What "Extras" Do We Get with Extracurriculars?
Technical Research Considerations

September 30, 2003

Duncan Chaplin
Michael J. Puma

The Urban Institute
2100 M Street NW
Washington, DC 20037

This report was prepared with support from the U.S. Department of Education, OERI Grant #R305T010557. We would like to gratefully acknowledge advice from Greg Acs, Avner Ahituv, Steve Bell, Avinash Bhati, Chris Bollinger, Mark Dynarski, Sharon Long, Jerome Lord, Chris Swanson, and Doug Wissoker and research assistance from Shannon McKay, Precious Jackson, and Ella Gao. Any opinions, observations, findings, conclusions, and recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the Urban Institute or the U.S. Department of Education.

Summary

Education has become one of the most important issues in American society, often dominating the political landscape in response to lagging academic achievement and persistent gaps in performance between advantaged and disadvantaged children. In response, a variety of educational reforms have been implemented ranging from increased accountability to parental school choice and efforts to improve educational "inputs" such as teacher skills.

An alternative strategy that has gained in popularity involves efforts to expand public and private investments in out-of-school-time programs, particularly those targeting the needs of disadvantaged children. Such programs provide opportunities for children and teens to have positive interactions with caring adults, improve socialization with other children, learn to avoid risky behaviors, and, last but not least, improve academic skills. However, students spend the majority of their time in these programs participating in extracurricular activities that are not designed specifically to improve academic skills. Consequently, this study asks two important research questions.

1. Does participation in out-of-school extracurricular activities improve academic achievement or behavior for elementary school children?
2. If so, are the impacts of participation related to the types of extracurricular activities that students pursue (e.g., music and arts, language, computer classes, sports)?

Initial analyses were conducted using relatively standard techniques to control for the fact that students who participate in out-of-school-time extracurricular activities differ in important ways from nonparticipants. These analyses found statistically significant and positive effects of such participation in arts, music, drama, and language classes. We were ready to write our report at this stage, but then tried to replicate our analyses using alternative statistical models. The results were not only highly dependent on the choice of analytical approach, *but the initial positive effects disappeared* when more appropriate analytical models were used.

We then examined the current literature that has been used to support investments in out-of-school-time extracurricular programs and found that the reported positive effects are likely a result of the same type of analytical error that led us to initially conclude that there were positive effects for elementary school children.

At a minimum, these results raise serious questions about the validity of many of the claims that out-of-school-time programs that do not directly target academic outcomes will improve such outcomes nonetheless. The results also suggest that more rigorous evaluation be conducted before further investments are made based on such claims, particularly where such expenditures are made as a trade-off against investments in other, possibly more effective, educational reforms.

Chapter 1 : Background

A challenge for our society is the persistent gap in academic performance between advantaged and disadvantaged children. Data from the National Assessment of Educational Progress (NAEP)—the nation's education "report card"—has, for example, shown enduring racial and ethnic disparities, with African-American and Latino 17-year-olds performing at the level of 13-year-old white students (Campbell, Hombo, and Mazzeo 2000). Although African-American and Latino students made significant progress toward closing the gap during the 1970s, subsequent NAEP achievement scores have stagnated for all ethnic groups (Campbell et al. 2000). This gap in academic achievement is associated with subsequent racial and ethnic differences in high school graduation rates, college attendance and completion rates, and, ultimately, employment and earnings.

As a consequence, investments in education have become a primary focus of federal, state, and local governments and the philanthropic community. But deciding how to improve educational outcomes, especially for disadvantaged children, has been a major challenge. For some, the answer lies in external systemic policy reform, including increased academic standards and expectations; greater accountability for school, teacher, and student performance; increased flexibility to accompany heightened accountability; sanctions for poor performance; and allowing parents to opt out of failing schools. Many of these themes characterize the recent No Child Left Behind Act of 2001 that requires states, districts, and schools receiving federal compensatory education funding (i.e., under Title I) to establish clear and measurable goals for what students are expected to learn and to use annual state assessments in math and reading in grades 3–8 to ensure that every child meets the goals. Schools that fail to make sufficient progress will receive special assistance, and if poor performance continues, students will be provided with the opportunity to attend better schools.

An alternative strategy focuses on internal reforms that seek to improve the quality of what happens inside school buildings. This strategy includes efforts to raise teacher skills through pre- and in-service training and increased applications of modern digital

technology (both included in the recent federal legislation) and a plethora of comprehensive school reforms that seek to change many aspects of the school environment (e.g., Success for All, HOTS, Core Knowledge).

Out-of-School-Time[1] Extracurricular Activities

Another alternative to these largely school-focused initiatives is to extend or augment the school experience by engaging students during their out-of-school time. Proponents view this as especially important for poor and disadvantaged students who lack access to the rich set of opportunities afforded more well-off students when they are not in school.

Students spend about 70 percent of their waking hours outside of school (Clark 1993; Miller et al. 1997), and this time "is seldom spent in activities that reinforce what they are learning in their classes" (Steinberg 1996). As a result, extending the learning day offers the potential to increase academic achievement by augmenting what takes place in school. In addition, there is a need to protect children from hazards when they are not in school and to deter them from experimenting with high-risk behaviors. The American family has changed dramatically in the past several decades and this has had consequences for the nation's children. Today, in about 68 percent of married-couple families with children age 6 to 17, both parents work outside the home; in single-parent families, 78 percent of female-headed families, and 84 percent of male-headed families, the custodial parent works outside the home (U.S. Bureau of Labor Statistics 2003). As a consequence, there are an estimated 4 million children age 5 to 12 who regularly spend time without adult supervision (Hofferth, Jankuniene, and Brandon 2000).

The time differential between when children leave school and when parents get home from work can amount to 20–25 hours per week (James and Jurich 1999). According to Vandell and Posner (1999), 44 percent of 3rd grade students spent at least some of their after-school time in unsupervised settings. Children without adult supervision are at significantly greater risk of poor school performance, risk-taking behavior, and substance

[1] In appendix D we describe how the activities covered in our study relate to those commonly covered by studies of out-of-school, after-school, and enrichment activities.

abuse, and the greater the amount of time spent in self-care, the higher the risk of poor outcomes (Pettit et al. 1997).

The Policy Agenda

In response to both lagging student achievement and concerns for the safety of school-age children, there is growing interest in expanding public and private investments in out-of-school-time programs, particularly those targeting the needs of disadvantaged children. Such programs are seen as an opportunity for children and teens to have positive interactions with caring adults, improve their socialization with other children, and improve their academic skills. In fact, nearly two-thirds of adults believe that after-school programs can improve outcomes for America's children and youth (Public Agenda 1997), and more than nine in ten agree that "there should be some type of organized activity or place for children and teens to go after school every day that provides an opportunity to learn" (Afterschool Alliance 2001). Seven out of ten voters would even support an $800 million annual expansion of federal funding for after-school programs (Afterschool Alliance 2001).

One of the most recent, and by far the largest, efforts to expand out-of-school-time programs is the 21st Century Community Learning Centers initiative funded by the U.S. Department of Education that has grown from a $1 million demonstration program in 1998 to an appropriation of almost $1 billion for fiscal year 2003. In addition, 26 states are reportedly increasing funding for after-school programs, and many others blend funds from private donors and/or child care, crime prevention, public safety, and recreation budgets (National Governors Association 1999).

Research on Out-of-School Time

Most of the research on the effects of out-of-school activities has focused on non-academic activities or on high school youth. Zill, Nord, and Loomis (1995), for example, examined the relationship between the way adolescents use their nonschool time and the incidence of high-risk behaviors such as drinking and substance abuse and early sexual activity, while others have studied the relationship between academic achievement and

participation in after-school activities by adolescents (Braddock 1981; Brown, Kohrs, and Lazarro 1991; Eide and Ronan 2000; Thomas and Moran 1991).

Although the research base on elementary and middle school students is more limited (Larner, Zippiroli, and Behrman 1999), there are indications that compared with nonparticipants, children who attend after-school programs display better peer relations and emotional adjustment (Baker and Witt 1996; Posner and Vandell 1994), improved social skills (Marshall et al. 1997), better schoolwork habits (Posner and Vandell 1994; Vandell and Pierce 1997, 1999), and higher school grades (Mayesky 1980a, b; Posner and Vandell 1994).

For example, Vandell and Posner (1999) report on evidence suggesting that students who spent more time **alone after school** while in 3rd grade exhibited a higher incidence of behavior problems both concurrently and for at least two additional years. More **unsupervised time with peers** was similarly associated with increased behavioral problems and with poor school adjustment in a number of grades. At the other end of the activities spectrum, the authors found that 3rd graders who spent more time in out-of-school enrichment activities were reported by their teachers to have better conduct in school, better work habits, and better relationships with their peers.

Using data from the nationally representative Early Childhood Longitudinal Study, Kindergarten Cohort, on 22,000 children enrolled in about 1,100 kindergarten programs (ECLS-K), Reaney, Denton, and West (2001) found that about two-thirds of the children were involved in at least one extracurricular activity and that having a greater number of "family risk factors" significantly correlated with lower participation in after-school activities. Most important, children with higher participation in after-school activities demonstrated higher scores on tests of reading, math, and general knowledge skills.

In addition, several evaluations of specific after-school programs have reported positive impacts on school grades, test scores, and classroom behavior (Whitaker, Gray, and Roole 1998). For example, Hamilton and Klein (1998) studied an after-school program in Philadelphia and found that 4th grade participants outperformed comparison

students in reading, language arts, and mathematics.[2] Huang and various colleagues have completed several studies of "LA's BEST," an after-school program that provides a variety of educational, recreational, and interpersonal skill-building activities for grade K–5 students. The most recent evaluation (Huang et al. 2000) followed 2nd through 5th grade participants over four years and found that long-term involvement (i.e., at least four years) in the program led to better school attendance, which in turn showed a significant correlation with higher achievement on standardized tests in mathematics, reading, and language arts.

More generally, many prominent researchers argue that evaluations of out-of-school programs have not yet demonstrated reliable impacts on youth outcomes, especially academic achievement (Fashola 1998). In most cases, the argument is that the research in this area has not involved adequate controls for "selection bias," caused by the fact that children who participate in out-of-school activities appear to be quite different from nonparticipants even before they participate, and these differences appear to be highly related to academic achievement and other developmental outcomes. Consequently, a lack of good controls may cause researchers to attribute these preexisting differences at least in part to participation.[3]

[2] Sample sizes for this study were, however, small: 213 in 1997–98 and 215 in 1998–99.

[3] Fashola (1998) favors the use of experiments, comparisons of students at different points on waiting lists, or comparisons of students in schools with and without programs. However, such data are generally hard to come by.

Chapter 2 : Methodology

This study addresses the questions of whether participation in out-of-school-time extracurricular activities by elementary school children increases academic achievement and improves school behavior and, if so, whether the types of activities that students pursue (e.g., music and arts, language, computer classes, sports) matter. By "extracurricular," we mean activities that are not designed to directly impact the academic skills targeted most heavily by current education reforms (math and reading). To the extent possible, this research effort tries to better control for the selection bias that has plagued much research in this area.

The Data

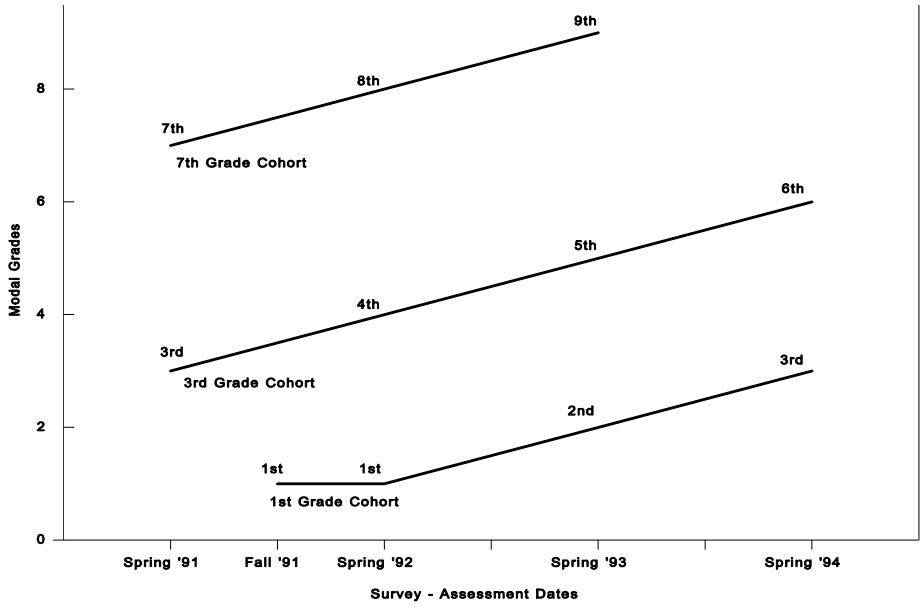
The analysis described in this paper is based on data collected as part of *Prospects*, the congressionally mandated study of the federal Title I program for disadvantaged children. The *Prospects* study, conducted between 1990 and 1997 by the U.S. Department of Education, is the largest assessment of educational opportunity since the 1966 Coleman Report and includes data collected from a nationally representative sample of more than 35,000 public school students (Puma et al. 1993, 1997).

These data are unique for several reasons. First, the data were collected from students covering ***multiple school grades***, and each student was followed ***longitudinally*** for up to four years (see exhibit 1). Three cohorts of students were included: (1) ***a 1st grade cohort*** consisting of 10,820 students who began 1st grade in the fall of 1991 and who were tracked from entry into school through completion of the 3rd grade in the spring of 1994; (2) ***a 3rd grade cohort*** consisting of 10,330 students enrolled in the 3rd grade in the 1990–91 school year, who were tracked from the end of their 3rd grade year (spring 1991) through completion of the 6th grade in the spring of 1994; and (3) ***a 7th grade cohort*** consisting of 7,215 students enrolled in the 7th grade during the 1990–91 school year, who were tracked from the end of 7th grade (spring 1991) through the completion of the 9th grade in the spring of 1993.

Second, the following consistent annual data (see exhibit 1) were collected from each student during the planned waves of data collection:

- ***Student Data.*** (1) Data were abstracted from **school records**; (2) teachers were asked to complete a **detailed profile** for each student including ratings of ability, motivation, attitudes, classroom behaviors, language skills, and health status; and (3) starting in the 3rd grade, **questionnaires** were administered to all students to collect information on current and past educational experiences, individual and family demographics, educational aspirations, perceived academic strengths, school grades and attendance, parent participation in their schooling, and participation in extracurricular activities within and outside of school.

Exhibit 1. Prospects Data Collection Points by Grade Cohort



- ***Parent Data.*** Information was collected on parent and family characteristics, parental attitudes and practices on student learning, and participation in out-of-school-time activities.
- ***Academic Achievement.*** All students were given the Comprehensive Test of Basic Skills (CTBS), 4th edition, a vertically equated test series designed to measure achievement status and gains in reading/language arts and mathematics. For students who could not be tested in English, the Spanish Assessment of Basic Education (SABE) was used.
- ***Teacher Data.*** Each student's reading/language arts and math teacher provided information on demographic characteristics (gender, race/ethnicity), teaching certification and experience, school climate ratings, classroom instructional practices,

class size and staffing in the classroom, school governance, and access to instructional resources.

- ***Principal and School Data.*** A principal questionnaire covered the principal's background and experience, school policies, administrative techniques, and school program features. A separate school characteristics survey focused on the organization of the school, staffing, enrollment and student demographics, and special programs offered.
- ***District Data.*** A survey of district administrators provided information on staffing, enrollment, daily attendance, length of the school day/year, provision of in-service staff training, overall student demographics, and information on the availability and structure of compensatory education programs.

Finally, the *Prospects* data are ***hierarchical*** in nature, with more than 80 students per school in the first year of the survey. As a consequence, the data can support the analysis of the rich relationships among student characteristics and in- and out-of-school experiences, family environment, and school and classroom characteristics and instructional processes.

Analysis Approach

As noted in chapter 1, a problem plaguing much of the research on out-of-school-time activities is what is commonly called "selection bias," that is, students who choose to participate in such programs are different from those who opt not to do so, and these differences are typically related to the outcomes of interest (e.g., academic and social skills, risk-taking behavior). The ideal solution to the selection problem is to use a true experiment. Since education researchers (and most other social scientists) can seldom conduct experiments,[4] they try to control for selection using nonexperimental methods, often relying heavily on the lagged value of the outcome as a control for selection bias—for example, controlling for a student's prior year test score when estimating the impacts of extracurricular activities on current test scores. This lag outcome variable enables researchers to, in theory, control for all unobserved factors that affected the outcome of interest before the intervention had its impact.

[4] Though we, like Fashola (1998) and others, certainly favor experiments when possible.

For this reason the lagged outcome is—correctly—considered a crucial and powerful control variable, and its use has become quite common in educational research.[5] There are two main strategies for including the lagged outcome variable as a statistical control. Some researchers use a **lag model** that adds the outcome value from the previous period as an additional independent variable. (See, for example, Chaplin 1998a, b; Gamoran and Hannigan 2000; Goldhaber and Brewer 2000; Jacob 2001; Ladd and Walsh 2002; Ludwig and Bassi 1999; Meyer 1996, 1999, 2002; and Phillips, Crouse, and Ralph 1998.) Alternatively, researchers use a **growth model** approach where the dependent variable is the change (or growth) in the outcome variable (e.g., Alexander, Entwisle, and Olson 2001; Becker and Powers 2001; Blau et al. 2001; Bryk and Raudenbush 1992; Entwisle and Alexander 1992; Gamoran et al. 1997; Mayer 1998; Morgan 2001; Snijders 1996; and Sunmonu et al. 2002).[6]

Meyer (1996) notes that in theory the growth model can be thought of as a restricted version of the lag model and we show this in chapter 3. However, even if the restriction is satisfied (that the coefficient on the lagged outcome equals one), measurement error in the lagged outcome variable can cause results to differ greatly between these two models. Measurement error can exist for at least two reasons. First, a student may happen to do better (or worse) than they normally would on a test at a given time because of variations in their mood or energy level. Second, the questions on a given test do not cover all material in a subject area. For both of these reasons a student's performance on a given test taken at a given time will not be a completely accurate measure of their true skills.[7] One can think of the difference between a student's true skills and their score on a specific test as measurement error. Measurement error in the outcome being used as a dependent variable will not bias coefficient estimates as long as it is uncorrelated with all

---

[5] We focus on papers that have lagged values of the outcome measured using the same metric as the outcome itself. Without this condition, the growth model, discussed below, would be difficult to justify.

[6] Similar methods have been used outside of education research, for instance by Long and Wissoker (1995). Growth models are also used when researchers have access to data on more than two points in time. The discussion here is limited to analyses of data on only two periods.

[7] The definition of true skills is somewhat unclear. One plausible definition would be how the student would score, on average, if they were given a large number of tests (using a different set of questions each time) at many different times.

other variables in the model. However, measurement error in a lagged outcome variable that is used as a control variable will cause bias even if it is totally random, as shown in appendix C.

Fortunately there are ways to control for measurement error in control variables. Once this is done, we find that the lag and growth model approaches generally yield similar results. But, as noted by Meyer (1996, 1999, and 2002), these results are often very different from results based on the lag model without controls for measurement error. Nevertheless, lag models without controls for measurement error are still quite common (Baker and Witt 1996; Chaplin 1998a, b; Gamoran and Hannigan 2000; Goldhaber and Brewer 2000; Ludwig and Bassi 1999; Marcus 1997; Mussoline and Shouse 2001; Smith 1996). In addition, many researchers omit even the control for the lagged value, at least in part because data on lagged outcomes are often difficult to obtain (Huang et al. 2000; Marshall et al. 1997; Posner and Vandell 1994; Reaney et al. 2001).[8]

In this paper, we estimate lag and growth models with and without controls for measurement error and models that omit the lagged outcome entirely (reflecting the variety of analyses in the current literature) and find some interesting and potentially important differences in results.

Measures of Out-of-School Activities

The key independent variables used in this analysis are the different types of "classes" that 4th graders take "outside of regular school."[9] Parents were asked about six types of out-of-school classes: (1) arts, music, or dance lessons, (2) language classes, (3) religious instruction, (4) computer classes, (5) sports, exercise, or gymnastics, and (6) history or culture lessons. Similarly, students were asked about the same set of activities, except for the "history or culture" category. We use the parent reports, as those

[8] Some researchers reporting benefits of after-school activities do so without a comparison group (e.g., Gregory 1996 and Riley et al. 1994). This type of research is not addressed here.

[9]The parents were asked, "Does your child attend classes outside of regular school to study any of the following?"

appear to be more accurate.[10] A summary variable was also created to measure if the parent reported that the child participated in **any** of the six types of activities; about 70 percent of the students reporting participated in at least one of these out-of-school-time activities.

The outcomes of interest include both academic test scores and classroom behaviors. Although the original intent was to examine effects for other years and grades, the analysis was restricted to the one-year follow-up of the grade 3 cohort because this is the only grade-cohort combination that has all of the necessary information.[11]

Outcome Variables

As noted above, this analysis focused on two categories of school-related student outcomes—academic achievement (standardized test scores) and teacher ratings of classroom behaviors. With regard to the academic outcomes, *Prospects* study staff administered the CTBS to each student in the spring of every school year (see exhibit 1). The CTBS scale-scores, designed to be compared across grades using Item Response Theory, cover four academic areas: (1) math concepts and applications (SSMA), (2) math computation skills (SSMC), (3) reading comprehension (SSRC), and (4) reading vocabulary (SSRV). The behavioral outcomes were based on teacher ratings of individual study students covering a wide range of domains. These ratings were factor analyzed to create three separate 3-point scales: student attention in class (ATTEN), student cooperation with classroom rules and procedures (COOP), and student classroom participation (PARTIC). See appendix B for details on how these variables were constructed.

---

[10] Evidence of this is discussed below in the section on measurement error.

[11] Decisions made by the U.S. Department of Education resulted in the dropping of specific questionnaire items from both the student and parent surveys as the *Prospects* study progressed.

Exhibit 2. Descriptive Statistics by Participation in Out-of-School-Time Lessons

	Participate?					
	No		Yes			
Variables	Mean	Std	Mean	Std	Absolute difference	Standardized difference
Outcome variables						
Math: concepts and applications, 1992	682.94	53.45	705.07	57.51	22.13	0.38
(SSMA) 1991	669.31	51.85	690.93	53.56	21.62	0.40
Math computation, 1992	686.63	40.50	698.79	42.89	12.16	0.28
(SSMC) 1991	663.46	47.48	677.45	47.30	13.99	0.30
Reading comprehension, 1992	687.42	51.32	706.81	53.68	19.39	0.36
(SSRC) 1991	665.51	57.08	687.91	59.60	22.40	0.38
Reading vocabulary, 1992	674.70	45.13	693.95	47.96	19.25	0.40
(SSRV) 1991	662.42	42.78	679.32	43.59	16.90	0.39
Teacher ratings						
Attentiveness, 1992	2.26	0.57	2.42	0.54	0.16	0.30
(ATTEN) 1991	2.31	0.56	2.44	0.52	0.13	0.25
Cooperativeness, 1992	2.59	0.43	2.67	0.40	0.08	0.20
(COOP) 1991	2.62	0.41	2.67	0.38	0.05	0.13
Participation, 1992	1.93	0.58	2.05	0.56	0.12	0.21
(PARTIC) 1991	1.97	0.57	2.07	0.54	0.10	0.19
Control variables						
Limited English proficient (ever)	0.15	0.36	0.09	0.28	-0.06	-0.21
Black	0.19	0.39	0.11	0.31	-0.08	-0.26
Hispanic	0.20	0.40	0.13	0.34	-0.07	-0.21
Female	0.57	0.49	0.52	0.50	-0.05	-0.10
Parent socioeconomic status	-0.20	0.83	0.31	0.91	0.51	0.56
Urban school	0.41	0.49	0.34	0.47	-0.07	-0.15
Rural school	0.34	0.47	0.34	0.47	0.00	0.00
"Educational load"	0.92	4.72	-1.04	4.36	-1.96	-0.45
Free/Reduced lunch percent	58.63	27.69	44.73	27.95	-13.90	-0.50
Change schools (ever)	0.15	0.36	0.09	0.29	-0.06	-0.21
Single-parent household	0.15	0.36	0.09	0.29	-0.06	-0.21
Differences between parent and child reports in 1992	0.89	0.99	1.27	1.01	0.38	0.38
Instrumental variable						
Differences between parent and child reports in 1991	1.13	1.04	1.28	1.01	0.15	0.15
Sample size	1,183		3,099			

Source: Prospects data, 3rd grade cohort, 1992. Sample used for Math Concepts and Applications.

Note: Standardized difference is absolute difference in means divided by standard error for the "yes" response.

Not surprisingly, as shown in exhibit 2,[12] participants do better than nonparticipants on all of the selected outcome measures.[13] For instance, participants score about 0.30 to 0.40 standard deviations higher than the nonparticipants on all of the academic tests and about 0.13 to 0.30 standard deviations higher on the behavioral scales. In addition, while the absolute point difference in test scores between participants and nonparticipants increase in two of the four subjects between 1991 and 1992, the standard deviation differences decrease for three of the subjects. Although the differences in test scores between participants and nonparticipants do not always move in the expected direction over time (upwards), changes in the teacher-reported behaviors do. However, the relative improvements in the behaviors of participants appears to be driven mostly by a deterioration over time in the behavior of the nonparticipants, rather than by absolute improvements for participants.

Control Variables

The independent, or control, variables used in the analyses are shown in the second part of exhibit 2. Again, not surprisingly, participants appear to be better off on most measures; that is, participants are less likely to be limited English proficient (LEP), be a minority, have parents with low socioeconomic status, attend a rural school, attend a school with a heavy "educational load,"[14] attend a school with a high fraction of poor students, have changed schools at least once by the end of 4th grade, or live in a single-parent household.

---

[12] In both exhibit 2 and appendix A, descriptive statistics are presented for the sample that has all of the following data: the Math concepts and applications variable, the summary measure of nonschool classes, and all of the control variables. The summary measure is set to missing only if all of the relevant activity variables are missing.

[13] Descriptive statistics for all variables are provided in appendix A.

[14] Educational load is a variable created from 11 school characteristics that tend to increase the demands on the instructional staff: teacher mobility; student mobility; percent of students in school who are limited English proficient; percent of students who are minority; percent of students living in poverty; percent of students needing special services; percent of students who are low-achieving (in the bottom quartile); whether the school has been designated as "needing improvement" under the Title I program; percent of students expelled or with disciplinary actions; percent of students in attendance; and percent of students repeating a grade. The measure was standardized across schools so that negative values represent schools with lower-than-average levels of educational demands.

Sample Sizes

The number of observations varies somewhat across the different regression models estimated below because of missing values. For instance, to estimate all outcomes jointly we can use only the 4,125 cases that had valid pre- and post-data for all of our outcomes. In contrast, when we estimate models for single outcomes we have as many as 5,907 observations.

Analytical Methods

As noted above, a number of models were estimated to reflect the different types of analyses typically found in the extant literature, including

- ordinary least squares (OLS) without controlling for the lagged value of the outcome,
- lag models without adjustment for measurement error,
- growth models, and
- lag models with adjustment for measurement error.

The basic OLS model can be written as

$$Y_{i1} = \alpha + X_{i1}'\beta + Z_{i1}'\zeta + e_{1i}, \quad (1)$$

where X_{i1} is a vector of out-of-school-time extracurricular activities variables, Y_{i1} is the outcome of interest (in this case, observed at the end of 4th grade), Z_{it} is a vector of control variables, and α, β, and ζ are parameters to be estimated. Adding in the lagged value of the outcome variable (observed at the end of 3rd grade) yields the lag model recommended by Meyer (1996),[15] which can be written as follows:

$$Y_{i1} = \alpha + X_{i1}'\beta + Z_{i1}'\zeta + Y_{i0}'\theta + e_{1i}, \quad (2)$$

where Y_{i0} is the lagged value of the outcome and θ is a parameter to be estimated.

[15] Meyer (2002) refers to this as a "Pre on Post" model.

This specification of the lag model contrasts with the widely used growth model,[16] described by Bryk and Raudenbush (1992) as having multiple levels, for example

Level 1:

$$Y_{it} = \alpha_i + t^*\delta_i + e_{2it},\ t= 0 \text{ and } 1$$

Level 2:

$$\alpha_i = \alpha_\alpha + X_{i1}'\lambda_\alpha + Z_{i1}'\zeta_\alpha + \varepsilon_{\alpha i}$$

$$\delta_i = \alpha_\delta + X_{i1}'\lambda_\delta + Z_{i1}'\zeta_\delta + \varepsilon_{\delta i}$$

Here, "t" represents time and α_α , λ_α , ζ_α, α_δ , λ_δ , and ζ_δ are parameters to be estimated.

In the first set of equations (level 1), there is one observation for each student and time period, the outcomes are test scores, and the independent variable is time. The second set of equations (level 2) uses the coefficient estimates on the intercepts and time from the level 1 equations as outcomes and these coefficient estimates are modeled as depending on the variables of interest (out-of-school activities) and other control variables. This model—referred to as a "two-level Hierarchical Linear Model (HLM)"[17]—allows the coefficients on time (the growth rate for student test scores) to vary with the student's participation in out-of-school-time activities. While the two levels are often presented separately, they are estimated jointly.[18]

Bryk and Raudenbush (1992) note that this two-level HLM model can be written as a single equation:

$$Y_{it} = (\alpha_\alpha + X_{i1}'\lambda_\alpha + Z_{i1}'\zeta_\alpha + \varepsilon_{\alpha i}) + t^*(\alpha_\delta + X_{i1}'\lambda_\delta + Z_{i1}'\zeta_\delta + \varepsilon_{\delta i}) + e_{2it}.$$

---

[16] Meyer (2002) refers to this as a "Growth Curve Model." Throughout I limit the discussion to two-period models.

[17] HLM models were estimated using SAS as suggested by Singer (1998). A standard addition to this model includes a third level that would allow for school-level clustering of observations. We estimated models allowing for this third level and found that the standard errors on the after-school activities variables were almost unchanged. This is probably because these variables vary a great deal within schools. Correcting standard errors for clustering within schools generally has larger impacts on the standard errors of school-level variables.

[18] One does not need to have more than one observation for every student to estimate growth models.

If the equation for t = 0 is subtracted from the equation for t = 1, all of the terms that do not depend on t are automatically dropped resulting in the following:[19]

$$Y_{i1} - Y_{i0} = \alpha_\delta + X_{i1}'\lambda_\delta + Z_{i1}'\zeta_\delta + \varepsilon_{\delta i} + e_{2i1} - e_{2i0}.$$

When Y_{i0} is added to both sides of the equation, we get

$$\begin{aligned} Y_{i1} &= \alpha_\delta + X_{i1}'\lambda_\delta + Z_{i1}'\zeta_\delta + Y_{i0} + \varepsilon_i + e_{2i1} - e_{2i0} \\ &= \alpha_\delta + X_{i1}'\lambda_\delta + Z_{i1}'\zeta_\delta + Y_{i0}'\Delta + \upsilon_i \end{aligned}$$

where $\Delta = 1$ and $\upsilon_i = \varepsilon_i + e_{2i1} - e_{2i0}$.

Now if we compare this to the lag model described earlier, that is,

$$Y_{i1} = \alpha + X_{i1}'\beta + Z_{i1}'\zeta + Y_{i0}'\theta + e_{1i},$$

we see that the two models are very similar except that the coefficient on the lagged value of the outcome is constrained to be one in the growth model. There are differences in the assumptions made about the error terms, but if the assumptions of the growth model are correct, and there is no measurement error in the lagged outcome, then the lag model should produce consistent estimates of the effects of out-of-school-time activities. Thus, the growth model can be thought of as a constrained version of the lag model. Because the lag model allows the coefficient on the lagged outcome to take on any value, it should be a more robust model to any violations of the assumption that the coefficient on the lagged outcome equals 1.

This assumption is important because the coefficient on the lagged value could be less than 1 for a number of reasons other than bias due to measurement error. Suppose, for instance, that the outcomes are student test scores. One might suspect that higher-scoring students forget some of the information they knew when they took the exam in the earlier period, especially if this information is not reinforced by additional interventions similar to the ones that helped them to score high in the first place. In addition, they might be misled by their lower-scoring peers. Similarly, lower-scoring

[19] This equation justifies an alternative method using the change in test scores as the outcome, which we refer to as the **differences growth model**.

students might acquire information or skills they lacked on earlier exams from higher-scoring peers, even in the absence of any other intervention. Therefore, students with lower test scores might be expected, on average, to have higher test score growth, in the absence of any other intervention, than students with higher pre-intervention scores. More generally, one might expect some "regression to the mean," which is commonly found in many situations.

One might argue that an expectation of "regression to the mean" by group is inconsistent with some students having higher test scores over very long periods of time (if not their entire lives). This phenomena of continued gaps in test scores could be explained, however, by continued inequality in the positive interventions that caused the initially observed differences, and that these continued supports more than make up for what might have otherwise been observed (a general moving of all students toward the middle of the distribution). The importance of continued interventions and peer effects might also help to explain the fade-out effect typically found for many interventions designed to help equalize outcomes for poor and nonpoor students.

It is also important to note that if the coefficient on the lagged outcome is equal to 1 then the variance of the outcome will probably increase over time. This increase in variance occurs because the variance for a given period is equal to the variance from the previous period plus any variance caused by additional inputs plus twice the covariance between past scores and current inputs (which is presumably positive).[20] We do see increased variances over time in all of our outcomes of interest except for the SSMC and SSRC tests. However, we note that the observed variance includes measurement error as well as variance in true skills. If measurement error is lower in the higher grade levels that could explain the patterns observed for SSMC and SSRC. Evidence of decreased measurement error could be found by looking at the R-squared statistics from regressions of these test score variables on the exogenous variables, such as race, gender, ethnicity, family background, and out-of-school-time activities. Such regressions were estimated

[20] We thank Rob Meyer for pointing out this implication.

and provided evidence of increasing R-squared statistics over time (i.e., a decrease in unexplained variance) for most of our outcomes, including the SSMC and SSRC tests.[21]

If we believe that the coefficient on the lagged outcome may be less than 1, then we need to estimate a lag model. A crucial assumption for the lag model, however, is that there is no measurement error in the lagged outcome variable. Unfortunately, measurement error is quite likely for our outcomes of interest. For instance, the math test score is designed to measure math skills expected of 4th graders. However, the test given may include questions that a particular student finds easier (or harder) than other questions that could have been asked about these skills. In addition, the student may be feeling worse (or better) than average at the time that they take the exam. For both of these reasons, a student's test score may be a noisy measure of their true underlying skills.

If measurement error does exist, then the coefficient on the lagged outcome variable will be biased (as stressed by Meyer 1996) and a standard form of measurement error (an error term added to the true value of the lagged outcome but uncorrelated with all other variables in the model) would bias the coefficient estimate on the lagged value downwards and thus reduce its efficacy as a control variable (Greene 2000; see appendix C). Thus, if the true coefficient on the lagged value were 1, a lag model without controls for measurement error would yield an estimated coefficient of less than 1 and the incorrect conclusion that the growth model was biased.[22]

---

[21] A plausible explanation for this pattern is presented by Yen (1985), who shows that measurement error is lower for more difficult questions.

[22] A number of researchers have estimated variations of the growth model controlling for lagged values of Y. In their simplest forms these are equivalent to the lag model. For example, Marcus (1997) controlled for lagged values of Y but did not control for measurement error. Therefore, based on our analyses, her estimates were probably biased. Yasumoto, Uekawa, and Bidwell (2001) also estimated a growth model with controls for lagged test scores but no controls for measurement error. They also allowed growth to change over time and allowed lagged test scores to affect both the growth and the change in growth. Consequently, their model would not simplify to the differences growth model discussed here. However, the results might still change considerably if measurement error in the lagged test scores were included. Blau et al. (2001) estimated a growth model controlling for previous performance using three categories of performance, again suggesting a different, but related, functional form to the models discussed here.

At the same time, however, measurement error in the lagged value does not bias estimates in the growth model[23] as long as it is uncorrelated with any of the other right-hand side variables, because the measurement error becomes part of the error term in the main equation. Thus, one method of controlling for measurement error is to assume that the coefficient on the lagged value of the outcome in a lag model is 1 and, based on this assumption, estimate a growth model. In our results below we estimate a variety of models and test whether the assumption of a coefficient of 1 on the lagged outcome can be rejected. Before presenting our results, however, we first describe how we control for measurement error.

Controlling for Measurement Error: Instrumental Variables

There are two standard methods of controlling for measurement error (Fuller 1987). The first, errors in variables (EV), involves using known information about the reliability of the variable in question to adjust the estimated parameters (Meyer 1996). However, this method is only as accurate as the known reliability information, which, in many practical situations, may be flawed. For example, adjusted-parameter estimates may be biased if reliability estimates are based on a nationally normed sample but the data being analyzed are for a select sample that has a very different reliability index.[24]

For this reason, we chose to control for measurement error using the alternative instrumental variable (IV) method, common in the econometrics field (Ashenfelter and Krueger 1994; Ladd and Walsh 2002; Meyer 1999).[25] The IV method involves regressing the mismeasured variable (e.g., the lagged test score) on all of the other independent variables and at least one additional "instrumental variable" that is assumed to affect the mismeasured variable (e.g., the prior year's test score), but to have no direct impact on

---

[23] Unless the lagged value is used as a control variable on the right-hand side.

[24] Alternative methods, based on item response theory, do allow reliability estimates to vary across individuals but are not feasible using the data we have.

[25] In one paper, Meyer (1999) used both methods and got very similar results. However, in a more recent paper (Meyer 2002), he argues in favor of the EV method over the IV one based on the concern that the IV method might produce too large a coefficient on the lagged value of Y and, consequently, underestimate the impact of the variables of interest (X) if there are unobserved factors that affect both test scores, the

the outcome of interest (the current test score). The results of this regression are then used to create a **predicted value** for the mismeasured variable that is then used as a control in the equation estimating the impact of out-of-school-time activities on the outcomes of interest.[26] (Appendix C presents a fuller discussion of the IV model.)

One common instrument to use in such situations is another mismeasured variable that is designed to measure the same underlying characteristic (lagged skills in our case). The new mismeasured variable will be valid as an instrument as long as its error is uncorrelated with the error in the original mismeasured variable. For the purposes of this analysis, we use an estimate of the child's ability to answer survey questions correctly. Because the questions are different and the survey was given at a different time (and possibly a different day) than the math and reading tests, we expect little correlation between the child's random mistakes on the survey and their random mistakes on the tests. This is a crucial assumption needed to use the IV model to correct for measurement error. To estimate mistakes the children made on the survey, we limit ourselves to a set of almost identical questions asked of the parents and children on the child's involvement in out-of-school-time extracurricular activities during the 3rd grade. We calculate the number of times the parent and child reported differently on these activities and use this as an instrumental variable for the lagged outcome in each of our equations describing factors that affect the 4th grade outcomes.[27]

An underlying assumption of this approach is that the parents are generally reporting accurately. We find a great deal of evidence in favor of this assumption. To start with, our "survey mistakes" variable is strongly and negatively correlated with standardized test scores; that is, low-achieving students were significantly more likely to report differently about out-of-school-time activities than their parents. This pattern would follow if low-skilled students make more mistakes when filling out surveys.

---

instrumental variable, and X. We have conducted simulations suggesting that if unobserved factors bias the estimates then it is not possible to choose between the IV and EV methods as both are likely to be biased.

[26] Both equations were estimated simultaneously in SAS using Proc SYSLIN. This adjusts the standard errors in the second stage for the uncertainty in the first stage.

However, this need not mean that the parent reports are accurate since parents who spend little time with their children may make more mistakes because they are not keeping careful track of their children's activities. These same parents may also provide less academic support for their children.

Thus, the strong association between the child/parent differences in reports and child test scores does not necessarily mean that the differences are due largely to child mistakes. However, we did find other evidence suggesting that the parent reports are more accurate than the child reports. First, the parent measures are more consistent over time. Second, the parent and child reports are in greater agreement in 4th grade than in 3rd grade, which is consistent with the children learning how to read the questions more accurately by the end of the 4th grade. Third, Vandell and Posner (1995) find that parents and researchers agree on children's regular after-school arrangements around 90 percent of the time. In contrast, parent-child agreement is only around 70 percent in 3rd grade and rises to 79 percent by 5th grade, when children's reading skills have likely improved. Thus, in total the evidence indicates quite strongly that differences between parent and child reports in these early grades are most likely due in large part to mistakes on the part of the children in answering the survey questions.

As shown in exhibit 2, our instrumental variable—the "survey mistakes" variable—has a slightly higher mean for participants than for nonparticipants. This result is not surprising given that participants have more opportunities to differ with their parents. However, this difference might cause some readers to worry that this instrument would under-predict lagged outcomes for the participants. Fortunately, this is not the case because the participation variables are also included in the first stage. Thus, the predicted lagged scores of participants remain higher than those of nonparticipants.[28]

---

[27] As shown in appendix A, this variable has a maximum value of 5 for students even though parents report on six different types of activities. This is because children were only asked about five of the six activities. We include all six activities that the parents report on when estimating impacts of nonacademic activities.

[28] As an additional test, we created a 4th grade version of our "survey mistakes" variable. Adding this control to our models had no impact on the results. The coefficients on the lagged outcome variables remained close to 1, and the estimated impacts of the out-of-school-time variables remained statistically insignificant, both when estimated as separate variables and when using the summary measure.

Chapter 3 : Findings

This chapter presents the results of the analyses using the different models discussed in chapter 2, the OLS model without a lag (exhibit 3), the lag model without controls for measurement error (exhibit 4), the growth model (exhibit 5), and the lag model controlling for measurement error (exhibit 6). As will be seen, the results are highly dependent on the choice of analytical approach.

Initial Analyses: The OLS Model without a Lag

As demonstrated in exhibit 3, the OLS models (without controlling for the lagged outcome) indicate statistically significant effects of student participation in art-related classes (art, music, or dance) and classes in religious instruction on all four standardized test scores and on all three of the behavioral scales. Less consistent results are found for participation in language classes. None of the 42 estimated coefficients is **negative** and significant,[29] and almost half (20) are, in fact, positive and statistically significant. Based on these results a careful researcher might conclude that participation in such out-of-school-time extracurricular activities during the elementary school years has important implications for academic achievement and classroom behavior, in addition to any beneficial impacts on the skills (e.g., music capability) that they are more directly designed to achieve.

These results are consistent with some previous research. Estimating similar models—ones that control for background characteristics other than the lagged value of the outcome, Huang et al. (2000), Marshall et al. (1997), Posner and Vandell (1994), and Reaney et al. (2001) all find positive relationships between participation in (or amount of participation in) after-school activities and outcomes for young children. In many cases, these positive relationships have been used to support investments in after-school programs.

---

[29] Unless stated otherwise, all results discussed in the text are statistically significant at the 5 percent level. Two of the coefficients on after-school lessons in exhibit 3 are negative and significant at the 10 percent level.

There is also research supporting the particularly strong results for arts activities. Many after-school programs include a focus on the arts (visual arts, music, dance, drama). These programs are designed to enrich the experiences of the children and serve as a "lure" to get them to participate more fully in other activities. In addition, they can potentially have a transference effect that can increase academic achievement. Although limited research has been done in this area, there are some suggestions of these programs' possible effects. Using small experimental studies of elementary school drama instruction, De la Cruz (1995)[30] found significantly higher oral expressive language and social skills for participants; Parks and Rose (1997) found significantly higher standardized test scores in reading comprehension; and Hetland (2000) found a significant relationship between piano instruction and spatial-temporal reasoning. Podlozny's (2000) meta-analysis of research on drama indicated a significant positive relationship between drama instruction and standardized reading test scores, oral expressive language, and writing skills. In addition, Vaughn (2000) screened more than 4,000 studies of music instruction and for those using experimental designs found an overall positive relationship between music instruction and math test scores.

While some experimental studies find positive effects, others do not. For instance, Moga et al. (2000) and Winner and Cooper (2000) conducted meta-analyses of experimental studies and found no significant effect of arts study on verbal or math skills.[31] Similarly, Costa-Giomi (1999) found no significant effects of three years of piano classes on verbal skills or spatial ability. Thus, the experimental results alone do not provide strong evidence of effects. In addition, as discussed below, results from the simple OLS models without lags do not hold up well to improvements in model specification.

---

[30] This study and many others on the impacts of arts are summarized in Deasy (2002).

[31] They did find effects on "creative thinking."**[[Please explain what "creative thinking" is.]]**

Exhibit 3. Estimated Effects of Out-of-School-Time Activities: OLS Model without Control for Lagged Value

	SSMA			SSMC			SSRC			SSRV			ATTEN			COOP			PARTIC		
Activity	Coef	Std Err		Coef	Std Err		Coef	Std Err		Coef	Std Err		Coef	Std Err		Coef	Std Err		Coef	Std Err	
Art, music, dance	11.83	2.12	* * *	7.53	1.69	* * *	10.22	1.96	* * *	8.79	1.75	* * *	0.12	0.02	* * *	0.07	0.02	* * *	0.08	0.02	* * *
Sports	-1.31	1.73		0.52	1.38		-2.90	1.60	*	-1.55	1.42		0.01	0.02		0.01	0.01		0.05	0.02	* * *
Language	15.91	4.65	* * *	8.70	3.71	* * *	9.13	4.30	* * *	5.24	3.83		0.10	0.05	* *	0.07	0.04	*	0.04	0.05	
Religion	10.38	1.74	* * *	5.13	1.39	* * *	9.73	1.61	* * *	8.49	1.43	* * *	0.09	0.02	* * *	0.04	0.01	* * *	0.07	0.02	* * *
History and culture	-2.46	4.18		5.61	3.33	*	-6.95	3.86	*	-2.28	3.44		0.00	0.04		0.00	0.03		0.01	0.05	
Computers	7.70	2.96	* * *	1.81	2.36		3.97	2.73		2.88	2.44		0.03	0.03		0.02	0.02		-0.02	0.03	
Parent-child differences in 1992	-9.29	0.88	* * *	-4.88	0.71	* * *	-7.46	0.82	* * *	-6.85	0.73	* * *	-0.06	0.01	* * *	-0.03	0.01	* * *	-0.02	0.01	* *

Source: Prospects data, grade 3 cohort, 1992.

Notes: The test scores in math (SSMA and SSMC) and reading (SSRC and SSRV) and the behavioral outcomes (ATTEN, COOP, and PARTIC) are defined in the text.

* Implies significantly different from 0 at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

The Lag Model

Exhibit 4 presents the results of the statistical models that add in controls for the lagged value of the outcome (without controls for measurement error). Once again, statistically significant effects are found for student participation in art-related classes and religious instruction, and less consistent results are noted for language classes. None of the estimated coefficients is negative and significant, and almost half (17) remain positive and statistically significant.[32] However, the results are very different from those in exhibit 3 as the estimated effects are much smaller. For instance, many of the coefficient estimates on arts and religion activities in exhibit 4 are about half as large as those shown in exhibit 3. Thus, controlling for the lagged value suggests much smaller impacts of out-of-school-time activities than those obtained using the OLS model.

Other researchers have also estimated models similar to those in exhibit 4. For example, Baker and Witt (1996) report positive impacts of participation, even after controlling for lagged outcomes. Vandell and Pierce (1997) did not find positive impacts of participation compared with nonparticipation, but did find positive impacts of high participation compared with low participation.

While controlling for the lagged values did reduce the estimated impacts of out-of-school-time activities in our models, we would have still had an important story to tell about the positive effects of participation in the arts and religious instruction on outcomes of 4th grade students if we had stopped our analysis here. Indeed, we were prepared to draw this conclusion until we estimated growth models.

[32] One is negative and significant at the 10 percent level.

Exhibit 4. Estimated Effects of Out-of-School-Time Activities: Lag Model without Control for Measurement Error

	SSMA			SSMC			SSRC			SSRV			ATTEN			COOP			PARTIC		
Activity	Coef	Std Err		Coef	Std Err		Coef	Std Err		Coef	Std Err		Coef	Std Err		Coef	Std Err		Coef	Std Err	
Art, music, dance	4.68	1.52	* * *	3.91	1.37	* * *	3.93	1.44	* * *	4.66	1.34	* * *	0.06	0.02	* * *	0.05	0.01	* * *	0.06	0.02	* * *
Sports	-0.67	1.24		-0.74	1.12		-1.99	1.17	*	0.34	1.09		0.00	0.01		0.00	0.01		0.04	0.02	* *
Language	10.74	3.33	* * *	3.81	3.00		5.21	3.15	*	0.17	2.93		0.07	0.04	*	0.06	0.03	*	0.04	0.05	
Religion	3.86	1.25	* * *	2.27	1.12	* *	3.73	1.18	* * *	2.58	1.10	* *	0.05	0.01	* * *	0.03	0.01	* * *	0.05	0.02	* *
History and culture	-4.09	2.99		3.52	2.70		-4.26	2.83		-0.91	2.64		-0.00	0.03		-0.02	0.03		0.01	0.04	
Computers	4.19	2.12	* *	-0.19	1.91		1.05	2.01		-0.69	1.87		0.01	0.02		0.01	0.02		-0.03	0.03	
Parent-child differences in 1992	-4.06	0.64	* * *	-2.56	0.57	* * *	-3.01	0.6	* * *	-2.92	0.56	* * *	-0.03	0.01	* * *	-0.02	0.01	* * *	-0.01	0.01	

Source: Prospects data, grade 3 cohort, 1992.

Notes: The test scores in math (SSMA and SSMC) and reading (SSRC and SSRV) and the behavioral outcomes (ATTEN, COOP, and PARTIC) are defined in the text.

* Implies significantly different from 0 at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

Expanded Analysis: The Growth Model

Exhibit 5 provides the estimated effects of the same set of out-of-school-time activities both on standardized test scores in reading and mathematics and on measures of classroom behavior. But in this case the estimation was done using the growth models discussed in chapter 2.[33] As shown, many of the coefficient estimates have been greatly reduced in size, but more important, **none is statistically significant at even the 10 percent level.** We also tried estimating models using a single summary measure of participation in any activity (about one-quarter of the sample had no participation) but still found no evidence of statistically significant impacts on any of the student outcomes.[34]

At the least, these results indicate that one should be uncertain about whether or not out-of-school-time extracurricular activities affect academic skills or student behaviors. In addition, these results leave a sense of uncertainty about which conclusions should be believed. On the one hand, the lag model allows the coefficient on the lagged outcome variable to differ from 1. On the other hand, the growth model results are not biased by measurement error. To choose between these two models, we needed to estimate a model that would both allow the coefficient on the lagged value to be less than 1 and, at the same time, control for measurement error. Fortunately, we were able to do this in our data set, as is discussed in the next section.

---

[33] As discussed earlier, models where the outcome is the change in test scores are similar to growth models estimated here. We estimated such difference growth models (without controlling for lagged values) and obtained results similar to the growth model, except that the difference growth models had smaller standard errors, perhaps because fewer parameters were being estimated.

[34] We estimated several variations of our models using these summary measures, with similar results.

Exhibit 5. Estimated Effects of Out-of-School-Time Activities: Growth Model

	SSMA			SSMC		SSRC		SSRV		ATTEN		COOP		PARTIC	
Activity	Coef	Std Err		Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err
Art, music, dance	2.70	2.45		-0.92	2.10	-0.23	2.47	1.98	2.00	0.04	0.03	0.02	0.02	0.03	0.03
Sports	-1.09	2.00		-1.64	1.71	-1.49	2.01	0.66	1.63	-0.01	0.02	-0.01	0.02	-0.02	0.02
Language	6.61	4.98		-2.45	4.26	2.77	5.02	-0.89	4.06	0.03	0.05	0.05	0.04	0.04	0.06
Religion	2.73	1.98		-0.06	1.70	0.97	2.00	0.74	1.62	0.01	0.02	0.02	0.02	0.00	0.02
History and culture	-4.37	4.38		-1.73	3.74	-2.88	4.42	-1.19	3.57	-0.01	0.05	-0.04	0.04	-0.00	0.05
Computers	1.76	3.17		-2.19	2.71	0.60	3.20	-0.86	2.58	-0.01	0.03	-0.00	0.03	-0.01	0.04
Parent-child differences in 1992	-2.48	0.98	* *	0.58	0.84	-0.23	0.99	-0.92	0.80	-0.01	0.01	-0.01	0.01	-0.00	0.01

Source: Prospects data, grade 3 cohort, 1992.

Notes: The test scores in math (SSMA and SSMC) and reading (SSRC and SSRV) and the behavioral outcomes (ATTEN, COOP, and PARTIC) are defined in the text.

* Implies significantly different from 0 at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

A Last Look: The Lag Model with Measurement Error Control

The results of our lag model controlling for measurement error are shown in exhibit 6[35] and are very similar to those of the growth model—only one of the estimated impacts of out-of-school-time classes is statistically significant. Thus, it appears that the control for measurement error is crucial to this analysis of the relationship between out-of-school-time activities participation and school outcomes. Without the control for measurement error one would draw the incorrect conclusion that there were many statistically significant impacts on test scores and classroom behavior.

We control for measurement error using instrumental variables. However, the use of instrumental variables is certainly not a panacea. Using instrumental variables has two main problems. First, instrumental variables are often very poor predictors and consequently produce imprecise results. Second, results based on one instrumental variable may not be robust when tested against alternative instruments. Fortunately, our instrumental variables passed both tests quite well. First, our main instrumental variable (survey mistakes) is a very strong predictor of the lagged outcomes in our models,[36] and we did have a second set of instrumental variables that also had strong predictive power, even after controlling for our main instrument and the other exogenous variables in our model. The additional instrumental variables are the lagged values of the parent reports on the after-school lessons that the children took.[37] Second, it turns out that when we added these instrumental variables to our models, the tests for overidentification were not rejected[38] and our substantive results remained unchanged from the growth model—the coefficients on the lagged values of the outcomes were not statistically different from 1

---

[35] This model also controls for measurement error in the parent reports on after-school classes using the lagged reports on classes as the instrumental variables. This correction had little impact on our results. In addition, we jointly estimate the models for all of the outcomes simultaneously, as suggested by Thum (1997). This also had little impact on our results.

[36] A joint test of significance of the coefficients on the instrumental variable (survey mistakes) across the seven outcomes yielded a p-value of 0.0001, controlling for all other exogenous variables in the model.

[37] The joint p-value for these variables was 0.003.

[38] We used the overidentification test recommended by Basmann (1960) as implemented in SAS Proc Syslin.

and the out-of-school-time activities were not jointly significant. Thus, our models appear robust to at least this set of alternative instrumental variables.

Not surprisingly, the statistical significance of the control variables is also reduced when we control for measurement error in the lagged outcome. Without controls for measurement error about one-third of the coefficient estimates on the control variables[39] were statistically significant at the 1 percent level in the lagged model. Controlling for measurement error reduces this by more than half. However, overall about 14 percent of the coefficient estimates on these control variables remain statistically significant at the 1 percent level after controlling for measurement error, and another 14 percent are significant at the 5 percent level. Thus, it appears that even after controlling for measurement error in the lagged outcome, we are obtaining reasonably precise estimates of the impacts of other variables.

[39] This excludes the intercept, the lagged outcome, the after-school activities, and the parent-child differences in 1992. There are 11 additional control variables in our model, as described in appendix A.

Exhibit 6. Estimated Effects of Out-of-School-Time Activities: Lag Model with Correction for Measurement Error

	SSMA			SSMC		SSRC		SSRV		ATTEN		COOP		PARTIC	
Lesson	Coef	Std Err		Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err
Art, music, dance	1.70	1.94		1.27	1.77	0.16	1.95	1.76	1.71	0.01	0.02	0.00	0.02	-0.01	0.04
Sports	-0.40	1.34		-1.66	1.30	-1.45	1.35	1.67	1.31	-0.00	0.02	-0.01	0.02	-0.01	0.03
Language	8.58	3.67	**	0.26	3.58	2.87	3.67	-3.39	3.52	0.05	0.05	0.05	0.04	0.06	0.08
Religion	1.14	1.64		0.19	1.44	0.14	1.68	-1.58	1.64	0.02	0.02	0.02	0.02	-0.05	0.04
History and culture	-4.77	3.24		2.00	3.09	-2.65	3.27	0.05	3.07	-0.01	0.04	-0.06	0.04	0.00	0.07
Computers	2.72	2.34		-1.64	2.21	-0.70	2.35	-3.21	2.26	0.00	0.03	0.01	0.03	-0.06	0.05
Parent-child differences in 1992	-1.87	1.02	*	-0.87	0.85	-0.34	1.01	-0.17	0.94	-0.01	0.01	-0.00	0.01	0.02	0.02

Source: Prospects data, grade 3 cohort, 1992.

Notes: The test scores in math (SSMA and SSMC) and reading (SSRC and SSRV) and the behavioral outcomes (ATTEN, COOP, and PARTIC) are defined in the text.

* Implies significantly different from 0 at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

Final Comments on Modeling Issues

An explanation for the differences in results across models can be seen in exhibit 7, which compares the coefficients on the lagged value of the outcome in each of the different models. In the first row, the lag model without controls for measurement error shows estimated coefficients on the lagged values between 0.26 and 0.74. In contrast, the growth model implicitly assumes coefficients of 1. Thus, if the lag model without a correction for measurement error were correct, it would provide strong evidence suggesting that the growth model was mis-specified. However, in the third row the estimated coefficients on the lagged outcomes from the lag model with controls for measurement error are presented and, as can be seen, none differ significantly from 1 at the 0.05 level and all are well over 0.74, the highest value observed in the model without measurement error. Thus, we cannot reject the growth model and had we lacked information sufficient to control for measurement error, we would have been better off relying on our growth model estimates than using the lag model without controlling for measurement error.

Exhibit 8 demonstrates how common this problem is by summarizing coefficients on lagged test scores from a number of previous studies with and without controls for measurement error. The coefficients on lagged values are generally substantially larger in those studies that control for measurement error and often not significantly different from 1. The smallest coefficient found with controls for measurement error is from a study by Girotto and Peterson (1999) that used the errors in variables (EV) method of adjusting the prior test score. One possible explanation for their result (a coefficient less than 1) is that they control for grade point average (GPA) in courses taken between the pre- and post-tests (as a measure of effort). If one believes that GPA is an alternative measure of their outcome (student skills) rather than an exogenous variable, then it might not be so surprising that controlling for this variable reduces the coefficient on the lagged outcome variable.

Exhibit 7. Estimated Effects of Lagged Outcome by Type of Model

	SSMA			SSMC			SSRC			SSRV			ATTEN			COOP			PARTIC		
Model	Coef	Std Err		Coef	Std Err		Coef	Std Err		Coef	Std Err		Coef	Std Err		Coef	Std Err		Coef	Std Err	
Lag model, no measurement error	0.74	0.01	***	0.52	0.01	***	0.60	0.01	***	0.70	0.01	***	0.58	0.01	***	0.50	0.01	***	0.26	0.02	***
Growth model	1.00			1.00			1.00			1.00			1.00			1.00			1.00		
Lag model, with measurement error	1.05	0.11		0.90	0.12		0.96	0.10		1.19	0.12		1.06	0.13		1.31	0.24		1.45	0.44	

Source: Prospects data, grade 3 cohort, 1992.

Notes: The test scores in math (SSMA and SSMC) and reading (SSRC and SSRV) and the behavioral outcomes (ATTEN, COOP, and PARTIC) are defined in the text.

* Implies significantly different from 1 at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

Exhibit 8. Coefficient Estimates on Lagged Outcome Variables

*A. **Not** Controlling for Measurement Error*

Study	Outcome	Coefficient estimates	Std Error	
Marcus (1997)[a]	AIDS knowledge	0.7900	0.0800	* * *
Chaplin (1998a)	12th grade math	0.7500	0.0084	* * *
	12th grade science	0.6400	0.0096	* * *
Jacob (2001)	12th grade math	0.6800	0.0110	* * *
	12th grade reading	0.4600	0.0140	* * *
Ludwig and Bassi (1999)	10th grade reading	0.8900	0.0100	* * *
	10th grade math	0.9500	0.0100	* * *
Goldhaber and Brewer (2000)	12th grade math	0.9100	0.0100	* * *
	12th grade science	0.7600	0.0100	* * *
Meyer (1999), non—college bound	12th grade math	0.6700	—	
Mussoline and Shouse (2001)	10th grade math	0.7400	0.0100	* * *
Vandell and Pierce (1997)[b]	Academic grades	0.40	—	* *
	Excused absences	0.44	—	* * *
	Aggression	0.53	—	* *
B. Controlling for measurement error				
Jacob (2001)	12th grade math	1.0580	0.0670	
	12th grade reading	1.0930	0.0850	
Meyer (1999), non–college bound	12th grade math	0.9470	0.0290	*
College bound	12th grade math	1.0240	0.0260	
Overall	12th grade math	0.9970	0.0180	
Ross and Broh (2000)	12th grade math and English grades and tests	1.0300	0.0127	* *
Girotto and Peterson (1999)	11th grade cognitive skills	0.6300	0.0200	* * *

Notes: — = not available.

a. Marcus (1997) estimated a model of growth on the lagged value, which had a coefficient of -0.21. This translates to a coefficient of 0.79 in a model of the outcome on the lagged value.

b. Vandell and Pierce (1997) reported on lagged outcomes for 17 variables in table 9, with coefficient estimates ranging from 0.09 to 0.53. All but four were significantly different from 0. Because standard errors were not reported I could only calculate statistical difference from 1 if the coefficient was significantly different from 0. All 13 of the coefficients that were statistically different from 0 at least at the 10 percent level were also statistically different from 1 at least at the 5 percent level.

* Implies significantly different from 1 at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

Excluded from exhibit 8 are results from a school-level analysis by Jones and Zimmer (2001) that used the lagged fraction of children scoring satisfactorily in a given subject in a given year. Their lagged coefficients ranged from 0.23 to 1.75 and were always significantly different from 1. Because their outcomes were measured at the school level, the measurement error is likely small. One plausible reason for their coefficients being very different than 1, however, is that these scores have clear floor and ceiling effects since schools cannot go below 0 or above 100 percent. In such cases we would argue that the growth model is less likely to be appropriate.

The bottom line for our research question is that the ideal nonexperimental studies should control both for lagged outcomes and for measurement error in these lagged variables, either by estimating growth models or by including the lagged outcome and controlling for measurement error directly. We were only able to locate two studies on the impacts of after-school activities for young children that satisfied these criteria, Hamilton and Klein (1998) and Mayesky (1980a). Both reported positive impacts. However, neither study controlled for any other background characteristics. To test for the potential importance of this omission, we reestimated our growth models without any controls for background characteristics. In these new models, the out-of-school-time activities variables were statistically significant for five of the seven outcomes (all but ATTEN and PARTIC). Out-of-school participation in arts positively impacted four outcomes, and out-of-school participation in language classes positively impacted two.[40]

To summarize, it appears that controlling for other background characteristics is important. Thus, our results indicate that the impacts reported by Hamilton and Klein (1998) and Mayesky (1980a) might have been much less clear had they also included such controls in their models.

[40] The coefficient estimates on religious participation were positive and significant at the 10 percent level for three outcomes and sports were negative and significant for two.

Chapter 4 : Substantive Summary of Results

As discussed in the previous chapter, we first approached this analysis using relatively widespread analytical methods to control for the inherent problem of selection bias, that is, the inclusion of the lagged outcome variable as a control variable. This model yielded positive effects of out-of-school-time participation in arts and music activities on academic achievement and classroom behavior of elementary school children. This seemed to be an important and interesting finding.

But these findings evaporated when we reanalyzed our data controlling for measurement error in the lagged outcome and when we estimated a growth model of student gains. While a great deal of literature indicates positive impacts of after-school activities on these outcomes, we could find no nonexperimental studies that estimated models that dealt with all of the issues we found to be important in our work (controlling for lagged values, dealing with measurement error, and including other background characteristics in the models). We did find some experimental evidence suggesting effects of participation in arts on academic skills. However, other experimental studies suggested less clear impacts. In summary, the lack of statistically significant findings in our better models and the unclear results from experimental studies lead to the conclusion that out-of-school-time extracurricular activities have no clear effects on the academic achievement or school behavior of elementary school children.

This interpretation of no clear effects based on our results may be too strong for several reasons. First, the statistical results are somewhat imprecise.[41] While the point estimates and standard errors are fairly small (generally less than 1/25th of a standard deviation for the test scores in exhibit 6), after-school activities are likely to be fairly small interventions compared with, say, regular classroom activities and what happens to children in their homes. Thus, effects that would help justify participation would also be fairly small. Second, nonparticipants may be engaged in other types of activities that also

[41] One reason for this imprecision could be that participation means very different things for different children. Indeed, Vandell and Pierce (1997) find a great deal of variation in the amount of time elementary school children participate in after-school activities.

promote their learning. If current participants did not have the option to participate in the activities reported here, they might not have alternatives as good as those of current nonparticipants. Third, we focused our analysis on classes that took place during the school year. Results by Alexander et al. (2001), Entwisle and Alexander (1992), Heyns (1978), and Phillips et al. (1998)[42] indicate that low-income children appear to do just as well as their higher-income counterparts in terms of test score gains during the school year, but during the summer they often experience test score drops both absolutely and relative to their higher-income counterparts. Thus, even if out-of-school-time extracurricular activities during the school year have relatively small impacts on the student outcomes measured in this paper, they may matter more during the summer, especially for low-income children.[43] Indeed, one of the few studies we found on out-of-school-time programs that did use a sound methodology (Sunmonu et al. 2002) found that a summer academic program appears to have had substantial impacts for the students participating regularly.[44]

It should also be noted that even if these out-of-school-time extracurricular activities have little positive impact on academic skills and classroom behavior, they are still likely to improve the skills they were designed to improve (i.e., arts, music, etc.). Indeed, one might take the lack of negative impacts as a good sign—that one can teach a child these additional skills and not harm them academically.

Finally, we note that we have little information on quality or intensity. Consequently, our estimates are relevant only for the average quality and intensity currently being experienced by youth in these types of activities. It is possible, though not obvious, that higher-quality programs would have larger impacts on the outcomes measured here,

---

[42] Phillips et al. (1998) only found evidence for this in math.

[43] To test for the possibility that after-school activities matter more for minority or low-socioeconomic status (SES) students, we ran our lag model with controls for measurement error and the summary variable for blacks and for low-SES students separately (the bottom 1,800 cases in our sample). In both cases none of the coefficient estimates on the out-of-school-time classes variables was statistically significant.

[44] Sunmonu et al. (2002) used a growth model with controls for race, gender, and economic status of the children. Another good study suggesting strong impacts of summer school is Jacob and Lefgren (2002). They used a regression discontinuity method to address the issue of selection bias discussed here.

though it is just as likely that they would only have larger impacts on the outcomes they were designed to impact.

In regards to intensity, one might surmise that students taking "classes" usually do so on a weekly basis. However, the total time per week, including practicing, could vary greatly. We do have a rough measure of intensity in that we have distinct variables for each of the six types of activities that a child could be engaged in. A child that is engaged in all six is probably (though not necessarily) engaged more intensively than a child engaged in only one or two activities. The variables describing the activities are not jointly significant in our results meaning that this measure of intensity does not indicate positive impacts.

Additional points could be made against these results. For instance, some might argue that out-of-school-time extracurricular activities help students not directly, but only indirectly by helping to keep the students engaged in school. While this is a compelling argument, it is not clear why such an effect would not show up in our results, both directly (as an increase in test scores) and indirectly, via teacher-reported student behavior. Since we found no evidence of effects on either outcome, it is hard to argue that these particular extracurricular activities had noticeable direct or indirect impacts.

Of course, after-school programs with strong academic components might have noticeable impacts on student skills, and a high fraction of after-school programs do offer academic activities. Unfortunately, it appears that they are not heavily used. Seppanen, deVries, and Seligson (1993) find that 80 percent of after-school programs offer some time for optional homework, just under half provide remedial academic help, and about one-third provide tutoring. In addition, they find that lower-income programs are even more academically oriented than the average program. However, Vandell and Posner (1999) report that children from low-income communities spend less than one quarter of their time on academics while in formal after-school program activities.[45] With such a low focus on academics, it would not be surprising to find little impact of such programs.

---

[45] This is still an improvement over the fraction (10 percent) of time spent on academics by low-income children not in formal after-care programs.

A related point is that extracurricular activities located at school can help improve academic outcomes by keeping children in school longer.[46] This argument is particularly compelling for older students who might drop out of school[47] or choose to leave school early and miss some classes. It is less plausible for 4th graders whose activities are controlled to a much greater degree by adults.

To summarize, while all of the caveats listed above are important, the ultimate conclusion is quite compelling. Our evidence indicates that the belief that out-of-school-time extracurricular activities provide extra academic benefits for young children is, at best, not supported by the available data. Further research and experimental studies may yield different results, but for now the evidence is quite clear. To improve the academic skills of young students, we probably need to teach them those skills directly.

[46] Because we could not distinguish between activities at school and those held elsewhere, we were not able to estimate such impacts.

[47] Indeed, we suspect that such impacts may be quite large, but that they will not likely be observed unless we can collect better data on dropouts.

References

Afterschool Alliance. 2001. *Afterschool Advocate*, 7 October. Washington, D.C.: Author.

Alexander, Karl L., Doris R. Entwisle, and Linda S. Olson. 2001. "Schools, Achievement, and Inequality: A Seasonal Perspective." *Education Evaluation and Policy Analysis* 23(2):171–91.

Ashenfelter, Orley, and Alan Krueger. 1994. "Estimates of the Economic Return to Schooling from a New Sample of Twins." *American Economic Review* 84(5):1157–73.

Baker, D., and P. Witt. 1996. "Evaluation of the Impact of Two After-School Recreation Programs." *Journal of Park and Recreation Administration* 14(3):23–44.

Basmann, R. L. 1960. "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics." *Journal of the American Statistical Association* (December):650–59.

Becker, William E., and John R. Powers. 2001. "Student Performance, Attrition, and Class Size Given Missing Student Data." *Economics of Education Review* 20(4):377–88.

Blau, Judith R., Vicki L. Lamb, Elizabeth Stearns, and Lisa Pellerin. 2001. "Cosmopolitan Environments and Adolescents' Gains in Social Studies." *Sociology of Education* 74(2):121–38.

Braddock, J. H. 1981. "Race, Athletics, and Educational Attainment: Dispelling the Myths." *Youth and Society* 12(3):335–50.

Brown, B. B., D. Kohrs, and C. Lazarro. 1991. "The Academic Costs and Consequences of Extracurricular Participation in High School." Paper presented at the annual meetings of the American Educational Research Association, Chicago, Ill., April 6.

Bryk, A. S., and S. W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, Calif.: Sage Publications.

Campbell, J. R., C. M Hombo, and J. Mazzeo. 2000. *NAEP 1999: Trends in Academic Progress (NCES Report No. 2000-469*). Washington, D.C.: U.S. Department of Education.

Chaplin, Duncan. 1998a. "Earnings Benefits of Math and Science Courses in High School." Urban Institute Working Paper**,** Washington, D.C.: The Urban Institute.

———. 1998b. "Raising Standards: The Effects of High School Math and Science Courses on Future Earnings." *Virginia Journal of Social Policy and the Law* 6(1):115–26.

Clark, R. M. 1993. "Homework-Focused Parenting Practices That Positively Affect Student Achievement." In *Families and Schools in a Pluralistic Society*, edited by N. F. Chavkin (pp. 85-105). Albany: State University of New York Press.

Costa-Giomi, E. 1999. "The Effects of Three Years of Piano Instruction on Children's Cognitive Development." *Journal of Research on Music Education* 47(3):198–212. Reviewed in Deasy, R. J. 2002. *Critical Links: Learning in the Arts and Student Academic and Social Development.* Washington, D.C.: Arts Education Partnership.

Deasy, R. J. 2002. *Critical Links: Learning in the Arts and Student Academic and Social Development.* Washington, D.C.: Arts Education Partnership.

De la Cruz, R. 1995. "The Effects of Creative Drama on the Social and Oral Language Skills of Children with Learning Disabilities." Doctoral diss., Illinois State University, Bloomington. Reviewed in Deasy, R.J. 2002. *Critical Links: Learning in the Arts and Student Academic and Social Development.* Washington, D.C.: Arts Education Partnership.

Eide, E., and N. Ronan. 2000. "Is Participation in High School Athletics an Investment or a Consumption Good? Evidence from High School and Beyond." Santa Monica, Calif.: RAND Corporation.

Entwisle, Doris R., and Karl L. Alexander. 1992. "Summer Setback: Race, Poverty, School Composition, and Mathematics Achievement in the First Two Years of School." *American Sociological Review* 57(1):72–84.

Fashola, Olatokunbo S. 1998. "Review of Extended-Day and After-School Programs and Their Effectiveness." Report No. 24. Baltimore, Md.: Johns Hopkins University Center for Research on the Education of Students Placed At Risk (http://www.csos.jhu.edu/CRESPAR/techReports/Report24.pdf accessed on 9/22/2003.)

Fuller, W. A. 1987. *Measurement Error Models.* New York: John Wiley and Sons.

Gamoran, Adam, and Eileen C. Hannigan. 2000. "Algebra for Everyone? Benefits of College-Preparatory Mathematics for Students with Diverse Abilities in Early Secondary School." *Education Evaluation and Policy Analysis* 22(3):241–54.

Gamoran, Adam, Andrew C. Porter, John Smithson, and Paula A. White. 1997. "Upgrading High School Mathematics Instruction: Improving Learning Opportunities for Low-Achieving Low-Income Youth." *Education Evaluation and Policy Analysis* 19(4):325–38.

Girotto, Jay R., and Paul E. Peterson. 1999. "Do Hard Courses and Good Grades Enhance Cognitive Skills?" In *Earning and Learning: How Schools Matter*, edited by Susan Mayer and Paul Peterson, pg. 205-30. Washington, D.C.: Brookings Institution Press and New York: Russell Sage Foundation.

Goldhaber, Dan D., and Dominic J. Brewer. 2000. "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement." *Education Evaluation and Policy Analysis* 22(2):129–45.

Greene, William H. 2000. *Econometric Analysis*. Upper Saddle River, N.J.: Prentice Hall.

Gregory, P. 1996. *Youth Opportunities Unlimited: Improving Outcomes for Youth Through After School Care.* Manchester: University of New Hampshire.

Hamilton, L. S., and S. P. Klein. 1998. *Achievement Test Score Gains Among Participants in the Foundations School-Age Enrichment Program.* Santa Monica, Calif.: RAND Corporation.

Hetland, Lois. 2000. "Learning to Make Music Enhances Spatial Reasoning." *The Journal of Aesthetic Education* 34(3–4):179–238.

Heyns, Barbara. 1978. *Summer Learning and the Effects of Schooling.* New York: Academic Press.

Hofferth, S. L., Z. Jankuniene, and P. Brandon. 2000. "Self-Care among School-Age Children." Paper presented at the biennial meeting of the Society for Research on Adolescence, Chicago, Ill., April.

Huang, D., B. Gribbons, K. S. Kim, C. Lee, and E. L. Baker. 2000. *A Decade of Results: The Impact of the L.A.'s BEST After-School Enrichment Program on Subsequent Student Achievement and Performance.* Los Angeles: UCLA Center for the Study of Evaluation, Graduate School of Education and Information Studies.

Jacob, Brian A. 2001. "Getting Tough? The Impact of High School Graduation Exams." *Education Evaluation and Policy Analysis* 23(2):99–121.

Jacob, Brian A., and Lars Lefgren. 2002. "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." National Bureau of Economic Research Working Paper 8918. Cambridge, Mass.: National Bureau of Economic Research.

James, D. W., and S. Jurich, eds. 1999. *MORE Things that DO Make a Difference for Youth: A Compendium of Evaluations of Youth Programs and Practices,* vol. 2. Washington, D.C.: American Youth Policy Forum.

Jones, John T., and Ron W. Zimmer. 2001. "Examining the Impact of Capital on Academic Achievement." *Economics of Education Review* 20(6):577–88.

Ladd, Helen F., and Randall P. Walsh. 2002. "Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right." *Economics of Education Review* 21(1):1–17.

Larner, Mary B., Lorraine Zippiroli, and Richard E. Behrman. 1999. "When School Is Out: Analysis and Recommendations." *The Future of Children* 9(2):4–20, (http://www.futureofchildren.org/usr_doc/vol9no2Art1done.pdf accessed 9/22/2003.)

Long, Sharon K., and Douglas A. Wissoker. 1995. "Welfare Reform at Three Years: The Case of Washington State's Family Independence Program." *The Journal of Human Resources* 30(4):766–90.

Ludwig, Jens, and Laurie J. Bassi. 1999. "The Puzzling Case of School Resources and Student Achievement." *Education Evaluation and Policy Analysis* 21(4):385–403.

Marcus, Sue M. 1997. "Using Omitted Variable Bias to Assess Uncertainty in the Estimation of an AIDS Education Treatment Effect." *Journal of Educational and Behavioral Statistics* 22(2):193–201.

Marshall, Nancy L., Cynthia Garcia Coll, Fern Marx, Kathleen McCartney, Nancy Keefe, and Jennifer Ruh. 1997. "After-School Time and Children's Behavioral Adjustment." *Merrill-Palmer Quarterly* 43(3):497–514.

Mayer, Daniel P. 1998. "Do New Teaching Standards Undermine Performance on Old Tests?" *Education Evaluation and Policy Analysis* 20(2):53–73.

Mayesky, M. 1980a. "Differences in Academic Growth as Measured in an Extended-Day Program in a Public Elementary School." Paper presented at the annual conference of the American Association of School Administrators, Anaheim, CA, February.

———. 1980b. "A Study of Academic Effectiveness in a Public School Care Program." *Phi Delta Kappan* December: 284–85.

Meyer, Robert H. 1996. "Value-Added Indicators of School Performance." In *Improving America's Schools: The Role of Incentives*, edited by Eric A. Hanushek and Dale W. Jorgenson (PAGE RANGE). Washington, D.C.: National Academy Press.

———. 1999. "The Effects of Math and Math-Related Courses in High School." In *Earning and Learning: How Schools Matter*, edited by Susan Mayer and Paul Peterson (PAGES). Washington, D.C.: Brookings Institution Press and New York: Russell Sage Foundation.

———. 2002. "An Evaluation of the Urban Systemic Initiative and Other Academic Reforms in Texas: Statistical Models for Analyzing Large-Scale Data Sets." In *Models for Analysis of NSF's Systemic Initiative Programs—The Impact of Urban Systemic Initiatives on Student Achievement in Texas, 1994–2000*, by Norman L. Webb, William H. Clune, Daniel Bolt, Adam Gamoran, Robert H. Meyer, Eric Osthoff, and Christopher Thorn (pp. 119-178). Technical Report to the National Science Foundation, Washington, D.C. Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison.

Miller, Beth M., Susan O'Connor, Sylvia W. Sirignano, and Pamela Joshi. 1997. *I Wish the Kids Didn't Watch So Much TV: Out-of-School Time in Three Low-Income Communities.* Wellesley, Mass.: National Institute on Out-of-School Time, Wellesley College.

Moga, E., K. Burger, Lois Hetland, and E. Winner. 2000. "Does Studying the Arts Engender Creative Thinking? Evidence for Near but Not Far Transfer." *Journal of Aesthetic Education* 34(3–4):91–104. Reviewed in Deasy, R. J. 2002. *Critical Links: Learning in the Arts and Student Academic and Social Development.* Washington, D.C.: Arts Education Partnership.

Morgan**,** Stephen L. 2001. "Counterfactuals, Causal Effect Heterogeneity, and the Catholic School Effect on Learning." *Sociology of Education* 74(October):341–74.

Mussoline, Lawrence J., and Roger C. Shouse. 2001. "School Restructuring as a Policy Agenda: Why One Size May Not Fit All." *Sociology of Education* 74(1):44–58.

National Governors Association. 1999. *Expanding Learning: Extra Learning Opportunities in the States.* Washington, D.C.: Author.

Parks, M., and D. Rose. 1997. *The Impact of Whirlwind's Reading Comprehension Through Drama Program on 4th Grade Students' Reading Skills and Standardized Test Scores.* Berkeley, Calif.: 3D Group. Reviewed in Deasy, R. J. 2002. *Critical Links: Learning in the Arts and Student Academic and Social Development.* Washington, D.C.: Arts Education Partnership.

Pettit, G. S., R. D. Laird, J. E. Bates, and K. A. Dodgem. 1997. "Patterns of After-School Care in Middle Childhood: Risk Factors and Developmental Outcomes." *Merrill-Palmer Quarterly* 43: 515–38.

Phillips, Meredith, James Crouse, and John Ralph. 1998. "Does the Black-White Test Score Gap Widen after Children Enter School?" In *The Black-White Test Score Gap*, edited by Christopher Jencks and Meredith Phillips (pp. 229-272). Washington, D.C.: Brookings Institute Press.

Podlozny, A. 2000. "Strengthening Verbal Skills Through Use of Classroom Drama: A Clear Link." *Journal of Aesthetic Education* 34(3–4):239–76. Reviewed in Deasy, R. J. (2002). *Critical Links: Learning in the Arts and Student Academic and Social Development.* Washington, D.C.: Arts Education Partnership.

Posner, J., and D. Vandell. 1994. "Low-Income Children's After-School Care: Are There Beneficial Effects of After-School Programs?" *Child Development* 65(2):440–56.

Public Agenda. 1997. *Kids These Days: What Americans Really Think About the Next Generation.* Washington, D.C.: Author.

Puma, M. J., C. Jones, D. Rock, and R. Fernandez. 1993. *Prospects: The Congressionally Mandated Study of Educational Growth and Opportunity: Interim Report.* Bethesda, Md.: Abt Associates.

Puma, M. J., N. Karweit, C. Price, A. Ricciuti, W. Thompson, and M. Vaden-Kiernan. 1997. *Prospects: Final Report on Student Outcomes.* Bethesda, Md.: Abt Associates.

Reaney, L., K. Denton, and J. West. 2001. "The World as Our Classroom: Enrichment Opportunities and Kindergartners' Cognitive Knowledge and Skills." Paper presented at the annual meeting of the Society for Research in Child Development, Minneapolis, Minn., April 19.

Riley, D., J. Steinberg, C. Todd, S. Junge, and I. McClain. 1994. *Preventing Problem Behaviors and Raising Academic Performance in the Nation's Youth.* Madison: University of Wisconsin.

Ross, Catherine E., and Beckett A. Broh. 2000. "The Roles of Self-Esteem and the Sense of Personal Control in the Academic Achievement Process." *Sociology of Education* 73(4):270–84.

Seppanen, Patricia S.; Love, John M.; deVries, Dianne Kaplan; Bernstein, Lawrence; Seligson, Michelle; Marx, Fern; & Kisker, Ellen Eliason. 1993. *National study of before and after-school programs. Final report.* Portsmouth, NH: RMC Research Corporation. (ERIC Document No. ED356043, http://ericeece.org/pubs/digests/ed-cite/ed356043.html accessed 9/22/2003).

Singer, Judith D. 1998. “Using SAS Proc Mixed to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models.” *Journal of Educational and Behavioral Statistics* 24(4):323–55.

Smith, Julia B. 1996. “Does an Extra Year Make Any Difference? The Impact of Early Access to Algebra on Long-Term Gains in Mathematics Attainment.” *Education Evaluation and Policy Analysis* 18(2):141–54.

Snijders, Tom. 1996. “Analysis of Longitudinal Data Using the Hierarchical Linear Model.” *Quality and Quantity* 30:405–26.

Steinberg, L. 1996. *Beyond the Classroom: Why School Reform Has Failed and What Parents Need to Do.* New York: Touchstone.

Sunmonu, Kola, John Larson, Yolanda Van Horn, Elizabeth Cooper-Martin, and Jennifer Nielsen. 2002. “Evaluation of the Extended Learning Opportunities Summer Program.” Rockville, Md.: Office of Shared Accountability, Montgomery County Public Schools.

Thomas, W. B., and K. J. Moran. 1991. “The Stratification of School Knowledge Through Extracurricular Activities in an Urban High School.” *Urban Education* 26(3):285–300.

Thum, Yeow Meng. 1997. “Hierarchical Linear Models for Multivariate Outcomes.” *Journal of Educational and Behavioral Statistics* 22(1):77–108.

U.S. Bureau of Labor Statistics. 2003. “Employment Characteristics of Families, Summary (Table 4).”(http://www.bls.gov/news.release/famee.t04.htm accessed 9/22/2003.)

Vandell, D. L., and K. M. Pierce. 1997. *Safe Haven Program Evaluation (1995–96).* Madison, Wisc.: Madison Metropolitan School District.

———. "Can After-school Programs Benefit Children Who Live in High-Crime Neighborhoods?" A Presentation at the Poster Symposium, Children's Out-of-School Time: The Next Generation of Research, Biennial Meeting of the Society for Research in Child Development, April 1999, in Albuquerque, New Mexico.

Vandell, D. L., and J. Posner. 1995. *An Ecological Analysis of the Effects of After School Care.* Chicago: Report to the Spencer Foundation**.**

———. 1999. “Conceptualization and Measurement of Children’s After-School Environments.” In *Assessment of Environments across the Lifespan*, edited by S. I. Friedman and T. D. Wachs (pp. 167-97) Washington, D.C.: American Psychological Association Press.

Vaughn, K. 2000. “Music and Mathematics: Modest Support for the Oft-Claimed Relationship.” *Journal of Aesthetic Education* 34(3–4):149–66. Reviewed in Deasy, R. J. (2002). *Critical Links: Learning in the Arts and Student Academic and Social Development.* Washington, D.C.: Arts Education Partnership.

Whitaker, G., K. Gray, and B. Roole. 1998. *After-School Program Handbook: Strategies and Effective Practices*. Chapel Hill: University of North Carolina Center for Urban and Regional Studies.

Winner, E., and M. Cooper. 2000. "Mute Those Claims: No Evidence (Yet) for a Causal Link Between Arts Study and Academic Achievement." *Journal of Aesthetic Education* 34(3–4):11–75. Reviewed in Deasy, R. J. (2002). *Critical Links: Learning in the Arts and Student Academic and Social Development.* Washington, D.C.: Arts Education Partnership.

Yasumoto, Jeffrey Y., Kazuaki Uekawa, and Charles E. Bidwell. 2001. "The Collegial Focus and High School Students' Achievement." *Sociology of Education* 74(3):181–209.

Yen, Wendy M. 1985. "Increasing Item Complexity: A Possible Cause of Scale Shrinkage for Unidimensional Item Response Theory." *Psychometrika* 50(4):399–410.

Zill, Nicholas, Christine Winquist Nord, and Laura Spencer Loomis. 1995. "Adolescent Time Use, Risky Behavior and Outcomes: An Analysis of National Data." Washington, D.C.: U.S. Department of Health and Human Services Office of Human Services Policy. (http://aspe.hhs.gov/hsp/cyp/xstimuse.htm accessed 9/22/2003.)

APPENDIX A
Detailed Descriptive Statistics

Exhibit A1. Descriptive Statistics for Variables Used in Analysis

Group	Label	N	Mean	Std Dev	Min	Max
TEST SCORES	Math: Concepts and Applications, 1992	4282	698.96	57.27	466.0	867.0
	(SSMA) 1991	4282	684.96	53.96	472.0	839.0
	Math Computation, 1992	4282	695.43	42.59	492.0	813.0
	(SSMC) 1991	4282	673.59	47.75	427.0	813.5
	Reading Comprehension, 1992	4282	701.46	53.73	521.0	847.0
	(SSRC) 1991	4282	681.72	59.75	547.0	843.2
	Reading Vocabulary, 1992	4282	688.63	47.97	516.0	835.0
	(SSRV) 1991	4282	674.65	44.02	555.0	835.0
Teacher ratings	Attentiveness, 1992	4282	2.38	0.55	1.0	3.0
	(Atten) 1991	4282	2.41	0.53	1.0	3.0
	Cooperativeness, 1992	4282	2.65	0.41	1.0	3.0
	(Coop) 1991	4282	2.66	0.39	1.0	3.0
	Participation, 1992	4282	2.02	0.57	1.0	3.0
	(Partic) 1991	4282	2.04	0.55	1.0	3.0
Parent reports, 1992	Art, music, or dance	4251	0.21	0.40	0.0	1.0
	Sports	4243	0.42	0.49	0.0	1.0
	Language	4197	0.04	0.19	0.0	1.0
	Religious instruction	4225	0.49	0.50	0.0	1.0
	History and culture	4215	0.04	0.20	0.0	1.0
	Computer	4227	0.09	0.28	0.0	1.0
CONTROLS	Limited English proficient (ever)	4282	0.11	0.31	0.0	1.0
	Black	4282	0.13	0.34	0.0	1.0
	Hispanic	4282	0.15	0.36	0.0	1.0
	Female	4282	0.53	0.50	0.0	1.0
	Parent socioeconomic status	4282	0.17	0.92	-1.8	2.6
	Urban school	4282	0.36	0.48	0.0	1.0
	Rural school	4282	0.34	0.47	0.0	1.0
	"Educational load"	4282	-0.50	4.54	-7.9	12.8
	Free/Reduced lunch percent	4282	48.57	28.56	1.0	100.0
	Change schools (ever)	4282	0.11	0.31	0.0	1.0
	Single-parent household	4282	0.11	0.31	0.0	1.0
INSTRUMENTAL VARIABLES	Differences between parent and child reports, 1992	4144	1.16	1.01	0.0	5.0
	1991	4282	1.24	1.02	0.0	5.0
Parent reports, 1991	Art, music, or dance	4282	0.18	0.38	0.0	1.0
	Sports	4282	0.37	0.48	0.0	1.0
	Language	4282	0.03	0.18	0.0	1.0
	Religious instruction	4282	0.48	0.50	0.0	1.0
	History and culture	4269	0.04	0.20	0.0	1.0
	Computer	4282	0.07	0.26	0.0	1.0
Child reports, 1991	Art, music, or dance	4282	0.28	0.45	0.0	1.0
	Language	4282	0.05	0.22	0.0	1.0
	Religious instruction	4282	0.19	0.39	0.0	1.0
	Computer	4282	0.13	0.34	0.0	1.0
	Sports	4282	0.56	0.50	0.0	1.0
Child reports, 1992	Art, music, or dance	4282	0.26	0.44	0.0	1.0
	Language	4282	0.04	0.20	0.0	1.0
	Religious instruction	4282	0.15	0.35	0.0	1.0
	Computer	4282	0.11	0.31	0.0	1.0
	Sports	4282	0.53	0.50	0.0	1.0
Parent reports	Any nonschool lessons, 1992	4282	0.72	0.45	0.0	1.0
	1991	4282	0.69	0.46	0.0	1.0

Source: Prospects Data, 3rd grade cohort. Sample used for Math Concepts and Applications Regressions.

APPENDIX B

Behavioral Scales Used in the Analysis

Child Behavior Ratings

All items used to create the behavior ratings were derived from the *Prospects* student profile that was completed annually by teachers for every student in the study. For the initial factor analyses, 1992 data were used as this was the first year for which data were available for all three grade cohorts. A total of 19 items from the rating scales were included in maximum likelihood factor analyses with oblique (promax) rotation, and we tested zero-, one-, two-, three-, and four-factor models. Based on both change in chi-square relative to change in degrees of freedom and the number of factors with eigenvalues greater than 1, the three-factor models were identified for further exploration. Similar, but not identical, factor patterns were found across the three cohorts. However, the structure found in the **3rd grade cohort** was the best fit and served as the bases for further estimation.

Data from the 1st and 7th grade cohorts were subjected to Procrustes rotation, which involves rotating to a target matrix, in this case the loadings from the 3rd grade cohort promax rotation. For both cohorts, this step resulted in reasonable factor loadings. Consequently, the 3rd grade cohort solution was selected as the standard for the behavioral scales.

Next, the 1991 data (available only for the for the 3rd and 7th grade cohorts) were subjected to maximum likelihood factor analysis with Procrustes rotation to the 3rd grade cohort three-factor solution. Again, this resulted in clean factors for both cohorts. This procedure was repeated for the 1993 and 1994 data, again resulting in virtually the same factors. Thus, we concluded that we could use the same three-factor solution across all three cohorts and all study years. The three factors appear to measure (1) cooperation and compliance, (2) attention and motivation, and (3) interest and participation. The items that loaded on each factor are listed in exhibits B1 through B3.

To create the final scale scores, items were coded to reflect the same directionality; for example, a high score on an item reflected that teachers rated a child as exhibiting the preferred behavior. Scale scores were created by averaging the items loaded on the scale, with a scale score range of 1 to 3.

Exhibit B1. Student Profile Items Loading on the Cooperation/Compliance Scale

Variable description	1991	1992	1993	1994
Gets along with teachers	11g	10f	10f	9f
Has respect for authority	11j	10i	10i	9h
Is honest most of the time	11d	10d	10d	9d
Is willing to follow rules	11b	10b	10b	9b
Can work with other students	11q	10p	10p	9n
Is happy most of the time	11l	10k	10k	9j
Does not disrupt the class	10d	9d	9d	8d
Makes friends easily	11f	10e	10e	9e
Enjoys school	11h	10g	10g	9g

Exhibit B2. Student Profile Items Loading on the Attention/Motivation Scale

Variable description	1991	1992	1993	1994
Attention span	9b	8b	8b	7b
Pays attention in class	10c	9c	9c	8c
Motivation to learn	9c	8c	8c	7c
Can concentrate for at least 1 hour	11n	10m	10m	—
Works hard at school	11a	10a	10a	9a
Cares about doing well	11c	10c	10c	9c
Is a creative person	11k	10j	10j	9i

Exhibit B3. Student Profile Items Loading on the Class Participation Scale

Variable description	1991	1992	1993	1994
Asks questions in class	10e	9e	9e	8e
Class participation	10f	9f	9f	8f
Asks for extra help	10g	9g	9g	8g

Note: Numbers and letters refer to questions in surveys.
— = not available.

APPENDIX C

Measurement Error and IV Models

Controlling for measurement error using an instrumental variable (IV) is quite common, especially for economists studying education. For example, Ashenfelter and Krueger (1994) uses IV when estimating the economic returns to education, and Chaplin (1998a,b) uses IV to control for measurement error when estimating the economic returns to skills.[48] Greene (2000) presents a good discussion if using instrumental variables to control for measurement error on pages 375–79.

The following is a brief explanation of why IV works to control for measurement error. First, let

$Y = X'\beta + e$, where
Y = a vector of current skills (one observation for each individual in the data),
X = [A L] (a matrix of independent variables),
A = a vector of after-school activities,
L = a vector of lagged skills, and
β is the vector of parameters to be estimated.

We assume that a simple ordinary least squares (OLS) model gives us an unbiased estimate of the coefficients on X,

$$\beta \text{ OLS1} = \text{inv}(X'X)(X'Y).$$

Now suppose we observe lagged skills with error:

Lu = L + u and u is a vector of random measurement errors uncorrelated with X and e.

---

[48] In the same papers, Chaplin also estimates the impacts of course-taking on test scores, controlling for lagged test scores without controls for measurement error. He has two years of test scores and can use the earlier test scores as instrumental variables for the later test scores when estimating the economic returns to skills. This method does not work when estimating the impacts of course-taking on skills since he only has two years of test scores—one for before the courses and one for after.

Let XU = [A Lu], and U = [0 u] so XU = X + U. Using OLS, we get

β OLS2 = inv(XU′ XU)(XU′ Y).

It turns out that the numerator of β OLS2 converges to the numerator of β OLS1 = limit (X′ Y):

limit (XU′ Y) = limit (X′ Y) + limit (u′ Y) = limit (X′ Y)

(as the number of observations approaches infinity), so the numerator is fine. This happens because cov(u,Y) = 0. However, the denominator of β OLS2does not do so well as it does not converge to the denominator of β OLS1 = limit (X′ X):

denominator β OLS2 = limit (XU′ XU) = limit (X′ X) + limit (U′ U) <> limit (X′ X)

because limit (U′ U) > 0.

Hence, the denominator of β OLS2 is biased. In general this will bias the estimated coefficient on lagged test scores down and, if lagged test scores are positively correlated with after-school activities (as they are in our data), then the coefficient on after-school activities will be biased up.

Now let Z = [A z] where z = an instrument for the lagged test scores and z = L + *e,* where *e* is another error term uncorrelated with L, U, and e. Also let E = [0,*e*] so Z = X + E. In our case, z is a vector of the differences between the parent and child reports on their after-school activities. The IV estimate of β is

β IV = inv(Z′ X)(Z′ Y).

Once again, it turns out that the numerator converges appropriately:

limit (Z′ Y) = limit((X + E)′ Y) = limit (X′ Y) + limit(E′ Y) = Limit (X′ Y)

because cov(E,Y) = 0.

In addition, in this case the denominator also converges well:

limit (Z′ X) = limit (X′ X) + limit(E′ X) = limit (X′ X)

because cov(E,X) = 0.

To summarize,

limit (β IV) = limit (inv(Z′ X)(X′ Y)) = limit (inv(Z′ X)) limit (X′ Y)
= limit(inv(X′ X)) limit (X′ Y) = limit (inv(X′ X)) limit (X′ (X′ β + e))
= limit(inv(X′ X))limit(X′ X) β = β

so β IV is a consistent estimate of β.

A careful reader might wonder how it is that we could use the same instrument ("survey misreports") for all of the outcomes in our analysis. This technique can be justified if the instrument is thought of as measuring an underlying skill that is a weighted average of all the outcomes in our analysis. Thus,

$$Z = \beta 1 * O1 + \beta 2 * O2 + \beta 3 * O3 + \beta 4 * O4 + \beta 5 * O5 + \beta 6 * O6 + \beta 7 * O7 + \varepsilon$$

where Oi is the lagged value of outcome i and βi is the coefficient on Oi.

Now, as long as these outcomes have no independent effects on each other, then Z can serve as an IV for all of them. To test this hypothesis we need to be able to instrument all of the lagged outcomes for measurement error. Unfortunately, we were not able to do this well because we lacked instruments that would reasonably be expected to

have differential impacts on each of these outcomes.[49] We did, however, have a set of six additional instruments (the lagged activities) that were found to be jointly significant predictors of the lagged outcomes even after controlling for all other exogenous variables in our analysis, including the main instrumental variable (survey mistakes).[50] When we estimated a model including all lagged outcomes in each model and controlling for measurement error in all, the additional lagged outcomes were not jointly significant. However, the standard errors on all variables in these models were very large.

A related point of interest: When we estimated lag models without controlling for measurement error, including the lagged values of all outcomes, the after-school activities variables remained jointly significant.[51] Thus, controlling for additional lagged outcomes does not produce the same result as controlling for measurement error in the main outcome. The lesson remains that controlling for measurement error in lagged outcomes appears to be key for obtaining unbiased results.

---

[49] One possibility would be to use outcomes lagged two years as the instrumental variables for the outcomes lagged one year. Such data are available for later years of the Prospects dataset, but in the later years we would not have the parent and child reports to create the instrumental variable used here.

[50] We estimated these models jointly in SAS using Proc Syslin. The p-value for the joint test of significance of all the lagged activities on all of the lagged outcomes was 0.003. One or two of the six activities was statistically significant at the 5 percent level in each model, except for the outcome PARTIC1991 and many others were statistically significant at the 10 percent level.

[51] The after-school activities were not jointly significant in the model that controlled for measurement error in all lagged outcomes, but as stressed above, the standard errors on all variables were fairly large in those models.

APPENDIX D

Defining After-School, Out-of-School, and Enrichment

We believe that the results of our paper can be applied to activities that are often described using the terms "after-school," "out-of-school," and "enrichment" as long as they are not academically focused. These terms are generally used to describe activities that children take part in when not in school and are often used interchangeably.[52] The term "enrichment" is probably the broadest of the three and can even refer to activities that take place during school.[53] It generally refers to activities that are not aimed directly at improving student academic skills. After-school and out-of-school activities do not occur during regular school hours but can include academic activities, such as homework, tutoring, and even extra classes or private lessons. "After-school" generally refers to activities taking place after school gets out and before children go home on regular school days (see, for instance, Whitaker et al. 1998), while out-of-school time can also include before-school, weekend, and holiday activities. All three terms can include extracurricular activities and overlap significantly with the activities covered in our study.

Most research in this area attempts to describe differences in the types of activities that youth engage in during their nonschool time. Vandell and Posner (1999) make distinctions based on the mode, quality,[54] and intensity of care.

---

[52] For instance, Larner et al. (2001) uses the terms "after-school programs" and "out-of-school programs" to refer to "programs in schools or community organizations that provide a range of activities in one place. . . . They offer supervised activities and a safe place to spend time when school is not in session (including holidays and summer vacations)." p. 5.

[53] Reaney et al. (2001) describe *enrichment* activities as nonacademic ones that occur outside of school, such as family outings and extracurricular activities. The outings include visits to a library, museum, zoo or aquarium, sporting event, play, or concert. Extracurricular activities include dance, athletics, clubs, music or drama lessons, arts, crafts, and non-English instruction outside of school. Although they focus on nonacademic activities, Reaney et al. (2001) do look for impacts on academic skills.

[54] Aspects of quality that are often discussed include child-staff ratios, class size, staff education and training, space, arrangement of space, availability of materials, health and safety, provisions for autonomy, child choice, privacy, consistency, and stability.

Modes can be defined in a number of ways. Common mode distinctions for young school-aged children include mother, other adult, sibling < 14, adolescent, self, after-school program. These distinctions probably derive from the child care literature, which employs fairly similar distinctions. These modes overlap with those used in our paper in the sense that our "classes" variables would probably fall partly under "after-school program" and partly under "other adult."

Distinctions more common in discussions of school-age care, as opposed to child care, include those made by Fashola (1998). She distinguishes between day care, after-school programs, and school-based academic extended-day programs. Day care programs are generally nonacademic but require licensing for sites and workers. After-school programs cover a similar range of activities but generally do not require licensed sites and staff. Finally, school-based academic extended-day programs typically occur at school, employ school staff, and focus primarily on academics. Our categories would presumably overlap with the nonacademic day care and after-school programs, at least somewhat.

As for other researchers, the distinctions we can make are determined to a large extent by the questions asked in the data we used. We focus our analysis on a subset of nonschool activities using answers to the question "Does your child attend classes outside of regular school to study any of the following?" The answer categories used in our analysis were

1. arts, music, or dance lessons,
2. language classes,
3. religious instruction,
4. computer classes,
5. sports, exercise, or gymnastics, and
6. history or culture.

We refer to these activities as extracurricular activities since none are designed to directly improve skills in the subjects we cover (math and reading). However, we note that tests could be designed to measure proficiency in the extracurricular activities we include in our analysis and that students participating in relevant activities would likely improve their skills as measured by such tests.