The consistency of class size effects: A meta-analytic approach

Spyros Konstantopoulos

Michigan State University

Background/context:
Class size reduction has been identified by some researchers as a promising school mechanism that can increase student achievement (Finn & Achilles, 1990; Krueger, 1999; Konstantopoulos, 2008). The effects of class size on student achievement have been of great interest to educational researchers and policy makers the last two decades. As a result, many states have introduced class-size reduction programs. California, for example, introduced a class size reduction program that provided financial incentives to schools that reduce class size in the early grades to twenty or fewer students per classroom. Wisconsin adopted a program that reduced class size to fifteen students per classroom in early grades in schools with high percentages of students from disadvantaged backgrounds. The best evidence about class size effects has been produced from Project STAR data and has suggested that on average smaller classes had positive effects on students' achievement in early grades (Krueger, 1999; Nye Hedges, & Konstantopoulos, 2000).

Purpose
Researchers have provided adequate evidence about the average effects of class size on student achievement. Although the estimation of the average treatment effect of class size reduction programs is important, the consistency of the treatment effect especially in large-scale experiments is also important. To date, the consistency of class size effects across schools has not been well discussed. Because class size effects can be computed for each school in large-sample studies one can evaluate the variability of class size effects across schools and understand how schools interact with the intervention and shape both the nature of the intervention as implemented and the effects expected. Significant variability of class size effects across schools would indicate differences in class size effects that are due to school context. The computation of class size effects for each school separately facilitates the construction of a distribution of class size effects across all schools in a sample. One can then use this distribution to identify the schools where the treatment was more (or less) successful and investigate school characteristics that contribute to the different degrees of success.

In this study I examined the consistency of the class size effects across schools using data from Project STAR. In particular, I computed the small class effect for each school and then used meta-analytic methods to compute the variability of the school-specific effects across schools. The analysis was conducted for each grade (i.e., kindergarten, first, second, or third) separately. In addition, because Project STAR intended to gauge the effects of having a full-time teacher aide in the classroom on student achievement, which represents the pupil teacher ratio in the classroom, I also examined the consistency of the effect of having a full-time aide in a regular classroom across all schools. Although the average effect of having a full time aide in the classroom has been shown to be small and non-significant (Nye et al., 2000), it is critical to examine whether the full time aide effect varies between schools and interacts with school context.

Setting/Design/Intervention/Sample:
Project STAR is a four-year large scale field experiment that involved students in seventy-nine elementary schools in forty-two districts in Tennessee. During the first year of the study, within each school, kindergarten students were assigned randomly to classrooms in one of three treatment conditions: smaller classes (with thirteen to seventeen students), larger classes (with twenty-two to twenty-six students), or larger classes with a full-time classroom aide. Teachers

were also assigned randomly to classes of different types. Some students entered the study in the first grade or subsequent grades, and were assigned randomly to classes at that time. Teachers at each subsequent grade level were also assigned randomly to classes as the experimental cohort passed through the grades. Districts had to agree to participate for four years and allow school visits for verification of class sizes, interviewing, and data collection, including extra student testing. They also had to allow research staff to assign pupils and teachers randomly to class types and to maintain the assignment of students to class types from kindergarten through grade three. Overall, more than 11,000 students participated in the experiment over the four-year period.

Project STAR has high internal validity because, within each school, students and teachers were assigned randomly to classes of different sizes. In addition, because Project STAR is a large-scale randomized experiment that includes a broad range of schools and districts (urban, rural, wealthy, and poor), it has higher external validity than smaller-scale studies. Moreover, the study was part of the everyday operation of the schools that participated and hence there is a lower likelihood that novelty effects affected the class size estimates.

Analysis:

Because random assignment in Project STAR was conducted within schools it is natural to compute class size effects within each school and then pool all estimates across schools to calculate an overall treatment effect. Therefore, meta-analysis is a natural method to analyze data from Project STAR. Conceptually Project STAR is a series of small-scale experiments that took place in each school throughout the State of Tennessee and therefore Project STAR data resemble meta-analytic data where each school contributes one class size effect. Treatment effects can be computed for each school separately, but since each school-specific estimate of class size effects is measured with different precision a weighted scheme is necessary to combine estimates and calculate the overall treatment effect across schools.

The computation of class size effects within each school is crucial because it adjusts for possible school effects or differences in achievement between schools (Kruger, 1999; Konstantopoulos, 2008). To compute class size effects within each school I used linear regression and regressed standardized mathematics or reading scores separately on two dummies that represent small class or regular class with a full-tile aide (regular class being the omitted category) and controlled for gender, race, and SES effects. Note that I computed intention to treat effects and not effects of class size as implemented or received because intention to treat effects are unbiased by design (see Friedman, 2006). The mean differences I computed for each school were in standard deviation units and indicated the standardized mean difference in achievement between small and regular or regular with full time aide and regular classes. Once the effect sizes for the class size effects were computed for each school I used mixed or random effects meta-analytic regression to combine the estimates (Konstantopoulos & Hedges, 2004). I used the inverse of the variance of each school-specific effect size as a weight in the weighted regression and I treated the school-specific estimates as random between schools. I employed proc mixed in SAS to analyze the data. The first model included only the intercept and therefore I computed the weighted mean across schools and the variance of the class size effects between schools. In subsequent models I used school characteristics as predictors to determine whether they explain variance between schools. In particular, I included in the regression equation school composition such as percent of minority and disadvantaged students in a school, percent of students who are present in school in a year, percent of teachers with graduate degrees and average teacher experience in each school, school urbanization such as urban, rural, or suburban school, school size per grade and number of classrooms per grade in each school. Finally, I also included in the model district fixed effects since it is plausible that districts may have contributed to the variability of the class size effects.

Findings / Results:

Small Class Effects

The range of small class effects across schools in each grade is presented in Table 1. In kindergarten the range was greater than 3 standard deviations in mathematics and nearly 3 standard deviations in reading. This suggests that the average student in a school that benefits the most from small classes is nearly 2 grades ahead than the average student in a school that benefits the least from small classes (see Hill, Bloom, Black, & Lipsey, 2008). Hill et al. estimated that the annual mathematics gain in kindergarten is nearly 1.3 standard deviations and the range of small class effects is more than twice as large. In reading the estimated annual gain in kindergarten is nearly 1.5 standard deviations and the range of small class effects is approximately twice as large. The range of small class effects in first grade was greater than 2.5 standard deviations in mathematics and 2 standard deviations in reading. In second grade the range of small class effects was greater than 2 standard deviations in mathematics and reading. Finally, in the third grade the range of small class effects was nearly 2 standard deviations in reading and smaller than 2 standard deviations in mathematics. Overall these results suggest that students in schools that benefit the most from small classes are at least 2 grades ahead in achievement than their peers in schools that benefit the least from small classes. This is not a trivial difference.

The results from the mixed effects meta-regression are reported in Table 2. In kindergarten the average small class benefit in mathematics was 0.19 standard deviations and significant, which suggests that across all schools students in small classes in kindergarten scored about one-fifth of a standard deviation higher than their peers in regular classes. The variance of the small class effects across schools was 0.21 and statistically significant. That is, in some schools the benefits of small class membership is more pronounced than in other schools. The average small class advantage in reading was 0.24 standard deviations and significant which suggests that across all schools students in small classes in kindergarten scored about one-fourth of a standard deviation higher than their peers in regular classes. The variability of the small class effect across schools was 0.21 and statistically significant which indicates a significant interaction between small classes and school context. In first grade the average small class benefit in mathematics was 0.28 standard deviations and significant. The variance of the small class effects across schools was 0.17 and statistically significant. The average small class advantage in reading was 0.25 standard deviations and significant. The variability of the small class effect across schools was 0.13 and statistically significant. In second grade the average small class benefit in mathematics was nearly 0.20 standard deviations and significant. The variance of the small class effects across schools was 0.18 and statistically significant. The average small class advantage in reading was 0.24 standard deviations and significant. The variability of the small class effect across schools was 0.11 and statistically significant. Finally, the average small class benefit in third grade mathematics was 0.16 standard deviations and significant. The variance of the small class effects across schools was 0.08 and statistically significant. The average small class advantage in reading was 0.23 standard deviations and significant. The variability of the small class effect across schools was 0.08 and statistically significant. Overall, the small class advantage across grades was one-fifth of a standard deviation or larger. The variance of small class effects was significant across schools and grades, however the variance estimates became smaller by the third grade.

Regular Class with Full Time Aide Effects
The range of regular class with full time aide effects across schools is presented in Table 3. In kindergarten the range was nearly 2.5 standard deviations in mathematics and reading. This suggests that the average student in a school that benefits the most from small classes is up to 2 grades ahead than the average student in a school that benefits the least from small classes (see Hill, Bloom, Black, & Lipsey, 2008). The range of regular class with full time aide effects in first grade was smaller than 1.5 standard deviations in mathematics and larger than 1.5 standard deviations in reading. This suggests that the average student in a school that benefits the most from small classes is at least one grade ahead than the average student in a school that benefits the least from small classes. In second grade the range of regular class with full time aide effects was greater than 1.5 standard deviations in mathematics and smaller than 1.5 standard deviations in reading. This suggests that the average student in a school that benefits the most from small classes is nearly 2 grades ahead in reading and one grade in mathematics than the average student in a school that benefits the least from small classes. Finally, in the third grade the range of regular class with full time aide effects was greater than 1.5 standard deviations in reading and mathematics. Again using the empirical benchmark by Hill et al. it appears that the average student in a school that benefits the most from small classes is nearly 3 grades ahead in mathematics and 4 to 5 grades ahead in reading and than the average student in a school that benefits the least from small classes. These differences are not trivial

The results from the mixed effects meta-regression are reported in Table 4. Across all grades the average full time aide effect was close to zero and statistically insignificant. That is, on average, reducing pupil teacher ratio in a classroom does not increase student achievement significantly or meaningfully. The estimates of the variance of the regular class with full time aide effects across schools were statistically significant only in kindergarten and second grade. In these two grades there was significant interaction between full time aide affects and school context. In other grades the between-school variance was not significant.

In order to identify the kinds of school characteristics that may be responsible for the inconsistency of the treatment effects I also used a meta-analytic regression that included several observed school characteristics. The results suggested that in kindergarten mathematics school characteristics and district effects explained 13 percent of the between-school variance of the small class effect. District effects were responsible for 10 of the 13 percent of the variance explained. School characteristics and district effects explained 42 percent of the full time aide effect across schools and district effects were responsible for 22 percent of the 42 percent. In kindergarten reading however, the school characteristics and district effects did not explain any between-school variance of the small class or regular class with full time aide effect. In grades 1 through 3 school characteristics and district effects did not explain any variance in mathematics or reading for small class or regular class with full time aid effects. These results indicate that the class size and the full time aide effects are more school dependent in mathematics than in reading, but only in kindergarten. Nonetheless, the remaining inconsistency of the class size effects was still significant at the .05 level in some grades. Most of the variability in the effects is unexplained and therefore it seems that both in reading and mathematics it is the unobserved school characteristics that are responsible for the inconsistency of the class size effects.

Conclusions:
This study examined the consistency of class size effects through third grade using data from Project STAR and meta-analytic methods. The findings provide additional support to the notion that the average small class effect is significant and important in early grades. In addition, the small class effects vary significantly across schools in all grades for both mathematics and reading. The inconsistency is larger in kindergarten and becomes smaller as students move through grades. Therefore, there is evidence that school context interacts with small class effects. The significant variation of small class effects across schools indicates that the treatment has low external validity or generality and does not scale up across the schools in the sample. The small class effect is inconsistent and it is positive and significant in some schools and negative and significant in others. The regular class full time effect was overall small and non-significant showing that on average decreasing pupil teacher ratio in the classroom does not effect student achievement positively. However, the full time aide effects vary significantly across schools in kindergarten and second grade. This suggests that school context interacts with full time aide effects and that reducing pupil teacher ratio is beneficial in some schools but not in others. Overall, it appears that some schools know how to make use of small classes or the full time aide more effectively than other schools. The schools that benefit most from small class or the full time teacher aide effects can be identified in Project STAR. Ideally, the next step would be to study these schools and determine the specific factors that helped maximize the benefit. In the same vein one could study the schools that benefited the least from the class size effects and identify the factors that hindered the success of the treatment. That process would help eventually with reconsidering the nature of the intervention and its implementation. Unfortunately such school data are not available from Project STAR. A new large-scale experiment would give us the opportunity to study such schools and understand how the class size mechanism is enacted.

References

Finn, J D., & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27,* 557-577.

Freedman, D. A. (2006). Statistical models for causation. What inferential leverage do They provide? *Evaluation Review, 30,* 691-713.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2,* 172-177..

Konstantopoulos, S. (2008). Do small classes reduce the achievement gap between low and high achievers? Evidence from Project STAR. *Elementary School Journal, 108,* 275-291.

Konstantopoulos, S., & Hedges, L.V. (2004). Meta-Analysis. In D. Kaplan (Ed.), *Handbook of Quantitative Methodology for the Social Sciences* (pp. 281-297). New York: Sage.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics, 114,* 497-532.

Nye, B., Hedges, L.V., & Konstantopoulos, S. (2000). Effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37,* 123-151.

Tables

Table 1. Range of Small Class Effects

	Minimum	Maximum
	Mathematics	
Kindergarten	-1.52	1.66
First Grade	-1.31	1.45
Second Grade	-0.97	1.26
Third Grade	-0.84	0.94
	Reading	
Kindergarten	-1.17	1.83
First Grade	-0.97	1.27
Second Grade	-0.99	1.23
Third Grade	-0.70	1.39

Table 2. Estimates of Average Small Class Effect and its Variability Across Schools

	Coefficient	SE	Variance	SE
		Mathematics		
Smalll Class in Kindergarten	0.190*	0.061	0.207*	0.045
Smalll Class in First Grade	0.280*	0.056	0.166*	0.039
Smalll Class in Second Grade	0.195*	0.060	0.181*	0.044
Smalll Class in Third Grade	0.158*	0.047	0.084*	0.028
		Reading		
Smalll Class in Kindergarten	0.241*	0.061	0.209*	0.046
Smalll Class in First Grade	0.247*	0.053	0.134*	0.033
Smalll Class in Second Grade	0.235*	0.051	0.113*	0.032
Smalll Class in Third Grade	0.227*	0.045	0.079*	0.027

* $p < 0.05$

Table 3. Range of Regular Class Full Time Aide Effects

	Minimum	Maximum
	Mathematics	
Kindergarten	-1.47	1.09
First Grade	-0.56	0.74
Second Grade	-0.97	0.88
Third Grade	-0.84	0.80
	Reading	
Kindergarten	-0.99	1.64
First Grade	-0.63	1.03
Second Grade	-0.59	0.70
Third Grade	-0.54	1.10

Table 4. Estimates of Average Regular Class Full Time Aide Effect and its Variability Across Schools

	Coefficient	SE	Variance	SE
		Mathematics		
Regular Aide Class in Kindergarten	0.022	0.061	0.214*	0.046
Regular Aide Class in First Grade	0.053	0.031	0.011	0.011
Regular Aide Class in Second Grade	0.031	0.039	0.045*	0.019
Regular Aide Class in Third Grade	-0.019	0.035	0.024	0.015
		Reading		
Regular Aide Class in Kindergarten	0.055	0.055	0.160*	0.037
Regular Aide Class in First Grade	0.056	0.033	0.021	0.013
Regular Aide Class in Second Grade	0.045	0.036	0.031*	0.015
Regular Aide Class in Third Grade	0.026	0.031	0.005	0.012

* $p < 0.05$