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.
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.

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

<table>
<thead>
<tr>
<th>Panel 1: Topics with WSIPP meta-analysis of impacts on student learning</th>
<th>Topic</th>
<th>Pg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charter schools</td>
<td>05</td>
<td></td>
</tr>
<tr>
<td>Expeditionary learning</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Instructional time (one additional day)</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>National Board for Professional Teaching Standards</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>National Guard Youth ChalleNGe Program</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Parent involvement in reading instruction</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Principals (school leadership)</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Project Lead the Way</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>School-wide positive behavior programs</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Teacher induction/mentoring</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Teacher professional development</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>Tutoring</td>
<td>51</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel 2: Topics WSIPP reviewed, but had too few rigorous evaluations to meta-analyze</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advancement Via Individual Determination (AVID)</td>
<td></td>
</tr>
<tr>
<td>Blended learning</td>
<td></td>
</tr>
<tr>
<td>High Schools that Work</td>
<td></td>
</tr>
<tr>
<td>Home schooling</td>
<td></td>
</tr>
<tr>
<td>International Baccalaureate</td>
<td></td>
</tr>
<tr>
<td>MicroSociety</td>
<td></td>
</tr>
<tr>
<td>Montessori</td>
<td></td>
</tr>
<tr>
<td>Professional learning communities</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel 3: Relevant topics WSIPP did not fully review (due to the complexity/weight of evidence)</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative schools for at-risk students</td>
<td></td>
</tr>
<tr>
<td>Extended instructional day</td>
<td></td>
</tr>
<tr>
<td>Project-based learning</td>
<td></td>
</tr>
<tr>
<td>School size (small schools)/personalization of learning</td>
<td></td>
</tr>
<tr>
<td>Technology-based innovations</td>
<td></td>
</tr>
<tr>
<td>Theme-based (A-STEM) or magnet schools</td>
<td></td>
</tr>
<tr>
<td>Wrap-around services</td>
<td></td>
</tr>
</tbody>
</table>
### 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**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Outcomes meta-analyzed</th>
<th>WSIPP meta-analytic result</th>
<th>Other outcomes examined in the research literature (not meta-analyzed for this report)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charter schools</td>
<td>Reading &amp; math test scores</td>
<td>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.</td>
<td>High school graduation; college enrollment; attendance; discipline and effects on nearby schools</td>
</tr>
<tr>
<td>Expeditionary learning</td>
<td>Reading, math, &amp; science test scores</td>
<td>Expeditionary learning does not have a consistent impact on student test scores (some evaluated programs have positive impacts, some negative).</td>
<td>Behavioral measures such as attendance and disciplinary incidents</td>
</tr>
<tr>
<td>Instructional time (one addtl. day)</td>
<td>Reading &amp; math test scores</td>
<td>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).</td>
<td>Labor market outcomes</td>
</tr>
<tr>
<td>NBPTS certification</td>
<td>Reading, math, &amp; other academic test scores</td>
<td>Students who have teachers with NBPTS certification have slightly higher test scores, on average.</td>
<td>Teacher recruitment and retention, self-reported impacts on teaching practices</td>
</tr>
<tr>
<td>National Guard Youth ChalleNGe Program</td>
<td>High school graduation</td>
<td>ChalleNGe appears to have a positive impact on high school graduation rates and mixed impacts on other outcomes.</td>
<td>Employment, housing, crime, health, substance abuse, GEDs.</td>
</tr>
</tbody>
</table>
### Exhibit E2
**Summary of WSIPP Meta-Analytic Results**

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

*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.” In the 2012-13 school year, an estimated 6,000 charter schools enrolled more than 2.3 million students across the country. 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).

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.

---

**Exhibit E3**

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

<table>
<thead>
<tr>
<th>Effect sizes included in the meta-analysis</th>
<th>Reading test scores</th>
<th>Math test scores</th>
<th>Other outcomes examined in the research literature (not meta-analyzed for this report)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>65</td>
<td>65</td>
<td>Other test scores (e.g. science and social studies); high school graduation; college enrollment; attendance; discipline; and effects on nearby schools</td>
</tr>
<tr>
<td>Average effect on academic outcomes</td>
<td>0.002 (0.007)</td>
<td>0.009 (0.011)</td>
<td></td>
</tr>
<tr>
<td>(standard error)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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.  

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).

---

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

Effect Size

Average effect = 0.002
(s.e. = 0.007)
Effect Sizes: Impacts on Student Academic Outcomes from Charter Schools
Study by Study Results

Math

Average effect = 0.009 (s.e. = 0.011)

Zimmer & Buddin, 2006(CA)
Bifulco & Ladd, 2006(NC)
Herman et al., 2012(CA)
Bettinger, 2005(MI)
Bets et al., 2006(CA)
CREDO, 2011b(PA)
Zimmer & Buddin, 2006(CA)
Zimmer et al., 2012(TX)
Zimmer et al., 2012(IL)
CREDO, 2009(OH)
Ni & Rorrer, 2012(UT)
Gleason et al., 2010(US)
CREDO, 2009b(TX)
CREDO, 2009(NM)
CREDO, 2009a(AZ)
Bets et al., 2006(CA)
CREDO, 2009a(IL)
CREDO, 2009c(CA)
CREDO, 2009(NC)
CREDO, 2009(MN)
Sass, 2006(FL)
Zimmer et al., 2012(PA)
Carruthers, 2012(NC)
Zimmer & Buddin, 2006(CA)
Zimmer et al., 2012(CT)
CREDO, 2009c(CA)
CREDO, 2009a(DC)
Zimmer et al., 2012(CA)
CREDO, 2009h(IL)
CREDO, 2009k(MO)
Zimmer et al., 2012(WI)
Solmon et al., 2001(AZ)
CREDO, 2009e(AR)
Booker et al., 2007(TX)
Witte et al., 2012(WI)
CREDO, 2009d(CO)
CREDO, 2008(LA)
CREDO, 2013b(AL)
Hoxby & Rockoff, 2005(IL)
Zimmer & Buddin, 2006(CA)
CREDO, 2013a(NJ)
CREDO, 2013(MA)
Imberman, 2011(NE)
CREDO, 2013(NJ)
Hoxby et al., 2009(NY)
Bets et al., 2006(CA)
Zimmer et al., 2012(CO)
Dobbie & Fryer, 2012(NY)
Dobbie & Fryer, 2012(NY)
CREDO, 2013c(NY)
Tuttle et al., 2013(US)
Woodworth et al., 2008(CA)
Angrist et al., 2012b(MA)
Nicotera et al., 2009(IN)
Woodworth et al., 2008(CA)
Angrist et al., 2012b(MA)
Woodworth et al., 2008(CA)
Ross et al., 2007(TN)
Angrist et al., 2011a(MA)
Abdulkadiroglu et al., 2011(MA)
Abdulkadiroglu et al., 2011(MA)
Woodworth et al., 2008(CA)
Woodworth et al., 2008(CA)
Supovitz & Rikoon, 2010(NY)
Woodworth et al., 2008(CA)
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.”

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

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

* 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

**Reading**
Average effect = 0.106 (s.e. = 0.048)

**Math**
Average effect = 0.273 (s.e. = 0.049)
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. Several recent studies have examined this assumption.

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).

---


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

<table>
<thead>
<tr>
<th>Effect sizes included in the meta-analysis</th>
<th>Reading test scores</th>
<th>Math test scores</th>
<th>Other outcomes examined in the research literature (not meta-analyzed for this report)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; year</td>
<td>14</td>
<td>1&lt;sup&gt;st&lt;/sup&gt; year</td>
<td>13</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; year</td>
<td>13</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; year</td>
<td>12</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; year</td>
<td>6</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt; year</td>
<td>5</td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; year</td>
<td>6</td>
<td>4&lt;sup&gt;th&lt;/sup&gt; year</td>
<td>5</td>
</tr>
<tr>
<td>5&lt;sup&gt;th&lt;/sup&gt; year</td>
<td>6</td>
<td>5&lt;sup&gt;th&lt;/sup&gt; year</td>
<td>5</td>
</tr>
<tr>
<td>6+ years</td>
<td>4</td>
<td>6+ years</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average effect on academic outcomes (standard error)</th>
<th>Reading test scores</th>
<th>Math test scores</th>
<th>Other outcomes examined in the research literature (not meta-analyzed for this report)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; year</td>
<td>-0.089 (0.025)</td>
<td>1&lt;sup&gt;st&lt;/sup&gt; year</td>
<td>-0.116 (0.044)</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; year</td>
<td>-0.084 (0.017)</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; year</td>
<td>-0.065 (0.037)</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; year</td>
<td>-0.046 (0.035)</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt; year</td>
<td>-0.076 (0.055)</td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; year</td>
<td>-0.035 (0.036)</td>
<td>4&lt;sup&gt;th&lt;/sup&gt; year</td>
<td>-0.066 (0.055)</td>
</tr>
<tr>
<td>5&lt;sup&gt;th&lt;/sup&gt; year</td>
<td>-0.016 (0.036)</td>
<td>5&lt;sup&gt;th&lt;/sup&gt; year</td>
<td>-0.055 (0.073)</td>
</tr>
<tr>
<td>6+ years</td>
<td>0.005 (0.046)</td>
<td>6+ years</td>
<td>-0.005 (0.120)</td>
</tr>
</tbody>
</table>

**Conclusion**

Charter schools in their initial years have, on average, negative impacts on student test scores. Charter schools in their 3<sup>rd</sup> 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

**Reading**

<table>
<thead>
<tr>
<th>Years School has been in Operation</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Year</td>
<td>-0.089</td>
</tr>
<tr>
<td>2nd Year</td>
<td>-0.084</td>
</tr>
<tr>
<td>3rd Year</td>
<td>-0.046</td>
</tr>
<tr>
<td>4th Year</td>
<td>-0.035</td>
</tr>
<tr>
<td>5th Year</td>
<td>-0.016</td>
</tr>
<tr>
<td>6+ Years</td>
<td>0.005</td>
</tr>
</tbody>
</table>

*Note: Error bars represent 95 percent confidence interval.*

**Math**

<table>
<thead>
<tr>
<th>Years School has been in Operation</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Year</td>
<td>-0.12</td>
</tr>
<tr>
<td>2nd Year</td>
<td>-0.065</td>
</tr>
<tr>
<td>3rd Year</td>
<td>-0.076</td>
</tr>
<tr>
<td>4th Year</td>
<td>-0.066</td>
</tr>
<tr>
<td>5th Year</td>
<td>-0.055</td>
</tr>
<tr>
<td>6+ Years</td>
<td>-0.005</td>
</tr>
</tbody>
</table>

*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**

#### Reading - 1st Year
- Average effect = -0.089
  - (s.e. = 0.025)
- Studies included:
  - Carruthers, 2012(NC)
  - Bifulco & Ladd, 2006(NC)
  - Nicotera et al., 2009(IN)
  - Zimmer et al., 2009(OH)
  - Zimmer et al., 2009(TX)
  - Ni & Rorrer, 2012(UT)
  - Zimmer et al., 2009(IL)
  - Ni & Rorrer, 2012(UT)
  - Sass, 2006(FL)
  - Zimmer et al., 2009(CA)
  - Zimmer et al., 2009(WI)
  - Zimmer et al., 2009(CO)
  - Zimmer et al., 2009(PA)
  - Booker et al., 2007(TX)

#### Reading - 2nd Year
- Average effect = -0.084
  - (s.e. = 0.017)
- Studies included:
  - Carruthers, 2012(NC)
  - Zimmer et al., 2009(OH)
  - Zimmer et al., 2009(TX)
  - Ni & Rorrer, 2012(UT)
  - Bifulco & Ladd, 2006(NC)
  - Zimmer et al., 2009(IL)
  - Zimmer et al., 2009(PA)
  - Sass, 2006(FL)
  - Zimmer et al., 2009(WI)
  - Zimmer et al., 2009(CA)
  - Ni & Rorrer, 2012(UT)
  - Booker et al., 2007(TX)
  - Zimmer et al., 2009(CO)

#### Reading - 3rd Year
- Average effect = -0.046
  - (s.e. = 0.035)
- Studies included:
  - Carruthers, 2012(NC)
  - Ni & Rorrer, 2012(UT)
  - Bifulco & Ladd, 2006(NC)
  - Sass, 2006(FL)
  - Ni & Rorrer, 2012(UT)
  - Booker et al., 2007(TX)

#### Reading - 4th Year
- Average effect = -0.035
  - (s.e. = 0.036)
- Studies included:
  - Carruthers, 2012(NC)
  - Bifulco & Ladd, 2006(NC)
  - Ni & Rorrer, 2012(UT)
  - Sass, 2006(FL)
  - Ni & Rorrer, 2012(UT)
  - Booker et al., 2007(TX)

#### Reading - 5th Year
- Average effect = 0.016
  - (s.e. = 0.036)
- Studies included:
  - Bifulco & Ladd, 2006(NC)
  - Carruthers, 2012(NC)
  - Booker et al., 2007(TX)
  - Sass, 2006(FL)
  - Ni & Rorrer, 2012(UT)
  - Ni & Rorrer, 2012(UT)

#### Reading - 6 or more Years
- Average effect = 0.005
  - (s.e. = 0.046)
- Studies included:
  - Carruthers, 2012(NC)
  - Ni & Rorrer, 2012(UT)
  - Ni & Rorrer, 2012(UT)
  - Booker et al., 2007(TX)
**Exhibit E9, continued**

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

**Math - 1st Year**
- Bifulco & Ladd, 2006 (NC)
- Zimmer et al., 2009 (OH)
- Carruthers, 2012 (NC)
- Zimmer et al., 2009 (IL)
- Zimmer et al., 2009 (TX)
- Ni & Rorrer, 2012 (UT)
- Sass, 2006 (FL)
- Zimmer et al., 2009 (WI)
- Nicotera et al., 2009 (IN)
- Zimmer et al., 2009 (PA)
- Booker et al., 2007 (TX)
- Zimmer et al., 2009 (CA)
- Zimmer et al., 2009 (CO)

Average effect = -0.116
(s.e. = 0.044)

**Math - 2nd Year**
- Carruthers, 2012 (NC)
- Zimmer et al., 2009 (OH)
- Bifulco & Ladd, 2006 (NC)
- Zimmer et al., 2009 (TX)
- Ni & Rorrer, 2012 (UT)
- Zimmer et al., 2009 (IL)
- Sass, 2006 (FL)
- Zimmer et al., 2009 (PA)
- Zimmer et al., 2009 (CA)
- Zimmer et al., 2009 (WI)
- Booker et al., 2007 (TX)
- Zimmer et al., 2009 (CO)

Average effect = -0.065
(s.e. = 0.037)

**Math - 3rd Year**
- Carruthers, 2012 (NC)
- Bifulco & Ladd, 2006 (NC)
- Sass, 2006 (FL)
- Ni & Rorrer, 2012 (UT)
- Booker et al., 2007 (TX)

Average effect = -0.076
(s.e. = 0.055)

**Math - 4th Year**
- Carruthers, 2012 (NC)
- Bifulco & Ladd, 2006 (NC)
- Sass, 2006 (FL)
- Ni & Rorrer, 2012 (UT)
- Booker et al., 2007 (TX)

Average effect = -0.066
(s.e. = 0.055)

**Math - 5th Year**
- Carruthers, 2012 (NC)
- Bifulco & Ladd, 2006 (NC)
- Booker et al., 2007 (TX)
- Sass, 2006 (FL)
- Ni & Rorrer, 2012 (UT)

Average effect = -0.055
(s.e. = 0.073)

**Math - 6 or more years**
- Carruthers, 2012 (NC)
- Bifulco & Ladd, 2006 (NC)
- Booker et al., 2007 (TX)
- Ni & Rorrer, 2012 (UT)
- Booker et al., 2007 (TX)

Average effect = -0.005
(s.e. = 0.120)
Urban Charter Schools

Charter schools have traditionally been located in cities; many are designed to serve minority students in high-poverty areas.\textsuperscript{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.”\textsuperscript{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.

\textit{Exhibit E10}

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

<table>
<thead>
<tr>
<th>Effect sizes included in the meta-analysis</th>
<th>Reading test scores</th>
<th>Math test scores</th>
<th>Other outcomes examined in the research literature (not meta-analyzed for this report)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
<td>39</td>
<td>Other test scores (e.g. science and social studies); high school graduation; college enrollment; attendance; and discipline (office referrals, suspensions, and expulsion)</td>
<td></td>
</tr>
<tr>
<td>Average effect on academic outcomes (standard error)</td>
<td>0.032 (0.016)</td>
<td>0.076 (0.018)</td>
<td></td>
</tr>
</tbody>
</table>

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.


\textsuperscript{10} Ibid.


**Exhibit E11**

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

**Study by Study Results**

- **Reading**

  - Average effect = 0.032 (s.e. = 0.016)

  - **Effects of Urban Charter Schools on Reading**
    - Zimmer & Buddin, 2006(San Diego)
    - Zimmer et al., 2012(Chicago)
    - Zimmer & Buddin, 2006(Los Angeles)
    - Woodworth et al., 2008(SF)
    - Betts et al., 2006(San Diego)
    - Betts et al., 2006(San Diego)
    - Zimmer & Buddin, 2006(Los Angeles)
    - Credo, 2009e(DC)
    - Betts et al., 2006(San Diego)
    - Credo, 2009h(Chicago)
    - Zimmer et al., 2012(Philadelphia)
    - Zimmer et al., 2012(Milwaukee)
    - Credo, 2009d(Denver)
    - Zimmer et al., 2012(San Diego)
    - Imberman, 2011(LUSD)
    - Zimmer et al., 2012(Denver)
    - Credo, 2013c(NYC)
    - Nicotera et al., 2009(Indianapolis)
    - Dobbie & Fryer, 2012(NYC)
    - Witte et al., 2012(Milwaukee)
    - Dobbie & Fryer, 2012(NYC)
    - Hoxby, 2009(NYC)
    - Credo, 2013a(MA)
    - Credo, 2013b(MI)
    - Zimmer & Buddin, 2006(San Diego)
    - Woodworth et al., 2008(SF)
    - Credo, 2012(NJ)
    - Angrist et al., 2012a(SF)
    - Angrist et al., 2012b(Boston)
    - Angrist et al., 2012b(MA)
    - Woodworth et al., 2008(SF)
    - Abdulkadiroglu et al., 2011(Boston)
    - Woodworth et al., 2008(SF)
    - Ross et al., 2007(Memphis)
    - Abdulkadiroglu et al., 2011(Boston)
    - Angrist et al., 2012b(MA)
    - Woodworth et al., 2008(SF)
    - Supovitz & Rikoon, 2010(NYC)
**Exhibit E11, continued**

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

Study by Study Results

![Effect Size Graph](image)

Math

- Average effect = 0.076
- (s.e. = 0.018)
Non-Urban Charter Schools

While charter schools traditionally operate in urban areas, there is a growing interest in charters “outside of central cities.” 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**

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

*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.

---

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

Study by Study Results

**Reading**
- Angrist et al., 2012b (MA)
- CREDO, 2012 (NJ)
- CREDO, 2013b (MI)

Average effect = 0.013
(s.e. = 0.046)

**Math**
- Angrist et al., 2012b (MA)
- CREDO, 2012 (NJ)
- CREDO, 2013a (MA)
- CREDO, 2013b (MI)

Average effect = 0.043
(s.e. = 0.027)
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**

**Reading**

Note: Error bars represent 95 percent confidence interval.

**Math**

Note: Error bars represent 95 percent confidence interval.
Studies Used in Any of the Meta-Analyses of Charter Schools on Student Academic Outcomes


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

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

*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

Studies Used in the Meta-Analysis of Expeditionary Learning Effects on Student Academic Outcomes


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

<table>
<thead>
<tr>
<th>Test scores (reading, math, &amp; general academic)</th>
<th><strong>Other outcomes examined in the research literature (not meta-analyzed for this report)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect sizes included in the meta-analysis</td>
<td>14</td>
</tr>
<tr>
<td>Average effect on academic outcomes (standard error)</td>
<td>0.001 (0.003)</td>
</tr>
<tr>
<td>Conclusion</td>
<td>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).</td>
</tr>
</tbody>
</table>

*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 E18**
Effect Sizes: Impacts on Student Academic Outcomes from One Additional School Day
Study by Study Results

---

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


---


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. 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

<table>
<thead>
<tr>
<th>Test scores (reading, math, &amp; general academic)</th>
<th>Other outcomes examined in the research literature (not meta-analyzed for this report)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect sizes included in the meta-analysis</td>
<td>14</td>
</tr>
<tr>
<td>Average effect on academic outcomes (standard error)</td>
<td>0.026 (0.004)</td>
</tr>
<tr>
<td>Conclusion</td>
<td>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.</td>
</tr>
</tbody>
</table>

*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 E20**

Effect Sizes: Student Academic Outcomes Associated with NBPTS Certification

Study by Study Results

<table>
<thead>
<tr>
<th>Test Scores</th>
<th>Effect Size</th>
<th>Average effect</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harris &amp; Sass, 2009 (FL)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goldhaber &amp; Anthony, 2007 (NC)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ladd et al., 2007 (FL)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clotfelter et al, 2006 (NC)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clotfelter et al., 2007 (NC)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chingos &amp; Peterson, 2011 (FL)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chingos &amp; Peterson, 2011 (FL)</td>
<td>0.026</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>Sanders et al., 2005 (NC)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clotfelter et al., 2010 (NC)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ladd et al., 2007 (NC)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cantrell et al., 2008 (CA)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stronge et al., 2007 (NC)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cavalluzzo, 2004 (FL)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vandevoort &amp; Berliner, 2004 (AZ)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Studies Used in the Meta-Analysis of NBPTS Certification and Student Academic Outcomes


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

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

*C 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.

Citations to the National Evaluation of the ChalleNGe Program


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

<table>
<thead>
<tr>
<th>Test scores (reading, math, &amp; general academic)</th>
<th>Other outcomes examined in the research literature (not meta-analyzed for this report)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect sizes included in the meta-analysis</td>
<td>9</td>
</tr>
<tr>
<td>Average effect on academic outcomes (standard error)</td>
<td>0.06 (0.10)</td>
</tr>
<tr>
<td>Conclusion</td>
<td>Elementary school-based programs that encourage parent involvement in reading instruction are associated with improved student reading outcomes, on average.</td>
</tr>
</tbody>
</table>

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

Test Scores

<table>
<thead>
<tr>
<th>Study</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powell-Smith et al., 2000</td>
<td>-0.03</td>
</tr>
<tr>
<td>Powell-Smith et al., 2000</td>
<td>-0.03</td>
</tr>
<tr>
<td>Miller &amp; Kratochwill, 1996</td>
<td>-0.03</td>
</tr>
<tr>
<td>Epstein, 1991</td>
<td>0.01</td>
</tr>
<tr>
<td>Mehran &amp; White, 1988</td>
<td>0.02</td>
</tr>
<tr>
<td>Erion, 1994</td>
<td>0.01</td>
</tr>
<tr>
<td>Rodick &amp; Henggeler, 1980</td>
<td>0.02</td>
</tr>
<tr>
<td>Heller &amp; Fantuzzo, 1993</td>
<td>0.30</td>
</tr>
<tr>
<td>Fantuzzo et al., 1995</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Average effect = 0.06 (s.e. = 0.10)
Studies Used in the Meta-Analysis of Parent Involvement in Reading Instruction
(School-based Programs)


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. 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.

---

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

<table>
<thead>
<tr>
<th>Effect sizes included in the meta-analysis</th>
<th>Reading test scores</th>
<th>Math test scores</th>
<th>Other outcomes examined in the research literature (not meta-analyzed for this report)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>6</td>
<td>High school graduation (1 study ES= 0.04); self-reported measures of principal effectiveness</td>
</tr>
<tr>
<td>Average effect on academic outcomes (standard error)</td>
<td>0.073 (0.025)</td>
<td>0.107 (0.032)</td>
<td></td>
</tr>
</tbody>
</table>

**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

Studies Used in the Meta-Analysis of Principal Impacts on Student Academic Outcomes


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.

<table>
<thead>
<tr>
<th>Effect sizes included in the meta-analysis</th>
<th>Reading test scores</th>
<th>Math test scores</th>
<th>Other outcomes*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average effect on academic outcomes (standard error)</td>
<td>0.01 (0.06)</td>
<td>0.11 (0.06)</td>
<td>0.00 (0.10)</td>
</tr>
<tr>
<td>Conclusion</td>
<td>PLTW improves student math scores but does not consistently impact reading or science test scores.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*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


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**

### Reading

- **Van Overschelde, 2013**
- **NWEA, 2010**

Average effect: 0.01 (s.e. 0.06)

### Math

- **Van Overschelde, 2013**
- **Rethwisch et al., 2012**
- **NWEA, 2010**

Average effect: 0.11 (s.e. 0.06)

### Science

- **Tran & Nathan, 2010 (U.S.)**
- **NWEA, 2010 (U.S.)**
- **Rethwisch et al., 2012 (IA)**

Average effect: 0.00 (s.e. 0.10)
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). 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). Those outcomes are measured in varied, non-standardized ways and thus are not meta-analyzed in this report.

---


**Exhibit E28**

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

<table>
<thead>
<tr>
<th></th>
<th>Reading test scores</th>
<th>Math test scores</th>
<th>Other outcomes examined in the research literature (not meta-analyzed for this report)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect sizes included in the meta-analysis</td>
<td>9</td>
<td>7</td>
<td>Attendance, grade retention, and discipline (office discipline referrals, suspensions, and expulsion)</td>
</tr>
<tr>
<td>Average effect on academic outcomes (standard error)</td>
<td>0.21 (0.05)</td>
<td>0.25 (0.07)</td>
<td></td>
</tr>
</tbody>
</table>

**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

- **Reading Test Scores**
  - Rimm-Kaufman et al., 2007 (U.S.)
  - Rimm-Kaufman et al., 2007 (U.S.)
  - Rimm-Kaufman et al., 2007 (U.S.)
  - Snyder et al., 2010 (HI)
  - Snyder et al., 2010 (HI)
  - Horner et al. 2009 (IL & HI)
  - Flay et al., 2001 (NV)
  - Flay et al., 2001 (NV)
  - Flay et al., 2001 (HI)

Average effect = 0.21 (s.e.) = 0.05

- **Math Test Scores**
  - Rimm-Kaufman et al., 2007 (U.S.)
  - Snyder et al., 2010 (HI)
  - Snyder et al., 2010 (HI)
  - Rimm-Kaufman et al., 2007 (U.S.)
  - Flay et al., 2001 (NV)
  - Flay et al., 2001 (HI)

Average effect = 0.25 (s.e.) = 0.07
Studies Used in the Meta-Analysis of K-12 School-wide Behavior Programs Effects on Student Academic Outcomes


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

<table>
<thead>
<tr>
<th>Reading and math test scores</th>
<th>Other outcomes examined in the research literature (not meta-analyzed for this report)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect sizes included in the meta-analysis</td>
<td>5</td>
</tr>
<tr>
<td>Average effect on academic outcomes (standard error)</td>
<td>0.07 (0.06)</td>
</tr>
<tr>
<td>Conclusion</td>
<td>For teacher induction programs, the results are mixed, but the average impact is positive.</td>
</tr>
</tbody>
</table>

*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.
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


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.

\textit{Exhibit E32}

\textbf{Effect Sizes: Impacts on Student Academic Outcomes from Teacher Induction/Mentoring}

<table>
<thead>
<tr>
<th>Effect sizes included in the meta-analysis</th>
<th>General PD, (reading &amp; math test scores)</th>
<th>Content-Specific PD, (reading &amp; math test scores)</th>
<th>Other outcomes examined in the research literature (not meta-analyzed for this report)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average effect on academic outcomes (standard error)</td>
<td>0.000 (0.00)</td>
<td>0.005 (0.00)</td>
<td>Teacher retention; self-reported measures of teacher outcomes</td>
</tr>
<tr>
<td>Conclusion</td>
<td>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.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*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.


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


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


E31. 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. One-on-one tutoring is an effective way to improve reading test scores.

**Exhibit E34**

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

*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 E35
Effect Sizes: Impacts on Student Reading Test Scores from One-on-One Tutoring
Study by Study Results

Reading Recovery

Peer Tutoring

Average effect = 0.34
(s.e. = 0.09)

Average effect = 0.22
(s.e. = 0.08)

Tutoring by Adults

Average effect = 0.12
(s.e. = 0.05)
Studies Used in the Meta-Analysis of Reading Recovery Effects on Reading Test Scores


---

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


---

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


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. 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, 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. 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. 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.
25 Ibid.
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 Difference 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:

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

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:

\[
ES\text{Var} = \frac{N_t + N_c}{N_tN_c} + \frac{ES^2}{2(N_t + N_c)}
\]

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:

\[
(3) \ ES = t \sqrt{\frac{N_t + N_c}{N_tN_c}}
\]

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 \( N_tN_c \) 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.

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,\textsuperscript{31} Lipsey and Wilson\textsuperscript{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.\textsuperscript{33} There are two types of studies, each requiring a different set of adjustments.\textsuperscript{34} First, for student-level studies that ignore the variance due to clustering, we make adjustments to the mean effect size and its variance,

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

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

where $\rho$ 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 $\rho$ for the school-level ($\rho = 0.114$) and the district level ($\rho = 0.052$). Class-level data were not available, so we use a value of $\rho = 0.200$ 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:

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

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


\textsuperscript{32} Lipsey & Wilson, 2001, equation 3.22, p. 49.

\textsuperscript{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.

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. 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.” The design effect is given by:

\[ D = 1 + (n - 1) \rho \]

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:

\[ N_{t(\text{eff})} = \frac{N_t}{D} \]

### 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:

\[ SE_T = \sqrt{\frac{N_t + N_c}{N_t N_c} + \frac{ES^2}{2(N_t + N_c)}} \]

Next, the inverse variance weight \( w \) is computed for each mean effect size with:

\[ w = \frac{1}{SE_T^2} \]

The weighted mean effect size for a group with \( i \) studies is computed with:

\[ \bar{ES} = \frac{\sum (w_i ES_T)}{\sum w_i} \]

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

\[ SE_{\bar{ES}} = \frac{1}{\sqrt{\sum w_i}} \]

Next, the lower, \( ES_L \), and upper limits, \( ES_U \), of the confidence interval are computed with:

\[ ES_L = \bar{ES} - z_{(1-a)} (SE_{\bar{ES}}) \]
\[ ES_U = \bar{ES} + z_{(1-a)} (SE_{\bar{ES}}) \]

---

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

\[
Q_i = \left( \sum w_i ES_i^2 \right) - \frac{\left( \sum w_i ES_i^2 \right)}{w_i} 
\]

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, \( v \):

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

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.


\(^{44}\) Ibid., p. 134.