year-olds in seven countries participating in PIRLS 2001. The design of the international study is described in detail in the PIRLS framework (Campbell, Kelly, Mullis, Martin, & Sainsbury, 2001) as well as in the PIRLS 2001 technical report (Martin, Mullis, & Kennedy, 2003).

The countries included in the present study were Bulgaria, France, Hong Kong SAR, Hungary, Italy, Norway, and Sweden. Variables indicating possession of cultural capital in the families were collected from

the home questionnaire used in PIRLS 2001, while the dependent variable was the total reading score for each student. Table 1 presents a summary account of these variables.

The countries included in the sample were chosen strategically with respect to ensuring variation across the features of the respective school systems, such as degree of decentralization. Another criterion for inclusion was that the home questionnaire had been administered, and yet another was a high respondent rate. The sample of students used in this study included those for whom a total achievement score had been estimated. Table 2 presents valid *N* and respondent rates. The table shows a high respondent rate across

items and countries. Because the Chinese language (relevant for Hong Kong SAR) does not have letters, we did not consider this variable in the analysis.

Method of analysis

The major aim of the present study was to explore the influence of parents' level of education on reading achievement, and to include and consider factors that mediate this influence. Our main method of analysis was structural equation modeling with latent variables. These variables reflect hypothetical constructs and are estimated from the relationships among two or more indicators. This is done with confirmatory factor analysis (CFA), which estimates and tests a

Table 1: Indicators of Reading Achievement and Socioeconomic Background of Students

Variable	Label	Information/Question/Statement	Source
<i>Parents' level of education</i>			
Hedum	Mother's education	Mother's educational level Eight alternatives: not finished compulsory school to Bachelor or Master exam	Parent
Heduf	Father's education	Father's educational level Eight alternatives: not finished compulsory school to Bachelor or Master exam	Parent
<i>Number of books at home</i>			
Hbook	Books at home	About how many books are there in your home? Five alternatives: 0–10 to more than 200	Parent
Hchbk	Children's books	About how many children's books are there in your home? Five alternatives: 0–10 to more than 100	Parent
<i>Early reading activities</i>			
Hacread	Read with child	Before your child began Grade 1, how often did you read books with him or her? Three alternatives: never/almost never/often	Parent
Hactell	Tell stories to child	Before your child began Grade 1, how often did you tell stories to him or her? Three alternatives: never/almost never/often	Parent
<i>Early reading abilities</i>			
Hablett	Recognize letters	How well could your child recognize most of the letters in the alphabet when he/she began Grade 1? Four alternatives: not at all to very well	Parent
Habword	Read words	How well could your child read some words when he/she began Grade 1? Four alternatives: not at all to very well	Parent
Habsent	Read sentences	How well could your child read some sentences when he/she began Grade 1? Four alternatives: not at all to very well	Parent
<i>Students' reading achievement</i>			
Read ach	Reading achievement	Mean of five plausible values of PIRLS total score	Student

Table 2: Valid N and Item Respondent Rate (%)

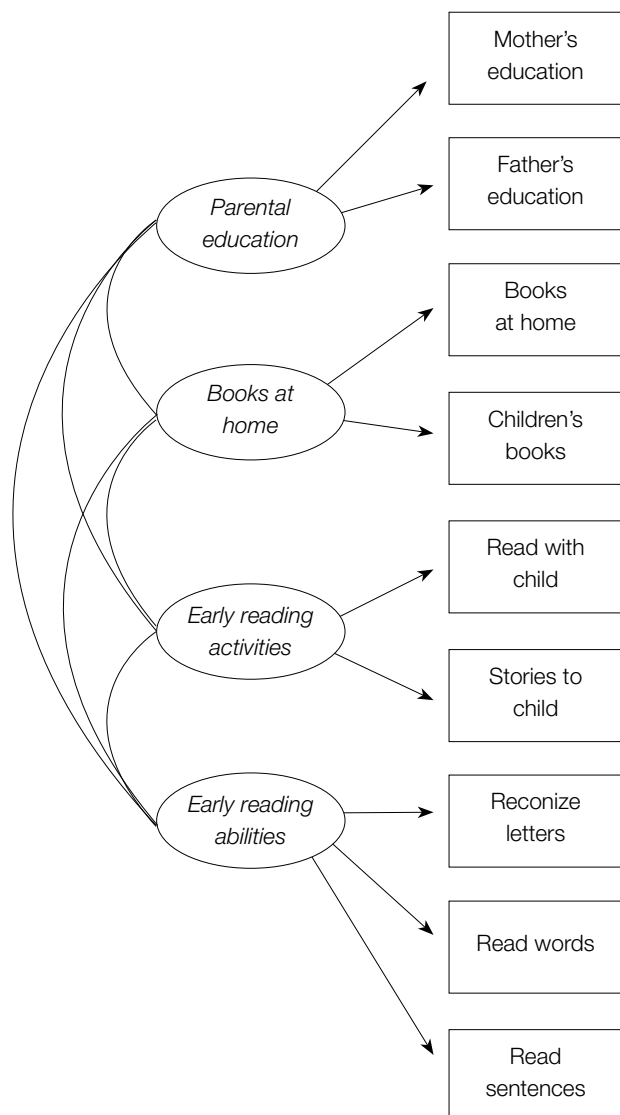
	Swe	Nor	Bgr	Fra	HKg	Hun	Ita
Reading achievement	6,044	3,459	3,460	3,538	5,050	4,666	3,502
Mother's education	88	87	91	73	85	88	93
Father's education	84	85	88	69	85	86	92
Books at home	91	91	95	89	94	95	96
Children's books	91	91	95	89	93	95	97
Read with child	90	90	95	88	92	94	91
Tell stories to child	90	90	95	88	91	94	94
Recognize letters	90	90	95	88	0	95	95
Read words	90	90	95	88	87	94	93
Read sentences	90	89	94	87	86	94	91
Complete data	78	80	82	59	68	80	80

hypothesized model against the observed covariances among observed variables. The latent variables have the advantage of being free from the non-systematic variance that causes underestimation of regression coefficients (Pedhazur & Pedhazur-Schmelkin, 1991).

We used case weights, the “complex sample” option, and the “missing data” option in Mplus (Muthén & Muthén, 2001) in order to account for stratification, cluster effects, and cases with incomplete data. Mplus was used under the modeling environment of STREAMS (Gustafsson & Stahl, 2001).

As we noted above, the questionnaires used in PIRLS included a home questionnaire addressed to parents. This questionnaire contained items indicating possession of cultural capital in the families. During the first step of our analysis, we set up a measurement model, and within it defined, as independent variables, four latent variables indicating possession of cultural capital. The relationships between manifest and latent variables are expressed by standardized loadings, which conceptually can be thought of as standardized regression coefficients. Hence, by squaring the standardized loadings, the amount of variance in the observed variables accounted for by the latent variables is obtained. We used a total achievement score as the outcome variable. In Figure 1, which presents the measurement model, the latent variables are independent variables that influence the observed variables, and are illustrated by the unidirectional arrows. The assumption in the model is that the latent variables are freely correlated. This is shown by the curved lines.

Figure 1: Measurement Model



During our second step, we fitted to the data a path model based on a conceptual model of the impact of parental education on students' reading achievement and estimates of the relationships among a set of constructs. This process allowed us to impose within a so-called structural relations model or path model with latent variables (Bollen, 1989; Jöreskog & Sörbom, 1993; Loehlin, 2004) a hypothesized structure of relationships among the latent variables. Our analysis thus traced the influence of parental education as mediated through the number of books at home, early reading activities with the children during the preschool years, and the children's early reading abilities at the beginning of formal schooling. The structural, or path, model presented in Figure 2 is to be read vertically from the top to the bottom, and the variables are arranged in a chronological and logical order. The starting point is the latent variable *parental education*.

Parental education is the only independent variable in the final model presented in Figure 2. Its causal sources are external to the path diagram. In a path diagram, the correlation between any two variables can be expressed as the sum of the compound paths connecting these two points. The numerical value of a compound path is equal to the product of the values of its constituent arrows. When a correlation between

two variables is due to the presence of one or more additional variables, the relationship is considered indirect (Loehlin, 2004).

Results

This section presents the results from the first step in the analysis, that is, the measurement mode, and then the results for the second step, that is, the structural equation model (path model). We first present the effects and then describe the relative importance of the direct versus the indirect effects. Finally, we exhibit the different paths from parents' level of educational attainment to students' reading achievement.

Results of the measurement model

The measurement model in Figure 1 defined four latent variables reflected in two or more observed indicators respectively. The figure shows the following: latent variable *parental education* measured by the two observed variables "father's education" and "mother's education;" *books at home* measured by "books at home" and "children's books;" *early reading activities* measured by "read with child" and "stories to child;" and *early reading abilities* measured by "recognize letters," "read words," and "read sentences." Reading achievement is measured by a single observed variable, "reading achievement."

Three indicators of model fit are presented here. Chi square and degrees of freedom is one. Because the estimation of the chi square can be inflated by a large sample size, we complemented it here with the two most commonly used descriptive indices indicated as providing a good fit between model and data—CFI and RMSEA. Table 3 presents the model fit for the measurement model.

According to the chi-square test, the model fit was not particularly good, with the chi square ranging from 69.15 in Hong Kong SAR ($df = 14$) to 378.67 in Bulgaria. The CFI, however, indicated a good fit in all countries, of at least .98. The RMSEA also indicated a good fit (about .03–.04) in all countries except Bulgaria, where it was estimated to be .07. According to, for example, Maruyama (1998), RMSEA should be less than .05 to indicate a good fit. The RMSEA for Bulgaria thus suggests that the model contains more information than is accounted for by the relationships set up between the variables.

Figure 2: Path Model

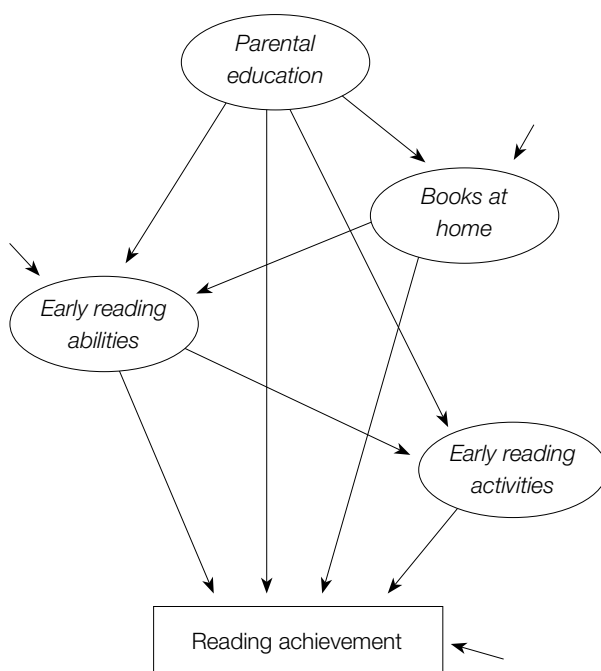


Table 3: Fit Indices for the Measurement Model

Model fit	Swe	Nor	Bgr	Fra	HKg	Hun	Ita
Chi square	231.567	141.491	378.67	219.033	69.152	125.609	129.236
Degrees of freedom	21	21	21	21	14	21	21
p-value	0	0	0	0	0	0	0
CFI	0.984	0.98	0.991	0.98	0.997	0.994	0.988
RMSEA	0.043	0.043	0.072	0.054	0.029	0.033	0.039

The measurement model was applied to each of the countries separately. Table 4 presents the results of the relationships between the observed and the latent variables. From the table it is apparent that the latent construct *parental education* measured by mother's and father's educational level had high loadings in all countries as had the latent construct *books at home*, as measured by the number of children's books at home and the total number of books at home. The relationship between the latent variable *early reading activities* and *read with child* was highest in Sweden (.86) and lowest in Italy (.63). *Stories to child* varied considerably across the countries in its loadings. It ranged from .33 in Sweden to .69 in Hong Kong SAR. The variation in this indicator may reflect varying meanings of telling stories to children in different countries and cultures. Finally, the latent variable *early reading abilities* was fairly well measured by the three indicators "recognize letters," "read words," and "read sentences." As noted above, because the Chinese language does not have letters, this variable derived from only two observed variables for Hong Kong SAR. Reading achievement was measured by a single observed variable, "total achievement," which had been estimated for every student.

In all countries, the observed variables accounted for about 60% of the variance. Because the latent variables were free from measurement errors, the strong relationships between observed and latent variables indicate that the latent constructs were fairly well measured and that reliability had thereby been improved.

Results of the path model

In this structural equation model, the relationships between the four latent variables were specified and tested against the data. The model was an attempt to unfold the mediating factors and mechanisms that explain why there is a relationship between *parental education* and "reading achievement"—a relationship that can be described as the focal relationship (Aneshensel, 2002). Also, by estimating the effects of a set of mediating variables, the model allowed a partial explanation for the relationship between *parental education* and "reading achievement." Within the SEM framework, this estimation is most easily done by decomposing the total effect of *parental education* on "reading achievement" into a set of direct and indirect effects.

Table 4: Standardized Estimates of Loadings of the Latent Constructs on Observed Variables in the Measurement Model

Std est for relationship	Swe	Nor	Bgr	Fra	HKg	Hun	Ita
Par educ → Heduf	0.62	0.58	0.74	0.74	0.82	0.77	0.75
Par educ → Hedum	0.81	0.74	0.85	0.83	0.79	0.85	0.82
Books → Hbook	0.84	0.85	0.90	0.83	0.80	0.86	0.83
Books → Hchbk	0.76	0.77	0.84	0.82	0.83	0.88	0.80
Early read act → Hacread	0.86	0.69	0.83	0.78	0.64	0.65	0.63
Early read act → Hactell	0.33	0.44	0.59	0.59	0.69	0.47	0.35
Early read abi → Hablett	0.76	0.73	0.77	0.61		0.74	0.66
Early read abi → Habword	0.96	0.98	0.97	0.99	0.89	0.99	0.97
Early read abi → Habsent	0.81	0.79	0.79	0.66	0.84	0.84	0.76

Note: All the relationships are statistically significant at the $p < .05$ level.

Again the chi-square test did not indicate an especially good fit. However, the other indices used indicated a good fit between model and data. The CFI ranged from 0.977 to 0.997 and the RMSEA from 0.024 to 0.062. As the indices of model fit differed little between the measurement model and the structural model, we can consider the latter as offering a good representation of the relationships between the variables. The fit indices for the path model are presented in Table 5.

The hypotheses formulated in the introduction about patterns of influence among variables representing cultural capital, literacy practices at home, and early reading abilities imply the following relationships:

1. *Parental education* affects *books at home*, *early reading activities*, *early reading abilities*, and "reading achievement."
2. *Books at home* affects *early reading activities*, *early reading abilities*, and "reading achievement."
3. *Early reading activities* affects *early reading abilities* and "reading achievement."
4. *Early reading abilities* affects "reading achievement."

This was the set of hypothesized relationships we tested in the structural model.

We also hypothesized there would be variation across countries in the total effects of parental education and the relative strengths of the direct and indirect effects. We subsequently found that not only did the total effect of parental education vary across countries but

so too did the relationships with respect to direct and indirect effects. These results are presented in Table 6.

As a starting point in regards to this table, we can see that the total effects of parental education on students' achievement vary quite considerably across the countries. The largest estimated influence is for Hungary (.58), followed by France (.51), Bulgaria (.47), Italy (.42), Norway (.40), Sweden (.38) and Hong Kong SAR (.14). Despite this variation, we can conclude that the effects of student background were substantial in this respect for all countries, but less so in Hong Kong SAR, which we discuss below. If we look a little closer at the estimates, we can also see that the relationships between direct and indirect effects are quite different across the countries. In Sweden, Norway, France, and Hungary, the direct effects of parental education are larger than the indirect effects. In Italy, the direct and indirect effects are much the same. Bulgaria shows much larger indirect than direct effects and Hong Kong SAR has almost no direct effects at all.

The path model specified the directions of connections. Table 7 presents the estimates for the relationships between all latent variables and their effects on reading achievement. The effects of all possible pathways presented in the table are estimated.

Effects of the number of books on early activities, early abilities, and reading achievement

As described previously, the number of books at home is an often-used indicator of cultural capital in students' families and has frequently been found to relate to

Table 5: Fit Indices of the Path Model

Model fit	Swe	Nor	Bgr	Fra	HKg	Hun	Ita
Chi square	249.132	180.53	367.865	257.639	68.361	167.612	173.642
Degrees of freedom	26	26	26	26	18	26	26
p-value	0	0	0	0	0	0	0
CFI	0.983	0.977	0.986	0.98	0.997	0.993	0.985
RMSEA	0.038	0.041	0.062	0.05	0.024	0.034	0.04

Table 6: Total, Direct, and Indirect Effects of Parents' Level of Education on Students' Reading Achievement

Direct and indirect effects on read ach	Swe	Nor	Bgr	Fra	HKg	Hun	Ita
Direct effect of parents' educ on read ach	0.23	0.29	0.13	0.30	0.01	0.40	0.22
Indirect effect of parents' educ	0.15	0.11	0.34	0.21	0.13	0.18	0.20
Total effect of parents' educ	0.38	0.40	0.47	0.51	0.14	0.58	0.42

Table 7: Standardized Estimates of Relationships between Latent Variables in the Path Model

Std est for relationship	Swe	Nor	Bgr	Fra	HKg	Hun	Ita
Par educ → Books	0.59	0.62	0.79	0.73	0.60	0.75	0.63
Par educ → Early read activities	0.11	0.11	0.12	0.2	0.26	0.12	0.02
Par educ → Early read abilities	0.09	0.02	0.12	-0.02	0.04	0.15	0.08
Par educ → Read ach	0.23	0.29	0.13	0.30	0.01	0.40	0.14
Books → Early read activities	0.43	0.46	0.55	0.47	0.42	0.49	0.62
Books → Early read abilities	0.02	-0.02	0.18	0.13	0.01	-0.13	-0.22
Books → Read ach	0.13	0.04	0.28	0.21	0.10	0.13	0.18
Early read act → Early read abilities	0.14	0.17	0.33	0.10	0.26	0.14	0.39
Early read act → Read ach	0.06	0.15	0.04	0.07	-0.02	0.13	0.20
Early read abi → Read ach	0.33	0.35	0.21	0.17	0.43	0.14	0.09

Note: Estimates below 0.05 are not statistically significant at the $p = < 0.5$ level.

student achievement in various school subjects. In this present study, the total effect of the *number of books* at home ranged from 0.04 in Italy to 0.38 in Bulgaria. A tentative suggestion is that this indicator might be an expression not only of cultural resources but also of economic resources in Bulgaria. In France, this variable also had quite a large effect on achievement (.27). In Sweden, the effect was .18, in Hungary it was .19, in Norway it was .13, and in Hong Kong SAR it was .14. The meaning of the availability of books at home is therefore likely to be different across countries.

The relationship between the indirect and direct effects of this variable is also interesting. While Norway and Italy showed no direct effects, it was apparent that the influence was mediated through activities with children during their early childhood, namely, reading aloud and telling stories. The effects emanating from number of books that run through early reading activities and then affect early reading abilities were small or negative in most countries. Only in France and Bulgaria was this relationship significant. This result suggests that, in most countries, activities with books in the home are essential for the positive development of children's early reading abilities. In other words, the mere existence of books in the home is not enough. Bulgaria had the largest direct effect (.28). In France, the effect was .21, in Sweden .13, in Hungary .10, and in Hong Kong SAR .10. The direct effects of books at home might also capture a dimension concerning parental aspirations for students and expectations of students' academic achievement.

Effects of early reading activities on early reading abilities and reading achievement

Previous research shows that informal reading activities in the homes of young children are of great importance for reading acquisition. However, our study found no direct effects of *early reading activities* in Hong Kong SAR and Bulgaria. Instead, all effects of early literacy activities were mediated through the students' early reading abilities at the time the children started school. For Hong Kong SAR, the effect was .09, and for Bulgaria it was .11. In Sweden, the direct effects (.06) and the indirect effects (.05) were of similar strength. In Norway, the direct effects were larger (.15) than the indirect effects (.06). This was also the case for France (direct .07 and indirect .02) and for Hungary (direct .13 and indirect .02). Italy had the largest discrepancy between direct and indirect effects (.27 and .04 respectively).

Effects of early reading abilities on reading achievement

Several previous studies have established a connection between emergent literacy skills during students' preschool years and students' later reading achievement. The effect of *early reading abilities* was largest in Hong Kong SAR, where the effect on reading achievement was 0.43. In Norway, it was .35, in Sweden .33, in Bulgaria .21, in Hungary .14, and in Italy .09. These findings indicate that ability to decode written language at the time of starting school had an impact in most of the countries on students' reading achievement four years after this time.

Discussion and conclusions

The seven countries in this comparative study were chosen according to certain criteria. The first was data quality (above all, not too many non-answered items on the home questionnaire). The second was the need to include countries that varied in terms of their policies on provision of preschool education, the features of their school systems (especially regarding degree of decentralization), and their historical and cultural contexts.

The estimates of direct and indirect effects in the path model in this study give further support to the often-stated hypothesis in the literature that level of parental education has an important influence on young school children's reading achievement. In some of the countries in the present study, the total effect of parental education was less pronounced than in others. This was the case for Hong Kong SAR, and our study is not the first one to produce this result in relation to Hong Kong.

This result for Hong Kong has previously been explained with reference to Confucian influence, with many researchers and commentators arguing that Chinese educational thought to a large extent is Confucian educational thought (Reagan, 2000). Here, considerable emphasis is placed on each student assuming responsibility for his or her own learning. Students' own efforts are paramount, and both personal talent and hereditary are considered of little significant importance (Bjerg, 2000). The responsibility for academic success is, however, seen as a mutual responsibility of parents and children. In addition, children are considered obligated to their parents to do their best (Reagan, 2000), and almost half of the parents in Hong Kong SAR send their fourth-graders to additional private tuition after a normal school day (Schwippert, *in press*).

For some countries, previous research has been unable to distinguish between possession of cultural capital and economic capital (Yang, 2003). Bulgaria and Hungary seem to be countries where this connection might be the case, given we found no relationships between the number of books at home and early reading abilities. Bulgaria and Hungary are examples of countries that, until quite recently, belonged to the East European sphere under the influence of the former Soviet Union, and where problems of poverty exist, especially in Bulgaria, which has high unemployment

and limited resources in schools (refer <http://www.landguiden.se>; OECD, 2001, 2006).

Some commentators have also proposed a decentralization hypothesis (see, for example, Yang, 2003). Decentralization as well as deregulation of responsibility for education, along with market-based principles, probably accentuate differences in quality across schools and thereby intensify the influence of students' social background. Support for this premise can be found in relation to Sweden, which has undergone a thorough development toward local decision-making and responsibility since the late 1990s. Over this time, the country has experienced growing variation in educational achievement across schools and declining reading performance among students. This same period has also seen an increase in autonomy at local and school levels in Norway, Hungary, and Italy.

The literature suggests that schooling has the potential to act as a compensatory and equalizing factor with respect to the impact of students' home background on educational achievement. According to Yang (2003), the effect of SES on students' achievement tends to be less evident in countries with more centralized school systems. However, several other (and interacting) factors are likely to be involved in this phenomenon. Some countries, although centralized, have a history of large across-school variance. Students might experience tracking (streaming) at an earlier point, and schools vary extensively in status and quality. France is an example of a country with such features. Its incidence of private funding for education also tends to be higher than that of most countries (OECD, 2006).

While the effect of literacy abilities at school-starting age were, among the countries in our analysis, of greatest importance in Hong Kong SAR, Sweden and Norway also showed relatively strong effects of children's emerging literacy skills. Sweden and Norway are two welfare states in which middle-class parents have good opportunities to take an active part in preparing their children for school, while parents in Hong Kong SAR are known to have high expectations of their children in terms of the effort they put into their learning and their school performance.

The variation across countries regarding early reading abilities does suggest, though, that the organization of preschool provision and the amount

of provision children experience and/or have access to might have an equalizing effect. Italy, for example, has a long tradition of providing high-quality child care, as has France (Schwippert, *in press*). In Italy, despite quite high poverty rates for children, we found that the effect of abilities was the lowest among the countries we analyzed. This result could be because of less variation in this dimension. Sweden and Norway are other countries where almost all children engage in preschool activities, although the emphasis in the preschool provision in these countries is more on social and creative development than on development of reading skills. Until recently, a common belief in Norway, and one reflected in the country's curriculum, was that children should not work with alphabetic activities until they reached seven years of age. This belief could be a contributor to the relative lack of mediating factors between parental education and reading achievement in Norway.

Enrolment in preschool, the organization of preschool provision, and the age at which children start school are other issues that may affect the impact of SES on school achievement. In all countries but Sweden and Bulgaria, primary school starts at the age of six. In Sweden, however, six-year-olds attend the so-called preschool class, which typically offers activities intended to prepare children for school.

In summarizing the results of this study, we can conclude that the effect of level of parental education on fourth-graders' reading achievement scores could be discerned as a somewhat unfolding story in the path analysis. The story indicates that better-educated parents in the participating countries not only have more books at home but also use their knowledge of books and written language to create an educational environment for their children where reading aloud and telling stories are important activities. These early activities not only influence the emergent literacy of first-graders but continue to exert an influence over the reading achievement of third-graders. Thus, the

overall pattern of results for the countries included in this study is that cultural reproduction starts in very early childhood in informal settings. Moreover, the knowledge of written language that children have at the time they start school influences their further reading acquisition. In short, the results presented here support the notion that cultural capital affects students' achievement.

That said, however, we need to note the substantial differences we found across the countries. First, it seems that, in some of the countries, it is not possible to separate the effects of cultural capital from effects related to possession of economic capital. Second, for a majority of the countries, we were unable to establish relationships between parental education and early reading abilities. The effects of parental education in these countries are instead direct and therefore, at least to some extent, reflect expectations and aspirations (Zhan, 2005). It appears that the specific cultures and contexts in countries may either encourage reading activities in families, or support a discourse that holds preschool and school responsible for educating children to read and write. There is also support here for the decentralization hypothesis, but further research on this matter is needed.

In this study, we found a number of similarities and differences across the countries that call for further exploration. Possible explanations for the variation involve interacting influences not considered here. Teacher competence is one such factor, as is student-teacher ratio and the relative levels of residential segregation.

Finally, when interpreting the total effects, we need to remember that the model presented in this paper considered effects at the individual level only. We are aware that the effects of students' possession of cultural capital can also be identified at group level. Nevertheless, the effects that we found in our analysis are considerable.

References

- Adams, J. M. (1990). *Beginning to read. Thinking and learning about print*. Cambridge MA: The MIT Press.
- Aneshensel, C. S. (2002). *Theory-based data analysis for the social sciences*. London: Sage.
- Bjerg, J. (2000) Berättelser om uppfostrandets och undervisningens historia. In J. Bjerg (Ed.), *Pedagogik-en grundbok [Pedagogy: A basic book]* (pp. 17–52). Stockholm: Liber.
- Bollen, A. K. (1989). *Structural equations with latent variables*. New York: Wiley & Sons.
- Bourdieu, P. (1984). *Distinction: A social critique on the judgement of taste*. Cambridge, MA: Harvard University Press.
- Bourdieu, P. (1990) *In other words: Essays towards a reflexive sociology*. Cambridge: Polity Press.
- Bourdieu, P. (2002). The forms of capital. In A. H. Halsey, H. Lauder, P. Brown, & A. Stuart Wells (Eds.), *Education: Culture, economy, society* (pp. 46–58). Oxford: Oxford University Press.
- Campbell, J. R., Kelly, D. L., Mullis, I. V. S., Martin, M. O., & Sainsbury, M. (2001). *Framework and specifications for PIRLS assessment 2001* (2nd ed.). Chestnut Hill, MA: Boston College.
- Coleman, J. S., Cambell, E. Q., Hobson, C. J., McPartlant, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington DC: United States Government Printing Office.
- Elley, W. B. (1992). *How in the world do students read?* The Hague: International Association for the Evaluation of Educational Achievement.
- Gustafsson, J.-E., & Rosén, M. (2007). *School, classroom and individual factors influencing reading performance in Grades 3 and 4 in Sweden*. Unpublished manuscript, University of Gothenburg.
- Gustafsson, J.-E., & Stahl, P.-A. (2001). *STREAMS user's guide, Version 2.5 for Windows 95/98/NT*. Mölndal, Sweden: Multivariate Ware.
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3), 1141–1117.
- Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher*, 18(4), 45–65.
- Härnqvist, K. (1992). Utbildningsreformer och social selektion. [Educational reforms and social selection]. In R. Åberg (Ed.), *Social bakgrund, utbildning, livschanser [Social background, education, life chances]* (pp. 71–78). Stockholm: Carlssons.
- Hecht, S. A., Burgess, S. R., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2000). Explaining social class differences in growth of reading skills from beginning kindergarten through fourth-grade: The role of phonological awareness, rate of access and print knowledge. *Reading and Writing: An Interdisciplinary Journal*, 12, 99–127.
- Jöreskog, K.-G., & Sörbom, D. (1993). *Lisrel 8: Structural equation modelling with the SIMPLS command language*. Chicago, IL: Scientific Software International, Inc.
- Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path and structural analysis* (4th ed). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lundberg, I. (1985). Longitudinal studies of reading and reading difficulties in Sweden. In G. E. Mackinnon & T.G. Waller (Eds.), *Reading research: Advances in theory and practice* (pp. 65–105). New York: Academic Press.
- Martin, M. O., Mullis, I. V. S., & Kennedy, A. M. (Eds.). (2003). *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.
- Maruyama, G. M. (1998). *Basics of structural equation modelling*. London: Sage.
- Moschovaki, E. (1999). Home background and young children's literacy development. *Early Child Development and Care*, 158, 11–19.
- Muthén, L. K., & Muthén, B. O. (2001), *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Myrberg, E., & Rosén, M. (2007). *Direct and indirect effects of parents' education on reading achievement among third graders in Sweden*. Manuscript submitted for publication.
- Organisation for Economic Co-operation and Development. (2001). *Knowledge and skills for life: First results from the OECD Programme for International Assessment (PISA) 2000*. Paris: Author.
- Organisation for Economic Co-operation and Development. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: Author.
- Organisation for Economic Co-operation and Development. (2006). *Starting strong II: Early childhood education and care*. Paris: Author.
- Paris, S. G., & Cunningham, A. E. (1996). Children becoming students. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 117–147). New York: Macmillan.
- Pedhazur, E. J., & Pedhazur-Schmelkin, L. (1991). *Measurement, design and data analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Reagan, T. (2000). *Non-western educational traditions*. London: Lawrence Erlbaum Associates.

- Schwippert, K. (Ed.). (in press). The impact of PIRLS in a number of countries. In *IEA studies in international comparative and multicultural education* (Vol. 7). Munster: Waxmann Publishing Co.
- Sénéchal, M., & LeFevre, J.-A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development*, 73(2), March/April, 445–460.
- Snow, C. E., Burns, M. S., & Griffin P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: United States National Research Council.
- Sonnenschein, S., & Munsterman, K. (2002). The influence of home-based reading interactions on 5-year-olds' reading motivations and early literacy development. *Early Childhood Research Quarterly*, 17, 318–337.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360–406.
- Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York: The Guilford Press.
- Statens Offentliga Utredningar (SOU). (1993). *Ursprung och utbildning: Social snedrekrytering till högre studier* [Origin and education: Social selection to higher education] (R. Erikson & J. O. Jonsson, Eds.). Stockholm: Department of Education.
- Yang, Y. (2003). *Measuring socio-economic status and its effects on individual and collective levels: A cross-country comparison*. Gothenburg: Göteborg Studies in Educational Studies, Acta Universitatis Gothoburgensis [University of Gothenburg].
- Zhan, M. (2005). Assets, parental expectations and involvement, and children's educational performance. *Children and Youth Services Review*, 28, 961–975.



The International Association
for the Evaluation of
Educational Achievement

Across almost 50 years and more than 20 research studies of cross-national achievements, IEA has contributed substantially to the development of a worldwide community of researchers in educational evaluation. The aim of the IEA International Research Conference-2006 (IRC-2006) was to provide an international forum for the exchange of ideas and information on important educational issues investigated through IEA research.

An IEA study typically takes four to five years from inception through to publication of the first round of international reports and release of the international database to a broad community of researchers. Some researchers working directly on projects have the chance to meet one another during the period of participation in a study. However, geography largely constrains further opportunities to exchange ideas, approaches, findings, and concerns. The biennial IEA IRC provides a forum for creative dialogue among researchers, leading to a greater understanding of the numerous roles that education plays in the development of nations and in shaping individuals.

The IRC-2006 featured researchers from six continents presenting results of their analyses of IEA data in the fields of mathematics and science (TIMSS), reading and literacy (PIRLS), civic education and citizenship (CivEd), and information technology in education (SITES). Papers in this volume and its companion volume cover a range of topics around these themes.