Using Large-Scale Research to Gauge the Impact of Instructional Practices on Student Reading Comprehension: An Exploratory Study

Harold Wenglinsky

Baruch College

Abstract

Small-scale research has identified classroom practices that are associated with high student performance in reading comprehension. It is not known, however, whether these findings generalize to larger samples and populations, as most large-scale studies of the impact of teaching on student performance do not include measures of classroom practices. Generalizing to larger populations is particulary important at a time when policies national in their scope are calling for "scientifically-based" instruction in reading. The current study explores the possibility of using large-scale data and methods to study classroom practices in reading comprehension. It finds that such studies are both feasible and necessary. They are feasible insofar as it proved possible to collect

and analyze data on classroom practices and student reading comprehension, and discern substantial effects of the one on the other. They are necessary insofar as the study confirmed the effectiveness of some classroom practices but not others. It therefore cannot be assumed that the findings of small-scale studies generalize to large populations.

Introduction

With the passage of the Federal Reading First program and parallel state efforts to improve student reading skills, policy makers and educators have been looking for "scientifically-based" materials, professional development and instruction. While the research base on early reading skills such as phonics has been found to be substantial, there is a lack of conclusive evidence on effective instruction in more advanced skills such as reading comprehension.

This research gap is, for the most part, attributable to the small-scale nature of research that seeks to identify effective techniques for teaching reading comprehension. Over the last forty years, researchers have conducted a host of small-scale studies identifying certain classroom practices as effective in teaching reading comprehension. The findings of these studies have been remarkably consistent, with the same set of classroom practices again and again appearing to be related to student reading comprehension performance. The strength of these studies lies in their high level of internal validity. Many use experimental designs, and many others are quasi-experimental. These studies also possess a common shortcoming, however. Because the development of a robust design is extremely labor intensive, such studies tend to be small in scale, limited to a few classrooms or schools. The degree to which these studies apply to large populations is therefore not known.

Large-scale research does not provide much information on classroom effects. The common method for gauging the impact of teaching on student performance using large-scale data is known as the production function. It involves collecting observational data on large numbers of teachers and students and then using the technique of regression analysis, which relates teacher characteristics to student performance. Most production function studies do not measure classroom practices, due to the difficulty of measuring them for large numbers of teachers. And most have not found a clear and consistent relationship to student outcomes for the teacher characteristics they do measure.

The present study addresses methodological problems common in the production function literature to demonstrate the possibility of using large-scale data to study classroom practices. The problems addressed here include the lack of measurement models, the low validity and reliability of teacher self-reports and the interdependence of many independent variables. This is accomplished by using national data on 7,194 fourth graders and their teachers from the 2000 National Assessment of Educational Progress (NAEP). The study relates teachers' classroom practices, as well as their background characteristics, to student performance on a reading comprehension assessment, taking into account student background characteristics.

The study finds that the addition of classroom practices to large-scale models of reading performance is vital to the successful isolation of teacher effects. Once such variables are introduced, teacher effects can prove quite substantial, nearly as large as student background effects. The study also finds that testing small-scale results with large-scale data is crucial for establishing the effectiveness of specific classroom practices. A link was confirmed between some but not all classroom practices and student performance. It would be premature, however, to draw substantiative conclusions about effective reading practices from these findings. Rather, the purpose of this paper is to draw methodological conclusions about the viability of large-scale research on instructional practice. Before addressing these conclusions, however, it is worth reviewing key findings from the prior literature and describing in some detail how the current study was conducted.

Background

Intensive study has been devoted to classroom effects on reading. The reports of the National Research Council (Snow et al., 1998) and the National Reading Panel (2000) have identified hundreds of studies of how classroom practices affect reading performance. Much of this research, however, focuses on the early stages of reading skill acquisition, such as phonemic awareness and word recognition, with much less on reading comprehension (Snow et al., 2000). Nonetheless, the body of research on reading comprehension is substantial enough to make it possible to identify seven kinds of practices that are consistently associated with improvements in student reading comprehension.

First, students perform better when explicitly taught metacognitive skills. Metacognitive skills are the ways in which readers glean meaning from texts. In a series of studies, Durkin (1978, 1981) found that teachers, in explicating texts, rarely instruct students in methods of explication. In the wake of this finding, a variety of approaches to teaching metacognitive skills were developed, including reciprocal teaching, questioning and direct instruction. Research on these techniques has generally found positive effects on reading comprehension (Cross & Paris, 1988; Rosenshine & Meister, 1994; Hansen & Pearson, 1983; Wharton McDonald et al., 1998; Kaniel et al., 2000; Mueller, 1997; Alfassi, 1998).

Second, students seem to perform better when reading and writing instruction are integrated. After years of these skills being taught separately, educators called for their integration in the 1980's. Later, reading and writing were combined under the single heading of language arts, as reflected in many of the recently promulgated state academic standards. Research on using writing to improve reading comprehension has generally supported the approach (Wharton-McDonald et al., 1998; Cantrell, 1999; Knapp et al., 1995).

Third, research on the texts students read has documented advantages to using trade books rather than basal readers. Basal readers tend to abridge texts to maximize accessibility, whereas trade books are real world texts, and are often chosen to convey content rather than just reading skills. The use of trade books has been found to increase student motivation as well as to improve reading comprehension skills (Popplewell & Doty, 2001; Guthrie, 2001; Guthrie et al., 2000; Guthrie, 2000; Guthrie et al., 1999; Guthrie, 1998; Guthrie 1996).

Fourth, students seem to benefit from time spent reading in class. Research on time on task has suggested that more time spent teaching reading is associated with improved performance. Particularly strong effects have been found for time spent in the act of reading, with oral reading appearing more beneficial than silent reading (Topping & Paul, 1999; National Reading Panel, 2000).

Three other instructional techniques supported by the literature are having students work in groups, involving parents and using authentic assessments to measure student progress. Supposedly, students learn more from one another and are more motivated when engaging in group work. A variety of parental involvement activities, from checking homework to reading together, have been found to be conducive to improved reading performance (Epstein, 2001; Epstein & Dauber, 1994). And measuring student performance through tasks that are as similar as possible to class work and homework seems to be more effective than having students take traditional multiple-choice or short-answer tests.

While small-scale research has succeeded in identifying numerous ways in which teachers affect student performance, large-scale research has generally not been able to confirm these findings. Beginning with the Equality of Educational Opportunity Study (Coleman et al., 1966), production function studies have found that most teacher effects are overwhelmed by student effects. Of the hundreds of production functions estimated in the wake of the Coleman Report, less than one-third could discover a link between student outcomes and teacher experience, less than one-quarter could do so for teachers salaries, and just one in ten could do so for educational attainment. Two types of teacher effect did prove more robust. Many studies have isolated modest effects for teachers' majoring in the subject they teach and teacher scores on basic skills tests. The relative lack of large-scale studies confirming teacher effects, however, has to led to meta-analysis of them coming to divergent conclusions, some accepting and others questioning the existence of teacher effects (Hanushek, 1997; Hanushek, 1996a; Hanushek, 1996b; Hanushek, 1989; Greenwald, Hedges & Laine, 1996; Hedges & Greenwald, 1996; Hedges, Greenwald & Laine, 1994).

The disappointing results of large-scale studies may stem from their various methodological shortcomings. First, these studies tend to focus on teacher effects that are relatively easy to measure with large-scale data, namely teacher background characteristics such as education level or college major. Such studies thus tend not to measure the effects that small-scale research has found to be substantial. Second, such studies tend to lack measurement models; they assume variables are perfectly measured and do not develop constructs from multiple indicators. Yet measurement error in teacher self-reports of behavior and background is substantial, and can be minimized through multiple indicators (Mayer, 1999). Third, such studies tend not to relate independent variables to one another, when small-scale research suggests that teacher variables are very much affected by student background and school context.

A few large-scale studies do relate classroom practice to student performance in mathematics and science, and reveal substantial classroom effects. The nationally representative National Educational Longitudinal Study (National Center for Education Statistics, 1996) found that an emphasis by teachers on conveying

higher-order thinking skills was positively associated with student performance in math but not in science. A study representative of the state of California (Cohen & Hill, 2000) found that reform-minded classroom practices were positively associated with student mathematics performance. And an analysis of the nationally representative 1996 National Assessment of Educational Progress in Mathematics (Wenglinsky, 2001) found that an emphasis on conveying higher-order thinking skills, engaging in hands-on learning activities, and receiving professional development to address special populations of students were all positively related to math scores. The fact that all three of these studies uncovered substantial teacher effects when classroom variables were included suggests the need for similar work in the area of reading. It is to this work that we now turn.

Research Questions, Data and Method

The exploratory study described here is designed to address two methodological research questions suggested by the prior literature. First, do the classroom practices identified as important by the small-scale literature prove to be uniformity related to student reading performance? If all are confirmed, large-scale research can be said to add little to what is already known. If it proves, however, that some practices are confirmed while others are not, this finding would suggest the need to conduct an independent program of studies using large-scale data. Second, does the addition of classroom practices to teacher effects models substantially increase the importance of these effects, compared to student background effects? If so, this finding would suggest the importance of including classroom practice variables in future large-scale studies of reading. As will be seen, difficulties encountered in doing this exploratory study indicate that large-scale studies of classroom practices, while vital, raise many methodological hurdles that subsequent research will need to overcome.

To answer these questions, it was necessary to obtain large-scale data that were representative of a large population and included measures of student reading comprehension, teacher background, classroom practices and student and school background. Fortunately, a recent administration of the National Assessment of Educational Progress met these criteria. NAEP was administered to a nationally representative sample of 7,194 fourth graders in 2000 to assess their reading comprehension skills. In addition to the assessment, questionnaires were administered to students and their reading teachers, generating information on their backgrounds and classroom practices. (For overview of the NAEP 2000 Reading Assessment, see National Center for Education Statistics, 2001).

The use of the NAEP, however, introduces certain methodological hurdles that the current study needed to overcome. First, the study needed to appropriately handle variability in the reading comprehension measure. To limit the amount of time students were assessed, each student answered only a limited number of test items; consequently, it is not possible to generate a single student score. Instead NAEP provides five scores based upon the items the student answered and student and school background information. The recommended procedure for conducting secondary analyses using these five scores, known as plausible values, is to estimate a separate model for each and then pool them. The unstandardized and standardized coefficients are pooled by calculating their means and variances through the following formula:

$$v = u + (1.2)B,$$

where v is the pooled variance, u is the average sampling variance and B is the variance among the five plausible values.

Second, the study needed to appropriately handle the sample design. Because NAEP is a clustered, stratified sample, student and teacher observations are not independent of one another. If treated as a simple random sample, these observations will underestimate standard errors. Consequently, standard errors need to be adjusted. One acceptable technique for doing so is using a design effect. NAEP provides weights, known as jackknife weights, that can be used to estimate the effect of the sample design on the standard error of each coefficient, known as the design effect. Because of the computationally expensive nature of estimating the effect for every coefficient in a model, it is appropriate to estimate effects for a subset of coefficients and then select one of these for the purpose of inflating the standard errors of all coefficients.

To relate the teacher and student characteristics to student reading comprehension, the statistical technique of structural equation modeling (SEM) was employed. Like regression analysis, SEM makes it possible to relate independent variables to dependent variables, taking into account both the independent variables and statistical controls. It has two advantages over regression. First, it can test the fit of entire path models, meaning that it can estimate the coefficients and overall goodness of fit of models that relate independent variables to one another as well as to the dependent variable. This makes it possible to incorporate intervening variables into the model. Second, it can construct its independent and dependent variables from observed variables through factor models. This makes it possible both to take into account measurement error and to reduce such error through the use of multiple indicators. (Note 1)

The current study estimated two sets of factor and path models: The first set consisted of five versions of a teacher background model, one for each plausible value, and the second set consisted of five versions of a classroom effects model, also one for each plausible value. These models were estimated using AMOS 3.6 (Arbuckle, 1996), an SEM package, and STREAMS 1.8 (Gustafsson & Stahl, 1997), a pre-and post-processor for SEMs. The factor model portion of the teacher background model constructed measures of four teacher background characteristics (major, education level, years of experience, and perceived preparedness to teach), two student background characteristics (socio-economic status (SES) and home reading behavior), one school characteristic (class size) and one student outcome (a plausible value for reading comprehension performance). SES was constructed from five measures, home reading behavior from four, and the rest from single measures. (Factor models for single measures fix factor loadings at 1 and error terms at 0. See Table 1 for full list of the measures employed.)

Table 1. Descriptive Statistics for Teacher, Student and School Background Characteristics

	M	SD	N

Teacher Background

Preparedness	-.03	.86	7914
Major in English (1=yes)	.19	.42	7914
Major in Reading/Language Arts (1=yes)	.22	.46	7914
Education Level (1=Master's or more)	.38	.44	7914
Years of Experience (From 1= low to 5=high)	3.33	1.22	7914

Student Socio-economic Status

Family Subscribes to Newspaper (1=yes)	.72	.43	7914
Family Subscribes to Magazines (1=yes)	.72	.41	7914
Family Owns Encyclopedia (1=yes)	.80	.37	7914
Family Owns More Than 25 Books (1=yes)	.93	.25	7914
Mother College Graduate (1=yes)	.71	.36	7914

Student Reading Background

Uses Libraries (From 1= never or hardly ever to 4=most every day)	2.60	.93	7914
Talks to Peers about Reading (From 1=never or hardly ever 4=almost every day)	2.64	1.10	7914
Reads for fun (From 1=a poor reader to 4= almost every day)	3.03	1.04	7914
Self-assessment of Reading Skills (From 1=a poor reader to 4=a very good reader)	3.20	.83	7914

School Background

Class Size (From 1=36 or more students to 5=1-20 students)	4.00	.86	7914

The path portion of the model then related the student outcome to the teacher background measures, taking into account the student background and school measures. The factor portion of the classroom practices model constructed measures of eight classroom practices (teaching metacognitive skills, integrating reading and writing, use of reading materials such as trade books and basal readers, time spent reading in class, working in groups, parental involvement, authentic assessment and traditional assessment), two students background characteristics (SES and home reading behavior), one school characteristic (class size) and one student outcome (a plausible value). Metacognitive skills was constructed from five measures, integrating reading and writing from four measures, reading materials from five measures, time spent reading from two measures, group work from two measures, authentic assessment from four measures and traditional assessment from three measures. (See Table 2 for full list).

Table 2. Descriptive Statistics for Classroom Practices

	M	SD
Metacognitive Skills		
Describe Author's Method (From 1=never or hardly ever to 4=almost every day)	2.61	.80
Explain Reading (From 1=never or hardly ever to 4=almost every day)	3.39	.61
Make Generalizations (From 1=never or hardly ever to 4=almost every day)	3.44	.59
Predict Outcomes (From 1=never or hardly ever to 4=almost every day)	3.48	.59
Learning New Words in Context (From 1=never or hardly ever to 4=almost every day)	3.78	.39
Writing in Service of Reading		
Writing about Literature (From 1=never or hardly ever to 4=almost every day)	.43	.46
Reading and Writing (From 1=never or hardly ever to 4=almost every day)	.70	.43
Writing about Reading (From 1=never or hardly ever to 4=almost every day)	3.18	.69
Answers Questions in Writing (From 1=never or hardly ever to 4=almost every day)	3.37	.61
Reading Materials		
Trade Books (From 1=never or hardly ever to 4=almost every day)	.20	.37
Basal Readers (From 1=never or hardly ever to 4=almost every day)	.17	.34
Reading Kits (From 1=never or hardly ever to 4=almost every day)	1.88	1.01
Children's Newspapers (From 1=never or hardly ever to 4=almost every day)	2.08	.79
Worksheets (From 1=never or hardly ever to 4=almost every day)	2.97	.82
Time Reading		
Reading Aloud (From 1=never or hardly ever to 4=almost every day)	3.59	.57
Reading Silently (From 1=never or hardly ever to 4=almost every day)	3.78	.45

Group Work

Work in Small Groups   .69   .42
(From 1=never or hardly ever to 4=almost every day)

Engage in Group Activities   2.34   .73
(From 1=never or hardly ever to 4=almost every day)

Parental Involvement

Parents Check Homework   2.33   1.01
(From 1=never or hardly ever to 4=almost every day)

Parents Sign Off   2.96   .92
(From 1=never or hardly ever to 4=almost every day)

Authentic Assessment

Portfolios   1.99   .97
(From 1=never or hardly ever to 4=once or twice a week)

Paragraphs   3.19   .80
(From 1=never or hardly ever to 4=once or twice a week)

Projects   2.65   .67
(From 1=never or hardly ever to 4=once or twice a week)

Oral   3.00   .88
(From 1=never or hardly ever to 4=once or twice a week)

Traditional Assessment

Multiple-Choice Test   2.87   .88
(From 1=never or hardly ever to 4=once or twice a week)

Short-Answer Test   3.17   .78
(From 1=never or hardly ever to 4=once or twice a week)

Tests   2.49   .67
(From 1=never or hardly ever to 4=once or twice a week)

For the final version of the models, some of the multiple indicator constructs were turned into single indicator constructs to identify which indicator was responsible for the classroom effect of a construct. Thus, integrating reading and writing and reading materials were divided into their constituent indicators. Student background, school characteristics, and student outcomes were measured as per the teacher background models. The path portion of the classroom effects model related the classroom practice constructs to the student outcome, taking into account student background and school characteristics, as well as relating the student background and school characteristics to each of the classroom practices, thus making it possible to gauge the extent to which classroom practices acted as intervening variables between student background, school characteristics and student outcomes. (Note 2)

Results

The factor models and goodness-of-fit statistics reveal that the models fit the data well. For all factor models, the constructs loaded substantially and on all of the corresponding indicators, and all loadings were statistically significant at the .05 level. (Factor models are not presented here, but are available upon request.). All ten of the factor and path models also had adequate goodness-of-fit statistics. For the teacher background models, the RMSEAs were at the .03 level, with normed goodness-of-fit indices of .92 and comparative goodness-of-fit indices at .92 and .93, depending upon the plausible value. For the classroom practice models, the RMSEAs were at the .05 level, with both normed and comparative goodness-of-fit indices at .98. These results suggest that the hypothesized models were confirmed by the observed data.

The path models for teacher background reveal only a modest effect of teaching on student reading comprehension (Table 3). The strongest effects come from students, with SES having the largest effect in the model (b=.37) followed by reading background (b =.14). The school control, class size, also had an effect, albeit a modest one (b =.03). Among the five teacher background variables, only one, years of experience, proved statistically significant, with an unstandardized coefficient of .05. This findings differs somewhat from the literature, in which teacher major tend tends to have an effect and teacher experience tends not to have one. This divergence may be attributable to the fact that this study is of fourth graders and their elementary school teachers, whereas most of the studies of teacher major are at the high school level. (Note 3)

Table 3. Structural Equation Model of Teacher Background Effects

	Reading Achievement
Student Socio-economic Status	140.04** .37
Student Reading Background	16.99** .14
Class Size	1.50** .03
Teacher Preparedness	.72 .02
Teacher Major	1.96 .02
Teacher Education Level	-2.57* -.03
Teacher Experience	1.48** .05
Error	.90

*p <.10
**p<.05

The classroom practice path models reveal much more substantial teacher effects (Table 4). As with the teacher background model, the strongest single effect is of SES (b=.43). This is followed, however, by two teacher effects, the positive effect

of metacognitive skill instruction (b= .31) and the negative effect of time spent reading in class (b=.30). Student reading background is next in importance, with students with stronger backgrounds scoring higher on the reading comprehension assessment (b=.13). Teachers' having students write about literature they are reading and using trade books as their primary reading materials had modest positive effects (b=.04 for each). Class size, as in the teacher background model, also had a statistically significant effect of that size.

Table 4. Structural Equation Models of Classroom Effects

	Reading Achievement
Student Socio-economic Status	138.66** .43
Student Reading Background	15.96** .13
Class Size	1.64** .04
Writing	.95** .04
Basal Reading	.19 .01
Trade Books	1.43** .04
Metacognition	1.67* .31
Time Reading	1.38* .30

*p <.10
**p< .05

In addition, the classroom practice path models indicate that students are exposed to very different practices depending upon their background characteristics and those of their schools. Affluent students are more likely to be exposed to metacognitive instruction (b=.04), writing about literature (b=.07) and reading trade books (b=.07) than their less affluent peers. There is, however, no difference in time spent reading in class or the use of basal readers between the two groups. Schools with smaller classes also differ from those with larger classes, with small class students more likely to be exposed to metacognitive instruction (b =.03) and writing about literature (b =.06) as well as to spend more time reading in class (b=.05). It thus appears that effective classroom practices act as intervening variables between student SES and reading comprehension performance, with higher SES students more likely to be exposed to those practices that are themselves associated with higher NAEP scores. With class size, the pattern is less clear, as the practices associated with smaller class sizes may or may not have a positive relationship to NAEP scores.

These findings answer the first research question in the negative and the second in the affirmative. The large-scale data do seem to confirm some of the findings from small-scale research but not others. Some practices, namely metacognition, using

trade books and a measure of integrating reading and writing, did prove positively related to reading comprehension. Other practices, however, such as having students work in groups, increasing parental involvement, and the use of authentic assessment, did not. And time spent reading in class actually had a negative relationship to student performance. The addition of classroom practices to large-scale models seems to make the overall impact of teachers comparable to that of student background. As with typical production functions, the teacher background model revealed only a single modest teacher effect. The classroom practice model, however, revealed multiple teacher effects, some of them quite strong. The total standardized effect for the four teacher variables (.70) is actually somewhat larger than the total standard effect of the two student background measures (.56).

Conclusions

These findings have significant methodological implications for research on teacher effects on reading comprehension. The finding that some of the classroom practices proved effective while others did not suggest the need for synergy between small-scale and large-scale research. The findings of small-scale, highly internally valid, studies should serve as the basis for large-scale, highly externally valid, studies. Only in this way can it be known if small-scale findings are applicable to large populations. (This does not rule out the possibility that small-scale research can, by itself, provide information about small populations.) The finding that the introduction of classroom practices leads to substantial teacher effects suggests the need for large-scale research to embrace such variables. Clearly, the failure of previous large-scale research to uncover substantial teacher effects is in large part due to its not including such variables. In addition, the other methodological advances of the current study over traditional production functions proved useful. The use of multiple indicators improved the quality of the measures employed, and the use of path models led to the finding that classroom practices act as intervening variables between student background and reading comprehension performance.

Yet while the current exploratory study does take some steps to improve the large-scale methodology for the study of teacher effects, much remains to be done. One shortcoming of the current study is the ad hoc manner in which it addressed problems with teacher self-reports. Because it relied on pre-existing data, the study made use of interaction effects to increase the likelihood that teachers reporting the use of certain practices were actually using them. Doing so, however, truncated the sample, and is based on the assumption that the more experienced, better prepared teachers are more likely to accurately assess and report what their practices are. This assumption may or may not hold true for a given teacher. A more effective technique for reducing problems with teacher self-reports would be to begin to design questionnaire items that make clearer what the practices are and minimize the social desirability effects. For instance, a questionnaire might include a scenario in a classroom and ask the respondent to describe how he or she would address it. Respondents could also be asked to rank order the effectiveness of the classroom practices of others, or to draw up a time budget for various practices. Such methods, often employed in small-scale research, need to be applied on a larger scale to enhance the reliability and validity of the large-sample teacher reports.

Another shortcoming of the study is its use of cross-sectional data. Because the data are cross-sectional, it is not clear whether particular practices enhance reading comprehension or high performing students are more likely to have teachers engaging in such practices. The study did address this problem in an ad hoc fashion by controlling for measures of student home reading behavior. Indeed, those controls may have resulted in underestimates of teacher effects, in that teachers may positively influence home reading behavior. Whatever the impact of the ad hoc procedure, it is no substitute for longitudinal data that follow student performance over time, and hence it is crucial for subsequent large-scale studies to collect such data. Indeed, the Early Childhood Longitudinal Study (ECLS), which will follow a national sample of students from kindergarten through fifth grade, testing their reading skills and measuring teacher classroom practices, may address this need.

A third shortcoming of this study is its failure to fully take into account the multilevel nature of its data. This study involved multiple levels of analysis in that it related teacher-level inputs to student-level outputs. Yet students are not selected at random, but are clustered within classrooms. The employment of design effects addressed this issue somewhat by increasing standard errors based upon clustering at the level for the school district. But it did not take into account the impact of classroom level non-independence on standard errors. It also did not distinguish between student-level and contextual effects; the influence of student SES, for instance, may be in part due to the average SES of that student's peers. To fully address all of these issues, multilevel techniques, such as Hierarchical Linear Modeling or multilevel versions of SEM (MSEMs) need to be employed.

Finally, rich data is needed on teacher background. This study used the same kinds of summary measures employed by production function studies, such as years of experience and education levels. It may be that teacher background is extremely important, but can only be fully gauged through learning in a more nuanced way about the background. The education level of the teacher may not be important, but the nature and extent of the teacher education curriculum may be. Perhaps certain kinds of induction experiences are more conducive to high student performance. And the nature and extent of professional development experiences may play a role in encouraging particular effective classroom practices. Thus, while the current study suggests that failure to consider classroom practices has led large-scale research to underestimate teacher effects, it may be that the effects of teacher background have also been underestimated. The only way to gauge the full impact of teachers is to collect as much information about them as is collected about their students, and see how the biographies and classroom actions of the two actors unfold together.

In sum, it should be possible to gauge the effectiveness of instructional practices in the area of reading comprehension using large-scale data. The ability to generalize from smaller to larger populations is critical given the existence of new policies that call for "scientifically-based" instruction in reading, that are national in scope. Without knowing the applicability of particular techniques to all populations, policy makers and educators run the risk of imposing that technique on an inappropriate population. Consequently, research should build upon the methods identified in this exploratory study to determine which instructional practices are help for all

students, and which are helpful for particular subpopulations of students.

Notes

1. SEM accomplishes this through three steps. First, factor and path models are specified by the researcher. Factor models indicate which observed variables load on which constructs. Path models indicate which constructs are permitted to be related to one another. Second, through an iterative process, the covariance matrix that these specifications imply (S) is matched with a covariance matrix of the observed data (S) to maximize their fit with one another. Finally, the resulting output consists, for each construct, of standardized factor loadings and standard errors for each indicator, standardized and unstandardized path coefficients and their standard errors for each relationships between constructs; and goodness-of-fit statistics including the root mean squared error of approximation (RMSEA) and indices such as the comparative and normed fit indices.

2. Because of issues in using teacher self-reports, the classroom practice measures had to be transformed for the five models. Research has found teacher self-reports of classroom practices to be frequently unreliable. Some teachers may misrepresent the practices in which they engage because they do not fully understand what the named practices are, and some may misrepresent the practices because they perceive the practices as socially desirable. The NAEP data indicated that most teachers claim to engage in most practices, and consequently, giving full weight to the responses of all teachers would make it difficult to distinguish between those that actually do and do not engage in a given practice. Instead, practices were weighted by teachers' years of experience and their perception of their preparedness, giving more weight to the responses of the more prepared, more experienced teachers. This was accomplished through calculating interaction effects between each classroom practice and the two teacher background measures, and substituting these for the classroom practice measures in the models.

3. Teachers' education level had a negative coefficient at the .10 level. This counterintuitive effect requires further exploration.

References

Alfassi, M. (1998). Reading for meaning: The efficacy of reciprocal teaching in fostering reading comprehension in high school students in remedial reading classes. *American Educational Research Journal, 35,* 2, 309-332.

Anderson, S. A. (2000). How Parental Involvement Makes a Difference. *Reading Improvement, 37*, 2, 61-86.

Arbuckle, J. L. (1997). *Amos users' guide. Version 3.6.* Chicago: Small Waters Corporation.

Bromley. K. & Modlo, M. (1997). Using cooperative learning to improve reading and writing in language arts. *Reading and Writing Quarterly, 13*, 1, 21-35.

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

Cantrell, S. C. (1999). The effects of literacy instruction on primary students' reading and writing achievement. *Reading Research and Instruction, 39,* 1, 3-26.

Cohen, D. K. & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics

reform in California. *Teachers College Record, 102,* 2, 294-343.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York. R. L. (1966). *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office.

Cross, D. R. & Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology, 80,* 2, 131-142.

Durkin, D. (1978). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly, 14,* 4, 481-533.

Durkin, D. (1981) Reading comprehension instruction in five basal reader series. *Reading Research Quarterly, 16,* 4, 515-544.

Epstein, J. L. (2001). *School, family and community partnerships: Preparing educators and improving schools.* Boulder, CO: Westview Press.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal of Legislation, 28*(2), 465-498.

Ferguson, R. F. & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. F. Ladd (Ed.), *Holding school accountable: Performance-based reform in education*(pp. 265-298). Washington, D.C.: The Brookings Institution.

Goldhaber, D. D. & Brewer, D. J. (1996) Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources, 32*(3), 505-520.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*, 3, 361-396.

Gustafsson, J. E. & Stahl, P. A. (1997). *STREAMS user's guide (Version 1.7).* Mölndal, Sweden: Multivariate Ware.

Guthrie, J. T. (1996). Growth of literacy engagement: Changes in motivations and strategies during concept-oriented reading instruction. *Reading Research Quarterly, 31*, 3, 306-332.

Guthrie, J. T., Schafer, W. D., Von Secker, C. & Alban, T. (2000). Contributions of instructional practices to reading achievement in a statewide improvement program. *Journal of Educational Research, 93,* 4, 211-225.

Guthrie, J. T., Schafer, W. D. & Huang, C.-W. (2001). Benefits of opportunity to read and balanced instruction on the NAEP. *Journal of Educational Research, 94,* 3, 145-162.

Hansen, J. & Pearson, P. D. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers. *Journal of Educational Psychology*, *75,* 6, 821-829.

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Research, 18,* 4, 45-51.

Hanushek, E. A. (1996a). A more complete picture of school resource policies. *Review of Educational Research, 66,* 3, 397-409.

Hanushek, E. A. (1996b). School resources and student performance. In G. T. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 43-73). Washington, D.C.: The Brookings Institution.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update*. Educational Evaluation and Policy Analysis, 19*, 2, 141-164.

Hedges, L. W. & Greenwald, R. (1996). Have times changed? The relation between school resources and student performance. In G. T. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success*. (pp. 74-92). Washington, D.C.: The Brookings Institution.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Research, 23*, 3, 5-14.

Kaniel, S., Licht, P., & Peled, B. (2000). The influence of metacognitive instruction of reading and writing strategies on positive transfer. *Gifted Education International, 15,* 1, 45-63.

Knapp, M. S. (1995).*Teaching for meaning in high-poverty classrooms*. New York, NY: Teachers College Press, Columbia University.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21*, 1, 29-45.

Mislevy, R. J. (1993). Should "multiple imputations" be treated as "multiple indicators"? *Psychometrika, 58,* 1, 79-85.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*(2), 125-145.

Mueller, M. E. (1997). Using metacognitive strategies to facilitate expository text mastery. *Educational Research Quarterly, 20,* 3, 41-65.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22,* 3, 399-420.

National Center for Education Statistics. (1996). *High school seniors' instructional experiences in science and mathematics.* Washington, D.C.: U.S Government Printing Office.

National Center for Education Statistics. (2001). *The nation's report card: Fourth-grade reading highlights*. Washington, DC.: U.S Government Printing Office.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Bethesda, MD: National Inst. of Child Health and Human Development (NIH),.

O'Reilly, P. E., Zelenak, C. A., Rogers, A. M., & Kline, D. L. (1996). *1994 trial state assessment program in reading secondary-use data files user guide.* Washington, D.C.: U.S. Department of Education.

Popplewell, S. R. & Doty, D. E. (2001). Classroom instruction and reading comprehension: A comparison of one basal reader approach and the four-blocks framework. *Reading Psychology, 22,* 2, 83-94.

Rosenshine, B. & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64,* 4, 479-530.

Snow, C. E., Ed.; Burns, M. Susan, Ed.; Griffin, Peg, Ed. (1998). *Preventing reading difficulties in young children*. Washington, D.C.: Committee on the Preservation of Reading Difficulties in Young Children.

Topping, K. J. & Paul, T. D. (1999). Computer-assisted assessment of practice at reading: A large scale survey using accelerated reader data. *Reading and Writing Quarterly, 15,* 3, 213-231.

Wenglinsky, H. (2001). The use of multilevel latent growth modeling to measure school effects. In R. Nata (Ed.). *Progress in Education*, 4 (pp. 167-195). Huntington, NY: Nova Science Publishers.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives, 10,* 12.

Wharton-McDonald, R., Pressley, M., & Hampton, J. M. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *Elementary School Journal, 99,* 2, 101-128.

About the Author

Harold Wenglinsky

Center for Educational Leadership
Baruch College School of Public Affairs

17 Lexington Ave, Box C-305
New York, NY 10010
Email: Harold_Wenglinsky@baruch.cuny.edu

Harold Wenglinsky is Associate Professor in the Baruch School of Public Affairs and Research Director of the Center for Educational Leadership. His research interests include the impact of teachers' instructional practices on student learning and maximizing the effectiveness of educational technology, about which he is currently completing a book.

The World Wide Web address for the *Education Policy Analysis Archives* is
epaa.asu.edu

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu