Appendix E

Innovative Schools in Washington: What Lessons Can Be Learned?

Meta-Analytic Procedures and Results

Click here to view the complete Innovative Schools report.

110 Fifth Avenue SE, Suite 214 • P.O. Box 40999 • Olympia, WA 98504-0999 • (360) 586-2677 • FAX (360) 586-2793 • www.wsipp.wa.gov

To estimate the effects of programs and policies on outcomes, we use statistical procedures researchers have been developing to facilitate systematic reviews of evaluation evidence. This set of procedures is called "meta-analysis" and we employ that methodology in this study.[1]

In meta-analysis, we pool the results of all credible evaluation studies we can locate on a similar topic. For example, one of our analyses estimates the average impacts on student reading ability from one-on-one tutoring programs. The combined results—a weighted average of the impacts measured in the national research literature—represent our best estimate to-date of the weight of the evidence on a specific topic.

In this appendix we describe our methods and results as follows:

E1. Topic Selection

E2. Results Summary

E3. Detailed Results

E4. Methodological Details

E1. TOPIC SELECTION

We meta-analyze research evaluating the strategies used in Washington State public schools designated as "innovative" described earlier in this report. These topics were selected based on what the schools identified as their innovative strategies; the approaches mentioned in legislation; consultation with the study advisory group; and the availability of high-quality evaluation studies.

The first panel of Exhibit E1 lists the topics for which we were able to locate and analyze enough studies that met our research criteria to draw conclusions about effectiveness. The topics are listed in alphabetical order.

The second panel of Exhibit E1 lists Washington innovative school strategies for which we searched, but could not find, a sufficient number of scientifically credible evaluations for meta-analysis.

The third panel of Exhibit 1 lists relevant topics for which such large and/or complex literatures exist that we could not fully analyze in time for this study.

[1] In general, we follow the meta-analytic methods described in: Lipsey, M.W., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Exhibit E1
Innovative School Strategies: Research Topics Reviewed for Meta-analysis

	Topic	Pg.
Panel 1: Topics with WSIPP meta-analysis of impacts on student learning	Charter schools	05
	Expeditionary learning	25
	Instructional time (one additional day)	27
	National Board for Professional Teaching Standards	30
	National Guard Youth ChalleNGe Program	33
	Parent involvement in reading instruction	34
	Principals (school leadership)	37
	Project Lead the Way	41
	School-wide positive behavior programs	43
	Teacher induction/mentoring	46
	Teacher professional development	48
	Tutoring	51
Panel 2: Topics WSIPP reviewed, but had too few rigorous evaluations to meta-analyze	Advancement Via Individual Determination (AVID)	
	Blended learning	
	High Schools that Work	
	Home schooling	
	International Baccalaureate	
	MicroSociety	
	Montessori	
	Professional learning communities	
Panel 3: Relevant topics WSIPP did not fully review (due to the complexity/weight of evidence)	Alternative schools for at-risk students	
	Extended instructional day	
	Project-based learning	
	School size (small schools)/personalization of learning	
	Technology-based innovations	
	Theme-based (A-STEM) or magnet schools	
	Wrap-around services	

E2. SUMMARY OF META-ANALYTIC RESULTS

Exhibit E2 summarizes the results of the meta-analyses presented in the appendix.

Exhibit E2
Summary of WSIPP Meta-Analytic Results

Topic	Outcomes meta-analyzed	WSIPP meta-analytic result	Other outcomes examined in the research literature (not meta-analyzed for this report)*
Charter schools	Reading & math test scores	Nationally, charter schools do not have a consistent impact on student test scores (some have positive impacts, some negative). Our analysis was unable to identify specific characteristics of charter schools that are associated with more positive outcomes. Knowledge is Power Program (KIPP) charter schools and charter schools located in urban areas have consistently positive impacts on student test score outcomes.	High school graduation; college enrollment; attendance; discipline and effects on nearby schools
Expeditionary learning	Reading, math, & science test scores	Expeditionary learning does not have a consistent impact on student test scores (some evaluated programs have positive impacts, some negative).	Behavioral measures such as attendance and disciplinary incidents
Instructional time (one addtl. day)	Reading & math test scores	One additional school day does not have a consistent impact on student test scores (there are some positive impacts and some negative; the effects may depend on how the time is used).	Labor market outcomes
NBPTS certification	Reading, math, & other academic test scores	Students who have teachers with NBPTS certification have slightly higher test scores, on average.	Teacher recruitment and retention, self-reported impacts on teaching practices
National Guard Youth ChalleNGe Program	High school graduation	ChalleNGe appears to have a positive impact on high school graduation rates and mixed impacts on other outcomes.	Employment, housing, crime, health, substance abuse, GEDs.

Exhibit E2
Summary of WSIPP Meta-Analytic Results

Topic	Outcomes meta-analyzed	WSIPP meta-analytic result	Other outcomes examined in the research literature (not meta-analyzed for this report)*
Parent involvement in reading instruction	Reading test scores	Elementary school-based programs that encourage parent involvement in reading instruction are associated with improved student reading outcomes, on average.	Parent, student, and teacher perceptions/ satisfaction with program
Principals (school leadership)	Reading & math test scores	School leadership affects student outcomes: a principal who is one standard deviation above typical principal effectiveness can improve student test scores.	High school graduation; self-reported measures of effectiveness
Project Lead the Way	Reading, math, & science test scores	Project Lead the Way improves student math scores but does not consistently impact student reading or science test scores.	GPA, enrollment in advanced math/ science courses or higher education
School-wide positive behavior programs	Reading & math test scores	School-wide interventions focused on encouraging positive behavior can improve academic outcomes (math and reading test scores).	Attendance, grade retention, and discipline (office discipline referrals, suspensions, and expulsion)
Teacher induction/ mentoring	Reading, math, & other academic test scores	For teacher induction programs, the results are mixed, but the average impact is positive.	Teacher retention; self-reported measures of teacher outcomes
Teacher professional development	Reading, math, & other academic test scores	Providing more quantity of general approaches to PD is not associated with improving student test scores. For content-specific PD, results are positive on average.	Teacher retention; self-reported measures of teacher outcomes
Tutoring	Reading test scores	One-on-one tutoring is an effective way to improve reading test scores.	Parent, student, and teacher satisfaction with program

*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

E3. DETAILED META-ANALYTIC RESULTS

This section presents our meta-analytic results for each topic listed in the top panel of Exhibit E1.

E3a. Charter Schools

A charter school is a public school governed under a legislative contract or state charter with state or local jurisdiction. Charter schools gain autonomy through exemptions from "selected state or local rules and regulations" and in return "must meet the accountability standards articulated in its charter."[2] In the 2012-13 school year, an estimated 6,000 charter schools enrolled more than 2.3 million students across the country.[3] In November 2012, Washington became the 42nd state (in addition to the District of Columbia) to authorize charter schools with the passage of Initiative 1240.

Like charters, Washington's designated innovative schools (the focus of the main body of this report) are assumed to be trying something new outside of a typical K-12 approach.

The studies included in this meta-analysis use a variety of research designs and statistical approaches to measure impacts on student outcomes.

- Some studies use a "lottery-based" approach. Here, the academic outcomes of students who won a lottery to an oversubscribed charter school are compared to the outcomes of students who did not win.
- Several studies use a student "fixed-effects" approach in an attempt to control for unobserved heterogeneity. Here, an individual student's gains while attending a charter school are compared to the same student's gains while attending a traditional school.
- Twenty-one of the 65 effect sizes included in both the reading and math analyses rely on the "virtual twin" method developed by the Center for Research on Education Outcomes (CREDO) at Stanford University. This method compares outcomes for charter school students to matched composites of up to seven students in traditional public schools that have similar observable characteristics (gender, race/ethnicity, special education designation, English language learner status, free or reduced priced lunch status, grade level, and prior achievement).[4]

The overall charter school results are presented in Exhibit E3 and the detailed results in Exhibits E4. The evidence is mixed (some positive, some negative), suggesting that charter schools do not, as a group, have a consistent impact on student test scores. Our analysis was unable to conclude which characteristics of charter schools are associated with more positive outcomes, because specific school characteristics are not commonly measured across studies.

[2] See http://nces.ed.gov/fastfacts/display.asp?id=30for more information.
[3] http://www.publiccharters.org/data/files/Publication_docs/NAPCS%202012-13%20New%20and%20Closed%20Charter%20Schools_20130114T161322.pdf
[4] For additional information on this method, see: Center for Research on Education Outcomes (CREDO) (2009). *Multiple choice: Charter school performance in 16 states* (Technical Appendix). Stanford, CA: Stanford University, Center for Research on Education Outcomes. Available at: http://credo.stanford.edu/

Exhibit E3

Effect Sizes: Impacts on Student Academic Outcomes from Charter Schools

	Reading test scores	**Math test scores**	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	65	65	Other test scores (e.g. science and social studies); high school graduation; college enrollment; attendance; discipline; and effects on nearby schools
Average effect on academic outcomes (standard error)	0.002 (0.007)	0.009 (0.011)	
Conclusion	Nationally, charter schools have no consistent impact on student test scores (some have positive impacts, some have negative). Our analysis was unable to conclude which characteristics of charter schools are associated with more positive outcomes.		

*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

The average effect of charter schools masks considerable differences in outcomes for particular schools. Some charter schools are associated with substantial positive impacts on student achievement, while other schools show negative or not significantly different impacts. The characteristics of high-performing charter schools are a subject of growing interest in the research literature. For example, Dobbie and Fryer (2012) find that five policies ("frequent teacher feedback, the use of data to guide instruction, high-dosage tutoring, increased instructional time, and high expectations") explain a substantial amount of school effectiveness.[5]

Too few studies have examined the characteristics of high-performing charter schools in a systematic way to be able to draw cause-and-effect conclusions regarding which characteristics are most important for student learning. However, we are able to examine the impact of other characteristics, including:

- use of the Knowledge is Power Program (KIPP) model;
- the number of years that a charter school has been in operation; and
- geographic location (urban and non-urban).

[5] Dobbie, W., & Fryer, R. (2012). *Getting beneath the veil of effective schools: Evidence from New York City*. Unpublished manuscript, Harvard University, Cambridge, MA.

Exhibit E4
Effect Sizes: Impacts on Student Academic Outcomes from Charter Schools
Study by Study Results

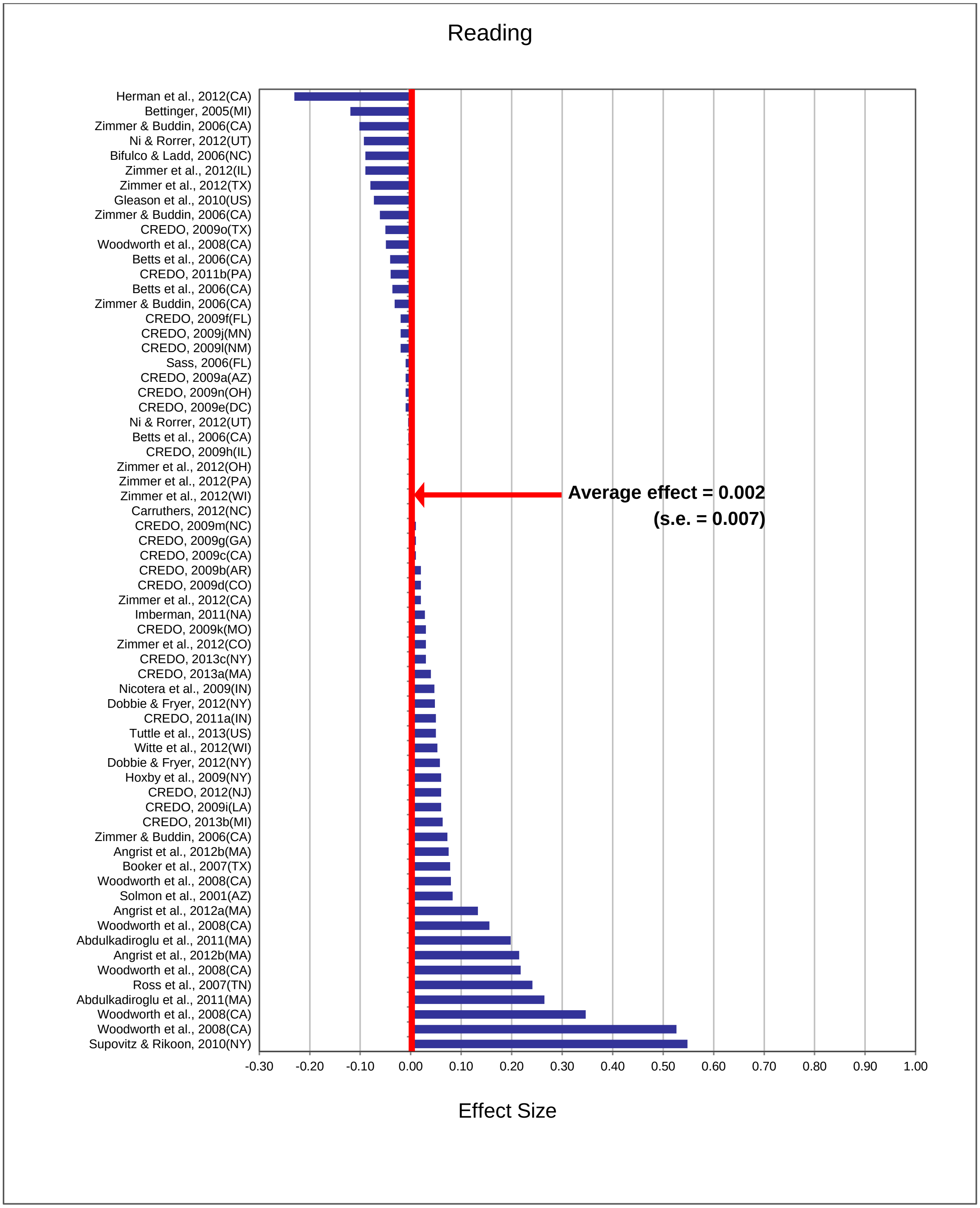
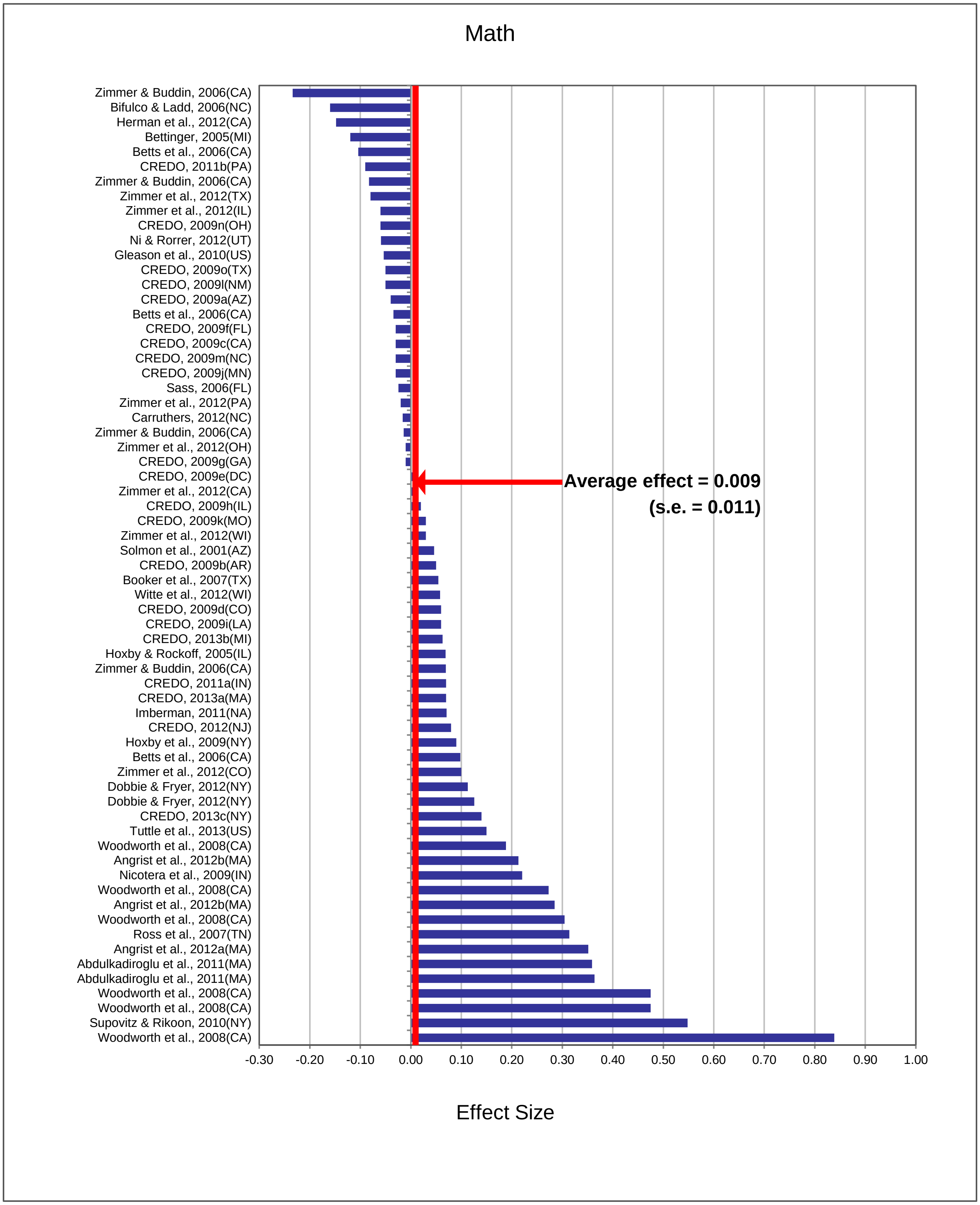


Exhibit E4, continued

Effect Sizes: Impacts on Student Academic Outcomes from Charter Schools

Study by Study Results



Knowledge is Power Program Charter Schools

The Knowledge is Power Program (KIPP) is a network of public charter schools serving more than 41,000 students in 20 states and the District of Columbia. The schools predominantly enroll low-income and minority students. The KIPP organization describes itself as being "committed to serving the students who need us most" and refusing to "accept anything less than an excellent college-preparatory education for students from low-income communities."[6] To achieve this goal, KIPP schools use the following set of operating principles called the "Five Pillars": (1) high expectations for academic achievement and conduct, (2) choice and commitment to "put in the time and effort required to achieve success," (3) more time, including extended days, weeks, and years, (4) the power to lead and control over the school budget and personnel by principals, and (5) a focus on results on standardized tests and other objective measures.

The studies included in this analysis are of KIPP middle schools around the country. Three studies report outcomes for individual KIPP schools, while the fourth study (Tuttle et al., 2013) uses the average impact of 41 schools from 14 states. One study (Angrist et al., 2012) uses a lottery-based research approach; the three other studies used a matched comparison design.

The overall results for KIPP charter schools are presented in Exhibit E5 and the detailed results in Exhibit E6. The evidence suggests that KIPP charter schools improve test scores in both reading and math more consistently than charter schools in general.

Exhibit E5
Effect Sizes: Impacts on Student Academic Outcomes from KIPP Charter Schools

	Reading test scores	**Math test scores**	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	9	9	Student engagement, educational aspirations, behavior, and satisfaction by subgroup
Average effect on academic outcomes (standard error)	0.106 (0.048)	0.273 (0.049)	
Conclusion	KIPP charter schools improve student reading and math test scores more consistently than charter schools in general.		

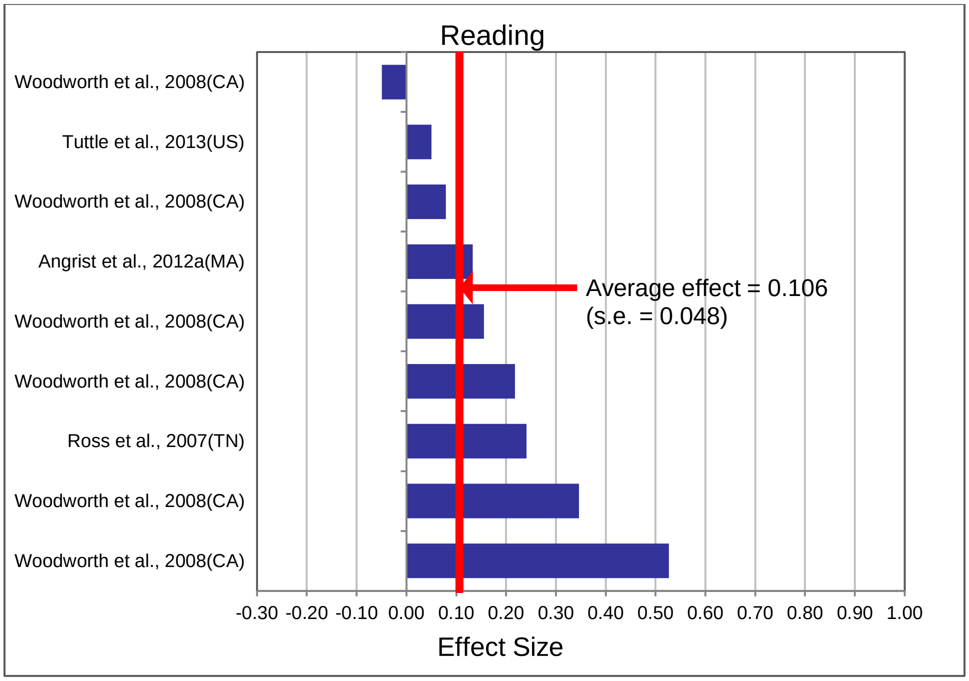
*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

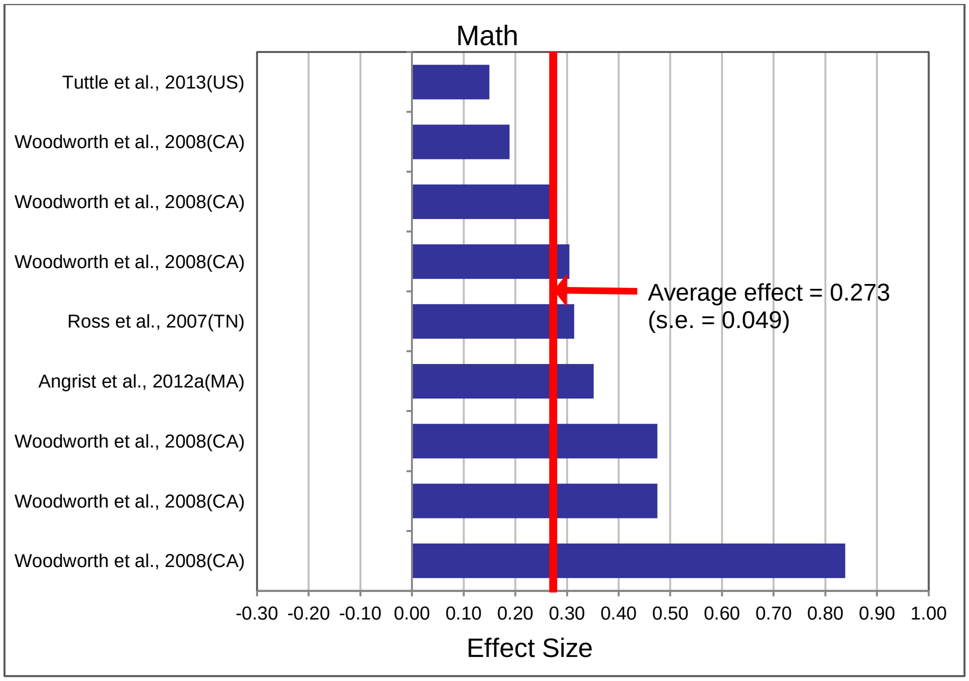
[6] See www.kipp.org for more information.

Exhibit E6

Effect Sizes: Impacts on Student Academic Outcomes from KIPP Charter Schools

Study by Study Results





Charter Schools by Years of Operation

Schools in their initial years of operation may struggle to produce positive outcomes, but these outcomes may improve over time due to a “maturation” effect. New charter schools are often in the process of developing an unfamiliar model and curriculum, frequently lack the support and resources from districts that are available to traditional public schools, and may employ a high percentage of inexperienced teachers. All of these factors may negatively impact student achievement.[7] Several recent studies have examined this assumption.[8]

The studies included in this analysis examine the effects of charter schools by the number of years of operation. The studies draw primarily on administrative data and use a student fixed effects approach. A few studies grouped some years together. For example, Zimmer et al., 2009, examines the effects of charters in operation for one, two, and three or more years.

We have fewer effect sizes available for meta-analysis of impacts by years of operation, because many studies did not examine this topic.

A related topic, the effect of “startup” charters that began from scratch compared to “conversion” charters that were once a traditional public school, were not included in this analysis.

The overall results for charter schools by years of operation are presented in Exhibits E7 and E8 and the detailed results in Exhibit E9. The evidence suggests that charter schools in their first two years of operation are associated with lower student test scores, while charter schools in operation for three or more years show no consistent impact (similar to the overall results from the national charter school literature presented earlier in this appendix).

---

[7] Carruthers, C. K. (2012). New schools, new students, new teachers: Evaluating the effectiveness of charter schools. *Economics of Education Review, 31*(2): 280-292.; Ni, Y., & Rorrer, A. K. (2012). Twice considered: Charter schools and student achievement in Utah. *Economics of Education Review, 31*(5): 835-849.

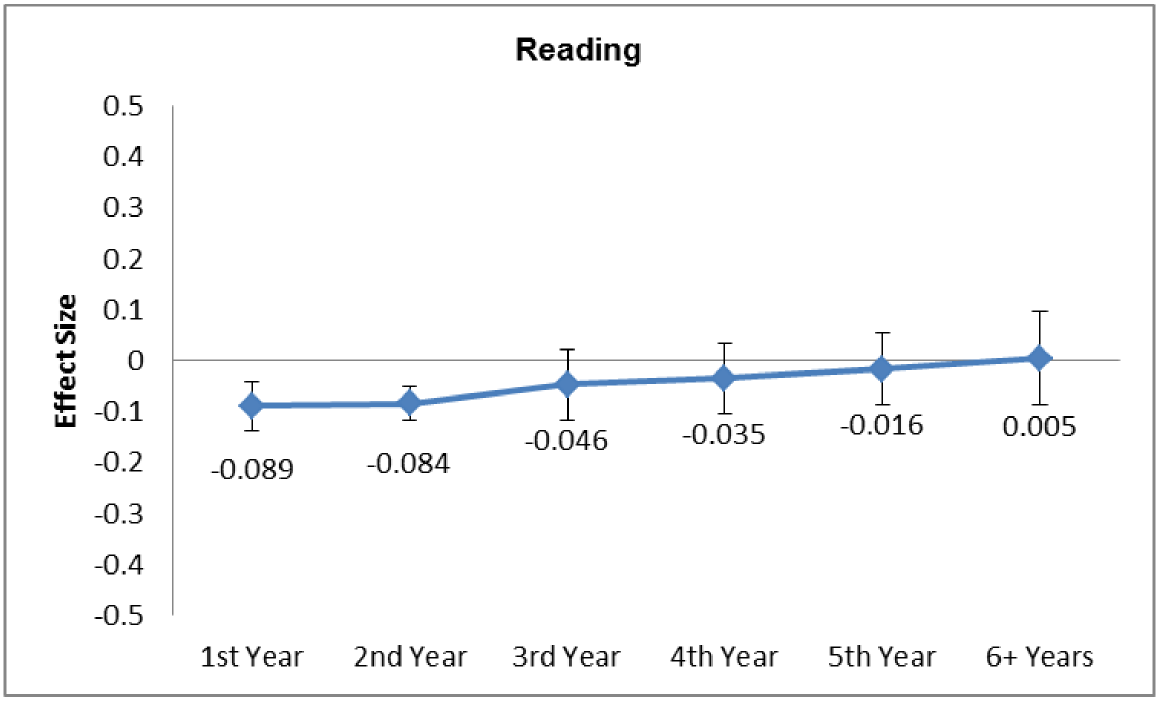
[8] Hanushek and colleagues examined the effect of charter school age in their analysis of Texas charter schools. The study was not included in this analysis due to the use of “academic” scores consisting of a composite of reading and math results, rather than reporting the results of each subject separately. For more information see: Hanushek, E.A., Kain, J.F., Rivkin, S.G., & Branch, G.F. (2007). Charter school quality and parental decision making with school choice. *Journal of Public Economics, 91*(5): 823-848.

Exhibit E7
Effect Sizes: Impacts on Student Academic Outcomes from Charter Schools by Years School has been in Operation

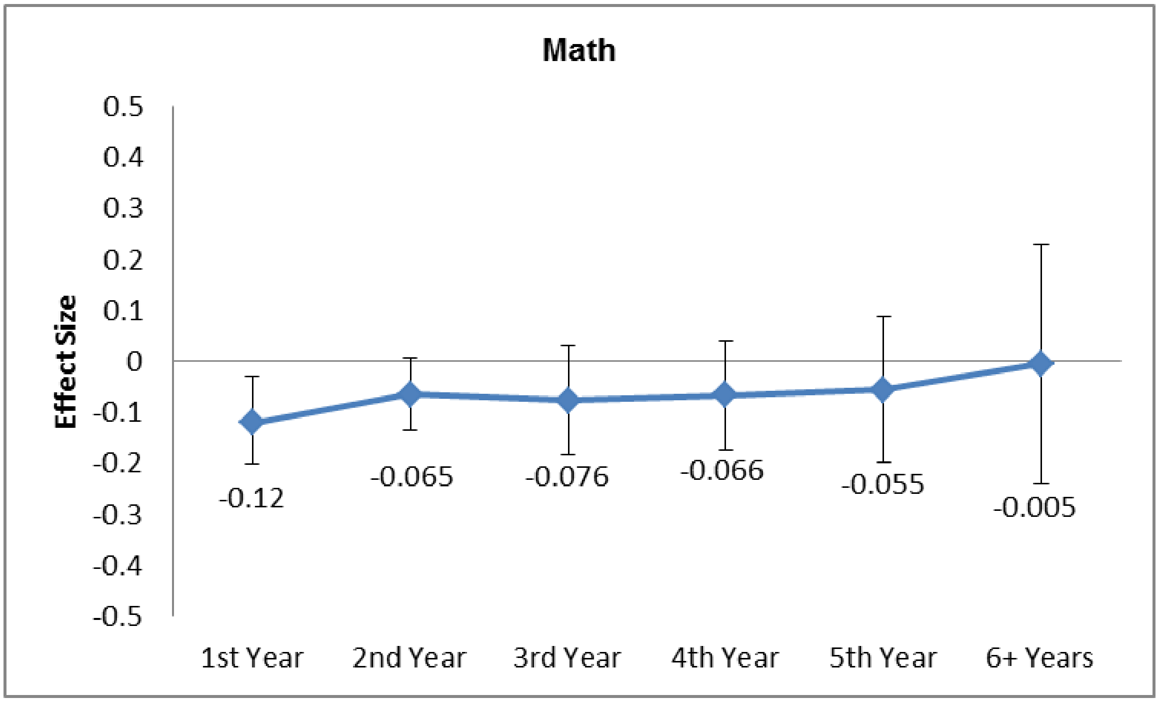
	Reading test scores		**Math test scores**		**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	1st year	14	1st year	13	NA
	2nd year	13	2nd year	12	
	3rd year	6	3rd year	5	
	4th year	6	4th year	5	
	5th year	6	5th year	5	
	6+ years	4	6+ years	3	
Average effect on academic outcomes (standard error)	1st year	-0.089 (0.025)	1st year	-0.116 (0.044)	
	2nd year	-0.084 (0.017)	2nd year	-0.065 (0.037)	
	3rd year	-0.046 (0.035)	3rd year	-0.076 (0.055)	
	4th year	-0.035 (0.036)	4th year	-0.066 (0.055)	
	5th year	-0.016 (0.036)	5th year	-0.055 (0.073)	
	6+ years	0.005 (0.046)	6+ years	-0.005 (0.120)	
Conclusion	Charter schools in their initial years have, on average, negative impacts on student test scores. Charter schools in their 3rd or subsequent years of operation are more varied in their impacts on student test scores (some have positive impacts, some have negative).				

*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

Exhibit E8
Effect Sizes: Impacts on Student Academic Outcomes from Charter Schools by Years School has been in Operation



Note: Error bars represent 95 percent confidence interval.



Note: Error bars represent 95 percent confidence interval.

Exhibit E9
Effect Sizes: Impacts on Student Academic Outcomes from Charter Schools, by Years School has been in Operation
Study by Study Results

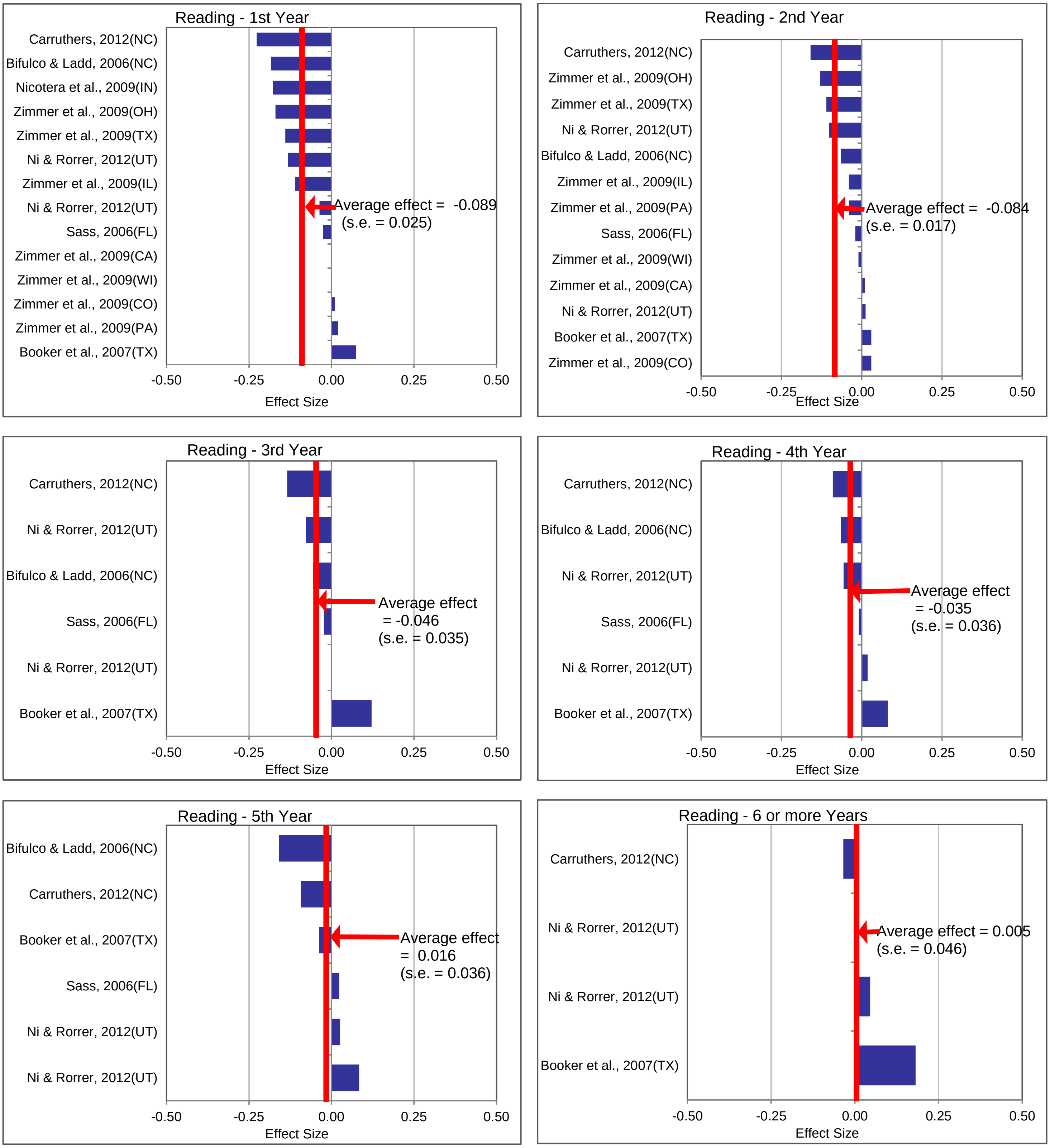
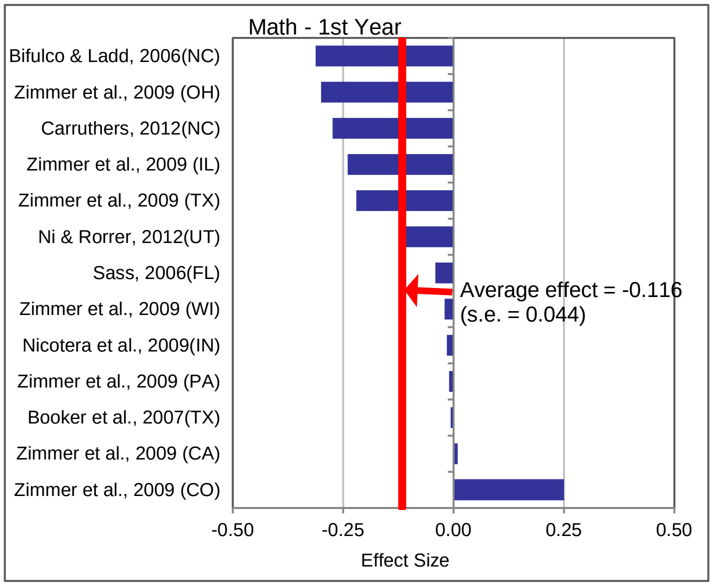
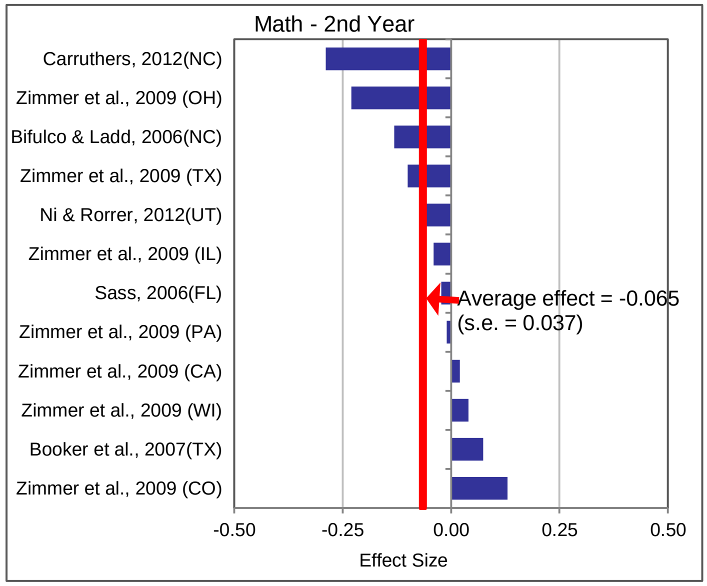
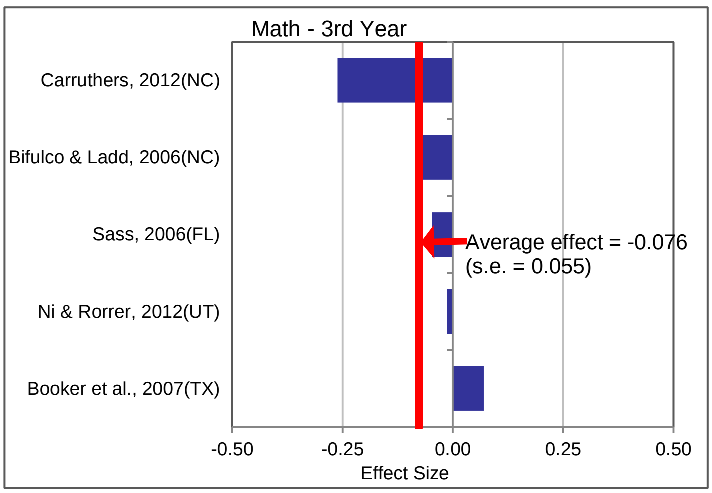
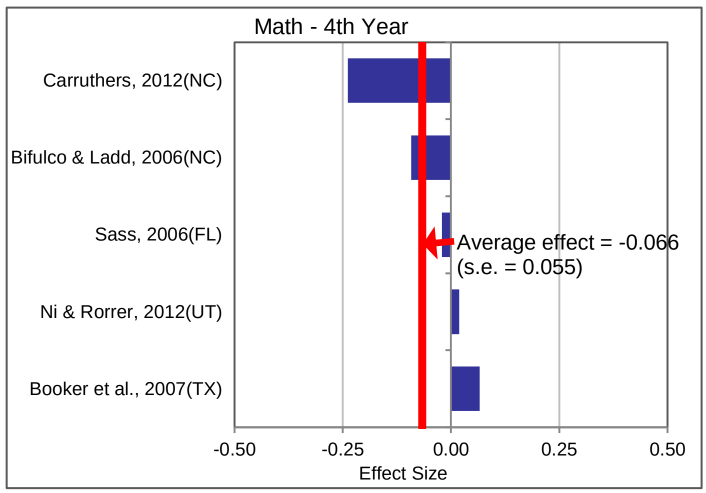


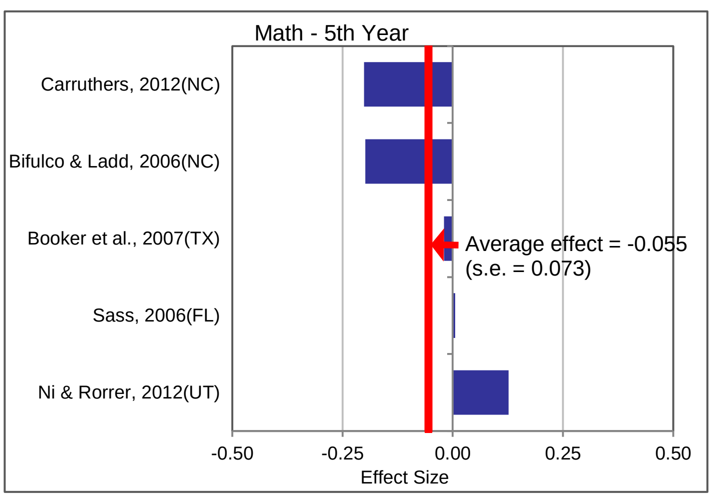
Exhibit E9, continued
Effect Sizes: Impacts on Student Academic Outcomes from Charter Schools by Years School has been in Operation
Study by Study Results

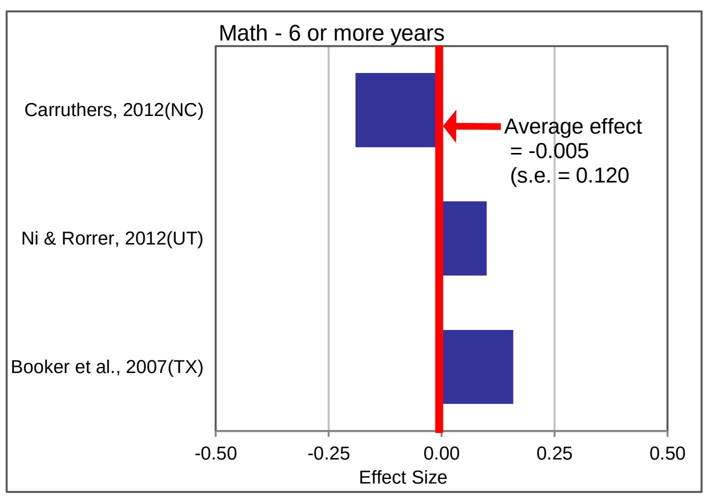












Urban Charter Schools

Charter schools have traditionally been located in cities; many are designed to serve minority students in high-poverty areas.[9] A body of literature suggests that charter schools located in urban areas may be more effective than charters located outside of the urban core. Possible explanations for stronger effects in urban areas include more competition from nearby schools; larger impacts for students who start from a lower achievement baseline; and more frequent use of a “No Excuses” model that “emphasizes instructional time, comportment, selective teacher hiring, and focuses on traditional math and reading skills.”[10] While this meta-analysis does not identify the reasons for urban charter school successes, we do find that charter schools located in urban areas more consistently improve student test scores than the impacts we found in our analysis of charter school effects in general.

The studies we use in this analysis include findings from specific cities (e.g. New York or Chicago), as well as statewide studies that examine impacts by urbanicity. The studies include a mix of lottery-based, fixed-effect, and matched comparison designs.

The overall results of the urban charter school analysis are presented in Exhibit E10 and the detailed results in Exhibit E11. The results show more consistent, and on average positive, impacts from charter schools located in urban areas on reading and especially math test scores, in comparison with our findings for charter schools in general.

Exhibit E10
Effect Sizes: Impacts on Student Academic Outcomes from Urban Charter Schools

	Reading test scores	**Math test scores**	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	38	39	Other test scores (e.g. science and social studies); high school graduation; college enrollment; attendance; and discipline (office referrals, suspensions, and expulsion)
Average effect on academic outcomes (standard error)	0.032 (0.016)	0.076 (0.018)	
Conclusion	Charter schools located in urban areas improve reading and math test scores more consistently than charter schools in general.		

*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

---

[9] Angrist, J.D., Pathak, P.A., & Walters, C.R. (2012). *Explaining Charter School Effectiveness* (Working Paper 12-11). Cambridge, MA: Department of Economics, Massachusetts Institute of Technology.
[10] Ibid.

Exhibit E11

Effect Sizes: Impacts on Student Academic Outcomes from Urban Charter Schools

Study by Study Results

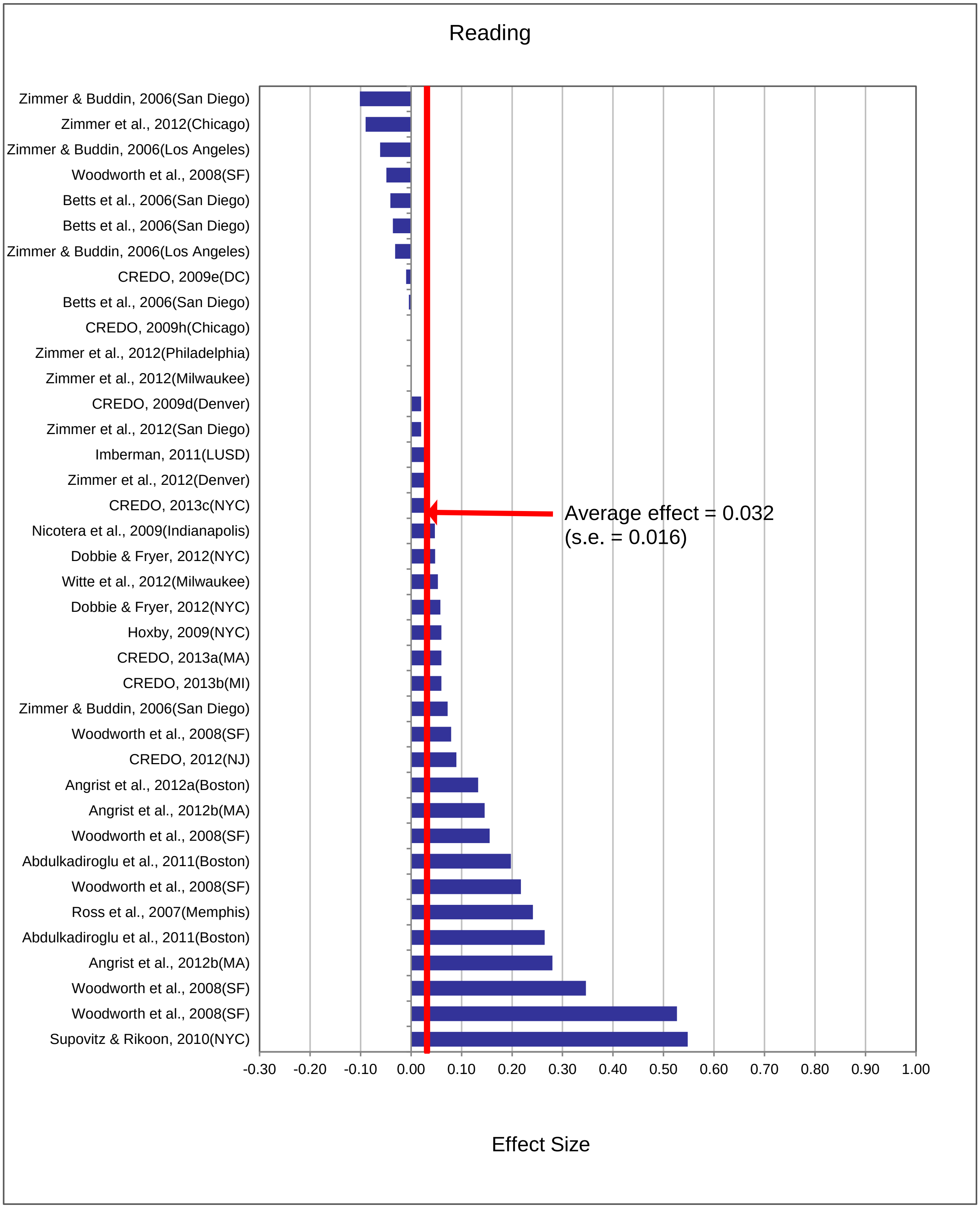
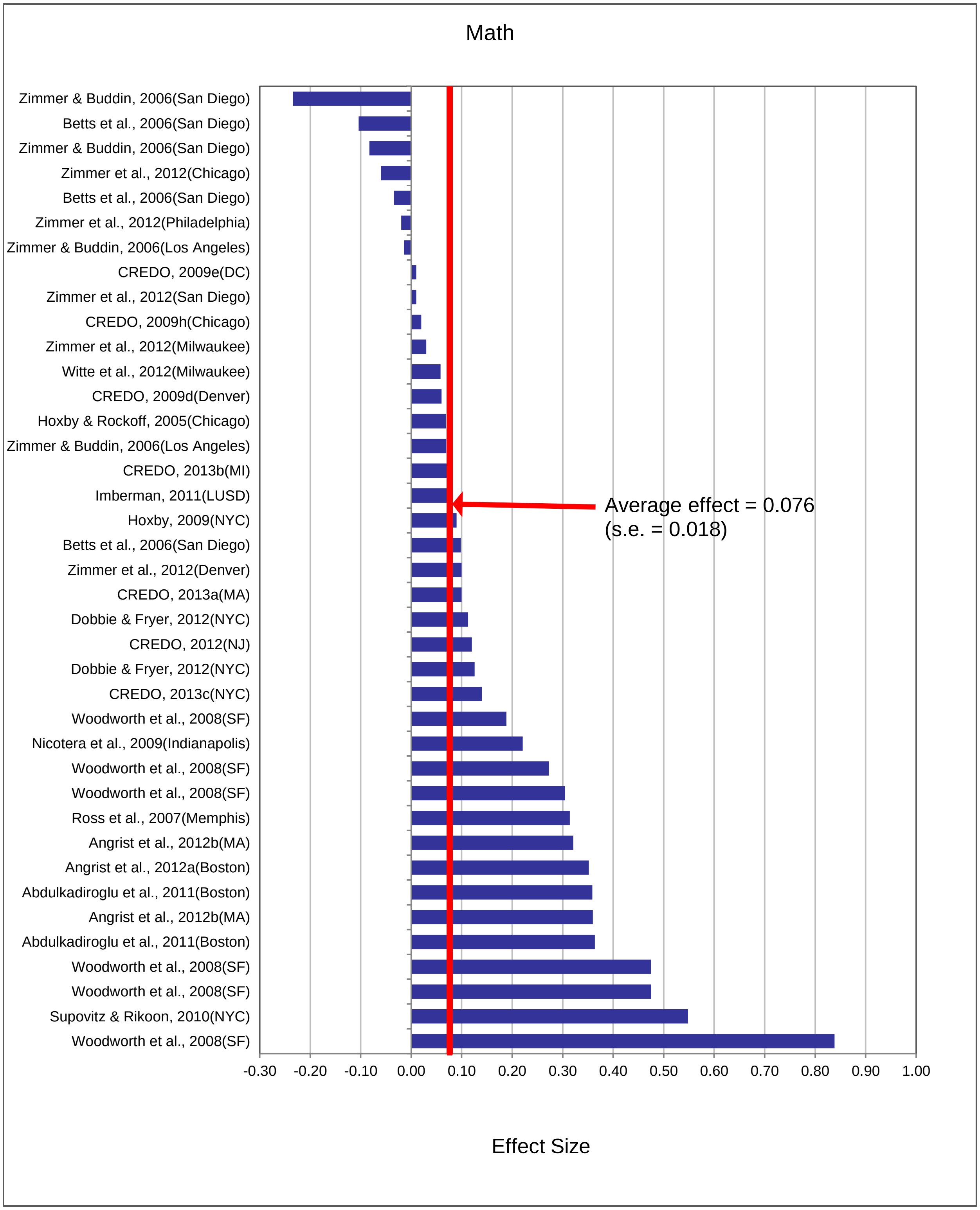


Exhibit E11, continued

Effect Sizes: Impacts on Student Academic Outcomes from Urban Charter Schools

Study by Study Results



Non-Urban Charter Schools

While charter schools traditionally operate in urban areas, there is a growing interest in charters "outside of central cities."[11] A few recent studies have begun to examine the impact of charters located outside of urban areas.

The effect sizes used in this analysis include only studies that conducted subgroup analysis to examine the impacts of charter schools located outside of urban areas. The effect sizes from the CREDO studies used in this analysis are weighted averages of the impacts of "suburban," "rural," and "town" charter schools.

The overall results of the non-urban charter school analysis are presented in Exhibit E12 and the detailed results in Exhibit E13. The evidence suggests that charter schools located outside of urban areas have no consistent impact on student test scores.

Exhibit E12
Effect Sizes: Impacts on Student Academic Outcomes from Non-urban Charter Schools

	Reading test scores	**Math test scores**	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	5	5	Attendance and discipline (in- and out-of-school suspensions)
Average effect on academic outcomes (standard error)	0.013 (0.046)	0.043 (0.027)	
Conclusion	Charter schools located outside of urban areas have no consistent impact on student test scores.		

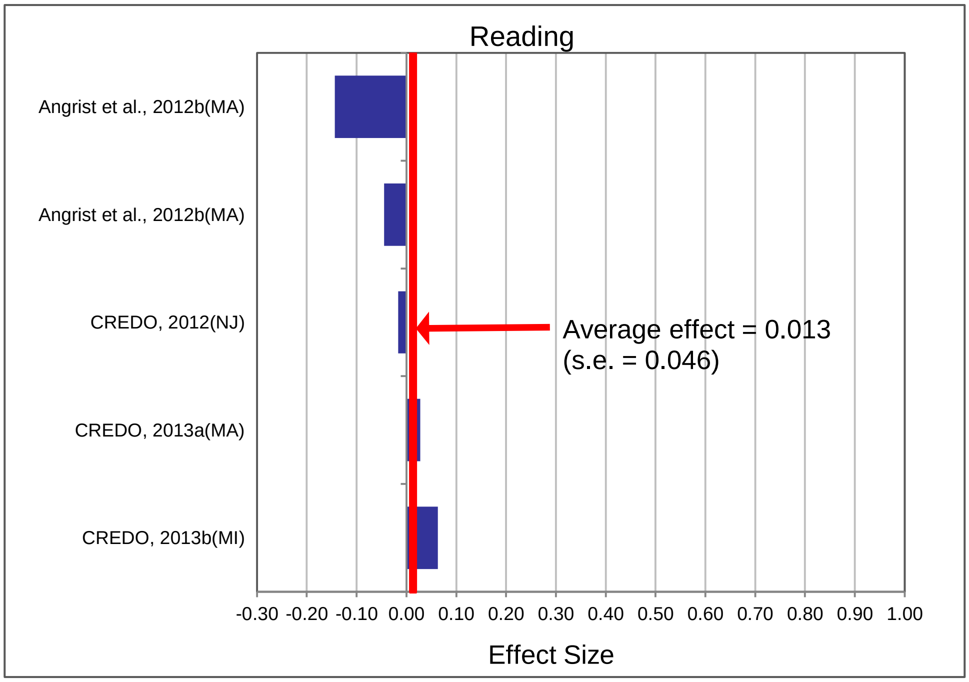
*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

[11] Angrist, J.D., Pathak, P.A., & Walters, C.R. (2012). *Explaining Charter School Effectiveness* (Working Paper 12-11). Cambridge, MA: Department of Economics, Massachusetts Institute of Technology.

Exhibit E13

Effect Sizes: Impacts on Student Academic Outcomes from Non-urban Charter Schools

Study by Study Results



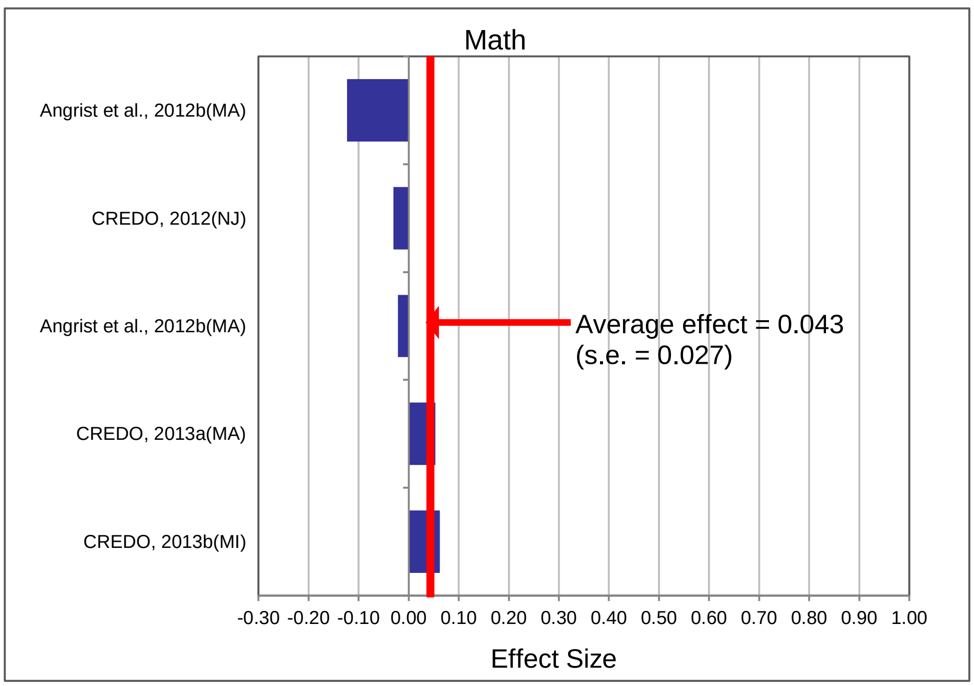
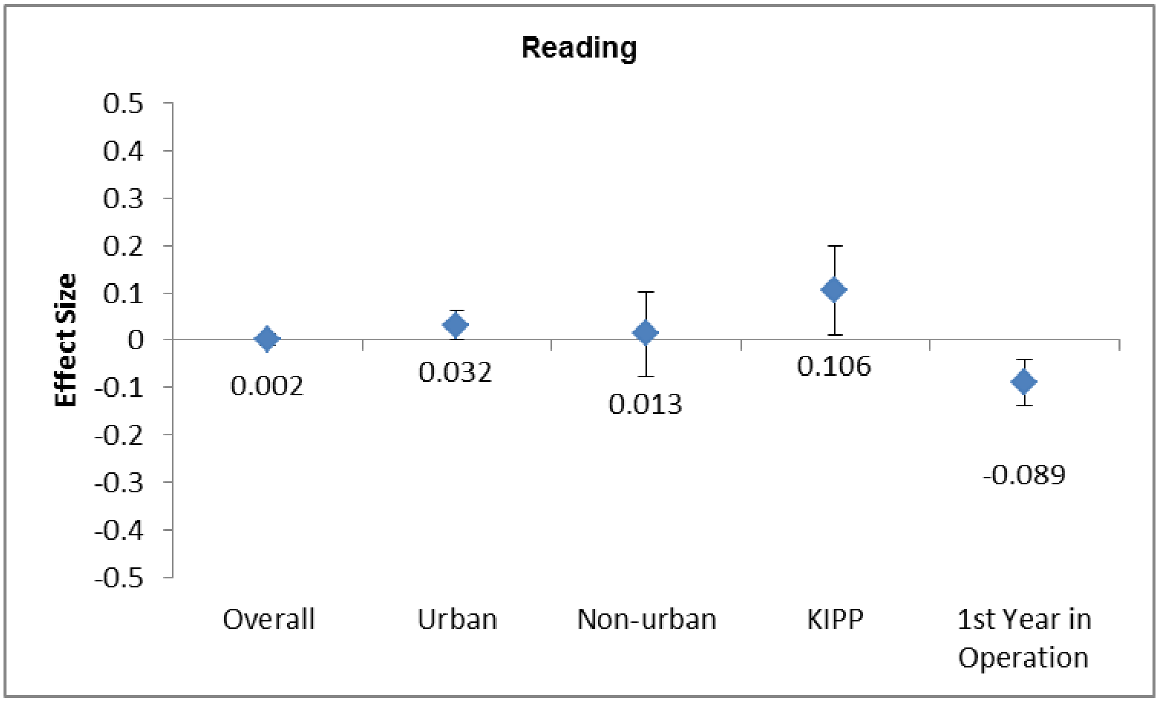
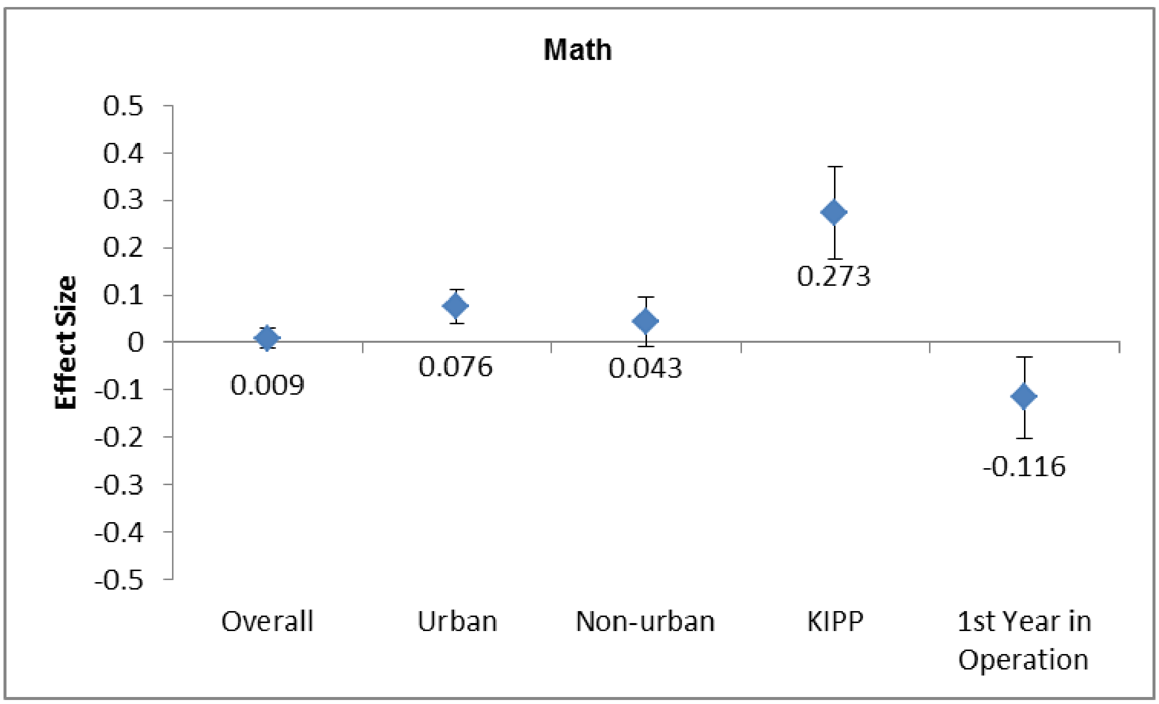


Exhibit E14 summarizes the charter school meta-analysis results across all of the sub-topics presented in this appendix.

Exhibit E14
Summary of Charter Meta-analyses Impacts on Student Academic Outcomes



Note: Error bars represent 95 percent confidence interval.



Note: Error bars represent 95 percent confidence interval.

Studies Used in Any of the Meta-Analyses of Charter Schools on Student Academic Outcomes

Abdulkadiroglu, A., Angrist, J.D., Dynarski, S.M., Kane, T.J., & Pathak, P.A. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *The Quarterly Journal of Economics, 126*(2): 699-748.

Angrist, J.D., Dynarski, S.M., Kane, T.J., Pathak, P.A., & Walters, C.R. (2012a). Who benefits from KIPP? *Journal of Policy Analysis and Management*. Advance online publication. doi: 10.1002/pam.21647.

Angrist, J.D., Pathak, P.A., & Walters, C.R. (2012b). *Explaining Charter School Effectiveness* (Working Paper 12-11). Cambridge, MA: Department of Economics, Massachusetts Institute of Technology.

Bettinger, E.P. (2005). The effect of charter schools on charter students and public schools. *Economics of Education Review, 24*(2): 133-147.

Betts, J.R., Rice, L.A., Zau, A.C., Tang, Y.E., & Koedel, C.R. (2006). *Does school choice work? Effects on student integration and achievement.* San Francisco, CA: Public Policy Institute of California.

Bifulco, R., & Ladd, H.F. (2006). The impacts of charter schools on student achievement: Evidence from North Carolina. *Education Finance and Policy, 1*(1): 50-90.

Booker, K., Gilpatric, S.M., Gronberg, T., & Jansen, D. (2007). The impact of charter school attendance on student performance. *Journal of Public Economics, 91*(5): 849-876.

Carruthers, C. K. (2012). New schools, new students, new teachers: Evaluating the effectiveness of charter schools. *Economics of Education Review, 31*(2): 280-292.

Center for Research on Education Outcomes (CREDO). (2009a). *Charter school performance in Arizona.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009b). *Charter school performance in Arkansas.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009c). *Charter school performance in California.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009d). *Charter school performance in Colorado.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009e). *Charter school performance in the District of Columbia.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009f). *Charter school performance in Florida.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009g). *Charter school performance in Georgia.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009h). *Charter school performance in Illinois.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009i). *Charter school performance in Louisiana.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009j). *Charter school performance in Minnesota.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009k). *Charter school performance in Missouri.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009l). *Charter school performance in New Mexico.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009m). *Charter school performance in North Carolina.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009n). *Charter school performance in Ohio.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2009o). *Charter school performance in Texas.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2011a). *Charter school performance in Indiana.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2011b). *Charter school performance in Pennsylvania.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2012). *Charter school performance in New Jersey.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2013a). *Charter school performance in Massachusetts.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2013b). *Charter school performance in Michigan.* Stanford, CA: Stanford University.

Center for Research on Education Outcomes (CREDO). (2013c). *Charter school performance in New York City.* Stanford, CA: Stanford University.

Dobbie, W., & Fryer, R. (2012). *Getting beneath the veil of effective schools: Evidence from New York City.* Unpublished manuscript, Harvard University, Cambridge, MA.

Gleason, P., Clark, M., Tuttle, C.C., Dwoyer, E., & Silverberg, M. (2010). *The evaluation of charter school impacts: Final report* (NCEE 2010-4029). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Herman, J.L., Wang, J., Rickles, J., Hsu, V., Monroe, S., Leon, S., & Straubhaar, R. (2012). *Evaluation of Green Dot's Locke transformation project: Findings for cohort 1 and 2 students* (CRESST Report 815). Los Angeles, CA: University of California, Los Angeles, Graduate School of Education & Information Studies, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Hoxby, C.M., Kang, J.L., & Murarka, S. (2009). *Technical Report: How New York City charter schools affect achievement.* Cambridge, MA: National Bureau of Economic Research.

Hoxby, C.M., & Rockoff, J.E. (2005). *The impact of charter schools on student achievement.* Cambridge, MA. Retrieved from http://www.rand.org/content/dam/rand/www/external/labor/seminars/adp/pdfs/2005hoxby.pdf.

Imberman, S.A. (2011). Achievement and behavior in charter schools: Drawing a more complete picture. *The Review of Economics and Statistics, 93*(2): 416-435.

Ni, Y., & Rorrer, A.K. (2012). Twice considered: Charter schools and student achievement in Utah. *Economics of Education Review, 31*(5): 835-849.

Nicotera, A., Mendiburo, M., & Berends, M. (2009). *Charter school effects in an urban school district: An analysis of student achievement gains in Indianapolis.* Paper presented at the National Center on School Choice Conference at Vanderbilt University, Nashville, TN.

Ross, S.M., McDonald, A.J., Alberg, M., & McSparrin-Gallagher, B. (2007). Achievement and climate outcomes for the knowledge is power program in an inner-city middle school. *Journal of Education for Students Placed at Risk, 12*(2): 137-165.

Sass, T.R. (2006). Charter schools and student achievement in Florida. *Education Finance and Policy, 1*(1): 91-122.

Solmon, L., Paark, K., & Garcia, D. (2001). *Does charter school attendance improve test scores?* The Arizona results. Phoenix, AZ: Goldwater Institute, Center for Market-Based Education.

Supovitz, J., & Rikoon, S. (2010). *Early achievement impacts of the Harlem Success Academy charter schools in New York City*. Unpublished manuscript, Graduate School of Education, University of Pennsylvania, Philadelphia, PA.

Tuttle, C.C., Gill, B., Gleason, P., Knechtel, V., Nicholas-Barrer, I., & Resch, A. (2013). *KIPP middle schools: Impacts on achievement and other outcomes.* Washington, DC: Mathematica Policy Research.

Witte, J.F., Wolf, P.J., Carlson, D., & Dean, A. (2012). *Milwaukee independent charter schools study: Final report on four-year achievement gains* (SCDP Milwaukee Evaluation Report #31). Fayetteville, AR: University of Arkansas, Department of Education Reform, School Choice Demonstration Project.

Woodworth, K.R., David, J.L., Guha, R., Wang, H., & Lopez-Torkos, A. (2008). *San Francisco Bay area KIPP schools: A study of early implementation and achievement. Final report*. Menlo Park, CA: SRI International.

Zimmer, R., & Buddin, R. (2006). Charter school performance in two large urban districts. *Journal of Urban Economics, 60*(2): 307-326.

Zimmer, R., Gill, B., Booker, K., Lavertu, S., & Witte, J. (2012). Examining charter student achievement effects across seven states. *Economics of Education Review, 31*(2): 213-224.

E3b. Expeditionary Learning

Two of Washington's Innovative Schools (Summit School and Thornton Creek Elementary) use the expeditionary learning (EL) approach. EL is a model of whole school reform that uses an approach of inquiry, project- and problem-based study (e.g., learning concepts and procedures then applying them to real-life contexts). Generally, EL does not have a prescribed curriculum. The studies included in this analysis use an Outward Bound-based approach. Teachers are trained to design curricular experiences that meet state and local standards. One of the instructional characteristics of EL Schools is the use of EL teacher-designed curriculum that can be a six-week to a year-long in-depth learning expedition.

Expeditions can be in the form of a fieldtrip, hands-on project in class, content-related guest speakers, live performances, or other active tasks. Unlike traditional schools, learning expeditions often integrate multiple subject areas in one expeditionary program. For example, a class exploring World War II (social studies/history) may also explore nuclear fission (science) at the same time. Teachers are required to play a much larger role in students' lives by setting up mentoring services, internships, and civic/community service activities, while maintaining open communication with parents through phone calls, newsletters, announcements, and open house nights.

The overall EL results are presented in Exhibit E15 and the detailed results in Exhibit E16. The evidence suggests that expeditionary learning has no consistent impact on student test score outcomes.

Exhibit E15
Effect Sizes: Impacts on Student Academic Outcomes from Expeditionary Learning

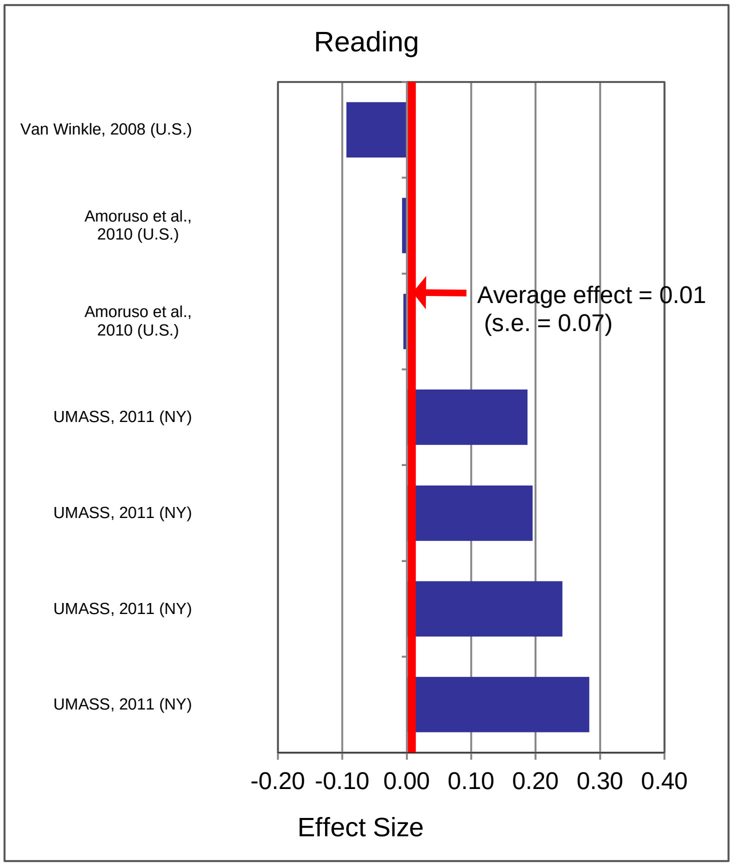
	Reading test scores	**Math test scores**	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	7	7	Behavioral measures such as attendance and disciplinary incidents
Average effect on academic outcomes (standard error)	0.01 (0.07)	0.00 (0.07)	
Conclusion	Expeditionary learning does not have a consistent impact on student test scores.		

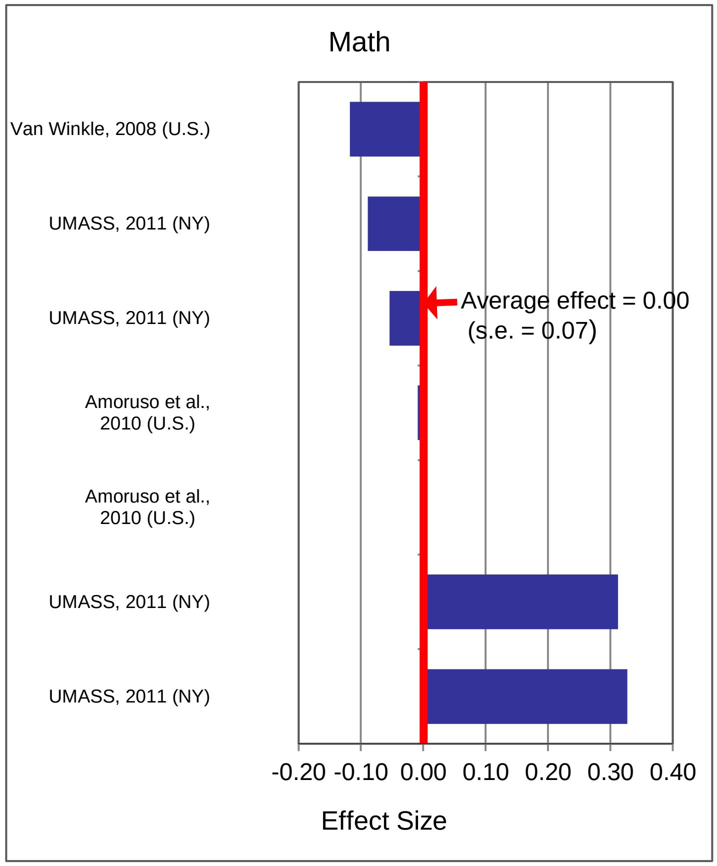
*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

Exhibit E16

Effect Sizes: Impacts on Student Test Scores from Expeditionary Learning Programs

Study by Study Results





<u>Studies Used in the Meta-Analysis of Expeditionary Learning Effects on Student Academic Outcomes</u>

Amoruso, M., Bontempo, B., & Wilson, D. (2010). *The relationship between ELS participation and academic growth.* Mountain Measurement, Inc.

UMASS. (2011). *Impacts of expeditionary learning model on student academic performance in Rochester, New York.* Amherst, MA: UMASS Donahue Institute Research and Evaluation Group.

Van Winkle, T. (2008). Expeditionary Learning Schools: The relationship between implementation gains and growth in student achievement (Doctoral dissertation. University of Wisconsin-Madison, Madison, WI).

E3c. Instructional Time (one additional day)

Some of Washington's innovative schools, such as Lincoln Center, provide students with additional instructional time (extended day) as one strategy to improve student learning. While we are unable to examine the full literature regarding additional learning time, we are able to present the results from a meta-analysis of the impact of increased instructional time in the form of one additional day per year.

The evaluations included in this analysis measure changes in the amount of instructional time in K-12 schools and subsequent impacts on student test scores and labor market earnings in adulthood. Some of the studies measure the effects of an average day and some measure the effects of additional time at the end of the year. We standardize those measures to approximate a change of one additional day.[12]

The overall results are presented in Exhibit E17 and the detailed results in Exhibit E18. The evidence suggests that one additional school day, while slightly beneficial on average, has no consistent impact on student test score outcomes.

Exhibit E17
Effect Sizes: Impacts on Student Academic Outcomes from One Additional School Day

	Test scores (reading, math, & general academic)	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	14	Attendance and self-reported perceptions of impacts of a longer school year
Average effect on academic outcomes (standard error)	0.001 (0.003)	
Conclusion	One additional school day does not have a consistent impact on student test scores (there are some positive impacts and some negative; the effects may depend on how the time is used).	

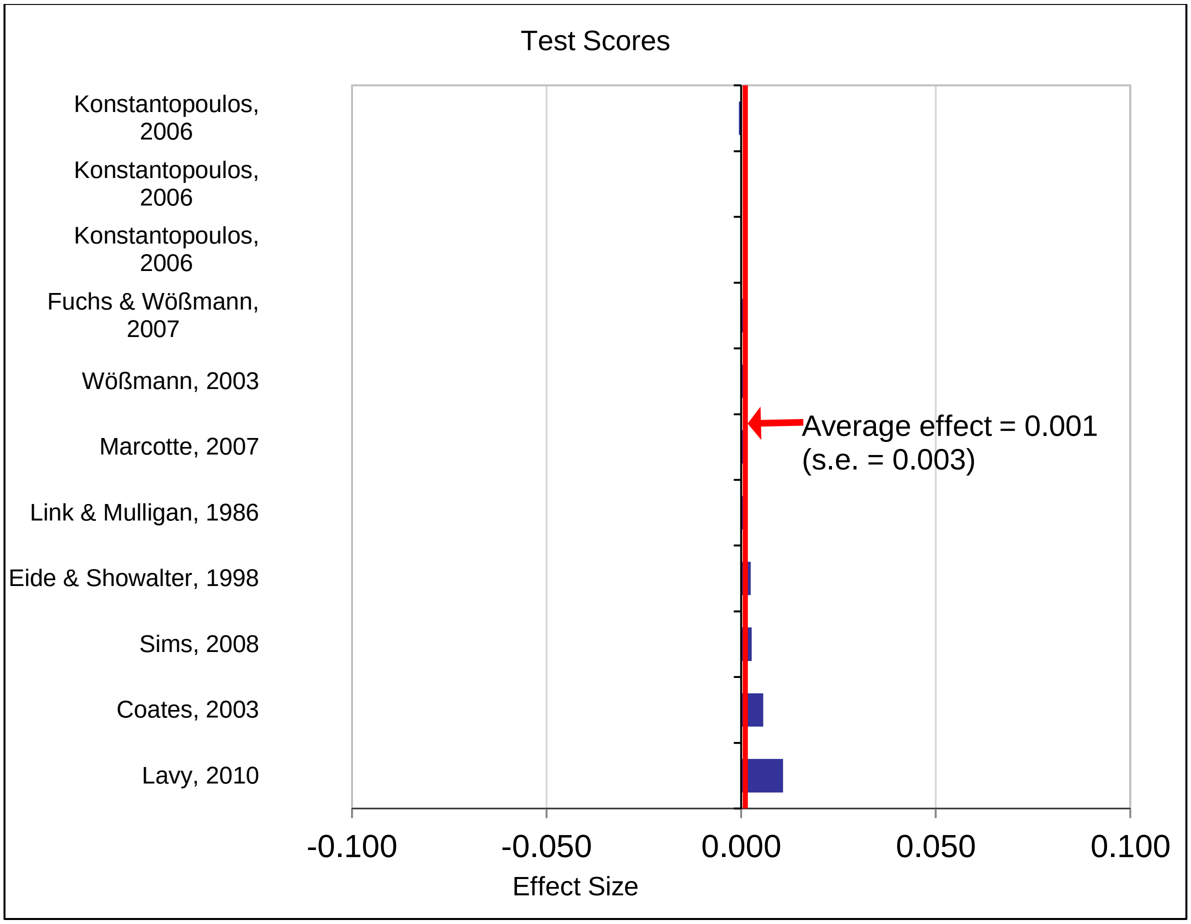
*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

[12] This meta-analytic review was last updated for an April 2011 Institute report. Pennucci, A. (2011) *The Economic Value of Learning Time in K-12 Schools: A Summary of Research Evidence and an Economic Analysis* (Document Number 11-04-2201). Olympia: Washington State Institute for Public Policy.

Exhibit E18

Effect Sizes: Impacts on Student Academic Outcomes from One Additional School Day

Study by Study Results



<u>Studies Used in the Meta-Analysis of Instructional Time (an Additional School Day)</u>
<u>Effects on Student Academic Outcomes</u>

Betts, J.R. (1996). Do school resources matter only for older workers? *The Review of Economics and Statistics, 78*(4): 638-652.

Card, D., & Krueger, A.B. (1992a). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy, 100*(1): 1-40.

Coates, D. (December 01, 2003). Education production functions using instructional time as an input. *Education Economics, 11*(3): 273-292.

Eide, E., & Showalter, M.H. (1998). The effect of school quality on student performance: A quantile regression approach. *Economics Letters, 58*(3): 345-350.

Fuchs, T., & Wößmann, L. (2007). What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics, 32*(2): 433-464.

Heckman, J., Layne-Farrar, A., & Todd, P. (1996). Human capital pricing equations with an application to estimating the effect of schooling quality on earnings. *The Review of Economics and Statistics, 78*(4): 562-610.

Konstantopoulos, S. (2006). Trends of school effects on student achievement: Evidence from NLS:72, HSB:82, and NELS:92. *Teachers College Record, 108*(12): 2550-2581.

Lavy, V. (2010, May). *Differences across and within countries in instructional time and achievements in math, science, and reading: A causal link?* Unpublished manuscript, University of London, Hebrew University and Royal Holloway, Department of Economics.

Link, C.R., & Mulligan, J.G. (1986). The merits of a longer school day. *Economics of Education Review, 5*(4): 373-381.

Loeb, S., & Bound, J. (1996). The effect of measured school inputs on academic achievement: Evidence from the 1920s, 1930s, and 1940s birth cohorts. *The Review of Economics and Statistics, 78*(4): 653-664.

Marcotte, D.E. (2007). Schooling and test scores: A mother-natural experiment. *Economics of Education Review, 26*(5): 629.

McHenry, P. (2011). The effect of school inputs on labor market returns that account for selective migration. *Economics of Education Review, 30*(1): 39-54.

Olson, C.A., & Ackerman, D. (2000). *High school inputs and labor market outcomes for male workers in their mid-thirties: New data and new estimates from Wisconsin* (Discussion Paper No. 1205-00). Madison, WI: University of Wisconsin-Madison, Institute for Research on Poverty.

Pischke, J.S. (2007). The impact of length of the school year on student performance and earnings: Evidence from the German short school years. *The Economic Journal, 117*(523): 1216-1242.

Rizzuto, R., & Wachtel, P. (1980). Further evidence on the returns to school quality. *The Journal of Human Resources, 15*(2): 240-254.

Sims, D.P. (2008). Strategic responses to school accountability measures: It's all in the timing. *Economics of Education Review, 27*(1): 58.

Wößmann, L. (2003). Schooling resources, educational institutions and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics, 65*(2): 117-170.

E3d. National Board for Professional Teaching Standards Certification

The National Board for Professional Teaching Standards (NBPTS) certification is an advanced teaching credential that complements state certification. Teachers earn NBPTS certification by successfully completing a one to three year assessment process. Washington State provides a $5,090 bonus to NBPTS-certified teachers. In the 2009-10 school year, 3,686 Washington teachers were NBPTS-certified. Baker Middle School, one of Washington's designated innovative schools, aims to have its entire teaching staff certified through NBPTS.

Overall test score outcomes are presented for reading and math combined because there is no systematic difference between the two sets of results.[13] We found that students who have teachers with NBPTS certification have slightly higher test scores, on average (see Exhibits E19 and E20). The available research does not answer the question of whether NBPTS identifies above average teachers or whether the process itself improves teaching practices.

Exhibit E19
Effect Sizes: Impacts on Student Academic Outcomes from NBPTS Certification

	Test scores (reading, math, & general academic)	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	14	Teacher recruitment and retention and self-reported impacts on teaching practices
Average effect on academic outcomes (standard error)	0.026 (0.004)	
Conclusion	Students who have teachers with NBPTS certification have slightly higher test scores, on average. The available evidence is inconclusive whether the certification recognizes already effective teachers or improves teaching practices.	

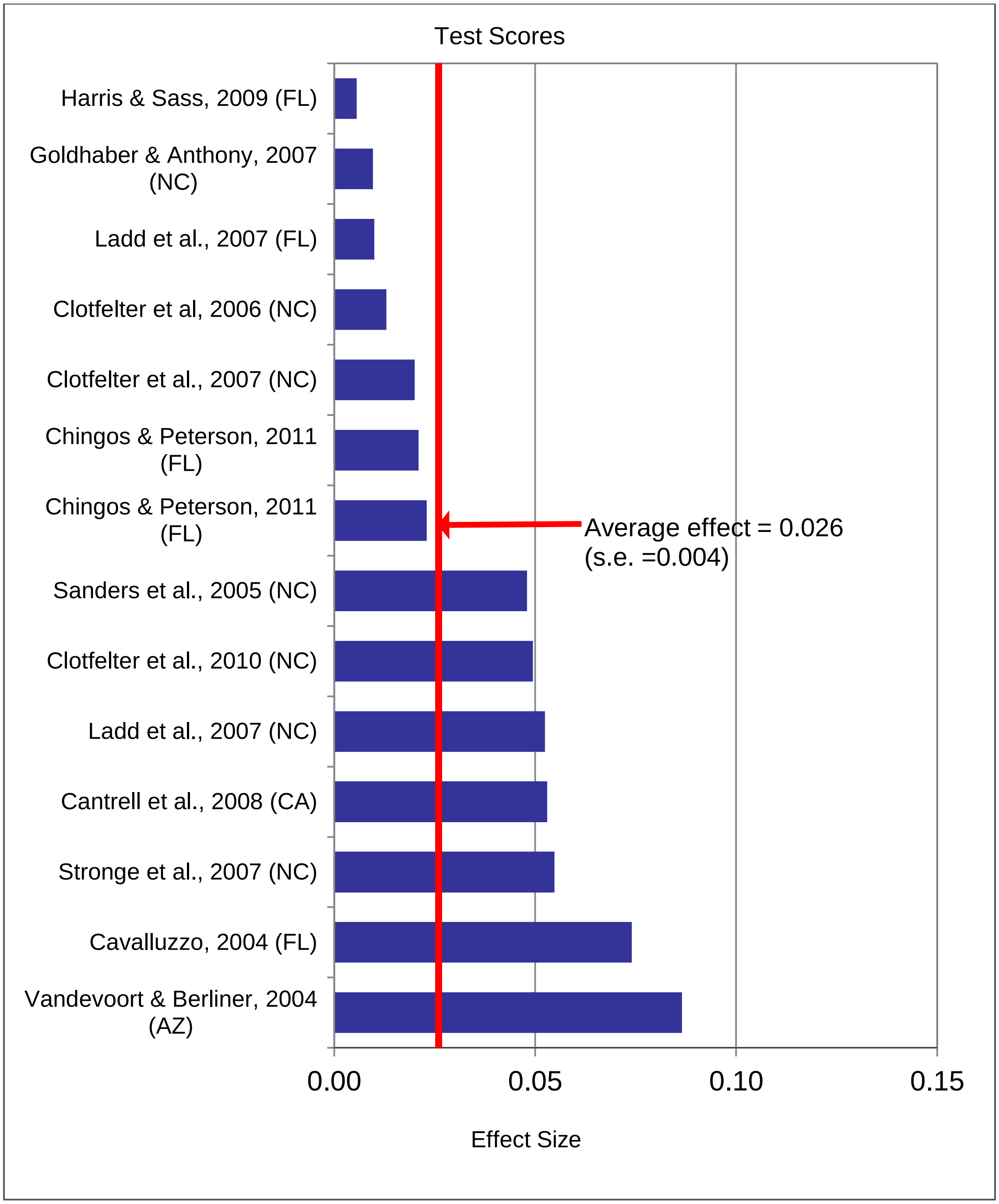
*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

[13] This meta-analytic review was last updated for an April 2012 Institute report. Lee, S., Aos, S., Drake, E., Pennucci, A., Miller, M., Anderson, L. (2012) *Return on Investment: Evidence-Based Options to Improve Statewide Outcomes April 2012 Update* (Document Number 12-04-1201). Olympia: Washington State Institute for Public Policy.

Exhibit E20

Effect Sizes: Student Academic Outcomes Associated with NBPTS Certification

Study by Study Results



Studies Used in the Meta-Analysis of NBPTS Certification and Student Academic Outcomes

Cantrell, S., Fullerton, J., Kane, T.J., & Staiger, D.O. (2008). *National board certification and teacher effectiveness: Evidence from a random assignment experiment* (Working Paper No. 14608). Cambridge, MA: National Bureau of Economic Research.

Cavalluzzo, L.C. (2004). *Is national board certification an effective signal of teacher quality?* Alexandria, VA: The CNA Corporation.

Chingos, M.M., & Peterson, P.E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review, 30*(3): 449-465.

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *The Journal of Human Resources, 41*(4): 778-820.

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review, 26(*6): 673-682.

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources, 45*(3): 655-681.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics, 89*(1): 134-150.

Harris, D.N., & Sass, T.R. (2007). *The effects of NBPTS-certified teachers on student achievement* (Working Paper 4). Washington, DC: The Urban Institute, National Center for Analysis of Longitudinal Data in Education Research.

Ladd, H.F., Sass, T.R., & Harris, D.N. (2007). *The impact of national board certified teachers on student achievement in Florida and North Carolina: A summary of the evidence prepared for the National Academies Committee on the evaluation of the impact of teacher certification by NBPTS*. Unpublished manuscript.

Sanders, W.L., Ashton, J.J., & Wright, S.P. (2005). *Comparison of the effects of NBPTS certified teachers with other teachers on the rate of student academic progress.* Final report. Retrieved from ERIC database.

Stronge, J.H., Ward, T.J., Tucker, P.D., Hindman, J.L., McColsky, W., & Howard, B. (2007). National Board Certified Teachers and Non-national Board Certified Teachers: Is there a difference in teacher effectiveness and student achievement? *Journal of Personnel Evaluation in Education, 20*(3-4): 185-210.

Vandevoort, L.G., Amrein-Beardsley, A., & Berliner, D.C. (2004). National Board Certified Teachers and their students' achievement. *Education Policy Analysis Archives, 12*(46).

E3e. National Guard Youth ChalleNGe

The Washington Youth Academy, one of Washington's designated innovative schools, operates under the National Guard Youth ChalleNGe program (ChalleNGe). This program was designed by the National Guard Bureau (Bureau) within the U.S. Department of Defense to help high school dropouts improve their long-term outcomes. The quasi-military residential program enrolls youths aged 16 to 18 who are unemployed, drug-free, and not heavily involved with the justice system. States can enter into "Master Cooperative Agreements" with the Bureau to operate ChalleNGe programs. Up to 75% of funding for ChalleNGe is provided by the federal government.

The Bureau and private organizations funded a random assignment evaluation of the ChalleNGe program in 12 states (not including Washington). The three-year follow-up results for selected outcomes measured in this multi-site, 2011 national study are presented in Exhibit E21. We did not perform a meta-analysis of the results because only a single study that met our coding criteria was identified; the effect size represents the national estimate for this program's high school graduation rate. Three citations are listed because a report was produced for each follow-up year during the evaluation.

Exhibit E21
Effect Sizes: Impacts on Student Academic Outcomes from ChalleNGe Programs

	High school graduation rates	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes estimated	1	Employment, housing, crime, health, substance abuse and GEDs
Average effect on academic outcomes	0.11	
Conclusion	ChalleNGe appears to have a positive impact on high school graduation rates and mixed impacts on other outcomes.	

*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

<u>Citations to the National Evaluation of the ChalleNGe Program</u>

Bloom, D., Gardenhire-Crooks, A., & Mandsager, C. (2009). *Reengaging high school dropouts: Early results of the National Guard Youth ChalleNGe Program Evaluation*. New York: MDRC.

Millenky, M., Bloom, D., Dillon, C. (2010). *Making the Transition: Interim Results of the National Guard Youth ChalleNGe Evaluation*. New York: MDRC.

Millenky, M., Bloom, D., Muller-Ravett, S., & Broadus, J. (2011). *Staying on Course: Three-Year Results of the National Guard Youth ChalleNGe Evaluation.* New York: MDRC.

E3f. Parent Involvement in Reading Instruction: School-based Programs

Parent engagement is a focus of many of Washington's designated innovative schools. We reviewed the research literature on school-based parent engagement programs but did not find sufficient credible evaluations to conduct a meta-analysis of this broader literature. We did, however, find evidence that elementary school-based programs that encourage parent involvement in reading instruction are associated with improved student reading outcomes, on average (see Exhibits 22 and 23).[14]

In a typical K-12 parent involvement program, teachers meet with parents in person and maintain contact over the phone to train and encourage parents to engage in planned, structured academic activities with their children at home, often in the form of tutoring.

Exhibit E22

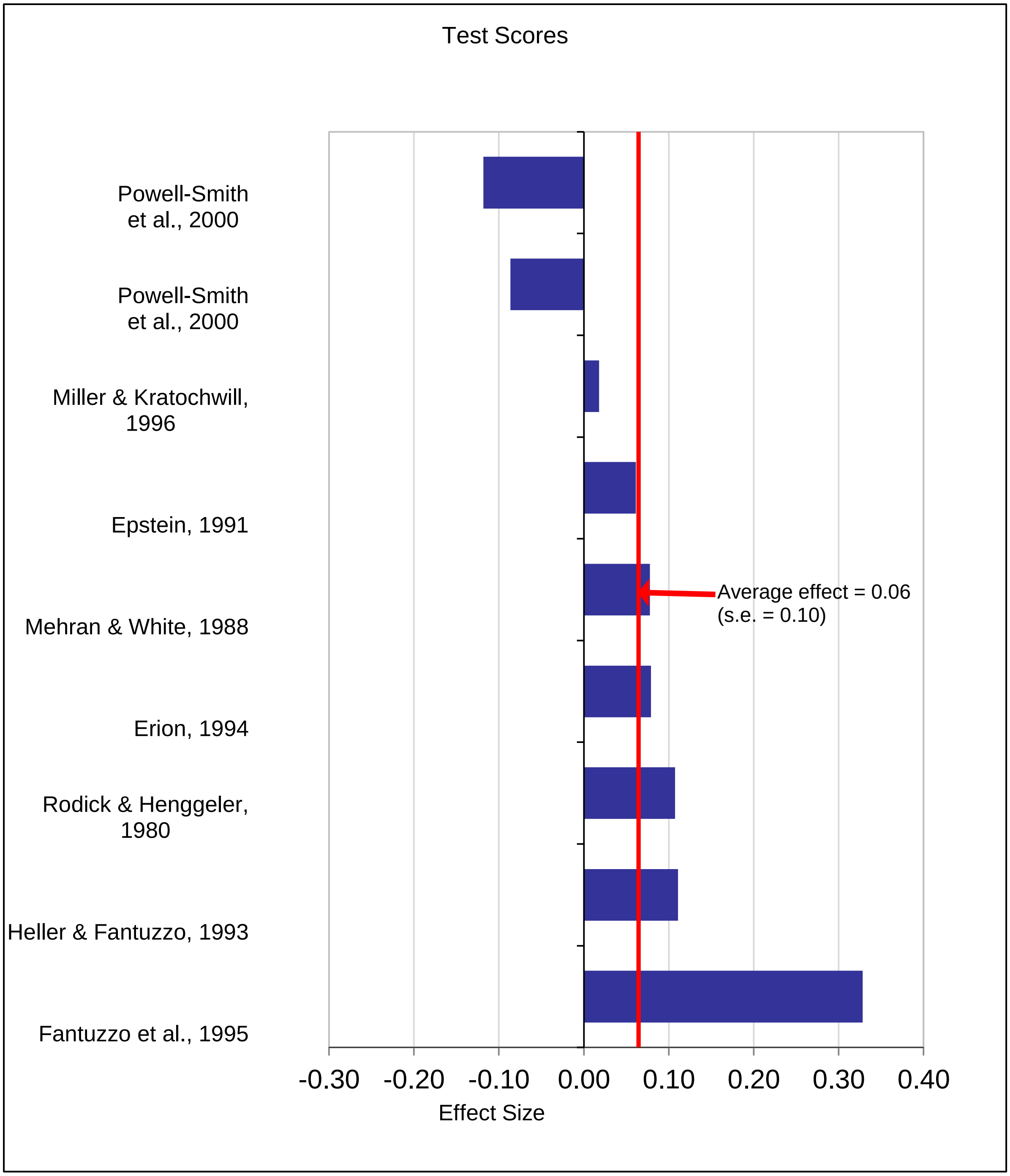
Effect Sizes: Impacts on Student Academic Outcomes from School-based Programs to Encourage Parent Involvement in Reading Instruction

	Test scores (reading, math, & general academic)	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	9	Parent, student, and teacher perceptions/satisfaction with program
Average effect on academic outcomes (standard error)	0.06 (0.10)	
Conclusion	Elementary school-based programs that encourage parent involvement in reading instruction are associated with improved student reading outcomes, on average.	

*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

[14] This meta-analytic review was last updated for an April 2012 Institute report. Lee, S., Aos, S., Drake, E., Pennucci, A., Miller, M., Anderson, L. (2012) *Return on Investment: Evidence-Based Options to Improve Statewide Outcomes April 2012 Update* (Document Number 12-04-1201). Olympia: Washington State Institute for Public Policy.

Exhibit E23
Effect Sizes: Impacts on Student Academic Outcomes from School-based Programs to Encourage Parent Involvement in Reading Instruction
Study by Study Results



Studies Used in the Meta-Analysis of Parent Involvement in Reading Instruction
(School-based Programs)

Epstein, J.L. (1991). Effects on student achievement of teachers' practices of parent involvement. In S.B. Silvern (Ed.), *Advances in Reading/Language Research (5)*: 261-276. Stamford, CT: JAI Press.

Erion, R.J. (1994). Parent tutoring, reading instruction and curricular assessment. *Dissertation Abstracts International, 54*(11): 4035A.

Fantuzzo, J.W., Davis, G.Y., & Ginsburg, M.D. (1995). Effects of parent involvement in isolation or in combination with peer tutoring on student self-concept and mathematics achievement. *Journal of Educational Psychology, 87*(2): 272-281.

Heller, L.R., & Fantuzzo, J.W. (1993). Reciprocal peer tutoring and parent partnership: Does parent involvement make a difference? *School Psychology Review, 22*(3): 517-534.

Mehran, M., & White, K.R. (1988). Parent tutoring as a supplement to compensatory education for first-grade children. *Remedial and Special Education, 9*(3): 35-41.

Miller, B.V., & Kratochwill, T.R. (1996). An evaluation of the Paired Reading Program using competency-based training. *School Psychology International, 17*(3): 269-291.

Powell-Smith, K.A., Shinn, M.R., Stoner, G., & Good, R.H., III. (2000). Parent tutoring in reading using literature and curriculum materials: Impact on student reading achievement. *School Psychology Review, 29*(1): 5-27.

Rodick, J.D., & Henggeler, S.W. (1980). The short-term and long-term amelioration of academic and motivational deficiencies among low-achieving inner-city adolescents. *Child Development, 51*(4): 1126-1132.

E3g. Principals (School Leadership)

In the site visits conducted for this study, researchers spent time observing and interviewing each school's principal—key implementers of the innovative strategies adopted by the schools. A small, but growing, research literature examines whether school principals directly affect student academic outcomes.

The studies in this analysis use a "fixed effects" statistical approach to examine variation in impacts on student test scores from different principals. The studies focus on principals that move from one school to another; variation in student outcomes can be estimated for different principals in the same school. The statistical models used in these evaluation studies typically include student, year, grade, and school fixed effects (in addition to principal fixed effects) in order to account for any achievement trends attributable to individual students, cohorts, grade levels, or schools (as opposed to the principals themselves). These methods allow us to quantify the variance, or distribution, of the impacts principals can have on student test score growth.

The overall results are presented in Exhibit E24 and the detailed results in Exhibit E25. The evidence confirms that school leadership affects student outcomes: a principal who is one standard deviation above typical principal effectiveness improves student test scores by about one-tenth of a standard deviation, on average.

Some principal impact research uses survey data or other methods to try to identify specific principal characteristics associated with greater school-wide achievement gains.[15] These studies include measures of years of administrative or teaching experience, teachers' perceptions of principals' leadership skills, focus on instructional time or quality, or a concept called "transformational leadership." These data are typically self-reported. Too few studies examine principal characteristics in a systematic way to be able to draw cause-and-effect conclusions regarding which characteristics are most important for student learning.

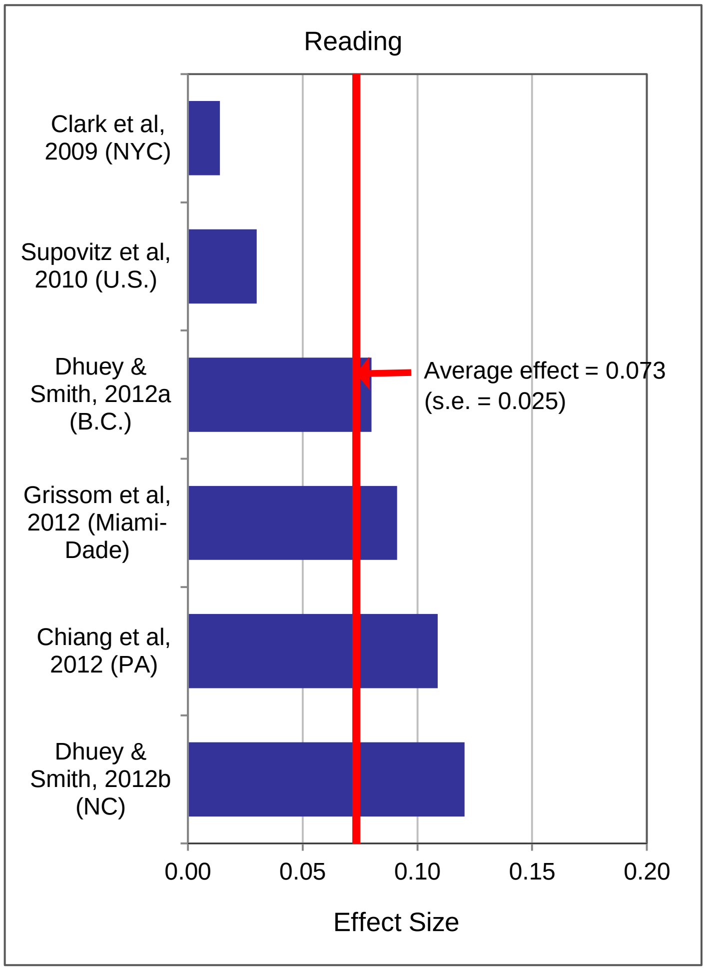
[15] Andrews, R., & Soder, R. (1987). *Principal Leadership and Student Achievement. Educational Leadership*, 44(6), 9.; Ballou, D., & Podgursky, M. (1995). What Makes a Good Principal? How Teachers Assess the Performance of Principals. *Economics Of Education Review*, 14(3), 243-52; Bell, L., Bolam, R. & Cubillo, L. (2003) *A systematic review of the impact of school leadership and management on student outcomes*. In: Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.; Bossert, S. T., & And, O. (1982). The Instructional Management Role of the Principal. *Educational Administration Quarterly*, 18(3), 34-64.; Brewer, D. J. (1993). Principals and Student Outcomes: Evidence from U.S. High Schools. *Economics of Education Review*, 12, 4, 281-92; Buck, S. (2012). *Estimating the effect of principal quality and experience*; Coelli, M., & Green, D. A. (2012). Leadership effects: school principals and student outcomes. *Economics of Education Review*, 31, 1, 92-109; Cullen, J. B., & Mazzeo, M. J. (2008). *Implicit performance awards: An empirical analysis of the labor market for public school administrators*. University of California, San Diego (December); Dean, J. (2012). *Leadership where it matters: Principal effectiveness and equity in Wisconsin*. University of Arkansas; Eberts, R. W., & Stone, J. A. (1988). Student achievement in public schools: Do principals make a difference? *Economics of Education Review*, 7(3), 291-299; Horng, E. L., Klasik, D., & Loeb, S. (2010). Principal's Time Use and School Effectiveness. *American Journal of Education*, 116, 4, 491-523; Human Capital Policies in Education: Further Research on Teachers and Principals. (2012). CALDER National Center for Analysis of Longitudinal Data in Education Research; Loeb, S., Kalogrides, D., & Horng, E. L. (2010). Principal Preferences and the Uneven Distribution of Principals Across Schools. *Educational Evaluation and Policy Analysis*, 32, 2, 205-229.

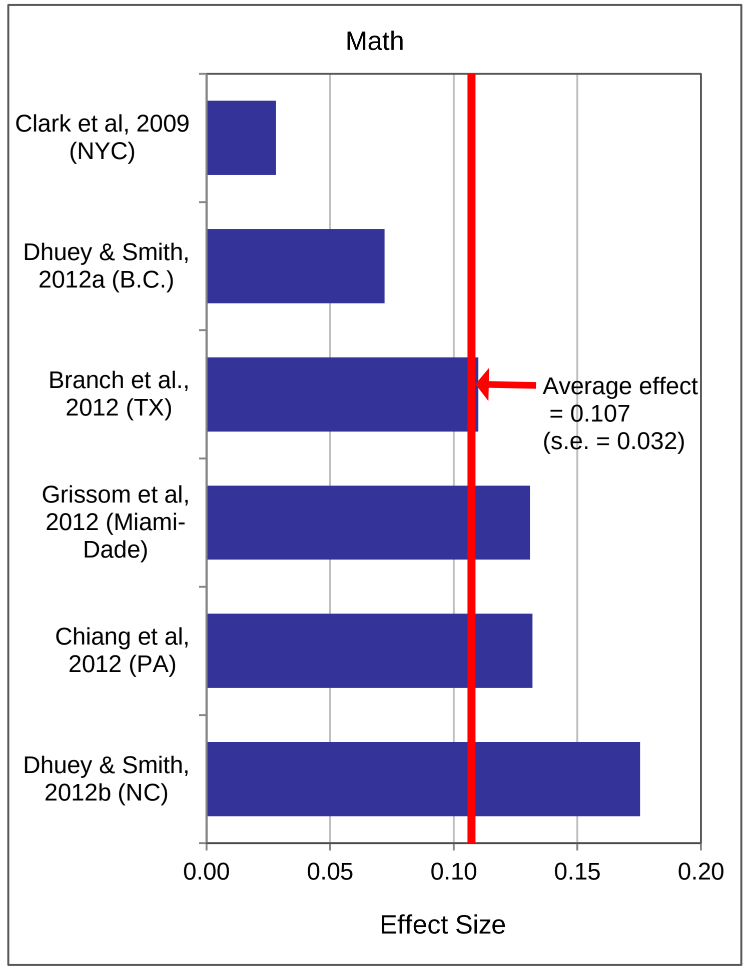
Exhibit E24
Effect Sizes: Impacts on Student Academic Outcomes from a Principal One Standard Deviation above Average

	Reading test scores	**Math test scores**	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	6	6	High school graduation (1 study ES= 0.04); self-reported measures of principal effectiveness
Average effect on academic outcomes (standard error)	0.073 (0.025)	0.107 (0.032)	
Conclusion	School leadership affects student outcomes: a principal who is one standard deviation above typical principal effectiveness can improve student test scores.		

*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

Exhibit E25
Effect Sizes: Impacts on Student Test Scores from a Principal One Standard Deviation above Average
Study by Study Results





<u>Studies Used in the Meta-Analysis of Principal Impacts on Student Academic Outcomes</u>

Branch, G.F., Hanushek, E.A., & Rivkin, S.G. (2012). *Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals* (Working Paper 17803). Cambridge, MA: National Bureau of Economic Research.

Chiang, H., Lipscomb, S., & Gill, B. (2012). *Is school value-added indicative of principal quality?* (Working Paper). Washington, DC: Mathematica Policy Research.

Clark, D., Martorell, P., & Rockoff, J. (2009). *School principals and school performance* (Working Paper 38). National Center for Analysis of Longitudinal Data in Education Research.

Dhuey, E., & Smith, J. (2012a). *How important are school principals in the production of student achievement?* Retrieved from The Society of Labor Economists website: http://sole-jole.org/13170.pdf.

Dhuey, E. & Smith, J. (2012b). *How school principals influence student learning* (Working Paper). Toronto, ON: University of Toronto.

Grissom, J.A., Kalogrides, D., & Loeb, S. (2012). *Using student test scores to measure principal performance* (Working Paper 18568). Cambridge, MA: National Bureau of Economic Research.

Supovitz, J., Sirinides, P., & May, H. (2010). How principals and peers influence teaching and learning. *Educational Administration Quarterly, 46*(1): 31-56.

E3h. Project Lead the Way

Project Lead the Way (PLTW) is an example of project-based learning focused on science, technology, engineering, and mathematics (STEM) education. PLTW is a nonprofit organization that develops engineering courses for high schools and middle schools and biomedical sciences courses for high schools. The curriculum is delivered through an online "virtual academy." Computer software and classroom materials for hands-on activities, as well as required teacher training, are the main costs related to the program. Toppenish High School, one of Washington's Innovative Schools, uses PLTW in its STEM program.

The overall PLTW results are presented in Exhibit E26 and the detailed results in Exhibit E27. The evidence suggests that PLTW has no consistent impact on student test score outcomes, although the average impact for math is positive.

Exhibit E26

Effect Sizes: Impacts on Student Academic Outcomes from Project Lead the Way

	Reading test scores	Math test scores	Other outcomes*
Effect sizes included in the meta-analysis	3	4	3 science test scores
Average effect on academic outcomes (standard error)	0.01 (0.06)	0.11 (0.06)	0.00 (0.10)
Conclusion	PLTW improves student math scores but does not consistently impact reading or science test scores.		

*Other outcomes not meta-analyzed for this report include student grade point averages and course-taking of advanced math and science courses or enrollment in higher education programs because those outcomes were not measured frequently enough or were measured in varied, non-standardized ways.

Studies Used in the Meta-Analysis of Project Lead the Way Effects on Student Academic Outcomes

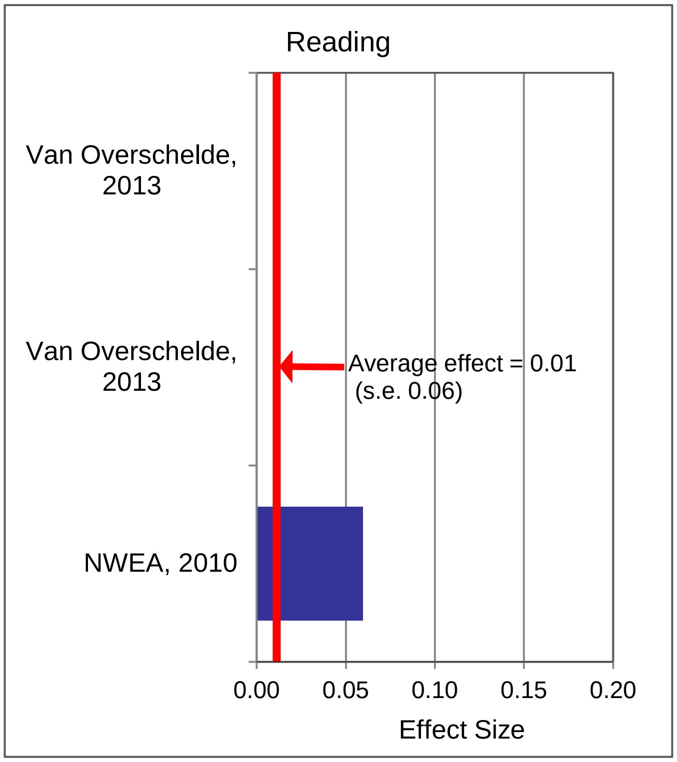
Northwest Evaluation Association. (2010). *Project Lead the Way - Initial Program Evaluation.* Portland, OR.

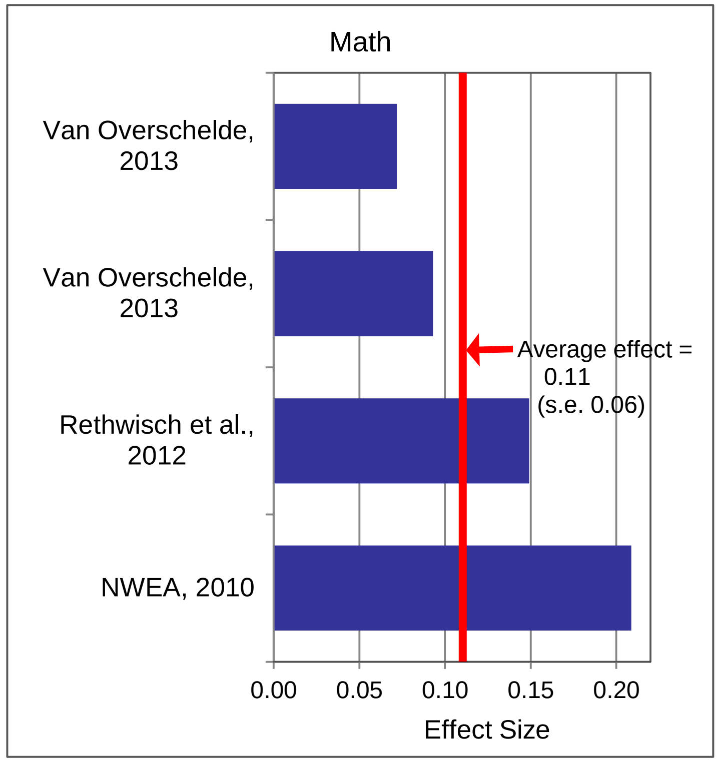
Rethwisch, D.G., Haynes, M.C., Starobin, S.S., Laanan, F.S., & Schenk, J.T. (2012). Proceedings from Asee Annual Conference and Exposition. *A study of the impact of Project Lead the Way on achievement outcomes in Iowa.* San Antonio, TX.

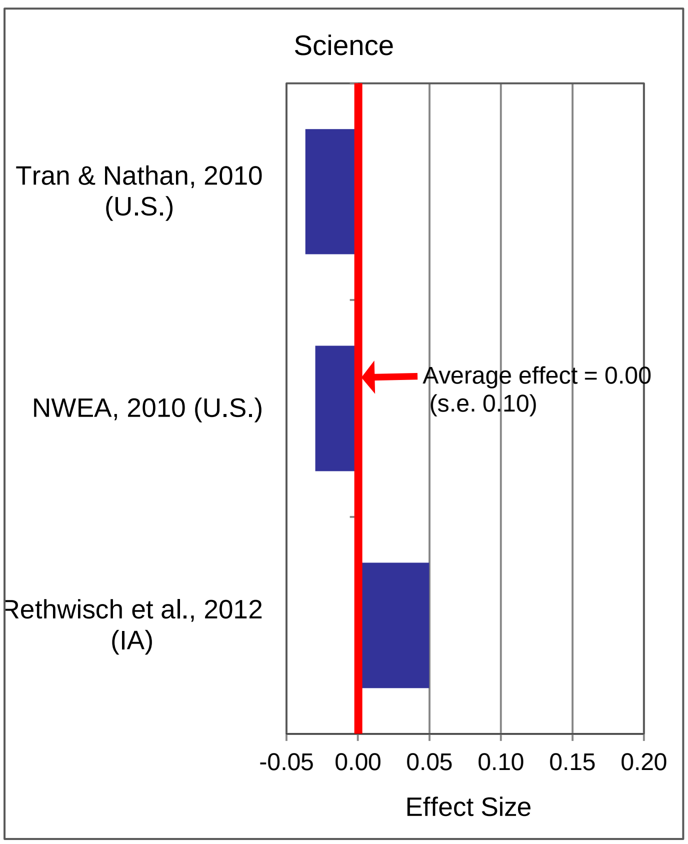
Tran, N.A., & Nathan, M.J. (2010). Pre-college engineering studies: An investigation of the relationship between pre-college engineering studies and student achievement in science and mathematics. *Journal of Engineering Education, 99(*2): 143-157.

Van Overschelde, J.P. (2013). *Project lead the way students more prepared for higher education.* San Marcos, TX: Texas State University.

Exhibit E27
Effect Sizes: Impacts on Student Test Scores from Project Lead the Way
Study by Study Results







E3i. School-wide Positive Behavior Programs

Some K-12 schools operate school-wide student behavior improvement programs as one way to focus the school environment on learning (rather than discipline or other issues). These programs are often described as "positive behavior" interventions or systems and include specific programs such as School-wide Positive Behavioral Interventions and Supports, Positive Action, and the Responsive Classroom. The programs encourage pro-social behavior for all students (in contrast, other interventions target problem behaviors among troubled students, who are not the focus of this analysis). School-wide behavior programs typically include a specialized curriculum, professional development for teachers and staff, and encouragement of and rewards for positive behaviors such as being on time and listening in the classroom.

The overall behavior program results are presented in Exhibit E28 and the detailed results in Exhibit E29. The evidence suggests that school-wide positive behavior programs can improve student academic outcomes.

Many evaluations of school-wide behavior programs also measure outcomes such as attendance and discipline (office discipline referrals, suspensions, and expulsions).[16] A related area of research examines students' social-emotional competencies and attitudes about themselves, other people, and school. Some studies found that school-wide behavior programs improve student outcomes in terms of externalizing (e.g., poor conduct) and internalizing (e.g., depression).[17] Those outcomes are measured in varied, non-standardized ways and thus are not meta-analyzed in this report.

---

[16] See, e.g., Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions*, 12, 3, 133-148; Caldarella, P., Shatzer, R.H., Gray, K.M., Young, K.R, & Young, E.L. (2011). The effects of schools-wide positive behavior support on middle school climate and student outcomes. *Research in Middle Level Education*, 35(4): 1-14; Forster, M., Sundell, K., Morris, R. J., Karlberg, M., & Melin, L. (2010). A Randomized Controlled Trial of a Standardized Behavior Management Intervention for Students With Externalizing Behavior. *Journal of Emotional and Behavioral Disorders*; Waasdorp, T. E., Bradshaw, C. P., & Leaf, P. J. (2012). The Impact of Schoolwide Positive Behavioral Interventions and Supports on Bullying and Peer Rejection A Randomized Controlled Effectiveness Trial. *Archives Of Pediatrics & Adolescent Medicine*, 166(2), 149-156.

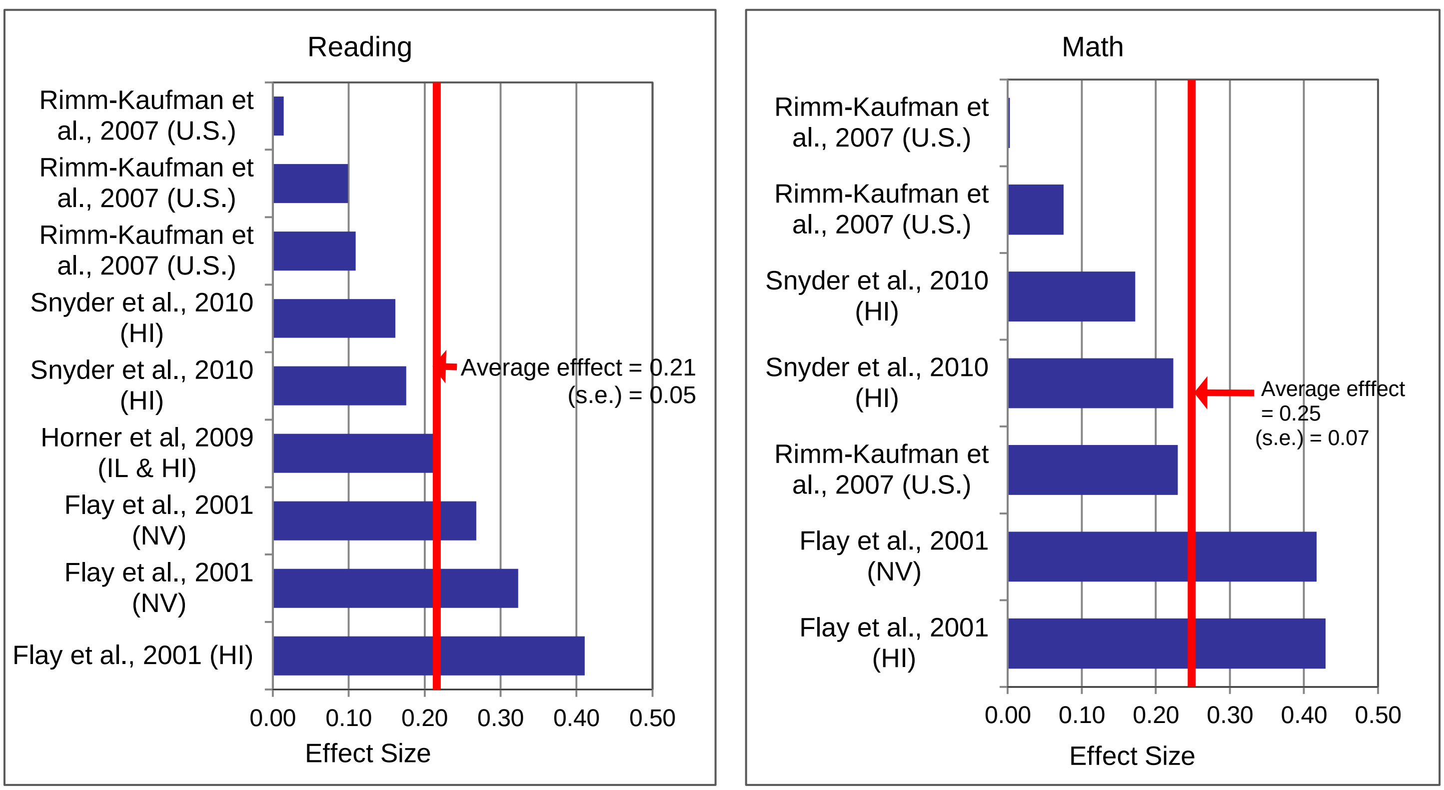
[17] Durlak, J., Weissberg, R., Dymnicki, A., Taylor, R., & Schellinger, K. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405-432.

Exhibit E28
Effect Sizes: Impacts on Student Academic Outcomes from School-wide Positive Behavior Programs

	Reading test scores	**Math test scores**	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	9	7	Attendance, grade retention, and discipline (office discipline referrals, suspensions, and expulsion)
Average effect on academic outcomes (standard error)	0.21 (0.05)	0.25 (0.07)	
Conclusion	School-wide interventions focused on encouraging positive behavior can improve academic outcomes (math and reading test scores).		

*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

Exhibit E29
Effect Sizes: Impacts on Student Test Scores from K-12 School-wide Behavior Programs
Study by Study Results



Studies Used in the Meta-Analysis of K-12 School-wide Behavior Programs Effects on Student Academic Outcomes

Flay, B.R., Allred, C.G., & Ordway, N. (2001). Effects of the Positive Action Program on achievement and discipline: Two matched-control comparisons. *Prevention Science 2*(2): 71-89.

Horner, R.H., Smolkowski, K., Todd, A.W., Esperanza, J., Sugai, G., Eber, L., et al. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions 11*(3): 133-144.

Rimm-Kaufman, S., Fan, X., Chiu, Y., & You, W. (2007). The contribution of the Responsive Classroom Approach on children's academic achievement: Results from a three year longitudinal study. *Journal of School Psychology 45*: 401-421.

Snyder, F., Vuchinich, S., Acock, A., Washburn, I., Beets, M., & Li, K. (2010). Impact of the Positive Action program on school-level indicators of academic achievement, absenteeism, and disciplinary outcomes: A matched-pair, cluster randomized, controlled trial. *Journal of Research On Educational Effectiveness, 3*(1): 26-55.

E3j. Teaching Induction/Mentoring

In many of the schools in this study, induction programs are provided to new teachers who have no prior classroom experience. In these programs, a veteran teacher mentors a novice teacher, offering guidance and support in the new teacher's first and often second years at the school. Some induction programs provide additional support such as professional development, structured peer group interaction, and observation of veteran teachers.

Washington State's Beginning Educator Support Team (BEST) program provides grants to districts and schools to help implement teacher induction programs. BEST grants were awarded to 28 districts (some as part of a consortium) in the 2011-12 school year.[18]

The overall teacher induction results are presented in Exhibit E30 and the detailed results in Exhibit E31.[19] The studies of teacher induction/mentoring compare more intensive programs to "induction-as-usual," because some form of mentoring (often informal) was typically already occurring in the schools studied. The evidence of effectiveness is mixed but positive on average.

Exhibit E30

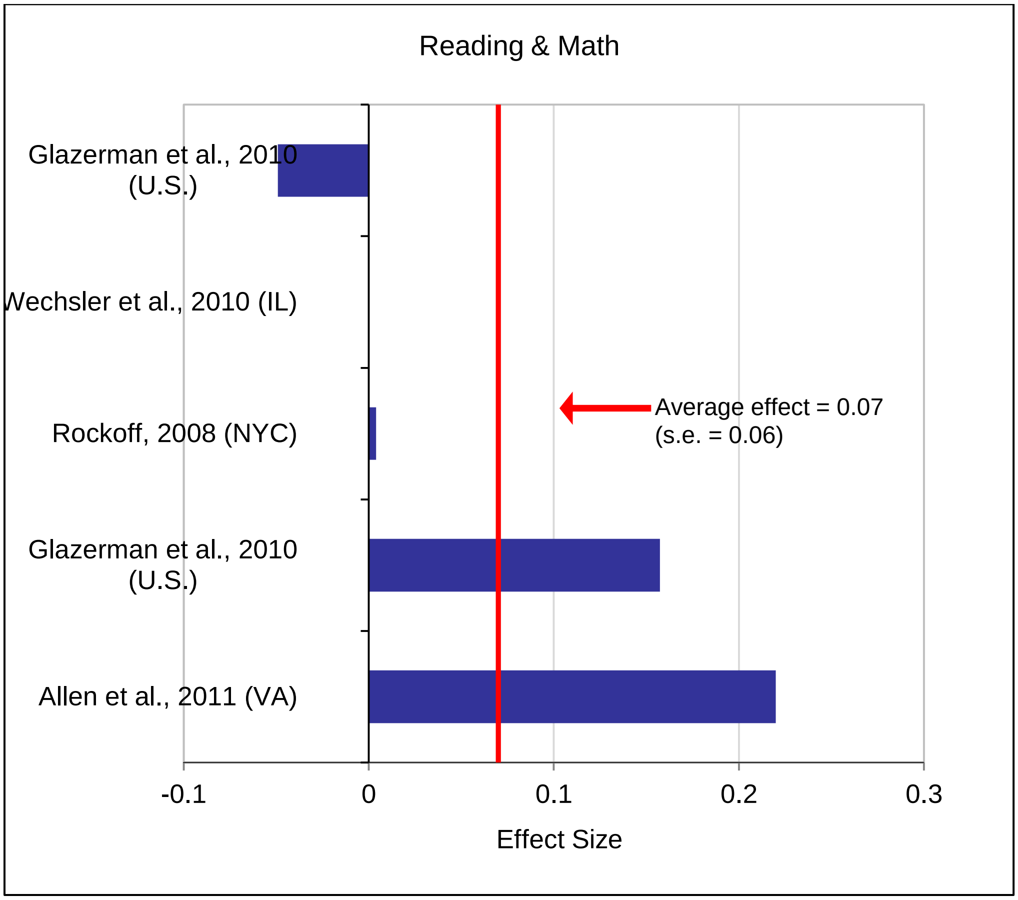
Effect Sizes: Impacts on Student Academic Outcomes from Teacher Induction/Mentoring

	Reading and math test scores	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	5	Teacher retention; self-reported measures of teacher outcomes
Average effect on academic outcomes (standard error)	0.07 (0.06)	
Conclusion	For teacher induction programs, the results are mixed, but the average impact is positive.	

*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

[18] For more information, visit: http://www.k12.wa.us/BEST/default.aspx.
[19] This meta-analytic review was last updated for an April 2012 Institute report Lee, S., Aos, S., Drake, E., Pennucci, A., Miller, M., Anderson, L. (2012) *Return on Investment: Evidence-Based Options to Improve Statewide Outcomes April 2012 Update* (Document No.12-04-1201). Olympia: Washington State Institute for Public Policy.

Exhibit E31
Estimates of the Effect of Teacher Induction Programs on Student Outcomes
Study by Study Results



Studies Used in the Meta-Analysis of Teacher Induction/Mentoring Program Effects on Student Academic Outcomes

Allen, J.P., Mikami, A.Y., Pianta, R.C., Gregory, A., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333*(6045): 1034-1037.

Glazerman, S., Isenberg, E., Dolfin, S., Bleeker, M., Johnson, A., Grider, M., et al. (2010). *Impacts of comprehensive teacher induction: Final results from a randomized controlled study.* Washington, DC: National Center for Education Evaluation and Regional Assistance.

Rockoff, J.E. (2008). *Does mentoring reduce turnover and improve skills of new employees?: Evidence from teachers in New York City* (Working Paper No. 13868). Cambridge, MA: National Bureau of Economic Research.

Wechsler, M.E., Caspary, K., Humphrey, D.C., & Matsko, K.K. (2010). *Examining the effects of new teacher induction.* Menlo Park, CA: SRI International.

E3k. Teacher Professional Development

In Washington, as in other states, teachers must complete certain professional development (PD) requirements in order to maintain certification and add endorsements.[20] Enhancing PD for educators is the focus of many of the schools in this study. We analyze research that examines impacts on student test scores from various approaches to teacher PD. The analysis addresses a basic question: What are the potential impacts from putting more resources into teacher training?[21]

The specific approaches studied are diverse. We organize the research literature into two categories—“general” and “content-specific”—broadly defined. Studies of general PD measure training in terms of time (variation in total in-service hours among teachers) or additional PD resources given to struggling schools to use at the schools’ discretion. Content-specific PD focuses on instructional strategies specific to a grade level and subject area. For both categories, the increased time or new approach is compared with professional development as-usual.

For this analysis, we standardize all measured impacts in terms of the effect of an additional day (eight hours) of training. Because teachers typically participate in more than one day of PD per year, the actual impacts are larger than shown in Exhibits E32 and E33. Many of the programs studied, particularly for content-specific PD, involve two-week summer institutes with follow-up sessions during the school year. Thus, these content-specific findings may also reflect the amount and structure of training.

Exhibit E32
Effect Sizes: Impacts on Student Academic Outcomes from Teacher Induction/Mentoring

	General PD, (reading & math test scores)	**Content-Specific PD, (reading & math test scores)**	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	4	9	Teacher retention; self-reported measures of teacher outcomes
Average effect on academic outcomes (standard error)	0.000 (0.00)	0.005 (0.00)	
Conclusion	Providing more professional development in general is not associated with improved student test scores. Content-specific professional development is associated with improved student test scores.		

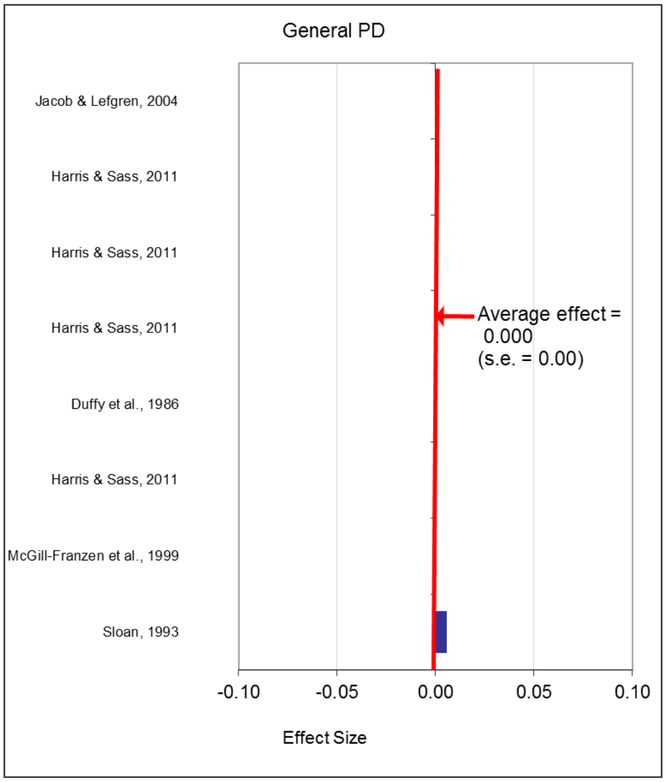
*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

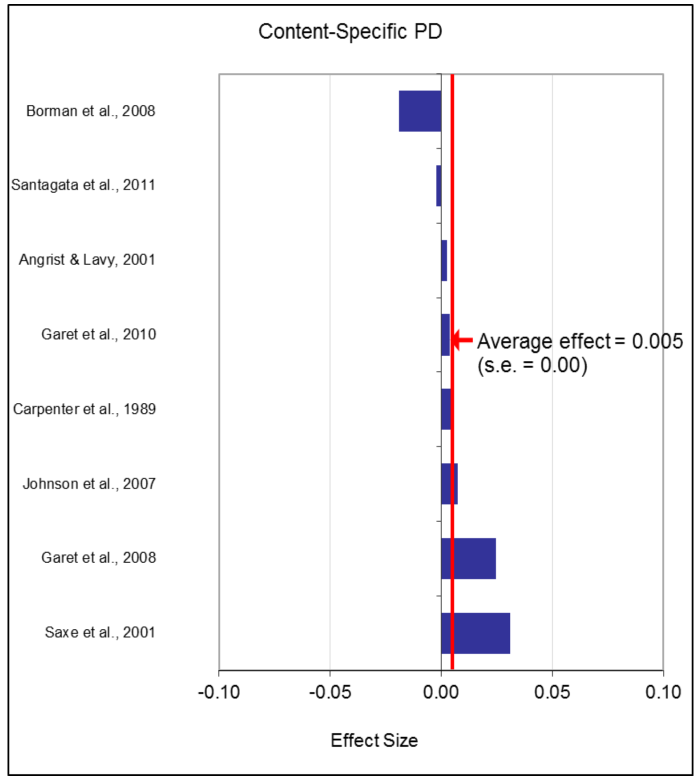
[20] For more information, visit: http://www.pesb.wa.gov and http://www.k12.wa.us/certification/TeacherMain.aspx.
[21] This meta-analytic review was last updated for an April 2012 Institute report: Lee, S., Aos, S., Drake, E., Pennucci, A., Miller, M., Anderson, L. (2012) *Return on Investment: Evidence-Based Options to Improve Statewide Outcomes April 2012 Update* (Document No. 12-04-1201). Olympia: Washington State Institute for Public Policy.

Exhibit E33

Estimates of the Effect of an Additional Day of Teacher Professional Development on Student Outcomes

Study by Study Results





<u>Studies Used in the Meta-Analysis of General Teacher PD Effects on Student Academic Outcomes</u>

Duffy, G. G., Roehler, L. R., Meloth, M. S., Vavrus, L. G., Book, C., Putnam, J., & Wesselman, R. (1986). The relationship between explicit verbal explanations during reading skill instruction and student awareness and achievement: A study of reading teacher effects. *Reading Research Quarterly*, 21(3), 237-252.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. J*ournal of Public Economics*, 95(7-8), 798-812.

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *The Review of Economics and Statistics*, 86(1), 226-244.

McGill-Franzen, A., Allington, R. L., Yokoi, L., & Brooks, G. (1999). Putting books in the classroom seems necessary but not sufficient. *The Journal of Educational Research*, 93(2), 67-74.

Sloan, H. A. (1993). Direct instruction in fourth and fifth grade classrooms. *Dissertation Abstracts International,* 54(08), 2837A.

Studies Used in the Meta-Analysis of Content-specific Teacher PD
Effects on Student Academic Outcomes

Santagata, R., Kersting, N., Givvin, K.B., & Stigler, J W. (2011). Problem implementation as a lever for change: An experimental study of the effects of a professional development program on students' mathematics learning. *Journal of Research on Educational Effectiveness, 4*(1): 1-24.

Johnson, C.C., Kahle, J.B., & Fargo, J.D. (2007). A study of the effect of sustained, whole-school professional development on student achievement in science. *Journal of Research in Science Teaching, 44*(6): 775-786.

Garet, M.S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., . . . Silverberg, M. (2008, September). *The impact of two professional development interventions on early reading instruction and achievement.* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences.

Garet, M.S., Wayne, A.J., Stancavage, F., Taylor, J., Walters, K., Song, M., et al. (2010, April). *Middle school mathematics professional development impact study: Findings after the first year of implementation.* Washington, DC: National Center for Education.

Angrist, J.D., & Lavy, V. (2001). Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools. *Journal of Labor Economics, 19*(2): 343-369.

Borman, G.D., Gamoran, A., & Bowdon, J. (2008). A randomized trial of teacher development in elementary science: First-year achievement effects. *Journal of Research on Educational Effectiveness, 1*(4): 237-264.

Carpenter, T.P., Fennema, E., Peterson, P.L., Chiang, C.P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*(4): 499-531.

Saxe, G., Gearhart, M., & Nasir, N. (2001). Enhancing students' understanding of mathematics: A study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education, 4*(1): 55-79.

E3I. Tutoring

We reviewed the evaluation literature on the impact of one-on-one tutoring programs in K-12 schools. The types of tutoring programs that have been rigorously evaluated focus on reading instruction for elementary school students. We group the evaluation studies into three categories: Reading Recovery, peer tutoring, and tutoring by adults.

Reading Recovery is a structured early literacy tutoring intervention for struggling readers, typically in the first grade. The program was developed in New Zealand and has been implemented and evaluated in other countries, including the United States. Teachers trained in Reading Recovery techniques provide the tutoring. We analyze this approach separately because there are a sufficient number of rigorous evaluations to do so.

Peer tutoring programs use students from the same classroom, or sometimes from higher grade levels, to provide one-on-one assistance to other students who are struggling to learn to read. Classroom teachers provide guidance and oversight.

Tutoring by adults programs typically use adult community volunteers, often pre-service teachers in training, to provide one-on-one assistance to first graders struggling to learn to read. Three studies examined the use of certified teachers as tutors, but we did not have sufficient evaluations to separately examine the impact of using teachers as tutors.

The overall tutoring results are presented in Exhibit E34 and the detailed results in Exhibit E35.[22] One-on-one tutoring is an effective way to improve reading test scores.

Exhibit E34

Effect Sizes: Impacts on Student Reading Test Scores from One-on-One Tutoring

	Reading Recovery	**Peer tutoring**	**Tutoring by adults**	**Other outcomes examined in the research literature** (not meta-analyzed for this report)*
Effect sizes included in the meta-analysis	6	9	28	Parent, student, and teacher satisfaction with program
Average effect on academic outcomes (standard error)	0.34 (0.09)	0.22 (0.08)	0.12 (0.05)	
Conclusion	One-on-one tutoring is an effective way to improve student reading test scores.			

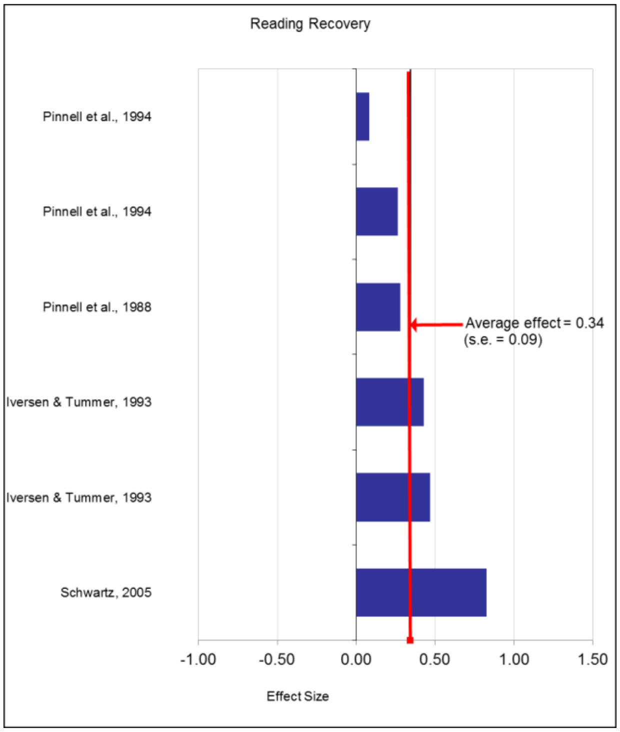
*We did not meta-analyze these outcomes because they were not the focus on this study, were not measured frequently enough to include in meta-analysis, or were measured in varied, non-standardized ways.

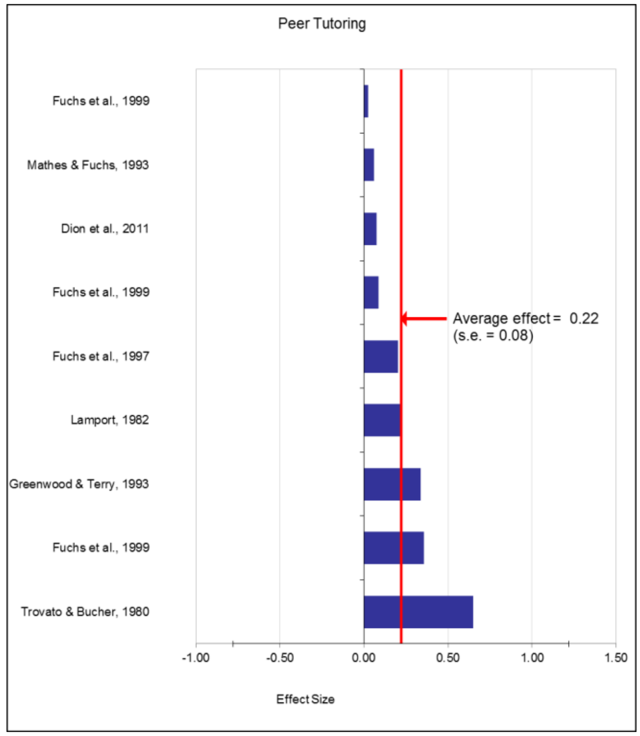
[22] This meta-analytic review was last updated for an April 2012 Institute report. Lee, S., Aos, S., Drake, E., Pennucci, A., Miller, M., Anderson, L. (2012) *Return on Investment: Evidence-Based Options to Improve Statewide Outcomes April 2012 Update* (Document No. 12-04-1201). Olympia: Washington State Institute for Public Policy.

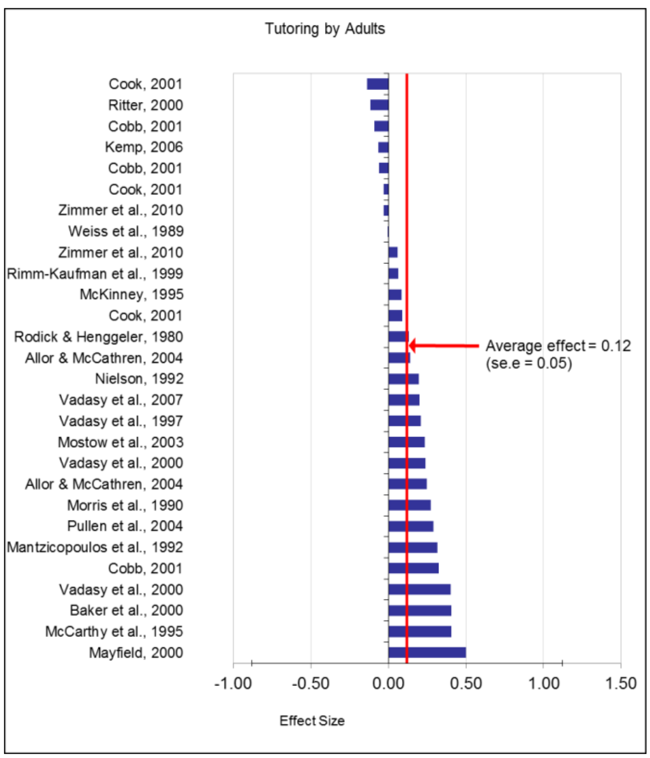
Exhibit E35

Effect Sizes: Impacts on Student Reading Test Scores from One-on-One Tutoring

Study by Study Results







Studies Used in the Meta-Analysis of Reading Recovery Effects on Reading Test Scores

Iversen, S., & Tunmer, W.E. (1993). Phonological processing skills and the Reading Recovery program. *Journal of Educational Psychology, 85*(1): 112-126.

Pinnell, G.S., DeFord, D.E., & Lyons, C.A. (1988). *Reading recovery: Early intervention for at-risk first graders.* Arlington, VA: Educational Research Service. (ERIC Document Reproduction Service No. ED 303790)

Pinnell, G.S., Lyons, C.A., DeFord, D.E., Bryk, A.S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly, 29*(1): 9-39.

Schwartz, R.M. (2005). Literacy learning of at-risk first-grade students in the reading recovery early intervention. *Journal of Educational Psychology, 97*(2): 257-267

Studies Used in the Meta-Analysis of Peer Tutoring Effects on Reading Test Scores

Dion, E., Roux, C., Landry, D., Fuchs, D., Wehby, J., & Dupere, V. (2011). Improving attention and preventing reading difficulties among low-income first-graders: A randomized study. *Prevention Science, 12*(1): 70-79.

Fuchs, D., Fuchs, L.S., Mathes, P.G., & Simmons, D.C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal, 34*(1): 174-206.

Fuchs, L., Fuchs, D., & Kazdan, S. (1999). Effects of peer-assisted learning strategies on high school students with serious reading problems. *Remedial and Special Education, 20*(5): 309-318.

Fuchs, L.S., Fuchs, D., Kazdan, S., & Allen, S. (1999). Effects of peer-assisted learning strategies in reading with and without training in elaborated help giving. *The Elementary School Journal, 99*(3): 201-219.

Greenwood, C.R., & Terry, B. (1993). Achievement, placement, and services: Middle school benefits of classwide peer tutoring used at the elementary school. *School Psychology Review, 22*(3): 497-516.

Lamport, K.C. (1983). The effects of inverse tutoring on reading disabled students in a public school setting. *Dissertation Abstracts International, 44*(03): 729A.

Mathes, P.G., & Fuchs, L.S. (1993). Peer-mediated reading instruction in special education resource rooms. *Learning Disabilities Research and Practice, 8*(4): 233-243.

Trovato, J., & Bucher, B. (1980). Peer tutoring with or without home-based reinforcement, for reading remediation. *Journal of Applied Behavior Analysis, 13*(1): 129-41.

Studies Used in the Meta-Analysis of Tutoring by Adults Effects on Reading Test Scores

Allor, J., & McCathren, R. (2004). The efficacy of an early literacy tutoring program implemented by college students. *Learning Disabilities Research and Practice, 19*(2): 116-129.

Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly, 35*(4): 494-519.

Cobb, J.B. (2001). The effects of an early intervention program with preservice teachers as tutors on the reading achievement of primary grade at risk children. *Reading Horizons, 41*(3): 155-173.

Cook, J.A. (2001). Every moment counts: Pairing struggling young readers with minimally trained tutors. *Dissertation Abstracts International, 62*(08): 2714A.

Kemp, S.C. (2006). Teaching to read naturally: Examination of a fluency training program for third grade students. *Dissertation Abstracts International, 67(*07A): 2447A.

Mantzicopoulos, P., Morrison, D., Stone, E., & Setrakian, W. (1992). Use of the SEARCH/TEACH tutoring approach with middle-class students at risk for reading failure. *Elementary School Journal, 92*(5): 573-586.

Mayfield, L.G. (2000). The effects of structured one-on-one tutoring in sight word recognition of first-grade students at-risk for reading failure. *Dissertation Abstracts International, 61*(02): 481A.

McCarthy, P., Newby, R.F., & Recht, D.R. (1995). Results of an early intervention program for first grade children at risk for reading disability. *Reading Research and Instruction, 34*(4): 273-294.

McKinney, A.D. (1995). The effects of an after-school tutorial and enrichment program on the academic achievement and self-concept of below grade level first and second grade students. *Dissertation Abstracts International, 56*(06): 2176A.

Morris, D., Shaw, B., & Perney, J. (1990). Helping low readers in grades 2 and 3: An after-school volunteer tutoring program. *Elementary School Journal, 91*(2): 133-150.

Mostow, J., Aist, G., Burkhead, P., Corbett, A., Cuneo, A., Eitelman, S., et al. (2003). Evaluation of an automated reading tutor that listens: Comparison to human tutoring and classroom instruction. *Journal of Educational Computing Research, 29*(1): 61-117.

Nielson, B.B. (1992). Effects of parent and volunteer tutoring on reading achievement of third grade at-risk students. *Dissertation Abstracts International, 52*(10): 3570A.

Pullen, P.C., Lane, H.B., & Monaghan, M.C. (2004). Effects of a volunteer tutoring model on the early literacy development of struggling first grade students. *Reading Research and Instruction, 43*(4): 21-40.

Rimm-Kaufman, S.E., Kagan, J., & Byers, H. (1999). The effectiveness of adult volunteer tutoring on reading among "at risk" first grade children. *Reading Research and Instruction, 38*(2): 143-152.

Ritter, G.W. (2000). The academic impact of volunteer tutoring in urban public elementary schools: Results of an experimental design evaluation. *Dissertation Abstracts International, 61*(03): 890A.

Rodick, J.D., & Henggeler, S.W. (1980). The short-term and long-term amelioration of academic and motivational deficiencies among low-achieving inner-city adolescents. *Child Development, 51*(4): 1126-1132.

Vadasy, P.F., Jenkins, J.R., Antil, L.R., Wayne, S.K., & O'Connor, R.E. (1997). The effectiveness of one-to-one tutoring by community tutors for at-risk beginning readers. *Learning Disability Quarterly, 20*(2): 126-139.

Vadasy, P.F., Jenkins, J.R., & Pool, K. (2000). Effects of tutoring in phonological and early reading skills on students at risk for reading disabilities. *Journal of Learning Disabilities, 33*(6): 579-590.

Vadasy, P.F., Sanders, E.A., & Tudor, S. (2007). Effectiveness of paraeducator-supplemented individual instruction: Beyond basic decoding skills. *Journal of Learning Disabilities, 40*(6): 508-525.

Weiss, J.A., Thurlow, M.L., Christenson, S.L., & Ysseldyke, J.E. (1989, March). Paired reading with adult volunteer tutors as a reading intervention for students with reading difficulties. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. Retrieved from ERIC database. (ED305606)

Zimmer, R., Hamilton, L., & Christina, R. (2010). After-school tutoring in the context of No Child Left Behind: Effectiveness of two programs in the Pittsburgh Public Schools. *Economics of Education Review, 29*(1): 18-28.

E4. META-ANALYSIS METHODOLOGY

E4a. Study Selection and Coding Criteria

A meta-analysis is only as good as the selection and coding criteria used to conduct the study.[23] Following are the key choices we made and implemented.

Study Selection. We use four primary means to locate studies for meta-analysis of programs: (1) we consult the bibliographies of systematic and narrative reviews of the research literature in the various topic areas; (2) we examine the citations in the individual studies themselves; (3) we conduct independent literature searches of research databases using search engines such as Google, Proquest, Ebsco, ERIC, PubMed, and SAGE; and (4) we contact authors of primary research to learn about ongoing or unpublished evaluation work. After first identifying all possible studies via these search methods, we attempt to determine whether the study is an outcome evaluation that has a valid comparison group. If a study meets the criterion, we secure a paper copy of the study for our review.

Peer-Reviewed and Other Studies. We examine all evaluation studies we can locate with these search procedures. Many studies are published in peer-reviewed academic journals while others are from reports obtained from the agencies themselves. It is important to include non-peer reviewed studies, because it has been suggested that peer-reviewed publications may be biased to show positive program effects. Therefore, our meta-analysis includes all available studies that meet our other criteria, regardless of publication source.

Control and Comparison Group Studies. Our analysis only includes studies that have a control or comparison group or use a quasi-experimental design such as regression discontinuity with multiple, sophisticated controls. We do not include studies with a single-group, pre-post research design. This choice was made because it is only through rigorous studies that causal relationships can be reliably estimated.

Random Assignment and Quasi-Experiments. Random assignment studies are preferred for inclusion in our review, but we also include non-randomly assigned comparison groups. We only include quasi-experimental studies if sufficient information is provided to demonstrate comparability between the treatment and comparison groups on important pre-existing conditions such as age, gender, and pre-treatment characteristics such as test scores.

Enough Information to Calculate an Effect Size. Following the statistical procedures in Lipsey and Wilson,[24] a study has to provide the necessary information to calculate an effect size. If the necessary information is not provided, and we are unable to obtain the necessary information directly from the study's author(s), the study is not included in our review.

Mean-Difference Effect Sizes. For this study, we code mean-difference effect sizes for continuous measures following the procedures outlined in Lipsey and Wilson.[25] For dichotomous measures, we use the d-Cox transformation to approximate the mean difference effect size, as described in Sánchez-Meca, Marín-Martínez, and Chacón-Moscoso.[26] We choose to use the mean-difference effect size rather than the odds ratio effect size because we frequently code both dichotomous and continuous outcomes (odds ratio effect sizes could also be used with appropriate transformations).

Outcome Measures of Interest. Our primary outcomes of interest include standardized, validated assessments of student learning. Most of the studies control for students' prior test scores using a value-

[23] All studies used in the meta-analysis are identified in the references to this paper. Many other studies were reviewed, but did not meet the criteria set for this analysis.
[24] Lipsey & Wilson, 2001.
[25] Ibid.
[26] Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, *8*(4), 448-467.

added model. Reading and math test scores are the most frequently measured outcomes. Some students also measure growth in science or general academic test scores. We also include measures of high school graduation and dropout rates when available.

E4b. Procedures for Calculating Effect Sizes

Effect sizes summarize the degree to which a program or policy affects an outcome. In experimental settings this involves comparing the outcomes of treated participants relative to untreated participants. There are several methods used by analysts to calculate effect sizes, as described in Lipsey and Wilson.[27] The most common effect size statistic is the standardized mean difference effect size, and that is the measure we use in this analysis.

Weighted Mean Different Effect Size. The mean difference effect size is designed to accommodate continuous outcome data, such as student test scores, where the differences are in the means of the outcome.[28] The standardized mean difference effect size is computed with:

$$(1)\ ES = \frac{M_t - M_c}{\sqrt{\frac{(N_t - 1)SD_t^2 + (N_c - 1)SD_c^2}{N_t + N_c - 2}}}$$

In this formula, ES is the estimated effect size for a particular program; M_t is the mean value of an outcome for the treatment or experimental group; M_c is the mean value of an outcome for the control group; SD_t is the standard deviation of the treatment group; and SD_c is the standard deviation of the control group; N_t is the number of subjects in the treatment group; and N_c is the number of subjects in the control group. The variance of the mean difference effect size statistic in (1) is computed with:[29]

$$(2)\ ESVar = \frac{N_t + N_c}{N_t N_c} + \frac{ES^2}{2(N_t + N_c)}$$

In some random assignment studies or studies where treatment and comparison groups are well-matched, authors provide only statistical results from a t-test. In those cases, we calculate the mean difference effect size using:[30]

$$(3)\ ES = t\sqrt{\frac{N_t + N_c}{N_t N_c}}$$

In many research studies, the numerator in (1), M_t - M_c, is obtained from a coefficient in a regression equation, not from experimental studies of separate treatment and control groups. For such studies, the denominator in (1) is the standard deviation for the entire sample. In these types of regression studies, unless information is presented that allows the number of subjects in the treatment condition to be separated from the total number in a regression analysis, the total N from the regression is used for the sum of N_t and N_c, and the product term $N_t N_c$ is set to equal $(N/2)^2$.

Pre/Post Measures. When authors report pre- and post-treatment measures without other statistical adjustments, we start by calculating two between-groups effect sizes: (a) at pre-treatment and, (b) at post-treatment. Then, we calculate the overall effect size by subtracting the post-treatment effect size from the pre-treatment effect size.

---

27 Lipsey & Wilson, 2001.
28 Ibid, Table B10, equation 1, p. 198.
29 Ibid, Table 3.2, p. 72.
30 Ibid, Table B10, equation 2, p. 198

E4c. Adjusting Effect Sizes for Small Samples

Since some studies have very small sample sizes, we follow the recommendation of many meta-analysts and adjust for this. Small sample sizes have been shown to upwardly bias effect sizes, especially when samples are less than 20. Following Hedges,[31] Lipsey and Wilson[32] report the "Hedges correction factor," which we use to adjust all mean-difference effect sizes, (where *N* is the total sample size of the combined treatment and comparison groups):

$$(4)\ ES'_m = \left[1 - \frac{3}{4N-9}\right] * \ ES_m$$

Adjusting Effect Sizes and Variances for Multi-Level Data Structures. Most studies in the education field use data that are hierarchical in nature. That is, students are clustered in classrooms, classrooms are clustered within schools, schools are clustered within districts, and districts are clustered within states. Analyses that do not account for clustering will underestimate the variance in outcomes at the student level (the denominator in equation 1 and, thus, may over-estimate the precision of magnitude on effect sizes). In studies that do not account for clustering, effect sizes and their variance require additional adjustments.[33] There are two types of studies, each requiring a different set of adjustments.[34] First, for student-level studies that ignore the variance due to clustering, we make adjustments to the mean effect size and its variance,

$$(5)\ ES_T = \ ES_m * \sqrt{1 - \frac{2(n-1)\rho}{N-2}}$$

$$(6)\ V\{ES_T\} = \ \left(\frac{N_t + N_c}{N_t N_c}\right)[1 + (n-1)\rho] + \ {ES_T}^2 \ \left(\frac{(N-2)(1-\rho)^2 + n(N-2n)\rho^2 + 2(N-2n)\rho(1-\rho)}{2(N-2)[(N-2) - 2(n-1)\rho]}\right)$$

where ρ is the intraclass correlation, the ratio of the variance between clusters to the total variance; *N* is the total number of individuals in the treatment group, N_t, and the comparison group, N_c; and *n* is the average number of persons in a cluster, *K*. In the educational field, clusters can be classes, schools, or districts. For this study, we used 2006 Washington Assessment of Student Learning (WASL) data to calculate values of ρ for the school-level (ρ = 0.114) and the district level (ρ = 0.052). Class-level data were not available, so we use a value of ρ = 0.200 for class-level studies.

Second, for studies that report means and standard deviations at a cluster level, we make adjustments to the mean effect size and its variance:

$$(7)\ ES_T = ES_m * \sqrt{\frac{1 + (n-1)\rho}{n\rho}} * \sqrt{\rho}$$

$$(8)\ v\{ES_T\} = \left\{\left(\frac{N_t - N_c}{N_t N_c}\right) * \left(\frac{1 + (n-1)\rho}{n\rho}\right) + \frac{[1 + (n-1)\rho]^2 * \ {ES_T}^2}{2n\rho(K-2)}\right\} * \rho$$

[31] Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107-128.
[32] Lipsey & Wilson, 2001, equation 3.22, p. 49.
[33] Studies that employ hierarchical linear modeling, or fixed effects with robust standard errors, or random effects models account for variance and need no further adjustment for computing the effect size, but adjustments are made to the inverse variance weights for meta-analysis using these methods.
[34] These formulas are taken from: Hedges, L. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32*(4), 341-370.

We do not adjust effect sizes in studies reporting dichotomous outcomes. This is because the d-Cox transformation assumes the entire normal distribution at the student level.[35] However, when outcomes are dichotomous, or an effect size is calculated from studies where authors control for clustering with robust standard errors or hierarchical linear modeling, we use the "design effect" to calculate the "effective sample size."[36] The design effect is given by:

$$(9)\ D = 1 + (n-1)\rho$$

The effective sample size is the actual sample size divided by the design effect. For example, the effective sample size for the treatment group is:

$$(10)\ N_{t(eff)} = \frac{N_t}{D}$$

Computing Weighted Average Effect Sizes, Confidence Intervals, and Homogeneity Tests. Once effect sizes are calculated for each program effect, and any necessary adjustments for clustering are made, the individual measures are summed to produce a weighted average effect size for a program area. We calculate the inverse variance weight for each program effect and these weights are used to compute the average. These calculations involve three steps. First, the standard error, SE_T of each mean effect size is computed with:[37]

$$(11)\ SE_T = \sqrt{\frac{N_t + N_c}{N_t N_c} + \frac{ES^2}{2(N_t + N_c)}}$$

Next, the inverse variance weight *w* is computed for each mean effect size with:[38]

$$(12)\ w = \frac{1}{SE_T^2}$$

The weighted mean effect size for a group with *i* studies is computed with:[39]

$$(13)\ \overline{ES} = \frac{\sum\left(w_i ES_{T_i}\right)}{\sum w_i}$$

Confidence intervals around this mean are then computed by first calculating the standard error of the mean with:[40]

$$(14)\ SE_{\overline{ES}} = \sqrt{\frac{1}{\sum w_i}}$$

Next, the lower, ES_L, and upper limits, ES_U, of the confidence interval are computed with:[41]

$$(15)\ \overline{ES_L} = \overline{ES} - z_{(1-\alpha)}\left(SE_{\overline{ES}}\right)$$

[35] Mark Lipsey (personal communication, November 11, 2007).
[36] Formulas for design effect and effective sample size were obtained from the Cochrane Reviewers Handbook, section 16.3.4, Approximate analyses of cluster-randomized trials for a meta-analysis: effective sample sizes. http://www.cochrane-handbook.org/
[37] Lipsey & Wilson, 2001, equation 3.23, p. 49.
[38] Ibid., equation 3.24, p. 49.
[39] Ibid., p. 114.
[40] Ibid.
[41] Ibid.

$$(16)\ \overline{ES_U} = \overline{ES} + z_{(1-\alpha)}\left(SE_{\overline{ES}}\right)$$

In equations (15) and (16), $z_{(1-\alpha)}$ is the critical value for the z-distribution (1.96 for α = .05). The test for homogeneity, which provides a measure of the dispersion of the effect sizes around their mean, is given by:[42]

$$(17)\ Q_i = \left(\sum w_i\, ES_i^2\right) - \frac{(\sum w_i\, ES_i^2)}{\sum w_i}$$

The Q-test is distributed as a chi-square with k-1 degrees of freedom (where k is the number of effect sizes).

Computing Random Effects Weighted Average Effect Sizes and Confidence Intervals. Next, a random effects model is used to calculate the weighted average effect size. Random effects models allow us to account for between-study variance in addition to within-study variance.[43] This is accomplished by first calculating the random effects variance component, *v:*[44]

$$(18)\ v = \frac{Q_i - (k-1)}{\sum w_i - \left(\sum wsq_i / \sum w_i\right)}$$

where wsq_i is the square of the weight of ES_i. This random variance factor is then added to the variance of each effect size and finally all inverse variance weights are recomputed, as are the other meta-analytic test statistics. If the value of Q is less than the degrees of freedom (*k-1*), there is no excess variation between studies and the initial variance estimate is used.

[42] Ibid., p. 116.
[43] Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97-111.
[44] Ibid., p. 134.