A Meta-Analysis of Supplemental Educational Services in Ohio: Implications for the Reauthorization of ESEA

Jane A. Beese
Youngstown State University

Abstract

The Supplemental Educational Services (SES) policy, under No Child Left Behind, requires underperforming Title I schools to offer academic tutoring to eligible students. A meta-analysis of findings from 44 external evaluations of district operated SES programs in Ohio showed that scores increased on the academic performance of at-risk students. When compared with an average annual gain in effect size from nationally normed tests, SES participants fell 0.69 standard deviations below. The subgroup analysis demonstrated greater effects in students who received services in reading and students in elementary and middle schools.

Keywords: Supplemental educational services (SES), No Child Left Behind (NCLB), Meta-analysis, subgroup analysis

Introduction

Prior to 2002, Title I programs required certain operational conditions and limited spending, but had no quality requirements. In fact, the federal government that provided the financial support for these programs did not audit the spending to ensure that the money was being used on compensatory instruction. The passage of NCLB changed all of that. For the first time, schools lost total autonomy when it came to expending Title I money. Specifically, if students in Title I schools were not meeting their state's educational standards, then the school had to make afterschool tutoring available to those students (using the Title I funds for this required expenditure). The Supplemental Educational Services (SES) policy was placed into the No Child Left Behind (NCLB) legislation as an option for parents and their children who were trapped in failing schools. The SES provision requires that perennially underperforming Title I schools that fail to meet adequate yearly progress (AYP) benchmarks for three consecutive years offer free academic tutoring to eligible students. The SES school option became a new legal requirement as of 2002 and by law must be monitored.

The U.S. Department of Education (2009) defines SES as "additional academic instruction designed to increase the academic achievement of students in schools in the second year of improvement, corrective action, or restructuring" and stipulates that these services must be offered outside the regular school day (p. 1). The SES school option has afforded students in failing schools access to tutors where they can get assistance in reading and math. Students who come from low-income families attending Title I schools, whether or not they performed at levels of proficiency, are eligible for SES. Each year, more than half a million children participate in SES (Ed.gov, 2013).

More than a decade later, the effects of SES programs on student academic achievement remain unknown. Legislation that will determine the fate of SES programs is currently pending. In order to make good decisions for our schools and children, we must fully understand the effects of the educational initiatives we have instituted before they are reformed yet again or abandoned. The first section of this paper recounts research on the effects of tutoring programs on student achievement will be reviewed. Next, a meta-analysis is used investigate the relationship between SES and student achievement gains. The final section considers findings within the context of empirical norms for student growth and suggestions for future research.

Review of the Literature

Past research provides evidence that tutoring programs can have positive effects on student performance. A meta-analysis by Cohen, Kulik, and Kulik (1982) found tutoring programs had a small effect on student performance in reading (0.21) and a large effect in math (0.60). A meta-analysis of volunteer tutoring programs conducted by Ritter, Barnett, Denny, and Albin (2009) found similar effect sizes for tutoring programs in reading (0.30) and math (0.27). These findings may have served as the theoretical framework for the SES policy.

Since the administration of SES, additional findings have been reported on the effects of tutoring. Berger et al. (2011) used a student fixed-effect approach that controlled for time-invariant student characteristics to study student achievement gains associated with SES participation relative to nonparticipation. The results indicated whether students experienced statistically significant achievement gains during periods of SES participation compared with periods of nonparticipation. Standardized z-scores were used to represent the differences in students' annual achievement gains. The effect between SES participation and achievement gains was statistically significant for mathematics (0.08 standard deviations) and reading (0.04 standard deviations), relative to nonparticipation. A Chicago Public Schools' (CPS) study of SES programs concluded that students receiving a minimum of 40 hours of tutoring had larger gains in reading and math skills than students who did not receive services (Ryan & Fatani, 2005). Findings were substantiated in research conducted in another CPS study (2007) and by Harding, Harrison-Jones, and Rebach (2012).

The effect of SES programs on student achievement may differ from other tutoring programs. SES is funded through the redirection of Title I monies. NCLB requires districts to set aside 20% of their Title I allocation for school choice transportation and SES. Schools are required to spend a minimum of 5% of the total set aside on SES. Schools may be resistant to spending these monies on such instruction because of the redirection of funds from already existing programs and the uncompensated increase in administrative work. Because SES is imposed on the district by federal requirements without the commitment of the schools, some of the conditions and circumstances of the programs could affect the delivery of such tutoring. Research on tutoring programs tends to show positive benefits when qualified tutors provide tutoring services in one to one or small group settings for a minimum of 40 hours (Morris, Tyner, & Perney, 2000; Ryan & Fattani, 2005; Slavin, Lake, Davis, & Madden, 2011).

Several researchers have suggested the effectiveness of tutoring programs might vary by grade levels and subject area. Cooper, Charlton, Valentine, and Muhlenbruck (2000) studied the effects of summer school and documented greater benefits for students in early elementary and secondary grades. Research by Lou et al. (1996), Erlbaum, Vaughn, Hughes, and Moody (2000), and Lauer et al. (2006) observed the largest effects in elementary and high school students. A study by Chappell, Nunnery, Pribesh, and Hager (2010) investigated 400 provider effects related to student achievement in a sample of 140,000 students and found a small effect in reading (0.17) and no effect in math. The purpose of this study is to provide a meta-analysis of a series of unpublished evaluations of tutoring programs in Ohio. The data-based results can add to the body of knowledge on educational reform to better understand the effectiveness of SES programs on student achievement in reading and math and provide information for future policy.

Method

With any policy change, it is important to monitor the impact of change and to evaluate the value of new programs. To determine the effects of SES programs on student achievement in math and reading, the researcher performed a meta-analysis. Meta-analysis is a statistical synthesis of the results from individual studies on the same topic that is used to calculate a summary of the effects of a program or policy of interest (Cohen, 1988; Glass, 1976; Lipsey & Wilson, 2001). The benefits of meta-analysis include improved precision and greater power (Deeks, Higgins, & Altman, 2008, p. 3). Where individual studies may be too small to detect statistical differences, the chance of detecting an effect increases when studies are combined, and the accuracy of the evaluation of an intervention is enhanced when based on more information. Meta-analysis provides a means to investigate "why the effects may be larger in some studies and smaller in others" (Denson & Seltzer, 2011, p. 3).

The researcher was part of an ongoing statewide evaluation project that investigated SES programs in Ohio. From 2006 to 2012, the Ohio Department of Education randomly selected and assigned district providers for evaluation. The providers represented in the evaluations varied in student to tutor ratio, number of hours tutored, instructional methods, delivery, years of operation, and structure but were considered representative of the programs in Ohio. Although the study population was limited to district and school providers, researchers have found few demographic or academic differences between students served by district providers and students served by private tutoring services (Berger et al., 2011). Research has found similar achievement gains associated with district providers and outside providers (Berger et al., 2011). National studies have characterized SES students as coming from low-income families, high poverty schools, and within the lower rankings on state assessments (U.S. Department of Education, 2005). This analysis is limited to evaluations of programs in Ohio and therefore limits the range of generalizations that can be made.

Data Collection

Full-text copies of the Ohio SES program evaluations that emphasized systematic data collection procedures and sound data analysis practices were obtained. Each year, the Ohio Department of Education randomly assigned SES providers for evaluation. The meta-analysis was limited to evaluations with single-subject designs (one-group pre-test post-test without random assignment). The researchers extracted data about student performance and added a comparison of reading and mathematics and a comparison between grade levels. The data included provider's names, evaluation year, number of SES participants, subject area tutored, and the statistical data needed for the meta-analysis. There were 4,408 students who participated in SES that constitute the study population (see Appendix A for a list of evaluations and sample sizes). SES providers were required to administer a pre- and post-assessment aligned with state content standards that were used to evaluate program effectiveness. The data that came from the administration of those assessments were used in the meta-analytic procedures and included means, standard deviations, *t*-test statistics, and effect sizes. For the evaluations that reported means and standard deviations, the effect size was calculated. For less reported studies, the effect size was calculated from summary statistics *t*-tests (Glass, 1976; Hedges & Olkin, 1985).

An effect size provides a standardized measure of the magnitude of study outcomes and allows us to compare effect sizes across studies that have used different scales of measurement (Field, 2005; Hattie, 2009). Hedges's *g* was selected as the primary index because it corrects for bias due to small sample size (Hedges & Olkin, 1985). This study examines change within students, and the observed effects will be larger than between group studies because individual differences are held constant. Therefore, Cohen's magnitude of effect is not labeled. The Comprehensive Meta-Analysis version 2.0 statistical software program was used to compute statistics.

Results

In this section we first describe the effects of SES on the children who received tutoring and the effects of tutoring programs on student achievement related to the subject area children were tutored and compared effect sizes between grade levels. In 43 of the 45 evaluations, the examination of the performance of students who received SES had improved when compared to their pretreatment performance. In two studies, there was a negative impact post treatment. Of the evaluations reported, 36 demonstrated statistically significant differences following participation in SES. The evaluations included no control groups or comparison groups; in other words, the "effects" that are being studied are the gains from pre- to post-assessment. This is a serious limitation because there is no way to determine whether the effects were due to the tutoring or to other efforts in the schools. Therefore, the discussion will include comparisons to empirical benchmarks (Hill, Bloom, Black, & Lipsey, 2008). The primary goals were to determine whether SES is an effective intervention in increasing student achievement and also to determine whether the effect differs among subjects tutored and by grade level.

Summary Effect

Random effects model. Based on the random effects model, the relationship between SES and student achievement had an effect size of 0.81 and the standard error of effect size was 0.07. The 95% confidence interval lower boundary is 0.67 and the upper boundary is 0.95. The null is not included in the parameters of the confidence interval, and the p value is < 0.001, which is statistically significant and indicates SES treatment has a true effect on student achievement.

Test of heterogeneity. Heterogeneity was examined to determine whether or not the dispersion of effects was due to random sampling error or to differences in effect size. There was some expectation that the true effects of SES treatment would vary due to the difference in program delivery. The p value is < 0.001. Therefore, we reject the null hypothesis that all studies share a common effect size and accept the alternative that the true effect is not the same and can be attributed to real differences in effect size from one evaluation to the next.

The results were significant. This indicates that 8% of the variance is due to random sampling error, and 92% of the variance is because of the true differences from study to study. Tau, the standard deviation of the true effect sizes, is 0.42 (Borenstein, Hedges, Higgins, & Rothstein, 2009). If we replace the observed effect sizes with the true effect sizes, the standard deviation of the true effect sizes will be 0.42 and the variance of those effect sizes would be Tau-squared, which is 0.18. The standard error and variance of Tau-squared tells us the accuracy of the estimate of Tau-squared. The standard error = 0.06 and the variance = 0.00. Residual weights were investigated, and none of the studies fell outside the parameters of 1.96. This indicates that the effects found in the individual studies are similar and can provide a combined estimate.

Fail-safe N. The Fail-safe *N* was calculated to assess the potential for publication bias. The Fail-safe N was 137 studies. Therefore, there would have to be 137 additional negative studies to render a significant meta-analysis insignificant. Because that number exceeds the critical number, no file drawer problem exists. It must be noted that none of the samples in this study have been published and in effect this study is the Fail-safe N.

Subgroup Analyses

To assess the relationship between study-level covariates and effect size, a meta-analysis was used to compare the mean effect for reading versus the math subgroups of studies and the summary effect versus grade levels. An effect was computed for each group to determine if effect size was related to the subject area or grade level. A random effects model was selected because variation among the studies was established. The true effect probably varies from study to study because among the studies differences exist in the instructional methods used in tutoring students, the level of expertise of the teachers, the details of the protocol, or other factors. The data were pooled to yield the within group estimates of T^2 and this common estimate was applied to all studies. According to Borenstein et al. (2009), "the increased accuracy that we get by pooling more studies is likely to exceed any real differences between groups in the true value of T^2" (p. 163). The formula for a random effects model with a pooled estimate of T^2 was used (Borenstein et al., 2009, p. 179).

Math and reading subgroups. There were 28 reading evaluations and 17 math evaluations. Based on the random effects model, at 95% significance level, the effect size for math was 0.68, and the standard error of effect size was 0.11 with a confidence interval of 0.46 to 0.90. The effect size for reading was 0.90, and the standard error of effect size was 0.09 with a confidence interval of 0.72 to 1.07. The variances are 0.01 for both subgroups. The difference between the two group's effect size is 0.22. These results indicate that SES treatment had a large effect on reading achievement and a medium effect on math achievement.

Comparing the effects. A *Z*-test was used to compare the two mean effects. For the random effects analysis, the two-tailed *p* value corresponding to $Z_{\text{Diff}} = 7.50$ is 0.001 and the *p* value for $Q_{between} = 2.23$ with $df = 1$ is < 0.13. These numbers tell us that the mean effect is not the same for the math studies as for the reading studies. There are true differences in effect between the math and reading studies.

The I^2 statistic for math is 91.50 with a *p* value < 0.001. Approximately 9% of the variance was within math studies, and 91% of the variance was between studies. The I^2 statistic for reading is 92.35 with a *p* value < 0.001. Approximately 7% of the variance was within reading studies, and 93% of the variance was between studies. Tau is 0.44 for math and 0.42 for reading. Tau^2 is 0.19 for math and 0.17 for reading. The standard error is 0.10 for math and 0.08 for reading with a variance of 0.01 for both. These results indicate that the effects found in the reading studies. and the effects found in the math studies are similar and can provide combined estimates for each group.

The statistical analysis for the student change appears larger in reading than in math. Therefore, the treatment effect may be higher for reading than in math. Based on the random effects model, the average effect size value related to the SES treatment effect on reading was 0.90, and the average effect size related to math was 0.68. An effect size of 0.90 in reading means that the mean score of students who participated in an SES program ranked in the 82nd percentile and an effect size of 0.68 in math means that the mean score of participants ranked in the 75th percentile. Students tutored in reading scored, on average, 0.22 standard deviations higher than those tutored in math.

Grade levels. Grade level information was available for 34 evaluations. There were 24 elementary, 8 middle school, and 2 high school evaluations. The evaluations were divided into three categories: elementary (K-5), middle (6-8), and high school (9-12). There was a large discrepancy in the sample sizes between the grade levels elementary 3,404, middle 189, high school 27.

Based on the random effects model, at 95% significance level, the effect size for elementary school was 1.92, the standard error was 0.17, and the variance was 0.03 with a confidence interval of 1.62 to 2.30; the effect size for middle school was 1.25, the standard error was 0.33, and the variance was 0.11 with a confidence interval of 0.61 to 1.92; the effect size for high school was 0.19, the standard error was 0.70, and the variance was 0.48 with a confidence interval of -1.19 to 1.58. The difference between the elementary and middle school groups effect size is 0.67. The findings for high school were not statistically significant. The statistical test was limited by the small sample size, n=2, (where a larger sample size would have ensured a more representative distribution of the population) and may have contributed to a Type II error.

Comparing the effects. A *Z*-test was used to compare the mean effects for the three groups. For the random effects model, the two-tailed *p* value corresponding to $Z_{\text{Diff}} = 11.48$ is 0.001 and the *p* value for $Q_{between} = 8.14$ with $df = 2$ is less than 0.02. This tells us that the mean effect is not the same between grade levels.

The I^2 statistic for elementary school is 89.02 with a *p* value < 0.001. Approximately 11% of the variance was within elementary studies, and 89% of the variance was between studies. The I^2 statistic for middle school is 76.84 with a *p* value <

0.001. Approximately 23% of the variance was within middle school studies, and 77% of the variance was between studies. Tau is 0.70 for elementary and 0.96 for middle school. Tau^2 is 0.50 for elementary and 0.95 for middle school. The standard error is 0.23 for elementary and 0.71 for middle with a variance of 0.05 for elementary and 0.50 for middle school.

Based on the random effects model, the average effect size related to elementary school was 1.88 and for middle school was 0.94. This indicates a large effect on student achievement at the elementary and middle school levels. SES had little to no effect on student performance at the high school level. The findings of this research are consistent with Lauer et al. (2006) and Grossman, Walker, and Raley (2001). An effect size of 1.88 for elementary students means that the score of the average student who participated in an SES program ranked in the 97th percentile and an effect size of 0.94 for middle school students means that the score of the average student ranked in the 83rd percentile.

Overall meta-analyses revealed that SES treatment had a true effect on student achievement. A large effect on student achievement in reading and a medium effect on student achievement in math was found. SES treatment on elementary, middle, and high school students resulted in large effects for both elementary and middle school students. No effect on high school students was found.

Discussion

The results of this study may suggest several implications related to SES outcomes and future policy. SES programs may have positive effects on reading and math achievement in at-risk students. According to the random effects model, the relationship between SES and student achievement has an effect size of 0.81. The findings of this research are consistent with the results of studies on the effects of tutoring including Cohen et al. (1982), Lauer et al. (2006), and studies related to SES including Borman, Rachuba, Fairchild, & Kaplan (2003), Ryan and Fatani (2005), CPS (2007), Berger et al. (2011), and Harding et al. (2012). However, without a control or comparison group, the effects of normal academic growth and school efforts are not controlled for and therefore the effects are inflated. While SES treatment appears to have a substantial effect on student achievement, to be certain, we must eliminate the normal growth that would occur during the academic year for this population of students.

Published benchmarks from Bloom, Hill, Black, and Lipsey's *Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions* were used to understand the effects of SES on student growth compared with the natural growth in academic achievement (2008). Comparisons show average annual growth tends to be similar across subject areas with a range of 0.02 to 0.38 difference between reading and math. The SES treatment effects appear higher than average annual growth for transition grade 2-3 with an effect size of 0.60 through transition grade 11-12 with an effect size of 0.06 in reading and for transition grade 3-4 with an effect size of 0.52 through transition grade 11-12 with an effect size of 0.01 in math (Bloom et al., 2008, p.16). Bloom et al. supports the findings of this study where higher SES treatment effects were found in reading than in math with a difference of 0.22 between subjects. The research-based learning strategies and instructional practices in reading may be a contributing factor in the effectiveness of SES programming. More research may be needed on effective instructional practices in math tutoring.

Statistical analysis indicates that the average student change is larger for elementary school than for middle school. Students tutored in the elementary grades, on average, scored 0.94 standard deviations higher than those tutored in middle school. Annual gains on standardized achievement tests vary substantially across grades with larger annual gains in the early elementary grades followed by gradually declining gains in later grades (Bloom et al., 2008). The effect of SES on student achievement declines as students move up in grade level however, it is important to interpret an intervention's effect within the context of expectations for the grades being treated (Bloom et al, 2008). Consequently, comparisons were made between the SES treatment effects by grade level and published average annual growth trends. For example, the SES treatment effect for elementary students (*es* = 1.88) relative to normal academic growth for the grade 1-2 transition in math (*es* = 1.03) compared to the effect for middle school students (*es* = 0.94) relative to normal academic growth for the grade 7-8 transition in math (*es* = 0.32) represents proportionally similar improvements (*es* = 0.85 compared to (*es* = 0.62) (Bloom et al., 2008). These findings demonstrate similar growth between average students without SES treatment and at-risk students with SES treatment. SES programs might be one way to support academic growth in at-risk students and assist efforts to close the achievement gap.

A limitation of this study was that data did not distinguish between regular academic growth and SES treatment. Hill, Bloom, Black, and Lipsey (2008) demonstrated that the average student change on standardized achievement tests from kindergarten to the first grade was about 1.5 standard deviation units in the absence of an intervention. When compared with the summary effect of the meta-analysis (*es* = 0.81) SES participants fell 0.69 standard deviations below the average annual gain in effect size from nationally normed tests. While it might be sensible to expect less change among at-risk students, it is difficult to quantify how much less, and also difficult to figure out how this mean might be affected by the highly aligned nature of tests in this case.

Schools with higher concentrations of at-risk students typically display poorer academic performance than schools with fewer numbers of these students (U.S. Department of Education, 2006). Students "at-risk" refers to the percentages of minority, economically disadvantaged, and English Language Leaner (ELL) students within a school (Kannapel & Clements, 2005). To fully understand the effects of SES on student outcomes, we must understand how these factors impact average school performance. Bloom et al. (2008) reasoned "researchers should tailor their effect size benchmarks to the contexts they are studying whenever possible" (p. 19). The achievement gap between minority and non-minority students is a widely recognized area for concern in the United States. On average, black fourth graders score 0.83 standard deviation lower in reading and 0.99 standard deviation lower in math than white fourth graders, with the difference decreasing slightly as students move up in grade level. A similar pattern exists between Hispanic and White students. The SES treatment effect for reading was 0.90 and signified a substantive change relative to the academic gap in effect size estimates between minority and non-minority students. The SES treatment effect for math was 0.68 and constituted a smaller change relative to the minority achievement gap. These comparisons support substantial growth in reading achievement for students who participate in SES program when compared with the performance trajectories for at-risk students who did not participate in the SES program.

Socio-economic status is the most widely recognized indicator of student risk. Researchers have continued to report large discrepancies between the achievement of high and low poverty students (Murnane, Willet, Bub, & McCartney, 2006; Reardon & Robinson, 2007). On average students eligible for free or reduced-price lunch score 0.74 standard deviation lower in reading and 0.85 in math than fourth grade students who are not eligible, with the difference decreasing slightly as students move up in grade level (Bloom et al, 2008). The effect size for SES was 0.90 for reading and 0.68 in math; both imply a substantive change relative to the academic gap in effect size estimates between students who are and are not eligible for a free or reduced-price lunch.

This study addressed significant issues about policy toward tutoring programs that can help to improve learning in at-risk populations. Without a comparison group, we can only infer what the actual effectiveness of SES might be, but we cannot be certain. Gains may be attributed to other factors, such as new technology, professional development, new curriculum, or changes in instructional practice. Chatterji, Kwon, and Sng (2006) argued that the program effects of SES are observed only in assessments that are aligned with the SES curriculum and that these effects are confounded by other efforts simultaneously being implemented (p. 30). Many researchers debate the overall effect of SES on student achievement declaring it negligible compared to other methods of reform (Chappell et al., 2010; Hattie, 2009). There is a continuing need to implement SES programs while further research is used to determine its effectiveness.

The relative effectiveness of these programs under many conditions or across many features should be considered. While this study looked at math and reading differences as well as grade level, it did not consider the size of the schools or districts, the levels of initial achievement performance for the districts, or racial/ethnic or socioeconomic status differences. Investigating specific variables associated with SES effectiveness may pinpoint viable strategies that can be introduced into the regular classroom and serve to inform the design and development of future programs and policy.

Legislation proposes the elimination of SES for students in failing public schools. The Department of Education’s blueprint for the reauthorization of ESEA, released in March 2010, recommends that chronically low-performing schools should no longer be required to fund SES but instead should be required to implement “data-driven interventions,” which could include “expanded learning time, supplemental education services, public school choice, or other strategies” (U.S. Department of Education, 2010, p. 10). Educators should address policy makers about the potential use of tutoring programs for increased student achievement.

We find ourselves on the brink of yet another phase of reform. But what lessons can we take away from the efforts educators toiled over under NCLB? While the data may suggest SES programs increase reading and math achievement in at-risk students when compared with normative expectations for academic growth in economically disadvantaged and minority subgroups, we cannot be certain without a control group. The effect of SES on reading was greater than in math so instructional strategies for math may be a topic for enhancement. The magnitude of the effect is still up for debate. Finally, evaluation requirements for educational programs should be more robust and include measures for comparison.

References

Berger, A., deSousa, J., Hoshen, G., Lampron, S., Le Flock, K. C., Petroccia, M., & Shkolnik, J. (2011). *Supplemental educational services and student achievement in five waiver districts.* Office of Planning, Evaluation and Policy Development, U.S. Department of Education. [ED518217].

Black, A. R., Doolittle, F., Zhu, P., Unterman, R., Grossman, J. B., & Warner, E. (2008). The evaluation of enhanced academic instruction in after-school programs: Findings after the first year of implementation. *U.S. Department of Education National Center for Education Evaluation Report,* 2008-4021.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *MDRC Working Papers on Research Methodology,* 1-51.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* West Sussex, United Kingdom: John Wiley.

Borman, G. D., Rachuba, L. T., Fairchild, R., & Kaplan, J. (2002). *Randomized evaluation of a multiyear summer program: Teach Baltimore*: *Year 3 report*. Retrieved from http://www.childcareresearch.org/location/3333

Burch, P. (2007). *Supplemental educational services under NCLB: Emerging evidence and policy issues.* Boulder: Educational Policy Research Unit. University of Colorado.

Chappell, S., Nunnery, J., Pribesh, S., & Hager, J. (2010). Supplemental educational services (SES) provision of no child left behind: A synthesis of provider effects. *Journal of Education for Students Placed at Risk*, *16*(1), 1-23. [EJ917516].

Chatterji, M., Kwon, Y. A., & Sng, C. (2006). Gathering evidence on an after-school supplemental instruction program: Design challenges and early findings in light of NCLB. *Education Policy Analysis Archives, 14*(12), 1-44.

Chicago Public Schools. (2007). *The 2007 supplemental educational services program: Year 4 summative evaluation.* Chicago: Chicago Public Schools Office of Extended Learning Opportunities, Research, Evaluation and Accountability.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*(2), 237-248.

Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs for the Society for Research in Child Development, 65*(1), 1-118.

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (Eds.). Chapter 9: Analysing data and undertaking meta-analysis. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions.* Version 5.0.1. [updated September 2008]. The Cochrane Collaboration, 2008. Retrieved from www.cochrane-handbook.org

Deke, J., Dragoset, L., & Bogen, K. Gill, B. (2012). *Impacts of title I supplemental educational services on student achievement.* Institute of Education Sciences

National Center for Education Evaluation and Regional Assistance. U.S. Department of Education. 1-12.
Denson, N., & Seltzer, M. (2011). Meta-analysis in higher education: An illustrative example using hierarchical linear modeling. *Research in Higher Education, 52,* 215-244. doi:10.1007/s11162-010-9196-x.
ED.gov. (2013). *Ed data express: Data about elementary & secondary schools in the U.S.* Retrieved from http://www.eddataexpress.ed.gov/
Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology, 92*(4), 605-619.
Field, A. P. (2005). Meta-analysis. In J. Miles & P. Gilbert (Eds.), *A handbook of research methods in clinical and health psychology* (pp. 295-308). Oxford, UK: Oxford University Press.
Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Research, 5*(10), 3-8.
Grossman, J. B., Walker, K., & Raley, R. (2001, April). *Challenges and opportunities in after-school programs: Lessons for policymakers and funders.* Philadelphia, PA: Public/Private Ventures.
Harding, H., Harrison-Jones, L., & Rebach, H. (2012). A study of the effectiveness of supplemental educational services for title I students in Baltimore city public schools. *The Journal of Negro Education, 81*(1), 52-66.
Hattie, J. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* London, UK: Routledge.
Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.
Heinrich, C. J., Meyer, R. H., & Whitten, G. (2010). Supplemental education service under No Child Left Behind: Who signs up, and what do they gain? *Educational Evaluation and Policy Analysis*, *32*(2), 273-298.
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(3), 172-177.
Kannapel, P. J. & Clements, S. K. (2005). Inside the black box of high-performing high-poverty schools: A report from the Prichard Committee for Academic Excellence. *Prichard Committee for Academic Excellence*; Lexington, KY.
Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research, 76*(2), 275-313.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.
Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*, 423-458.

Munoz, M., Chang, F., & Ross, S. H. (2011). *No child left behind and tutoring in reading and mathematics: Impact of supplemental educational services on large scale assessment.* Paper presented at the annual meeting of the Consortium for Research in Educational Accountability and Teacher Evaluation 2011 National Evaluation Institute.

Murnane, R. J., Willet, J. B., Bub, K. L., & McCartney, K. (2006). Understanding trends in the black-white achievement gaps during the first years of school. *Brookings-Wharton Paper on Urban Affairs.*

Reardon, S. F. & Robinson, J. (2007). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. In Ladd, H. & Fiske, E. (Ed.). *Handbook of research in education finance and policy*. New York, NY: Prentice Hall.

Ritter, G. W., Barnett, J. H., Denny, G. S., & Albin, G. R. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A meta-analysis. *Review of Educational Research, 79*(1), 3-38.

Ryan, S., & Fatani, S. (2005). *SES tutoring programs: An evaluation of the second year—Part one of a two part report* (Policy report). Chicago: Office of Research, Evaluation and Accountability, Chicago Public Schools.

U.S. Department of Education. (2005). *No Child Left Behind: Supplemental educational services non-regulatory guidance.* Washington, DC: Department of Education.

U.S. Department of Education. (2006). The condition of education 2006. *National Center for Education Statistics.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Education. (2009). *Supplemental educational services: Non-regulatory guidance.* Washington, DC: Author. Retrieved from http://www2.ed.gov/policy/elsec/guid/suppsvcsguid.doc

U.S. Department of Education. (2010). *A blueprint for reform: The reauthorization of the elementary and secondary education act*. Washington, DC: Author. Retrieved from http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf

Appendix

Study	*N*	Subject	Grade	t	Dir. of effect	*p*	ES
Toledo (2007) [1]	358	Reading	Elementary	19.455	+	0.00	1.03
Toledo (2008) [1]	353	Reading	Elementary	24.150	+	0.00	1.29
Toledo (2009) [1]	664	Reading	Elementary	34.249	+	0.00	1.33
Akron Digital Acad (2009) [1]	11	Reading	Middle	1.453	+	0.17	0.44
International Acad. (2009) [1]	73	Reading	--	8.096	+	0.00	0.95
Ravenna (2010)	92	Reading	Elementary	14.370	+	0.00	1.50
Akron Digital Acad. (2010) [1]	9	Reading	--	2.075	+	0.06	0.69
Columbus (2010)	150	Reading	Middle	4.428	+	0.00	0.36
Mt. Healthy (2010)	26	Reading	Elementary	4.821	+	0.00	0.95
Timberline (2010)	17	Reading	Elementary	3.645	+	0.00	0.88
Cincinnati (2010) [1]	17	Reading	--	4.546	+	0.00	1.10
Springfield (2010)	69	Reading	--	1.520	+	0.13	0.18
Toledo (2010)	202	Reading	Elementary	13.579	+	0.00	0.96
Canal Winchester (2011)	10	Reading	Elementary	6.042	+	0.00	1.91
Cincinnati (2011)	62	Reading	Elementary	2.305	+	0.02	0.29
Columbus (2011)	72	Reading	Middle	6.299	+	0.00	0.74
Elida (2011)	5	Reading	Elementary	4.180	+	0.01	1.87
Hamilton (2011) [1]	20	Reading	Elementary	4.750	+	0.00	1.06
Maple Hts. (2011) [1]	20	Reading	Elementary	3.681	+	0.00	0.82
Maple (2011)	9	Reading	Elementary	0.456	-	0.66	-0.15
Mt. Healthy (2011)	11	Reading	Elementary	4.352	+	0.00	1.31
Timberline (2011)	22	Reading	Elementary	29.042	+	0.00	6.19
Northwest (2011) [1]	83	Reading	Elementary	8.167	+	0.00	0.90
Orville (2011) [1]	63	Reading	Middle	11.653	+	0.00	1.47
Ravenna (2011) [1]	49	Reading	Elementary	10.023	+	0.00	1.43
Springfield (2011)	202	Reading	--	5.015	+	0.00	0.35
Toledo (2011)	329	Reading	Elementary	14.952	+	0.00	0.82
Zanesville (2011) [1]	42	Reading	High School	2.247	+	0.03	0.35
Toledo (2007) [2]	340	Math	Elementary	9.950	+	0.00	0.54
Toledo (2008) [2]	238	Math	Elementary	22.450	+	0.00	1.46
Toledo (2009) [2]	102	Math	Elementary	13.796	+	0.00	1.37
Akron Digital Acad. (2009) [2]	23	Math	Middle	1.531	+	0.14	0.32
International Acad. (2009) [2]	65	Math	--	1.597	+	0.11	0.20
Akron Digital Acad. (2010) [2]	22	Math	Middle	1.446	+	0.16	0.31
Cincinnati (2010) [2]	17	Math	--	3.847	+	0.00	0.93
Akron Digital Acad. (2011) [2]	16	Math	High School	2.224	+	0.04	0.56
Hamilton (2011) [2]	23	Math	Elementary	4.898	+	0.00	1.02
Maple Hts. (2011) [2]	22	Math	Elementary	3.620	+	0.00	0.77
Northwest (2011) [2]	86	Math	Elementary	7.138	+	0.00	0.77
Orville (2011) [2]	63	Math	Middle	10.843	+	0.00	1.37
Ravenna (2011) [2]	49	Math	Elementary	3.379	+	0.00	0.48
Zanesville (2011) [2]	49	Math	--	4.384	+	0.00	0.63
Cincinnati (2011) [2]	62	Math	Elementary	1.173	+	0.24	0.15
Toledo (2010) [2]	159	Math	Elementary	12.512	+	0.00	0.99
Columbus (2010) [2]	32	Math	Middle	0.500	-	0.62	-0.09