EXAMINATION OF THE INFLUENCE OF SELECTED FACTORS ON PERFORMANCE ON ALBERTA LEARNING ACHIEVEMENT TESTS

W. Todd Rogers, Xin Ma, Don A. Klinger, Teresa Dawber, Laurie Hellsten, Denise Nowicki, & Joanna Tomkowicz

In this study, we identified factors that influence the academic performance of grade-6 students in a large city school system in language arts and mathematics. For language arts, five student variables, seven class variables, and two school level variables accounted for approximately 50 per cent, 75 per cent, and 90 per cent of the initial variability at the corresponding levels. For mathematics, seven student, five class, and three school level variables accounted for approximately 40 per cent, 60 per cent, and 90 per cent of the initial variability at the school level. Taken together, the results reveal that the majority of variability is at the student level, and additional variables need to be identified to better explain the variability at this level.

Key words: school effectiveness, hierarchical linear multi-level modeling, large-scale testing

Dans cette étude portant sur le système scolaire d'une grande ville, les auteurs identifient les facteurs ayant une incidence sur le rendement scolaire d'élèves de 6[e] année en langue (anglais) et en mathématiques. Pour le domaine de la langue, cinq variables ayant trait aux élèves expliquent environ la moitié de la variabilité initiale quant aux élèves, sept variables ayant trait à la classe expliquent environ 75 % de la variabilité initiale quant aux classes et deux variables ayant trait à l'école, environ 90 % de la variabilité initiale quant aux écoles. En mathématiques, sept variables ayant trait aux élèves, cinq ayant trait aux classes et trois aux écoles expliquent respectivement environ 40 %, 60 % et 90 % de la variabilité initiale. Pris en bloc, les résultats révèlent que la majeure partie de la variabilité se situe au niveau des élèves et que d'autres variables doivent être identifiées afin de mieux expliquer la variabilité à ce niveau.

Mots clés : efficacité de l'école, régression hiérarchique multiple, tests communs.

Without doubt, large-scale achievement testing is being used today to monitor the quality of education in schools throughout Canada. At the provincial level, nine of the ten provinces have large-scale achievement testing programs. At the national level, all the provinces and territories participate in the national Pan-Canadian Education Indicators Program (PCEIP, formerly SAIP) and the *Programme for International Student Assessment* (PISA). At the same time, increasing attention has been paid to determining student, classroom, and school factors that influence the performance of students (Fitz-Gibbon, 1998; Ma, 2001; Mandeville & Anderson 1987; Raudenbush & Byrk, 2002; Rumberger, 1995; Willms, 1992).

Relevant student-level factors include gender, prior performance, and family characteristics such as socio-economic status. For example, gender differences appear to be subject specific: males outperform females in mathematics and science, while females outperform males in reading and writing (Battistich, Solomon, Kim, Watson, & Schaps, 1995; Beller & Gafni, 1996; Sammons, West, & Hind, 1997). Willms (1992) pointed out the essential roles of both prior performance and socio-economic status when studying factors that influence student and school performance. Rogers, Wentzel, and Ndalichako (1997) found that prior performance accounted for 40 per cent to 50 per cent of the variance in performance in language arts and mathematics at the grades-3 and -6 levels. Lytton and Pyryt (1998) described socio-economic status as the "most ubiquitous and significant influence on achievement found in almost all investigations" (p. 282) of student achievement. They found that between 35 per cent and 50 per cent of the variability in the achievement of elementary school students was attributable to socio-economic status.

At the classroom and school levels, students perform better when there is more parental involvement (Goldring & Shapira, 1996; Ho & Willms, 1996), academic success is emphasized (Lytton & Pyryt, 1998; Zigarelli, 1996), and the disciplinary climate is conducive to teaching and learning (DeBaryshe, Patterson, & Capaldi, 1993; Ma & Klinger, 2000; Ma & Willms, 1995). In addition to these *practice* factors (Raudenbush & Willms, 1995; Willms, 1992), *school context* factors such as class and school size and average SES are often considered in studies of factors influencing achievement. For example, Rogers et al. (1997) found that, at the school level, between 30 per cent and 40 per cent of the variability among mean levels of performance in language arts and mathematics was accounted for by average SES for the school (see also, Ho & Willms, 1996). In contrast, despite theoretical expectations, school size has not been found to influence consistently academic achievement (Griffiths, 1996; Luyten, 1994).

Unfortunately the analyses employed in much of the previous research have not adequately accounted for the heirarchical nature of the data analyzed. Students belong to classes that in turn are in schools. If these level effects are not separated, then findings at the student level could be confounded by relationships with variables at one or both of the class and school levels. Heirarchical linear multilevel model (HLM) analyses can systematically estimate the separate influences or effects at the student- class-, and school-levels, thereby addressing this concern (Luke, 2004; Raudenbush & Byrk, 2002; Raudenbush & Willms, 1995).

There is need to consider simultaneously comprehensive sets of student-level variables along with practice and context variables at the class and school levels to better understand what influences student achievement. HLM analyses of the corresponding data for these variables will help provide a more complete understanding of what these influences are. Consequently, the purpose of the present study was to systematically examine, using HLM, the influence of an expanded set of student and class and school practice and context variables on the performance of grade-6 students in language arts and mathematics in a large city school system. Grade 6 was selected because it marks the end of elementary school and because the students' reading level was sufficient to respond to the student questionnaire.

METHOD

The study was conducted in a large city school system. Student performance on the Alberta Provincial Language Arts and Mathematics Achievement Tests administered to grade-6 students were the two dependent variables. Willms (1992, p. 58) suggested if the HLM model/analyses does not include measures of prior performance, the estimates of effects at the class and school levels would likely be biased.

Therefore, reading performance as measured by Highest Level of Achievement Test (HLAT) for grade 5 was used.[1] The school system provided the scores on the two achievement tests and the HLAT as well as special needs designations for each special needs student and a combined socio-economic index for each school. The combined socio-economic index was based on enumeration data provided by Statistics Canada and student mobility (percentages of students who transfer in, transfer out, withdraw, or suspended) for the school. The remaining data were collected by means of questionnaires completed by students, their teachers, and their principals.[2]

Survey Questionnaires

Student questionnaire. The items for the student questionnaire were adopted from questionnaires used in similar studies in British Columbia, New Brunswick, the National Longitudinal Study of Children & Youth (Human Resources Development Canada, 1997), and from the student questionnaires used in the advocate-adversary program evaluation of the Hawaii 3-on-2 Program (teams of three teachers instructing a class formed from two classes of students at consecutive grade levels; Northwest Regional Educational Laboratory, 1977). The items were arranged in four sections. In the first section, *About Me*, the students described themselves in terms of selected demographic characteristics (e.g., gender, family composition), the activities they did after school and on weekends, and things that they used at home to do their homework. In the second section, *About Me and My Feelings*, the students described how they felt about themselves. Self-esteem, reading and mathematics self-concept, and locus of control were assessed in this section. The students' feelings about school, including academic press, sense of belonging, enjoyment of school, and discipline and safety, were obtained in the third section, *About Me and My School*. In the fourth section, *My Parents, Me and My School*, the students indicated the activities they did with members of their family and how their parents (or guardian) were involved in their education.

The response options used for items related to activities were "yes" and "no." A five- or six-point Likert type scale was used for the remaining items (e.g., false, mostly false, sometimes false/sometimes true, mostly true, true; or every day, a couple of times a week, once a week, several times a month, at least once, never). Cartoon figures were inserted at different points in the questionnaire to encourage the students to continue.

Teacher and principal questionnaires. The teacher and principal questionnaires were developed to be parallel to each other as much as possible to allow the analysis of the same variables, but with different referents – class and school. The same items were used where it was appropriate to do so, with some differences. For example, the teacher questionnaire contained six items relevant to the how they organized and applied instruction the classroom. These items were not included in the principal's questionnaire.

The class- and school-level variables were organized into two sets. *Class and school practice* (Willms, 1992; Raudenbush & Willms, 1995) contained variables endogenous to a school such as leadership, curricular content, utilization of resources, classroom instruction, and opportunity to learn. *Class and school context* contained variables exogenous to the practices of administrators and teachers such as the social and economic composition and characteristics of a school's community, and the demographic composition of the student body.

The items for these two questionnaires were selected from questionnaires used in similar studies in British Columbia, New Brunswick, and the National Longitudinal Study of Children & Youth (Human Resources Development Canada, 1997) or constructed by the research team (e.g., classroom process items). The response options varied according to the nature of the item. For example, the teachers and principals were asked to provide numerical information for some items. For other items, a two point "yes – no" format was appropriate, while for other items Likert scales with varying numbers of scale points were appropriate.

Data Collection

The three questionnaires were distributed in sufficient numbers to each school principal for distribution within the school. This was done at the end of May, with completed questionnaires to be returned by the end of June. A teacher other than the students' regular teacher administered the student questionnaire during a class period. The grade-6 teachers and the principal separately completed their own questionnaires and placed them in the self-addressed envelopes provided. Completed questionnaires were returned to the Centre for Research in Applied Measurement and Evaluation at the University of Alberta.

Response Rates

Principals. There was at least one grade-6 class in 155 schools in the school system. In the case of one school, the students received their instruction through distance education and, therefore, did not meet as a class. Therefore, the number of eligible schools was 154. The principals of five schools did not allow their teachers or students to participate. Of the remaining 149 schools, no principal questionnaire was received from three despite follow-up telephone calls. One principal's questionnaire was not useable due to extensive non-response. Therefore, the final number of useable principal questionnaires was 145. Of this number, the student data for two classes in one school were lost in transit and teacher data for 15 schools were not available. Consequently, the final number of schools included in the analysis was 129.

Teachers. A total of 250 teacher questionnaires were received, of which 223 were useable. Seven of the 25 non-useable questionnaires were received from teachers who believed that the responses would not be treated confidentially; data for two teachers could not be used because the student questionnaires for their classes were lost in transit; and 18 teachers did not provide their room and school identification numbers making it impossible to match their data with the data of their students.

Students. The total student enrollment at grade 6 was estimated to be 5,135.[3] The total number of useable questionnaires returned was 4,864. Approximately 40 questionnaires were returned for students who had moved from the school; the parent(s) of three students asked that their child not participate; and, in the case of one school, the student data were lost.

Final sample sizes. Only classes of students for whom student, teacher, and principal data were available were included in the final samples for language arts and for mathematics. One hundred twenty-one students did not have a language arts score; 131 did not have a mathematics score.

A further 1,054 students were lost from each sample because of teacher and/or principal non-response. Lastly, an additional 65 students in 25 classes were removed because the classes they were in had fewer than five students, the minimum number considered necessary for the statistical analyses performed. Thus the final student sample sizes were 3,624 students for language arts and 3,643 students for mathematics. These students were in 198 classes in 129 schools.

Non-response bias check. To assess the influence of the non-response by principals and teachers, the geographical distribution of the 26 missing schools was assessed. First, they were located throughout the district, with no concentration in any one area of the city. Second, using HLM analyses, two variables at the class level were found that were significantly different ($p < 0.01$) between the two groups of classes/schools. Both were related to the presence of mild and severely impaired special needs students in the classroom; the percentages of these students were greater in those classes/schools with no teacher and/or principal data. No other differences were found.

Analysis

The responses to all the questionnaires were entered with 100 per cent verification. The error rate for out-of-range entries was less than 0.001 per cent.

Formation of variable item sets. The items included in the questionnaires were of two types. The first type included count data (such as the number of students with special needs or the number of students in the school) and demographic data (such as gender of the students, teachers, and principals, or the highest degree achieved by the teachers and principals). The second item type included measured variables like self-concept toward mathematics or academic emphasis given to language arts. Factor analysis (Gorsuch, 1983) was used to form the final set of items for each measured variable. To determine the number of factors, the Kaiser-Guttman rule (Guttman, 1954), the Scree test (Cattell, 1966), and the number of common image factors yielded by varimax rotation of the components yielded by an image analysis (Kaiser, 1962) were employed. Often these three procedures identified different, but close numbers of factors. A principal axis extraction followed by a varimax rotation and an oblimin transformation with $\delta = 0$ was then performed for the different number of factors identified. In each case the varimax solution yielded the best simple structure and clearest interpretation. Only the items that measured or assessed the variable of interest were retained. For example, while the students responded to six items that dealt with their own academic press, the final set included four items. The remaining two items were not consistent with these four; the inclusion of these two items led to an increase in the error of measurement. Lastly, factor scores were computed by adding the item scores for each item with a factor pattern coefficient greater than or equal to $|0.30|$ on that factor (Morris, 1979).

Identification of influential variables. Altogether, there were 26 variables (factors and individual items) at the student level, 62 variables at the class level, and 59 variables at the school level. A three-level Hierarchical Linear Model (HLM) (Luke, 2004; Kreft & de Leeuw, 1998; Raudenbush & Byrk, 2002) was used to determine the influence of these variables on language arts and mathematics performance at the student, class, and school levels. The random-intercepts model with fixed slopes (Raudenbush & Byrk, 2002) was employed. All the variables were grand mean centred. Except for the dichotomous variables (e.g., gender, mild/moderate disabling condition), the variables were standardized (mean zero and standard deviation one). The analyses were completed in two sequential steps using the HLM/3L computer program (Bryk, Raudenbush, & Congdon, 1996):

1. null model analyses to obtain an initial partitioning of the total variation into three components corresponding to the three levels of the analyses.
2. full model analyses which involved the prediction of a) achievement at Level 1 by the student variables, b) the Level 2 intercept, which is the class mean adjusted for the predictor variables at Level 1 by the class variables, and c) the Level 3 intercept, which is the school mean adjusted for the predictor variables adjusted for the predictor variables at Level 2, by the school variables. The estimation procedure takes into account the covariation among the Level 1, 2, and 3 predictor variables.

In the case of the Level 1 analysis, the number of students was sufficient to allow all the predictor variables to be entered at the same time. The variables initially retained at Level 1 were those variables that were significantly different from zero at the 0.01 level of significance and had a coefficient of at least $|0.08|$ in the case of language arts and $|0.05|$ in the case of mathematics. These two cut-off values corresponded to natural breaks in the distribution of the absolute values of the coefficients for each subject area.

The number of classes and schools was not sufficient to allow all the Level 2 and Level 3 variables to be entered simultaneously. In such cases, it is common to consider subsets of variables that are conceptually related (Raudenbush & Byrk, 2002). Consequently, a series of analyses were completed at Levels 2 and 3 in which the variables were entered in sets. For example, at Level 2 one set at the class level was "teacher background." The seven variables in this cluster were entered simultaneously. The variables retained were those that were significantly ($p < 0.01$) different from zero and had a coefficient of at least 0.05 in absolute value. Once each cluster had been analyzed, the variables retained were then analyzed simultaneously to identify from among this reduced set the final set of Level 2 variables that were significantly ($p < 0.01$) different from zero and had a coefficient of at least $|0.08|$ in the case of language arts and $|0.05|$ in the case of mathematics. The same two-stage procedure was repeated at Level 3. At each step, variables at the preceding levels that were initially selected were retained only if they continued to be significantly different from zero and satisfied the minimum value. Using this strategy, the model with the greatest parsimony was found, given the full set of variables at the three levels and the relationships among them.

HLM analyses require that there be complete data at all levels above Level 1. Therefore before beginning the analyses as described above, it was first necessary to address the issue of missing data due to non-response to individual items in the teacher and principal questionnaires. In the case of the principals, those who missed an item were telephoned and asked to provide their responses. It was not possible to call the teachers because their names were not known. In this case, the mean of the set of items for each variable was based on only the items completed. For those cases where all the items in a set were not answered, the mean for the remaining teachers was substituted. The numbers of missing responses (and the number of clusters) in which mean imputation was used were 11 (1 cluster), 6 (1), 5 (1), 3 (2), 2 (2), and 1 (4). Hence, the effect of the mean imputation upon the cluster variances was negligible.

RESULTS

The final results of the HLM analyses are summarized in Table 1 for language arts and Table 2 for mathematics. The values of the coefficients reported in these tables are interpreted as follows. As shown in Table 1, the coefficient for prior reading performance is 0.598. First, this value indicates that students who performed well on the prior reading test also performed well on the language arts test. Second, the value indicates that by holding all the other variables constant except prior reading performance, a change of one standard deviation in prior reading is associated with an improvement of 0.598 of a standard deviation in language arts. In the case of like gender, a dichotomous variable, the value of 0.269 indicates that females scored, on average, 0.269 standard deviations higher on the language arts test than the males, holding all other variables constant. It is important to note that there is no cause-and-effect claim. Instead these results reflect a relational interpretation that, for example, indicates that high scores on the language arts test tend to go with high scores on the prior reading test and that females tend to score higher on this test.

Influential Variables: Language Arts

Student level. The first four of the five predictors retained at the student level for language arts are context variables and, as such, are not amenable to manipulation (see Table 1). The strongest of these variables is prior reading achievement as measured by the Grade 5 HLAT. Students who performed well on this test tended to perform well on the language arts test. The second strongest predictor is gender: girls outperformed outperformed boys by 0.269 standard deviations.[4]

Table 1: Language Arts 3 Level Hierarchical Linear Model

Variable	Coefficient	Standard Error
Level 1: Student		
Prior reading achievement	.598	.016
Gender of student	.269	.021
Designated with a mild/moderate disabling condition	- .198	.067
Number of parents	.138	.026
Reading self-concept	.105	.013
Level 2: Class		
Gender of teacher	.119	.020
% Academic challenge	.076	.021
% with speech, hearing, vision, mobility or other health-related problem	- .072	.025
Teaching variety	- .068	.020
Parent involvement in child's Education	.061	.020
% Repeating grade	- .055	.013
Variety of assessment methods	.041	.016
Level 3: School		
Frequency of severe Discipline problems	- .071	.023
Frequency of academic recognition	.055	.019
Variances	Initial	Final (% Reduction)
Level 1: Student	.771	.371 (51.8%)
Level 2: Class	.153	.042 (72.8%)
Level 3: School	.101	.010 (90.5%)

Students designated with a mild/moderate disabling condition achieved at a level 0.198 standard deviations below students with no designated impairment.[5] Students with two parents or equivalent living at home scored 0.138 standard deviations higher than students with only one parent or equivalent living at home. The fifth variable, reading self-concept, was positively related to language arts performance. Students with higher reading self-concept outperformed students with lower reading self-concept by 0.105 standard deviations.

Class level. Of the seven variables that were retained at the class level for language arts, three were related to student composition of the class and a fourth was related to parent(s)/guardian(s) involvement in their children's education. The remaining three variables were related to characteristics of the teacher. As might be expected, classes with a higher percentage of students designated as academic challenge (i.e., gifted) had greater adjusted language arts class means than classes without such students. A classroom in which the proportion of academic challenge students is one standard deviation greater than the average classroom could be expected to have a class language arts mean that is 7.6 per cent greater than the average classroom mean. In contrast, the adjusted means for classes with a higher percentage of students with speech, hearing, vision, mobility, or other health-related problems were 0.072 class standard deviations lower than the adjusted means for the average classroom. As well, the adjusted performances of classes with greater than average numbers of students repeating grade 6 were 0.055 class standard deviations lower than the performances of classes with fewer repeating students. Lastly, as the proportion of students who reported that their parent(s)/guardian were involved in their (the students') education, held high expectations for them, and believed school was important increased, the classroom language arts score increased by 0.061 class standard deviations.

Turning to the teacher variables, the adjusted mean of classes with female teachers was 0.119 class standard deviations higher than the adjusted language arts mean of classes with male teachers. Teaching experience at a variety of grade levels was negatively related to class performance; the adjusted mean of classes with teachers who taught at multiple levels (Grades K-3; 4-6; 7-9; 10-12) was 0.068 class standard deviations less than the adjusted means of classes with teachers who had less variety. In contrast, the adjusted class mean for teachers who used a greater variety of assessment methods was 0.041 standard deviations higher than the adjusted mean for teachers who used a smaller variety of assessment methods.

School level. Two variables were retained at the school level. The adjusted school mean of schools with more severe discipline problems was 0.071 standard deviations lower than the adjusted school means of schools with less severe problems. In contrast, schools in which academic achievement was more frequently recognized at the school level outperformed schools in which academic achievement was less frequently recognized by 0.055 school standard deviations.

Influential Variables: Mathematics

Student level. Five of the seven predictors retained at the student level for mathematics are student context variables (see Table 2). Although not as strong as for language arts, prior reading achievement as measured by the HLAT administered to the students at the end of grade 5 was a significant predictor at the student level. For every increase of one standard deviation in the HLAT score, a student could be expected to score 0.476 standard deviations higher on the mathematics test. Likewise, gender was retained, but with less strength than that observed for language arts. However, in the case of mathematics, boys outperformed girls by 0.074 standard deviations. Both students designated with a mild/moderate disabling condition and students designated with a severe disabling condition tended to perform less well, 0.542 and 0.327 standard deviations, than students with no designated disabling condition. Students with two parents or equivalent living at home scored 0.128 standard deviations higher than students with only one parent or equivalent living at home.

The sixth and seventh variables at the student level were mathematics self-concept and school enjoyment. Mathematics self-concept was positively related to mathematics performance. Students with higher mathematics self-concept scores one standard deviation above the mean could be expected to score 0.277 standard deviations higher than the mean. Somewhat surprisingly, school enjoyment was negatively related to mathematics performance; students who enjoyed school scored 0.088 standard deviations lower than students who enjoyed school less.

Table 2: Mathematics 3 Level Hierarchical Linear Model

Variable	Coefficient	Standard Error
Level 1: Student		
Designated with a Mild/moderate disabling condition	-.542	.062
Prior reading achievement	.477	.016
Designated with a severe disabling condition	-.327	.160
Math self-concept	.277	.013
Number of parents	.128	.026
School enjoyment	-.088	.013
Gender of student	-.074	.024
Level 2: Class		
Math taught in the morning	.120	.046
% Repeating grade	- .106	.013
% Academic challenge	.067	.021
Students independently working alone or in small groups	.058	.021
Parent involvement in child's education	.058	.026
Level 3: School		
% Chronically late	-.076	.031
Frequency of severe Discipline problems	- .069	.028
% ESL	.050	.022
Variances	Initial	Final (% Reduction)
Level 1: Student	.751	.420 (44.0%)
Level 2: Class	.157	.061 (61.4%)
Level 3: School	.113	.012 (89.1%)

Class level. Two of the five variables that were retained at the class level were related to the student composition of the class and a third was related to the involvement of the students' parent(s)/guardians in their child's education. The remaining two were related to when mathematics was taught and the proportion of time students worked independently of the teacher. The adjusted mean level of mathematics performance for classes with a higher percentage of students designated as academic challenge students (i.e., gifted students) was 0.067 class standard deviations higher than the adjusted mean for classes with fewer such students. As for language arts, the mean levels of mathematics performance for classes with a greater percentage of students who were repeating grade 6 tended to be lower, by 0.106 class standard deviations, than classes with a smaller percentage or no repeaters. Classes with students whose parent(s)/guardian(s) were involved in their children's education, held high expectations for their children, and for whom school was important outperformed classes where this type of parent involvement was not as strong or prevalent by 0.058 class standard deviations.

Classes in which mathematics was taught in the morning outperformed classes in which mathematics was taught in the afternoon or in both the morning and afternoon by 0.120 class standard deviations. Further, classes in which the students spent a greater proportion of class time in mathematics working either alone or in small groups outperformed classes in which a greater proportion of time was spent on direct teaching to the whole class or small group instruction by 0.058 class standard deviations.

School level. Three variables were retained at the school level. Schools with higher percentages of chronically late students performed 0.076 school standard deviations below schools for which the percentages were lower. Similarly, the mathematics performance of schools with more severe discipline problems was 0.069 school standard deviations lower than the mathematics performance of schools with less severe problems. In contrast, schools with higher percentages of students for whom English was a second language scored 0.050 school standard deviations higher than schools with lower percentages of English-as-a-Second Language students.

Explained Variance

For both language arts and mathematics, most of the initial variation in achievement was among students within classes: 77.1 per cent in language arts and 75.1 per cent in mathematics. The next largest variation was at the class level: 15.3 per cent for language arts and 15.7 per cent for mathematics. Lastly, the initial variation at the school level was 10.1 per cent for language arts and 11.3 per cent for mathematics. In general, the amounts of variance to be accounted for at the class and school levels are rather modest in comparison to the amount of variance to be accounted for at the student level (see also Yair, 1997).

The five student level, seven class level, and two school level variables retained in the final model for language arts accounted for 51.8 per cent of the initial variability among students, 72.8 per cent of the variability among adjusted class means, and 90.5 per cent of the variability among adjusted school means (see Table 1). For mathematics, the seven student level, five class level, and three school level variables retained in the final model accounted for 44.0 per cent of the variability at the student level, 61.4 per cent of the variability at the class level, and 89.1 per cent of the variability at the school level (see Table 2).

The variances at all three levels of the final models for language arts and mathematics are considerably less than the initial values found for both subjects. However, there is still some unexplained variability, particularly at the student and class levels. Further, comparison of the variance components for language arts with the variance components for mathematics reveals that although the initial values of variance to be explained were quite comparable, the amounts explained for language arts at the student and class levels are greater than the corresponding amounts for mathematics.

DISCUSSION

Although a comprehensive set of variables was considered in the present study, there likely are other variables that were not fully captured in the questionnaires used. Two illustrative findings of the present study that support this likelihood are both related to gender. At the student level, although girls tended to outperform males in language arts, boys tended to outperform girls in mathematics. However, the difference in performance was greater for language arts than for mathematics, suggesting greater gender equity in mathematics than in language arts. This finding may be attributable to the emphasis placed on equity issues in mathematics education (Gambell & Hunter, 1999). However, degree of equity was not considered in the present study. At the class level, although the adjusted language arts mean for classes taught by female teachers was greater than the adjusted language arts mean for classes taught by male teachers, this difference did not appear for mathematics. Although the differences between boys and girls have been found in other similar studies and many ideas have been put forward to explain gender differences (Battistich, Solomon, Kim Watson, & Schaps, 1995; Gambell & Hunter, 1999; Ma & Klinger, 2000; Sammons, West & Hind, 1997), the differences found for teachers appear not to have been found in previous studies. Further, there are no ready explanations in the data in the present study that can be used to explain the teacher gender differences.

The failure of the indirect measures of SES used in the present study and by others (e.g., educational possessions such as a computer, internet access, and other educational resources) and involvement in social-culture activities (e.g., fine-arts lessons, participation in sports, attending sports events) to predict both language arts and mathematics performance is likely attributable to the medium to strong relationship between SES and achievement in general (Lytton & Pyryt, 1998; Sirin, 2005; Willms, 1992). The effects of these indirect measures of SES disappeared when prior achievement was considered. However, other indirect measures of SES, such as family structure (e.g., number of parents, number of siblings), did not disappear in the presence of prior achievement. For both language arts and mathematics, students with the equivalent of two parents at home tended to outperform students with one parent at home (see also Sammons et al., 1997). These findings point to the complex nature of SES, its relationship to achievement, and the care that needs to be taken when interpreting findings using indirect or surrogate measures (Sirin, 2005).

Student's self-concept in reading was positively related to performance in language arts but not in mathematics. Likewise, self-concept in mathematics was positively related to performance in mathematics but not in language arts. These subject specific findings are consistent with those reported by Marsh, Byrne, and Shavelson (1988) and Marsh (1992). What is not known is the direction of influence between these two variables. Does the development of positive self-concept influence motivation, sustained effort, and persistence, which in turn leads to improved academic performance and academic self-concept?

Students designated with a mild/moderate disabling condition tended to perform less well than students without a disabling condition in both language arts and, especially, mathematics. In addition, students designated with a severe disabling condition tended to perform less well than students without a disabling condition in mathematics. The difference in the strength of the relationships between language arts and mathematics is likely due to the preeminent role reading has in society. This emphasis translates into greater emphasis upon language arts in preservice education courses and in-service professional development. If the students with a disabling condition are required to write the same tests as non-disabled students, renewed attention needs to be paid to determine how best to help these students so that they may acquire the knowledge and skills measured by the tests and/or how their scores on these tests are interpreted (Guideline B. II. 2, *Principles for Fair Student Assessment Practices for Education in Canada*, 1993, p. 17).

The influence of student characteristics carried over to the class level. The performance of classes was dependent on the percentages and types of special needs children within the class. Expectedly, classes with higher percentages of gifted children outperformed classes with lower percentages of gifted students in both language arts and mathematics. In contrast, classes with higher percentages of students with speech, hearing, vision, mobility, or other health problems and classes with a greater number of repeaters performed less well than classes with lower percentages. An important observation to make here is that if these student characteristic variables at the class level were deleted from the analysis, class size entered as a predictor, with classes having a smaller number of students outperforming classes having a greater number of students. When these variables were included, class size did not enter the final equation (see Griffith, 1996; Luyten, 1994). Classes containing students with learning disabilities; speech, hearing, vision, mobility, or other health problems; or students who have or are repeating a grade, may be more successful if the class size is reduced.

Parental involvement played a different role at the class level than at the student level. Although parent(s)/guardian(s) involvement in their children's education was not significantly related to performance at the student level, it was positively related to class performance for both language arts and mathematics. This suggests that the absence or presence of parental/guardian involvement was essentially constant within each class, but varied between classes, with those classes having more involvement outperforming classes having less involvement (see also Sammons, et al., 1995). As pointed out by Lytton and Pyryt (1998), the findings for parental involvement "might reasonably also be viewed as a school [class] characteristic since the school [class] sets the motivation for and the climate in which volunteer effort flourishes or withers" (p. 293). Students learn more and perform better in classes (schools) that have strong positive parental involvement (Goldring & Shapira, 1996; Ho & Willms, 1996). Hence it is possible that attempts to encourage parent(s)/guardian(s) not already involved in their children's education to become more involved may result in higher overall achievement for the students in that class and school.

In addition to these two class context variables, class practice variables influenced performance. However, they were subject dependent. In the case of language arts, classes with female teachers outperformed classes with male teachers; classes with teachers who had taught at a greater number of grade levels during their teaching career performed less well than classes with teachers who had taught at fewer grade levels; and classes in which teachers used a greater variety of assessment methods outperformed classes where teachers used a more restricted set of assessment methods. For mathematics, classes in which mathematics was taught in the morning outperformed classes with mathematics taught in the afternoon or in both the morning and afternoon and classes in which students spent a greater proportion of time working either alone or in small groups outperformed classes in which a greater amount of time was devoted to direct teaching. Some of these findings were expected. For example, many teachers will say that "reading and 'rithmetic'" should be taught in the morning. Other findings, like the influence of the gender of the teacher and the teaching experience of teachers, are in need of further research to gain greater understanding. For example, how does the performance of male in-service teachers compare to their female counterparts? Are there differences in attendance patterns across subject areas at teacher professional days?

At the school level, one variable – disciplinary climate – was influential for both language arts and mathematics. For both subjects, the performance of schools in which more severe disciplinary problems (e.g., bullying, possession of weapons, use of drugs, and theft of school property) were more prevalent was lower than the performance of schools in which these problems were less prevalent. The remaining school level variables that influenced performance were subject dependent. Language arts performance was greater in schools that more frequently recognized academic success while mathematics performance was lower in schools with a greater proportion of chronically late students. Although others have shown that students learn more in schools that have a disciplinary climate conducive to teaching and learning (DeBaryshe, Patterson, & Capaldi, 1993; Ma & Klinger, 2000; Ma & Willms, 1995) and emphasize academic success (Lytton & Pyryt, 1998; Zigarelli, 1996), it is not clear why differences for the two variables were not consistent across the two subjects. Lastly, schools with a greater proportion of students for whom English was a second language had higher average mathematics scores than schools with a smaller proportion of English-as-Second-Language students. This finding is likely attributable to differences between the language requirements associated with the two subject areas.

CONCLUSION

Comprehensive sets of student variables at the student level and practice and context variables at the class and school levels were considered and analyzed in the present study. For language arts, the student variables retained were, in order of strength of prediction, prior reading performance, student gender, student designated with a mild/moderate disabling condition, number of parents, and reading self-concept. The class variables included teacher gender, per cent of students in class who were in academic challenge (i.e., gifted) programs, per cent of students in class who have a speech, hearing, vision, or other health related problem, use of a variety of teaching methods, parent(s)/guardian involvement in child's education, per cent of students in class who were repeating a grade, and use of a variety of assessment methods, while the school variables included frequency of severe discipline problems and frequency of academic recognition. These variables accounted for 51.8 per cent of the initial variability among students, 72.8 per cent of the initial variability among classes, and 90.5 per cent of the initial variability among schools.

For mathematics, the student variables were student designated with a mild/moderate disabling condition, prior reading achievement, student designated with a severe disabling condition, mathematics self-concept, number of parents, school enjoyment, and student gender. The class variables were mathematics taught in the morning, per cent of students in the class who were repeating grade 6, per cent of students in the class who were gifted, students working independently or in small groups, and parent(s)/guardian involvement in child's education, while the school variables included per cent of students in school who were chronically late, frequency of severe discipline problems, and per cent of students in school who had English as a second language were the three variables at the school level. These variables accounted for 44.0 per cent of the variability at the student level, 61.4 per cent of the initial variability at the class level, and 89.1 per cent of the initial variability at the school level.

Some of the results found in the present study correspond to findings present in the literature. For example, the analyses revealed the strong relationship between prior and current performance. Girls outperformed boys in language arts while boys outperformed girls in mathematics. Student self-concept was influential and, as expected, subject specific. Students with special needs and health problems performed less well. Parental involvement, positive disciplinary climate, and academic emphasis were associated with higher academic performance.

But other, unexpected and as yet unexplained results were found. There is no ready explanation in the current data that can be used to explain the finding linking language arts achievement to teacher gender. Nor is there a ready explanation of why certain class practice variables influenced performance in language arts and other class practice variables influenced performance in mathematics. Further, comparison of the number of variables retained for language arts and mathematics with the number of variables considered reveals that several variables were not retained. Although there are two reasons that may explain why some variables were not retained, reasons why others did not are not so readily available. First, predictor variables that are highly related to the academic performance of interest and at the same time to each other will not all enter the model because of the common variance they share. This reason advanced earlier to explain why some, but not all, of the indirect measures of SES disappeared in the presence of prior achievement. Likewise, the effect of class size at the class level and school size and SES at the school level disappeared when the composition of the class was considered. Findings like these may help explain why class and school size have not shown a consistent effect on academic achievement (Griffiths, 1996; Luyten, 1994). Other variables reported to effect performance in other studies of school effectiveness, such as principal leadership, were not found in the present study. Once again, the common or shared variance and/or lack of variance may explain why these variables were not in the final model. But it may be that the total reliance on survey methods is a shortcoming (Willms, 1992). Replicated case studies in which classes are observed over an extended period of time and teachers and principals are interviewed at different time points during the school year may yield the data and information needed to clarify issues like these and to identify other variables that will reduce further the amount of unexplained variability at the student and class levels.[6]

ACKNOWLEDGEMENTS

The authors of this report wish to thank the members of the Advisory Committee for their advice and the support they provided to us. In addition, special thanks are extended to Alberta Learning, the Alberta Teachers' Association, Edmonton Public Schools, and the Applied Research Branch, Human Resources

Development, Canada, who provided the necessary funding to complete the study. We especially want to express our sincere and deep appreciation to the grade-6 students, parents, teachers, and principals who completed the survey questionnaires. Their cooperation at a busy time of the year reflects, in our opinion, their commitment to education and its improvement.

NOTES

1. The HLAT is part of the school district's assessment program. Unfortunately there was no measure of prior mathematics achievement. In light of the positive correlation between language arts and mathematics performance on the grade-6 provincial examinations, the prior reading measure, which was administered the year before, was used for both language arts and mathematics.
2. A questionnaire was also developed and sent to parents. However, the number of parent(s)/guardian questionnaires returned was insufficient to include the parent(s)/guardian in the analyses. Further, the sample received was not representative of the full population of parents, particularly parents living in lower income areas within the city. Consequently, no further attention was given to the parent component of the study.
3. This estimate is based on the official enrollment count for the school district in which the study was conducted.
4. The interpretation of the standard deviation here and in all that follow is the same as that discussed above. The proportion indicates that by holding all the other variables in the model for language arts constant except prior reading, a change of one standard deviation in prior reading is associated with an improvement of 0.269 standard deviations in language arts.
5. As explained on page 518, the presence of both mildly and severally impaired children differed between the classrooms for which the teacher and/or principal data were available and the classrooms for which the teacher and/or principal data were missing. Had these data been available, it may have been that the severely impaired variable would have also been retained in the final model.
6. Copies of the student, teacher, and principal questionnaires used in this study may be obtained from the first author at the Centre for Research in Applied Measurement and Evaluation, 6–110 Education Centre North, University of Alberta, Edmonton, AB, T6G 2G5

REFERENCES

Battistich, V., Solomon, D., Kim, D., Watson, M., & Schaps, E. (1995). Schools as communities, poverty levels of student populations, and students' attitudes, motives, and performance: A multilevel analysis. *American Educational Research Journal,* 32(3), 627-658.

Beller, M., & Gafni, N. (1996). The 1991 International Assessment of Educational Progress in mathematics and sciences: The gender differences perspective. *Journal of Educational Psychology,* 88, 365-377.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

Byrk, A. S., Raudenbush, S. W., & Congdon, Jr., T. T. (1996). *HLM*. Chicago, IL: Scientific Software International, Inc.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.

DeBaryshe, B. D., Patterson, G. R., & Capaldi, D. M. (1993). A performance model for academic achievement in early adolescent boys. *Developmental Psychology,* 29(5), 795-804.

Fitz-Gibbon, C. (1998). Indicator systems for schools: Fire fighting it is! In D. Shorrocks-Taylor (Ed.), *Directions in Educational Psychology* (pp. 189-205). London, UK: Whurr Publications Ltd.

Gambell, T. J., & Hunter, D. M. (1999) Rethinking gender differences in literacy. *Canadian Journal of Education,* 24(1), 1-16.

Goldring, E. B., & Shapira, R. (1996). Principals' survival with parental involvement. *School Effectiveness and School Improvement*, 7, 342-360.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Griffiths, J. (1996). Relation of parental involvement, empowerment and school traits to student academic performance. *Journal of Educational Research,* 90(1), 33-41.

Guttman, L. (1954). Some conditions for common factor analysis. *Psychometrika*, *19*, 149-161.

Ho, S., & Willms, J. D. (1996). Effects of parental involvement on eighth-grade achievement. *Sociology of Education*, 69(2), 126-141.

Human Resources Development Canada. (1997). *National longitudinal survey of children & youth cycle 2 survey instruments book 2 – education, 10-11 and 12-13 year olds*. Ottawa, ON: The Author.

Kaiser, H. F. (1962). Image analysis. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 156-166). Madison, WI: The University of Wisconsin Press.

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Newbury Park, CA: Sage Publications.

Luke, D. A. (2004). Multilevel modeling. Sage university paper series. Quantitative applications in the social sciences, 07-143. Beverly Hills: Sage Publications.

Lytton, H., & Pyryt, M. (1998). Predictors of achievement in basic skills: A Canadian effective schools study. *Canadian Journal of Education,* 23(3), 281-301.

Luyten, H. (1994). School size effects on achievement in secondary education: Evidence from the Netherlands, Sweden, and the USA. *School Effectiveness and School Improvement, 5*, 75-99.

Ma, X. (2001). Stability of school academic performance across subject areas. *Journal of Educational Measurement*, 38(1), 1-18.

Ma, X., & Klinger, D. A. (2000). Hierarchical linear modeling of student and school effects on academic achievement. *Canadian Journal of Education*, *24*, 41-55.

Ma, X., & Willms, J. D. (1995, April). *The effects of school disciplinary climate on eighth grade achievement.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Mandeville, G. K. & Anderson, L. W. (1987). The stability of school effectiveness indices across grade levels and subject areas. *Journal of Educational Measurement*, 24(3), 203-216.

Marsh, H. (1992). *The content specificity of relations between academic self-concept and achievement: An extension of the Marsh/Shavelson model.* ERIC NO: ED349315.

Marsh, H. W., Byrne, B. M., and Shavelson, R. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, 80(3), 366-380.

Morris, J. D. (1979). A comparison of regression prediction accuracy on several types of factor scores. *American Educational Research Journal*, 16(1), 17-24.

Northwest Regional Educational Laboratory. (1977). *3 on 2 evaluation report*. Portland, OR: Author.

Principles for Fair Student Assessment Practices for Education in Canada. (1993). Edmonton, Alberta. (Mailing Address: Joint Advisory Committee, Centre for Research in Applied Measurement and Evaluation, 6-110 Education Building North, University of Alberta, Edmonton, Alberta, T6G 2G5).

Raudenbush, S. W., & Byrk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Second edition). Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20(4), 307-335.

Rogers, W. T., Wentzel, C., & Ndalichako, J. (1997). *Examination of the influence of selected factors on performance on Alberta Education Achievement Tests within Edmonton Public Schools* (Tech. Rep.). Edmonton, AB: University of Alberta, Centre for Research in Applied Measurement and Evaluation.

Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal*, 32(3), 583-625.

Sammons, P., West, A., & Hind, A., (1997). Accounting for variations in pupil attainment at the end of Key Stage 1. *British Educational Research Journal,* 23(4), 489-511.

Sirin, S. R. (2005). Socioeconomic status and achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417-453.

Willms, J. D. (1992). *Monitoring school performance: A guide for educators*. Lewes: Falmer.

Yair, G. (1997). When classrooms matter: Implications or between classroom variability for educational policy in Israel. *Assessment in Education*, 4(2), 225-249.

Zigarelli, M. A. (1996). An empirical test of conclusions from effective schools research. *Journal of Educational Research,* 90(2), 103-110.