The Cost-Effectiveness of Five Policies for Improving Student Achievement

Stuart S. Yeh
University of Minnesota

Comparisons of student achievement effect sizes suggest that systems in which student performance in math and reading is rapidly assessed between 2 and 5 times per week are 4 times as effective as a 10% increase in per pupil expenditure, 6 times as effective as voucher programs, 64 times as effective as charter schools, and 6 times as effective as increased accountability. Achievement gains per dollar from rapid assessment are even greater—193 times the gains that accrue from increasing preexisting patterns of educational expenditures, 2,424 times the gains from vouchers, 23,166 times the gains from charter schools, and 57 times the gains from increased accountability. Two sensitivity analyses suggest that the relative advantage for rapid assessment is not sensitive to the particular parameter estimates.

Keywords: cost-effectiveness; voucher programs; charter schools; accountability; assessment; educational productivity

In recent years, policymakers and educational researchers have focused on four policy alternatives for improving student achievement—increased educational spending, voucher programs, charter schools, and increased accountability—based on evidence that these approaches may be effective in improving student performance (Carnoy & Loeb, 2002; Greene, Peterson, & Du, 1999; Greenwald, Hedges, & Laine, 1996; Hanushek & Raymond, 2005; Howell, Wolfe, Campbell, & Peterson, 2002; Hoxby, 2002; Hoxby & Rockoff, 2005; Rouse, 1998). However, the research suggests very modest effect sizes in comparison with a fifth approach, which involves the implementation of rapid assessment systems that provide testing feedback to teachers and students regarding student performance in math and reading 2 to 5 times per week (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Fuchs & Fuchs, 1986; Kluger & DeNisi, 1996). What is needed is a comparison of relative effect sizes and relative cost-effectiveness so that policymakers have the information needed to identify the most promising policy alternatives. This article compares the relative cost-effectiveness of the five policies, using best-evidence estimates drawn from available data regarding the effectiveness and costs of rapid assessment, increased spending, voucher programs, charter schools, and accountability, using a conservative methodology for calculating the relative effectiveness of rapid assessment. Two sensitivity analyses suggest that the basic conclusion—that rapid assessment is much more cost-effective than the alternatives—is robust to changes in the underlying assumptions.
ies for

Cost-Effectiveness Analysis

Cost-effectiveness analysis systematically compares the costs and outcomes of alternative policies to aid decisions about the most efficient course of action (Levin, 1988; Levin & McEwan, 2001). In the area of education, cost-effectiveness analysis assesses outcomes in educational terms, such as student achievement, whereas cost-benefit analysis assesses outcomes in terms of monetary value. It is important to note that cost-effectiveness analysis provides a standard framework and techniques that permit policymakers to compare disparate interventions (Levin & McEwan, 2001). Although systemic education reforms have multiple goals, they can be compared with classroom-level interventions because they aim to improve student achievement. Policymakers wish to know which approach was most cost-effective, regardless of whether it was a systemic or a classroom-level intervention.

The basic strategy in cost-effectiveness analysis is to derive results for educational effectiveness by using standard evaluation procedures or studies and to combine that information with cost data that are derived from the ingredients approach (Levin, 1988; Levin & McEwan, 2001). It is important to note that the evaluation of effectiveness is separable from the evaluation of costs (Levin, 1988). Thus, estimates of effectiveness can be derived from published evaluations and then combined with estimates of costs derived through the ingredients approach: (a) identification of ingredients, (b) determination of the value or cost of the ingredients, and (c) analysis of the costs in an appropriate decision-oriented framework (Levin, 1988; Levin & McEwan, 2001). The costs of an intervention are defined as the value of the resources that are given up by society to implement the intervention, regardless of how those costs are reported by government entities (Levin, 1988; Levin & McEwan, 2001). These resources are the ingredients of the intervention, and the social value of the ingredients constitute its overall cost (Levin, 1988; Levin & McEwan, 2001).

Rapid Assessment 2

Rapid assessment may be defined as systems that provide nonjudgmental testing feedback, immediately after each test, 2 to 5 times per week to students and teachers, regarding student performance in subjects such as math or reading. The concept of rapid assessment is embodied by Reading Assessment, a popular program designed to encourage students to read books at appropriate levels of difficulty while alerting teachers to learning difficulties and encouraging teachers to provide individualized tutoring or small-group instruction. This is achieved through a system of frequently assessing each student’s reading comprehension and monitoring each student’s reading level. First, books in the school’s library are labeled and shelved according to reading level. Second, students select books to read based on their interests and their reading levels, according to the results of the STAR Reading test, a norm-referenced, computer-adaptive test (Renaissance Learning, n.d.). This helps students to avoid the frustrating experience of choosing a book that is too difficult. After finishing a book, the student takes a computer-based quiz, unique to the book, that is intended to monitor basic reading comprehension (Rapid Assessment Corporation has created more than 100,000 quizzes). Similarly, Math Assessment is a popular program that provides individualized, printed sets of math problems, a system of assessing student performance on those problems, and a scoring system in which students and teachers receive rapid, frequent feedback on student performance upon completion of every set of problems.

Three meta-analyses have been conducted regarding the effect of feedback on student achievement, involving studies that experimentally compared the achievement of students...
who were frequently tested with a group of similar students who received the same curriculum but who were not frequently tested. A meta-analysis of 21 experimental studies involving testing found that students who were tested 2 to 5 times per week outperformed students who were not frequently tested, with an average effect size of 0.7 standard deviations (SD; Fuchs & Fuchs, 1986), equivalent to raising the achievement of an average nation such as the United States to the level of the top five nations (Black & Wiliam, 1998). When teachers were required to follow rules about using the assessment information to change instruction for students, the average effect size exceeded 0.9 SD, and when students were reinforced with material tokens in addition to the frequent testing, the average effect size increased even further, exceeding 1.1 SD (Fuchs & Fuchs, 1986). A second meta-analysis of 40 feedback studies (Bangert-Drowns et al., 1991), which included studies involving nontesting feedback (e.g., praise or criticism) as well as studies involving testing feedback, found that feedback was more effective when it involved testing (effect size = 0.6 SD) and was presented immediately after a test (effect size = 0.7 SD). A third meta-analysis of 131 studies involving nontesting feedback, as well as studies involving testing feedback, found that praise or criticism attenuated the effectiveness of feedback (Kluger & DeNisi, 1996). Emotionally neutral (i.e., testing) feedback that is devoid of praise or criticism “is likely to yield impressive gains in performance, possibly exceeding 1 SD” (Kluger & DeNisi, 1996, p. 278)—much higher than the average effect size of 0.4 SD when all types of feedback studies were lumped together. A recent review of research summarized the results of previous meta-analyses regarding feedback and found an average effect size of 0.79 SD (Hattie & Timperley, 2007).

These results suggest the nature of effective feedback systems: nonjudgmental, involving frequent testing (2 to 5 times per week), presented immediately after a test. Under these conditions, the three meta-analyses of feedback interventions (Bangert-Drowns et al., 1991; Fuchs & Fuchs, 1986; Kluger & DeNisi, 1996) suggest that the effect size for testing feedback is no lower than 0.7 SD. However, the meta-analyses generally involved short implementations of rapid assessment (the average duration across all studies in the three meta-analyses was only 3.4 weeks), often with students in special education who may not be representative of the general student population, and the effectiveness of Reading and Math Assessment in large-scale field trials may differ. Thus, it is useful to examine the best controlled field trials of Reading and Math Assessment.

Two randomized experiments evaluated the effectiveness of the Reading Assessment program (Nunnery, Ross, & McDonald, 2006; Ross, Nunnery, & Goldfeder, 2004). The first experiment, involving 1,665 Memphis students (a district in which 71% of all students are eligible for free/reduced-price lunch), found an average effect size of 0.270 SD per grade in grades kindergarten through 6 during a 9-month school year (Ross et al., 2004). Using hierarchical linear modeling (HLM), the second experiment, involving 978 students (89.9% African American and 83% eligible for free/reduced-price lunch) found an average effect size of 0.175 SD per grade in grades 3 through 6 during a 9-month school year (Nunnery et al., 2006). After averaging the grade-level effect sizes across the two studies (where they overlap, i.e., in grades 3 through 6), the effect size for Reading Assessment averaged 0.279 SD per grade in grades kindergarten through 6, with an unusually disadvantaged population of students.

The only randomized study of Math Assessment, involving 1,880 students in grades 2 through 8 in 80 classrooms and 7 states, found an effect size of 0.324 SD during a 7-month period, after controlling for treatment integrity (Ysseldyke & Bolt, in press). The only national, peer-reviewed, quasi-experimental evaluation of Math Assessment, involving 2,202 students in grades 3 through 10 in 125 classrooms in 24 states, found that students in the treatment group gained an average of 0.392 SD per grade during 1 semester (18 weeks), compared with students not receiving Math Assessment (at pretest, the scores of treatment and comparison students were not:

Despite the re

The best studi

In Milwaukee, mated a random Parental Choice F

The results su

Milwaukee. The

programs are imp

in which the choi

program is imp

in which the choi

vouchers per year in

and 2.

for the 4th grade,
the same curricular studies involved students who demonstrated students who showed deviations (SD) from the norm such as the average. Teachers were reinforced with additional instruction for students who showed deviations and feedback studies (e.g., Ysseldyke & Tardrew, in press). The mean effect size across these two studies is 0.358 SD. Averaging the effect size estimates for Reading and Math Assessment produces an overall effect size of 0.319 SD.

**Spending**

Despite the research suggesting that rapid assessment may be effective in improving student achievement, several other proposals have dominated policy discussions, including proposals to increase educational expenditures. Estimates of the effects of increased spending on student achievement are bounded by two meta-analyses (and associated follow-up studies). Hanushek's (1986, 1989, 1997) meta-analyses suggest that there is only a weak relationship between educational spending and student achievement, whereas Greenwald et al. (1996) drew a more optimistic conclusion based on their meta-analysis and estimated that a 10% increase in per-pupil expenditure would increase student achievement in math and reading by 0.083 SD per year. Thus, 0.083 SD represents an upper-bound estimate of the effect of a 10% increase in preexisting patterns of educational spending.

**Voucher Programs**

The best studies of voucher programs used random assignment or lottery-based assignment designs. At most, these studies suggest small, positive effects on student achievement for students who received vouchers.

Three randomized field trials in New York City, Washington, D.C., and Dayton, Ohio, suggest that vouchers boosted achievement by a statistically insignificant 1.8 national percentile rank (NPR) points in “total achievement” (math and reading) during the course of 2 years (0.9 NPR points or 0.572 normal curve equivalent [NCE] points per year, equal to an effect size of 0.027 SD; Howell et al., 2002).

In Milwaukee, Wisconsin, where the use of a lottery system to allocate vouchers approximated a random assignment design, Rouse (1998) estimated that students selected for the Parental Choice Program showed zero gains in reading, but gained between 1.5 and 2.3 NCE points per year in math (an average of 1.9 NCE points, equal to an effect size of 0.090 SD) compared with comparison group students; whereas Greene et al. (1999) estimated that after 4 years, reading scores were 5.84 NCE points higher (an annual gain of 1.460 NCE points or an effect size of 0.069 SD), whereas math scores were 10.65 NCE points higher (an annual gain of 2.663 NCE points or an effect size of 0.126 SD).

The results suggest that the effects of voucher programs on voucher recipients are small when implemented in urban districts such as New York, Washington, D.C., Dayton, and Milwaukee. There is no evidence to suggest that effect sizes are larger when voucher programs are implemented on a larger scale or if the focus is on the effects of increased competition on students who remain in traditional public schools, rather than the effects on voucher recipients. Hoxby (2002) analyzed the effects of the Milwaukee voucher program, which she selected along with two charter school initiatives, “because they are the only ones in which the choice schools can, legally, garner a large enough share of enrollment to provide a nonnegligible amount of competition for the regular public schools” (p. 142). Hoxby reported annual gains of 2.1 NPR (1.176 NCE) points per year in reading (an effect size of 0.056 SD) and 2.8 NPR (1.568 NCE) points per year in math (an effect size of 0.074 SD; see Wisconsin Department of Public Instruction, 2007, Tables 1 and 3, 1996 norm conversions for the 4th grade, for the 25th through 50th national percentiles).
Table 1

Effect Sizes in Standard Deviation Units for Key Studies of Voucher Programs

<table>
<thead>
<tr>
<th>Study</th>
<th>Reading</th>
<th>Math</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Howell, Wolf, Campbell, &amp; Peterson, 2002</td>
<td>0.027</td>
<td>0.027</td>
<td>0.027</td>
</tr>
<tr>
<td>Rouse, 1998</td>
<td>0.000</td>
<td>0.090</td>
<td>0.045</td>
</tr>
<tr>
<td>Greene, Peterson, &amp; Du, 1999</td>
<td>0.069</td>
<td>0.126</td>
<td>0.098</td>
</tr>
<tr>
<td>Average</td>
<td>0.032</td>
<td>0.081</td>
<td>0.057</td>
</tr>
</tbody>
</table>

Although Greene and Winters (2004) also found positive effects of competition from vouchers, their study confounded the effect of vouchers with the branding effect of Florida’s school accountability system (Carnoy, 2001; Appendix A). In Florida, students attending schools that receive a grade of F twice during any 4-year period are eligible to receive vouchers to attend another public school or a private school. During a 1-year period, students eligible for vouchers gained 5.9 NPR (3.3 NCE) points in math on the Stanford Achievement Test, compared with students in all other Florida public schools (Greene & Winters, 2004). However, independent analyses strongly suggest that much of this effect may be attributable to the galvanizing effect on school staff of being branded an F school, rather than any effect due to vouchers (Carnoy, 2001). Subtracting the spurious branding effect results in a pure voucher effect size of only 0.02 SD (Carnoy, 2001).

This evidence suggests that it is appropriate to focus attention on the direct effect of voucher programs on voucher recipients, rather than any indirect effect on neighboring public schools. The average effect size across the three key studies of the direct effects of voucher programs is 0.032 SD in reading and 0.081 SD in math, or 0.057 SD overall, one-fifth of the 0.319 SD impact of rapid assessment (Table 1).

**Charter Schools**

Given that charter schools are extremely heterogeneous, it would be useful to differentiate the types of charter schools that are most effective. Although it is possible that researchers will be able to provide this information in the future, the evidence that is presently available fails to demonstrate that any particular type of charter school is significantly more effective than the others. Thus, policymakers are forced to fall back on a more basic question, namely whether charter school programs are, in general, an efficient use of public resources.

Ideally, large-scale randomized experiments comparing the achievement of students who were randomly accepted into charter schools with the performance of students who were randomly rejected and attended traditional schools would be used to estimate the effect of charter schools on student achievement. However, no true random-assignment studies have been completed thus far, and the two lottery-based studies that approximated a random-assignment design only involved a total of four charter schools, which is inadequate for estimating the average impact of charter schools. Averaging the effect sizes from these two studies suggests an effect of approximately 0.018 SD, but no generalizations can be made to the larger population of charter schools (analysis available on request from the author).

As with voucher programs, the indirect effect of charter school competition on public school performance is uncertain. Holmes, Desimone, and Rupp (2003, 2006) estimated the galvanizing effect of the presence of North Carolina’s charter schools on public school performance and concluded that charter school competition increased traditional school performance by about 1%. However, Bifulco and Ladd (2005, 2006), using a stronger methodological approach,
Table 2

<table>
<thead>
<tr>
<th>Study</th>
<th>All Years of Operation</th>
<th>Fifth Year of Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reading</td>
<td>Math</td>
</tr>
<tr>
<td>Bifulco &amp; Ladd, 2006</td>
<td>-0.095</td>
<td>-0.160</td>
</tr>
<tr>
<td>Sass, 2006</td>
<td>0.000b</td>
<td>-0.019b</td>
</tr>
<tr>
<td>Hanushek, Kain, Rivkin, &amp; Branch, 2005</td>
<td>-0.170</td>
<td>-0.170</td>
</tr>
<tr>
<td>Average</td>
<td>-0.088</td>
<td>-0.116</td>
</tr>
</tbody>
</table>

a. Bifulco and Ladd note that charter schools in operation for 5 years exhibited an anomalous dip in performance; therefore, estimated impacts for charter schools in operation for 4 years were substituted.
b. Using coefficients from the restricted value-added model.

concluded that North Carolina's charter schools exerted a negative effect on charter school student performance—an effect that was not offset by positive effects due to increased competition and was not explained by observable or unobservable differences in the students who attend charter schools (Bifulco & Ladd, 2005).

A prime concern when random assignment is not used is controlling selection bias (i.e., the threat that differences in outcomes are attributable to observable or unobservable differences between students who attend charter schools and comparison students who attend traditional schools). However, this threat can be controlled through the use of panel data with multiple observations of performance for each student who switches between a traditional school and a charter school (or the reverse). Thus, students serve as their own controls, and the benefit of charter schools is estimated by comparing achievement gains while attending a charter school to achievement gains while attending a regular public school.

Three large-scale studies used this method, which controls for student maturation and history effects as well as selection bias and, therefore, has many of the key strengths of a random-assignment design. These studies are used to derive the estimate of charter school effects used in this article's cost-effectiveness analysis. Despite differences in the charter school programs, the studies consistently found negative effect sizes. Bifulco and Ladd (2006) evaluated the effect of North Carolina's charter schools on student achievement, finding a negative effect size of 0.095 SD on reading scores and a negative effect size of 0.160 SD on math scores. Sass (2006) evaluated the effect of Florida's charter schools on student achievement, finding zero effect on reading scores and a negative effect size of 0.019 SD on math scores. Hanushek, Kain, Rivkin, and Branch (2005) evaluated the effect of Texas's charter schools on student achievement, finding zero effect on reading scores and a negative effect size of 0.019 SD on math scores. Hanushek, Kain, Rivkin, and Branch (2005) evaluated the effect of Texas's charter schools on student achievement, finding a negative effect size of 0.17 SD on combined reading and math scores. The effect of competition from charter schools on the performance of traditional public schools was inconsequential (Bifulco & Ladd, 2006; Sass, 2006). The average effect size across these key studies of charter schools was −0.088 SD in reading and −0.116 SD in math, or −0.102 SD overall. However, in all three studies, negative effects were largest in the first year of charter school operation and diminished with time. The average effect size for charter schools after 5 years in operation was 0.005 SD (Table 2).

The overall conclusion—that charter school effect sizes are small—does not change even if the larger group of studies using less rigorous methods is considered. An analysis of every study published in the period from 2000 to 2005 regarding the link between charter schools and academic achievement concluded that: "Regardless of the methods used, the results are mixed, some positive about charters and some negative, with null or mixed findings the most common" (Hill, Angel, & Christensen, 2006, p. 140).
Accountability

Studies by Amrein and Berliner (2002a, 2002b) concluded that there is no evidence that states that implemented high-stakes tests demonstrated improved student achievement on various measures, such as the Scholastic Achievement Test (SAT), American College Test (ACT), Advanced Placement (AP) exams, or the National Assessment of Educational Progress (NAEP). However, both their methodology and their findings have been challenged (Braun, 2004; Rosenshine, 2003).

Two studies have escaped methodological criticism (Carnoy & Loeb, 2002; Hanushek & Raymond, 2005). Both studies found positive effects of increased educational accountability pressure. Carnoy and Loeb (2002) modeled gains in student achievement as a function of differences in degree of accountability pressure and found that significantly stronger accountability was associated with modest increases in student achievement at the eighth-grade level, but effects of increased accountability on fourth-grade scores were not significant. An even more sophisticated study employed a longitudinal panel design with controls for fixed state effects as well as measures of parental education, school spending, and race (Hanushek & Raymond, 2005). This study, which provides the best available evidence for the impact of accountability policies, found that high stakes testing, where consequences are linked to student achievement, leads to larger growth in achievement than would have occurred in the absence of such tests, but effects were small, 0.2 SD over 4 years, or 0.05 SD per year (Hanushek & Raymond, 2005).

Cost Analysis

The next step is to integrate the preceding synthesis of the literature regarding effect sizes for each of the five alternative interventions with information about the cost of each intervention. The resulting cost-effectiveness analysis is intended to illustrate the relative cost-effectiveness of each intervention. The standard method of calculating a cost-effectiveness ratio for a given intervention is to divide the intervention’s costs by the intervention’s effect size (ES; Levin, 1988; Equation 1).

\[
\text{(1) Cost-Effectiveness Ratio} = \frac{\text{Cost}}{\text{ES}}
\]

An alternative is to calculate the inverse, an effectiveness/cost ratio (Equation 2).

\[
\text{(2) Effectiveness-Cost Ratio} = \frac{\text{ES}}{\text{Cost}}
\]

Finally, a relative effectiveness-cost ratio may be calculated to compare the effectiveness-cost ratio of rapid assessment with the effectiveness-cost ratio of an alternative intervention (Equation 3).

\[
\text{(3) Relative Effectiveness-Cost Ratio (REC)} = \frac{\frac{\text{ES}_{RA}}{\text{Cost}_{RA}}}{\frac{\text{ES}_{Alt}}{\text{Cost}_{Alt}}}
\]

Where \( \text{ES}_{RA} \) is the effect size (in standard deviation units) of rapid assessment, \( \text{ES}_{Alt} \) is the effect size (in standard deviation units) of a given alternative intervention, \( \text{Cost}_{RA} \) is the cost of rapid assessment, and \( \text{Cost}_{Alt} \) is the cost of the alternative intervention. Costs include opportunity costs (i.e., the foregone benefits of the best alternative use of the resources needed to implement an intervention).
is no evidence that achievement on variance College Test of Educational have been challenged 2002; Hanushek & accountability as a function of different accountable eighth-grade level significant. An even roles for fixed state race (Hanushek & for the impact of occurs are linked to ve occurred in the 0.05 SD per year

garding effect sizes cost of each intervention’s effect size
cost-effectiveness-cost effectiveness of each intervention

technology. ES_{AE} is the cost ion. Costs include resources needed

The cost analysis was conducted with regard to two widely implemented variants of rapid assessment, the characteristics of which match the previously cited characteristics of effective feedback systems. Reading Assessment and Math Assessment, described above, provide immediate testing feedback in reading and math to each student, 2 to 5 times per week. Rapid assessment programs have been implemented in classrooms in more than 65,000 schools (Northwest Regional Educational Laboratory, 2006), including statewide implementation of Reading Assessment in Idaho (Renaissance Learning, 2002).

Table 3 lists the cash costs associated with program implementation. For the purpose of the cost analysis, it is assumed that both Reading Assessment and Math Assessment are purchased and implemented for each student receiving the rapid assessment intervention, along with two diagnostic assessments (STAR Reading and STAR Math) and mark-scan devices for each classroom. For the purpose of calculating the costs per student, it is assumed that initial one-time costs are averaged over an enrollment of 500 students in 25 classrooms per building. Initial fees, teacher and administrator training, and the cost of the scanners are amortized over 7 years, which is (arbitrarily) assumed to be the life of the program (schools that choose to continue using the programs for a longer period of time would effectively reduce the annual cost). The costs include access to 100,000 book quizzes for every student, plus access to Math Assessment grade-level libraries tagged to state standards for grades 1 through 7 as well as multiple subject-area libraries for the secondary grades (prealgebra, algebra 1, algebra 2, geometry, probability and statistics, precalculus, calculus, basic math, chemistry, physics). The assessment programs are simple to implement; thus, an administrator could instruct each teacher regarding the use of the software. However, the Rapid Assessment Corporation offers full-day training sessions costing $149 per teacher, and the cost analysis assumes that every classroom teacher and one administrator (for every 500 students) undergoes 2 full-day training sessions, 1 day for Math Assessment and 1 day for Reading Assessment. In addition, the cost analysis assumes a 50% teacher turnover rate during the 7-year implementation period and assumes that each new teacher receives 2 full-day training sessions.

Implementation requires that each classroom of students has access to one computer and one printer (math problems are printable so that students can work individually without using a computer). Based on a nationally representative survey, 93% of all instructional classrooms were online by the year 2003, implying that students in those classrooms had access to at least one classroom computer, and a linear extrapolation of recent trends in online access suggests that 100% of classrooms will have access to a classroom computer by the year 2006 (Parsad & Jones, 2005). In addition, researchers note that available computer resources are frequently underused (Cuban, 2001). Although most classrooms that have a computer also have a printer, the printer may be cheap and unreliable. However, given that internet-connected computers can be linked through a local area network (LAN) to print from any printer in the same building, it is feasible to use high-capacity printers in the school’s media center (Reilly, n.d.). It is also possible to print to a Xerox machine in the same building if the Xerox is equipped with a LAN card (Reilly, n.d.). Thus, it is feasible to implement the rapid assessment programs without purchasing additional computers or printers.

The researcher verified (through classroom observation of operating procedures as well as teacher and administrator interviews in eight elementary, middle, and high schools) that cost figures reflect the actual operating experience of schools in a typical district. Although a school of 500 students taking 2 to 5 assessments per week in math and reading suggests that 2,000 to 5,000 assessments are processed weekly, the burden on teachers is minimal because
Table 3
Costs to Implement Rapid Assessment

<table>
<thead>
<tr>
<th>Item</th>
<th>Fixed Cost (Per School)</th>
<th>Annual Variable Cost (Per Student)</th>
<th>Annual Variable Cost (500 Students)</th>
<th>Total Annual Cost (Per Student)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading Assessment</td>
<td>$1,499.00</td>
<td>$4.00</td>
<td>$2,000.00</td>
<td>$4.43</td>
</tr>
<tr>
<td>STAR Reading</td>
<td>$1,499.00</td>
<td>$0.39</td>
<td>$195.00</td>
<td>$0.82</td>
</tr>
<tr>
<td>Reading Assessment training</td>
<td>$5,736.50</td>
<td>—</td>
<td>—</td>
<td>$1.64</td>
</tr>
<tr>
<td>Math Assessment</td>
<td>$1,499.00</td>
<td>$4.00</td>
<td>$2,000.00</td>
<td>$4.43</td>
</tr>
<tr>
<td>STAR Math</td>
<td>$1,499.00</td>
<td>$0.39</td>
<td>$195.00</td>
<td>$0.82</td>
</tr>
<tr>
<td>AccelScan scanners and cards</td>
<td>$7,875.00</td>
<td>$8.10</td>
<td>$4,050.00</td>
<td>$10.35</td>
</tr>
<tr>
<td>Math Assessment training</td>
<td>$5,736.50</td>
<td>—</td>
<td>—</td>
<td>$1.64</td>
</tr>
<tr>
<td>Total</td>
<td>$25,344.00</td>
<td>$16.88</td>
<td>$8,440.00</td>
<td>$24.13</td>
</tr>
<tr>
<td>After 10% discount</td>
<td>$22,809.60</td>
<td>$15.19</td>
<td>$7,596.00</td>
<td>$21.72</td>
</tr>
</tbody>
</table>

Source: Rapid Assessment Corporation, August 8, 2006.

a. Fixed costs are spread over 500 students and averaged over a 7-year implementation period.
b. $149/full-day training x (37.5 teachers + 1 administrator).
c. Assuming 25 classrooms per school x $315 per scanner.
d. 180 instructional days per 9-month year x $45 per mark card.

Given that rapid assessment is designed to supplement, rather than supplant, normal reading and math activities, the opportunity costs of using rapid assessment primarily involve the time required by teachers to monitor students. During designated periods of the day devoted to rapid-assessment-related activities, the majority of students read books selected from the school library or work on printed sets of math problems (Yeh, 2006). Students who complete a book sit at the classroom computer to take a brief comprehension quiz. Students who complete a set of math problems scan their bubble sheets. Teachers typically tutor individual students or small groups of students. To the extent that these activities do not displace the reading and math activities that may be expected in the absence of rapid assessment, and given that the assessments are self-administered by students and scoring and reporting is handled by computer software, the opportunity costs of implementing the program primarily involve the time required by teachers to ensure that students select and read appropriate books, take the comprehension quiz without assistance from other students, and complete and scan their answers to assigned math problems.

According to teachers who were interviewed, the time saved by the program’s scoring and student progress-monitoring features (which replace the time-consuming conventional tasks of grading math homework and assessing reading comprehension) more than offset the opportunity costs of helping students to select books and monitoring student use of the classroom computer (Yeh, 2006). Thus, the costs of rapid assessment are primarily the cash costs listed in Table 1, equal to $21.72 per student, per year, in 2006 dollars. However, large fixed costs ($22,809.60) incurred at start-up create opportunity costs equal to the income that would otherwise be earned if this amount were instead expended in seven equal annual installments (the arbitrary lifetime of the program) and the remaining funds were invested in an interest-bearing account. Assuming a real interest rate of 3% and a discount rate of 3.5%, the foregone income is $1,928.82, or $0.55 per student per year in a school with 500 students during the 7-year lifetime of the program. The social cost of rapid assessment is the cash cost ($21.72) plus opportunity costs ($0.55), or a total of $22.27 per student, per year.

The most recent dollars equals $1C equals $1,046.40, and the August 20 per pupil expenditure 0.01432420 and a Equation 3 indicates the gains t: expenditures.

Significant costs at random from p schools and distri switch to private numbers of studie the District of Co pulled from the 1, of Columbia, and within each scho students per grade ecutations suggest

As of January (Belfield, 2006). Cleveland’s public 1,204 of the 5,7 Center for Educ students per grad (The present cost schools, rather th: attended public sc alternative ways o

The most rece Milwaukee’s pub Department of Pt pulled from each or an average of 2 per school).

Florida had en 2005 (Florida De per school, assun schools in Flori Florida Supreme Scholarship Prog
Cost of a 10% Increase in Educational Expenditures

The most recent available data show that total expenditure per pupil in constant 2004–2005 dollars equals $10,464 (National Center for Education Statistics, 2005c). A 10% increase equals $1,046.40, or $1,118.83 in 2006 dollars, using the January 2005 price deflator (190.7) and the August 2006 price deflator (203.9), meaning that it would cost $1,118.83 to increase per pupil expenditure by 10%. Using an effectiveness-cost ratio for rapid assessment of 0.01432420 and an effect size of 0.083 SD for a 10% increase in per pupil expenditure, Equation 3 indicates that the achievement gains per dollar from rapid assessment are 193 times the gains that accrue through a 10% increase in preexisting patterns of educational expenditures.

Costs of Voucher Programs

Significant costs to society would arise to the extent that voucher programs pull students at random from public school classrooms, making it difficult (if not impossible) for public schools and districts to reduce expenditures at the same rate that students use vouchers to switch to private schools. For example, in the study by Howell et al. (2002), very small numbers of students were offered vouchers in the baseline year: 1,300 in New York, 809 in the District of Columbia, and 515 in Dayton, Ohio. If these 2,624 students were randomly pulled from the 1,427 public schools in those three cities (1,213 in New York, 165 in District of Columbia, and 49 in Dayton; National Center for Education Statistics, 2000, 2001) and, within each school, were randomly pulled from an average of four grade levels, only 0.46 students per grade level, per school used vouchers to transfer to private schools. Similar calculations suggest that the same issue arises in each of the major voucher programs.

As of January 2006, 5,734 students participated in Cleveland’s (Ohio) voucher program (Belfield, 2006). Based on the proportion (21%) of students who transferred from Cleveland’s public schools during the 2000–2001 school year (Schiller, n.d), I estimate that 1,204 of the 5,734 students transferred from Cleveland’s 122 public schools (National Center for Education Statistics, 2005a), an average of 9.87 students per school or 2.47 students per grade level, per school (assuming an average of four grade levels per school). (The present cost-effectiveness analysis focuses on students who transferred from the public schools, rather than recipients of vouchers who were previously homeschooled or who never attended public schools, because the purpose of the analysis is to determine the effectiveness of alternative ways of improving the achievement of students currently enrolled in public schools.)

The most recent available data show that 3,074 students used vouchers to transfer from Milwaukee’s public schools to private schools in the 2004–2005 school year (Wisconsin Department of Public Instruction, 2006), suggesting that an average of 13.08 students were pulled from each of Milwaukee’s 235 public schools (U.S. Department of Education, 2005a), or an average of 3.27 students per grade, per school (assuming an average of four grade levels per school).

Florida had enrolled 734 students into the Opportunity Scholarship Program by November 2005 (Florida Department of Education, 2007), or an average of 0.05 students per grade level, per school, assuming an average of four grade levels per school in each of the 3,700 public schools in Florida (U.S. Department of Education, 2005a). (However, on January 5, 2006, the Florida Supreme Court issued a ruling declaring the private school option of the Opportunity Scholarship Program unconstitutional [Florida Department of Education, 2006].)
On average, therefore, very few students (between 0.05 and 3.27 students per grade level, per school) used vouchers to transfer to private schools in the city of New York, the District of Columbia, Dayton, Cleveland, Milwaukee, and the state of Florida. This would not have allowed any of the public school–building principals to eliminate teaching positions in response to the transfers, and there would have been no decrease in overhead, facilities, or administrative costs. At the same time, there are far fewer private schools, and thus voucher programs funnel public school students into a small number of classrooms that quickly fill up and require more buildings, teachers, and administrators in proportion to the ratio of public to private school students (i.e., 9:1; see National Center for Education Statistics, 2005b). If public school students enroll in private schools, the private school classrooms fill up 9 times faster than public school classrooms are depleted. Thus, voucher programs typically create a need for private schools to add teachers, administrators, and facilities without corresponding reductions in public school costs. These inefficiencies increase the total costs of teaching the same number of students (both public and private).

Although the real cost of educating the remaining public school students barely declines, state funding declines by a fixed amount every time a student uses a voucher to transfer to a private school (Miles & Roza, 2004). By law, school districts must balance their budgets: When revenues are reduced, expenditures must also be reduced. When building principals are unable to save costs by consolidating classrooms and reducing teaching positions, the reduction in revenue associated with declining enrollment is typically offset by reducing other services—art, music, extracurricular programs and other support services (e.g., see Horn & Miron, 2000, who identified the same issue when students transfer to charter schools). These reductions in expenditures can be mistakenly attributed to cost savings, for example, when expenditures are regressed on factors including student enrollment in order to calculate the reduction in expenditure associated with a decline in enrollment (see, e.g., Gottlob, 2004). However, expenditures decline because of the forced reduction in services, not because of savings generated through reduced enrollments.

The most reliable way to determine the amount of any cost savings would be to analyze actual financial records from local school districts, rather than imputing these savings by using the per-pupil expenditure amounts used by state governments to allocate funding to local districts. This type of analysis of local district records has apparently not been performed with regard to voucher programs, (see U.S. General Accounting Office, 2001, regarding the Cleveland and Milwaukee voucher programs). However, the cost savings that may accrue if the largest voucher program (in Milwaukee) is tripled in size may be estimated from the cost savings generated as a consequence of the 9,870 Philadelphia public school students who transferred to charter schools by the 2000–2001 school year (Pennsylvania Economy League Eastern Division, 2001). The school district was able to recover 17.38% of the costs to educate the transfer students ($10,100,000 of the $58,120,000 costs, including the $3,220,000 share of transportation and administrative costs attributable to the students who transferred), implying that 82.62% of the costs are additional costs to society. However, there is no evidence that take-up rates for large-scale voucher programs would approach the take-up rate of Philadelphia’s charter school program, and thus, there is no evidence that savings would approach Philadelphia’s 17.38% figure.

The real social cost of educating large numbers of students in private schools (who are currently educated in public schools) is difficult to estimate for several reasons: Private school tuition figures exclude costs that are offset by corporate and noncorporate subsidies (U.S. General Accounting Office, 2001), as well as the cost of services that would be required by many students (and, by law, are currently provided by public schools, but not private schools), including transportation, free and reduced-price meals, special education, vocational education, and services for Levin, 1998; Le private schools, Charter schools resulting in teac (Nelson et al., 2 are more than $ adjusted dollars the June 1998 p school cost info switching to pri systems to mon services th estimate of the transfer to priv If the effecti size for vouche dollar from rapi

When studer identical to the plest case, if a expenditures (v well as enrollm savings) to soci However, to instead are ne teachers, and f society pays fo tive units, facil voucher progr public schools with very little rarely be able taining the ne required to tea (2000) found t there are no co loss of this rev art, music, and reduced by cu charter schools For the pur able to student mated to be 17 transfer studer
nts per grade level, the District would not have enough positions in head, facilities, or, and thus voucher that quickly fill up the ratio of public statistics, 2005b). Ifoms fill up 9 times typically create a corresponding costs of teaching thens barely declines, shool to transfer to once their budgets: building principals hing positions, the offset by reducing services (e.g., see transfer to charter o cost savings, for enrollment in order rollment (see, e.g., action in services,ould be to analyze these savings by allocate funding to nly not been per-ffice, 2001, regard-t savings that may be estimated from sic school students sylvania Economy 7.38% of the costs ts, including the the students who-ty. However, there approach the take-eidence that savings tools (who are cur-ons: Private school te subsidies (U.S.uld be required by ot private schools), cational education, and services for students with disabilities and limited English proficiency (Belfield, 2006; Levin, 1998; Levin & Driver, 1997). Although comprehensive cost data are not available for private schools, such data are available for charter schools (Nelson, Muir, & Drown, 2003). Charter schools are subject to much of the same competitive cost pressures as private schools, resulting in teacher salaries that are substantially lower than traditional public school salaries (Nelson et al., 2003). For example, in Texas and Minnesota, charter school teacher salaries are more than $13,760 lower, and in Pennsylvania, more than $17,513 lower, in inflation-adjusted dollars, than are comparable host-district teacher salaries (Nelson et al., 2003), using the June 1998 price deflator (163) and the August 2006 price deflator (203.9). Thus, charter school cost information was combined with estimates of cost savings resulting from students switching to private schools, and cost estimates for transportation, record keeping, information systems to monitor and assess voucher eligibility of both students and schools, and adjudication services that would be required by a voucher system, to obtain an inflation adjusted final estimate of the per-pupil cost to society when typical public school students use vouchers to transfer to private schools and receive comparable services: $9,646.01 (Table 4).

If the effectiveness-cost ratio for rapid assessment is 0.01432420, and the average effect size for voucher programs is 0.057 SD, Equation 3 indicates that the achievement gains per dollar from rapid assessment are 2,424 times the gains that accrue through voucher programs.

Costs of Charter Schools

When students attend public charter schools, many of the costs to educate the students are identical to the costs of educating the same students in traditional public schools. In the simplest case, if a regular public school is converted to a charter school, and if revenues and expenditures (which are measures of resources expended by society on charter schools), as well as enrollment and educational services, remain identical, there are no additional costs (or savings) to society as a result of implementing the charter school.

However, to the extent that charter schools are not simply converted public schools, but instead are newly created schools that require new facilities, new administrators, and new teachers, and pull students from widely scattered classrooms in traditional public schools, society pays for the creation and maintenance of a larger number of school-level administrative units, facilities, and teachers (Nelson, 1997). As illustrated by the calculations regarding voucher programs, only a fraction of these costs are offset through savings in the sending public schools and districts. The cost to hire charter school teachers will largely be a new cost with very little offset due to reduced need for teachers in the sending schools, which will rarely be able to reduce their own costs in proportion to the costs of establishing and maintaining the new charter schools. In essence, more facilities, administrators, and teachers are required to teach the same total number of students (Nelson, 1997). Thus, Horn and Miron (2000) found that when regular public schools lose scattered students to a charter school, there are no cost savings. However, each lost student took away $6,000 of state funding. The loss of this revenue was absorbed by the local district, typically through reduced funding of art, music, and extracurricular programs and support staff and services. Expenditures were reduced by cutting services, rather than through savings attributable to student transfers to charter schools.

For the purpose of the charter school cost analysis, the public school cost savings attributable to students transferring to charter schools, based on Philadelphia's experience, was estimated to be 17.38% of the costs (including administrative and transportation costs) to educate transfer students (see above). The annual, per-pupil, inflation-adjusted cost to society was
Table 4
Annual Per Pupil Cost to Society of a Voucher Program

<table>
<thead>
<tr>
<th>Cost Category</th>
<th>2006 Dollars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction, administration, facilities, and operations</td>
<td>7,098.88a</td>
</tr>
<tr>
<td>Food service</td>
<td>323.00b</td>
</tr>
<tr>
<td>Special education</td>
<td>1,082.55a</td>
</tr>
<tr>
<td>Transportation</td>
<td>2005.57c</td>
</tr>
<tr>
<td>Cost savings (17.38%)</td>
<td>(1,826.64)</td>
</tr>
<tr>
<td>Voucher record keeping and monitoring</td>
<td>68.19</td>
</tr>
<tr>
<td>Information services</td>
<td>50.81</td>
</tr>
<tr>
<td>Adjudication services</td>
<td>49.82</td>
</tr>
<tr>
<td>Other</td>
<td>793.83b</td>
</tr>
<tr>
<td>Total</td>
<td>9,646.01</td>
</tr>
</tbody>
</table>

b. Adjusted for the additional costs of mandatory services provided by public schools but not charter schools.
c. Levin and Driver (1997). Adjusted for inflation using the June 1995 price deflator (152.5) and the August 2006 price deflator (203.9).
d. Students frequently travel further to reach private schools and on routes and schedules incompatible with traditional public school bus transportation. In Cleveland, for example, some voucher students' homes are too remote to be served efficiently by buses, forcing the Cleveland Municipal School District to pay an estimated $1,372-$2,058 per pupil per year to transport these students by taxi. This is an extra $800-$1,486 above the $572 cost of traditional school bus transportation (Neas, 2001), adjusted for inflation using the September 2001 price deflator (178.3) and the August 2006 price deflator (203.9).
e. Including compensatory, vocational, bilingual, and career and technology education.

Estimated using a methodology similar to that used with regard to voucher programs, except that costs specific to voucher programs (voucher record keeping, monitoring, information and adjudication services) were dropped, and the actual costs for student transportation incurred in Philadelphia were used, resulting in a figure of $8,086.30 per transfer student (Table 5).

If the effectiveness-cost ratio for rapid assessment is 0.01432420 and the average effect size for charter schools is 0.005 SD, Equation 3 indicates that the achievement gains per dollar from rapid assessment are 23,166 times the gains that accrue through charter schools.

Costs of Accountability

Estimating the costs of implementing stronger educational accountability is difficult because there is no standard set of activities associated with increased accountability. However, the core attribute is the implementation of standardized tests where results are linked with consequences for students or schools. A majority of states have implemented high school exit exams that students must pass to graduate from high school. By the year 2009, these requirements will affect 70% of students nationwide (Gayler, Chudowsky, Hamilton, Kober, & Yeager, 2004). With few exceptions, the exams require mastery of material at the high school level (Center on Education Policy, 2005). Thus, the analysis of the costs of implementing increased educational accountability focuses primarily on two costs: the costs of implementing the standardized achievement tests and the incremental social costs of denying diplomas to students who would otherwise receive them (due to the incremental increase in the number of students who are denied diplomas solely because exit exams raise the graduation bar). Costs are calculated as deviations from the preexisting status quo (before the implementation of exit exams and higher graduation standards). If no additional students are denied diplomas as a result, cost-effectiveness costs does not appear in the analysis.

Although some teacher profession (NCLB) mandate enhancements may exceed the level of achievement gain (Hanushek & Raymond, 2004; McGee, 2003), they remain a standard set of activities associated with increased accountability. As a result, the costs of implementing exit exams and higher graduation standards are projected to be $3.9 billion (Gayler, 2006) and to exceed the level of achievement gain (Hanushek & Bay, 2004). In 2003, the cost of multiple-choice testing is $3.9 billion if incurred by schools excepting disabilities, incurred by schools (Center on Education Policy, 2005). Thus, the analysis of the costs of implementing increased educational accountability focuses primarily on two costs: the costs of implementing the standardized achievement tests and the incremental social costs of denying diplomas to students who would otherwise receive them (due to the incremental increase in the number of students who are denied diplomas solely because exit exams raise the graduation bar). Costs are calculated as deviations from the preexisting status quo (before the implementation of exit exams and higher graduation standards). If no additional students are denied diplomas as a result, cost-effectiveness costs does not appear in the analysis. However, the core attribute is the implementation of standardized tests where results are linked with consequences for students or schools.
Table 5
Annual Per-Pupil Cost to Society of a Charter School Program

<table>
<thead>
<tr>
<th>Cost Category</th>
<th>2006 Dollars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction</td>
<td>4,549.51a</td>
</tr>
<tr>
<td>Administration</td>
<td>1,404.01a</td>
</tr>
<tr>
<td>Facilities and operations</td>
<td>1,445.36c</td>
</tr>
<tr>
<td>Food service</td>
<td>323.00a</td>
</tr>
<tr>
<td>Special education</td>
<td>1,082.55c</td>
</tr>
<tr>
<td>Transportation</td>
<td>322.09a</td>
</tr>
<tr>
<td>Cost savings (17.38%)</td>
<td>1,534.05</td>
</tr>
<tr>
<td>Otherd</td>
<td>793.83a</td>
</tr>
<tr>
<td>Total</td>
<td>8,086.30</td>
</tr>
</tbody>
</table>

b. Adjusted for the additional costs of mandatory services provided by public schools but not charter schools.
d. Including compensatory, vocational, bilingual, and career and technology education.

Table 5 Annual Per-Pupil Cost to Society of a Charter School Program

<table>
<thead>
<tr>
<th>Cost Category</th>
<th>2006 Dollars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction</td>
<td>4,549.51</td>
</tr>
<tr>
<td>Administration</td>
<td>1,404.01</td>
</tr>
<tr>
<td>Facilities and operations</td>
<td>1,445.36</td>
</tr>
<tr>
<td>Food service</td>
<td>323.00</td>
</tr>
<tr>
<td>Special education</td>
<td>1,082.55</td>
</tr>
<tr>
<td>Transportation</td>
<td>322.09</td>
</tr>
<tr>
<td>Cost savings</td>
<td>1,534.05</td>
</tr>
<tr>
<td>Other</td>
<td>793.83</td>
</tr>
<tr>
<td>Total</td>
<td>8,086.30</td>
</tr>
</tbody>
</table>

b. Adjusted for the additional costs of mandatory services provided by public schools but not charter schools.
d. Including compensatory, vocational, bilingual, and career and technology education.

diplomas as a result of raising the graduation bar, the only costs that would be included in the cost-effectiveness analysis would be the costs of administering the exams. The calculation of costs does not assume there are no benefits of raising the graduation bar: Those benefits are measured in terms of increased student achievement.

Although some researchers assert that extraordinary increases in remediation services and teacher professional development will be necessary to meet the federal No Child Left Behind (NCLB) mandate that all students meet proficiency standards by the year 2014, the cost of these enhancements is not included in the current cost analysis because the enhanced level of services exceeds the level of educational services associated with the 0.05 $SD$ estimate of student achievement gains, derived from achievement data from the 1992 to 2002 period, used here (Hanushek & Raymond, 2005). Presumably, the implementation of these enhancements would be associated with an impact estimate for educational accountability that exceeds 0.05 $SD$.

In 2003, the General Accounting Office estimated that if every state used its existing mix of multiple-choice and open-response question types, the cost of the assessments would be $3.9 billion over 7 years, excluding the costs of alternative assessments for students with disabilities, expenditures for English language proficiency testing, and expenditures incurred by school districts that could increase the total to $7 billion (National School Boards Association, 2003). Thus, the lower figure is a conservative estimate of the costs of implementing the state-mandated tests, equaling $4.3 billion over 7 years (or $618,407,341 per year) in 2006 dollars after adjusting for inflation using the June 2003 price deflator (183.7) and the August 2006 price deflator (203.9), or $12.73 for each of the 48,574,000 students that are projected to be enrolled in 2006 (see U.S. Department of Education, 2005d).

Warren, Jenkins, and Kulick (2006) provide the current best estimate of the incremental increase in the fraction of students who potentially would be denied diplomas as a consequence of implementing stronger educational accountability nationwide, defined as the implementation of exit exams that require mastery of high school level material. The achievement of 28 graduating classes between 1975 and 2002 was modeled based on data from each of the 50 states and the District of Columbia, yielding 1,428 state-years as the units of analysis for models estimating the effect of the implementation of exit exams. After adjusting for an array of
covariates, high school completion rates were 2.1% lower in states with this type of exit exam, controlling for state and year fixed effects, implying that an extra 89,908 students would be denied diplomas annually (0.021 \times \text{national public student 9th-grade enrollment of 4,281,345 in academic year 2004–2005}) (U.S. Department of Education, 2005b).

Flores-Lagunes and Light (2004) estimated four models of the "sheepskin" (signaling) effect—that is, the difference in real earnings for an individual who obtains a high school diploma compared to a similar individual who does not, controlling for highest level of education (exactly 12 years), actual work experience, and age at expected graduation date. The mean hourly wage of an individual with a high school diploma was $13,932 ($12,425 + $1,507). For the two models that best fit the data, estimates of the sheepskin effect ranged from 7.37% to 8.15%, or between $94,514 and $104,517 (an average of $99,516) of the $1,282,419 earned by a high school graduate over the course of a lifetime, in present value terms (Day & Newburger, 2002), assuming that real earnings grow at a rate equal to a discount rate of 3.5% (see Sullivan, 1997), adjusted for inflation using the March 1999 price deflator (165.0) and the August 2006 price deflator (203.9).

The cost of raising the graduation bar ($8,947,284,528) was calculated by multiplying the increased number of students denied diplomas every year (89,908) by the discounted present value of the lifetime difference in income (a measure of the economic output lost to society) between an individual who graduates from high school (but completes no further schooling) and an individual who does not graduate but is identical with regard to years of schooling, actual work experience, and age ($99,516), then divided by the comparable 2004 figure for national public student prekindergarten to 12th (PK–12) grade enrollment (i.e., 48,359,697; U.S. Department of Education, 2005b) to derive the per pupil cost of $185.01.

The total cost of increased educational accountability is the cost of the assessments ($12.73), plus the discounted present value of the cost to society of lost economic output as a result of the shift toward standards-based exit exam requirements ($185.01), or a total of $197.74. Multiplied by projected 2006 PK–12 enrollment of 48,574,000 students that are projected to be enrolled in 2006, the total cost is $9,605,022,760 per year in 2006 dollars. Using the adjusted 0.05 SD effect size estimate derived from Hanushek and Raymond (2005), Equation 3 indicates that the achievement gains per dollar from rapid assessment are 57 times the gains that accrue through accountability testing.

**Summary**

For each intervention, Table 6 summarizes the effect size, annual cost per student, total annual cost to implement the intervention for every student in U.S. public schools enrolled in PK–12, and the effectiveness-cost ratio. Rapid assessment is substantially more effective and less costly than the alternative interventions, resulting in an effectiveness-cost ratio that is between 57 and 23,166 times larger than the ratios for the alternatives. This implies that a dollar spent on rapid assessment raises student achievement 23,166 times as fast as a dollar spent on charter schools.

Although the magnitude of the relative effectiveness-cost ratios may appear to be improbable, they mainly reflect the very small effect sizes and large costs of increased spending, voucher programs, charter schools, and educational accountability. To draw an analogy, if "Investment A" returns $231.66 per year (equivalent to 23,166 cents), that investment is 23,166 times as profitable as "Investment B," which only returns 1 cent per year. The cost-effectiveness of rapid assessment only seems improbable because it is being compared with highly ineffective and costly alternatives.

### Effect Size, Am

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Am</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapid assessment</td>
<td>57</td>
</tr>
<tr>
<td>Increased spending</td>
<td></td>
</tr>
<tr>
<td>Vouchers</td>
<td></td>
</tr>
<tr>
<td>Charter schools</td>
<td></td>
</tr>
<tr>
<td>Accountability</td>
<td></td>
</tr>
</tbody>
</table>

### Sensitivity Analysis

A key assumption is whether all classrooms have access to computers and software. The cost of rapid assessment remains 1,828 times as cost-effective as charter schools, and 43 times as cost-effective as vouchers, plus school's media center requires $103.33 per student, remains 42 times as cost-effective as charters, and 15 times as cost-effective as increased spending.
s type of exit exam, 
3 students would be 
llment of 4,28L345 
epskin" (signaling) 
tains a high school 
ighest level of edu-
uation date. The 
13.932 ($12,425 + 
pskin effect ranged 
of $99,516) of the 
le,in present value 
rate equal to a dis-
put lost to society) 
(annual cost of further schooling) 
ble 2004 figure for 
at (i.e., 48,359,697; 
0.01), or a total of 
students that are pro-
2006 dollars. Using 
1 Raymond (2005), 
sment are 57 times 
}

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Effect Size (SD)</th>
<th>Cost Per Student (Dollars)</th>
<th>Total Cost (Billions of Dollars)</th>
<th>Effectiveness-Cost Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapid assessment</td>
<td>0.319</td>
<td>22.27</td>
<td>1.1</td>
<td>0.01432420</td>
</tr>
<tr>
<td>Increased spending</td>
<td>0.083</td>
<td>1,118.83</td>
<td>54.3</td>
<td>0.00007419</td>
</tr>
<tr>
<td>Vouchers</td>
<td>0.057</td>
<td>9,646.61</td>
<td>468.5</td>
<td>0.00000591</td>
</tr>
<tr>
<td>Charter schools</td>
<td>0.005</td>
<td>8,086.30</td>
<td>392.8</td>
<td>0.00000062</td>
</tr>
<tr>
<td>Accountability</td>
<td>0.050</td>
<td>197.74</td>
<td>9.6</td>
<td>0.00025286</td>
</tr>
</tbody>
</table>

a. Standard Deviation / Cost Per Student, in dollars.
b. If every student used a voucher or enrolled in a charter school, there might be additional cost savings after a period of adjustment involving the closure (or conversion) of public schools. However, there is no evidence that universal availability of voucher or charter school programs would substantially alter per-pupil costs because existing take-up rates are low.
c. Using alternate estimate of the effect size (0.005 SD) after initial shakedown period.

Sensitivity Analysis I

A key assumption is that schools do not need to purchase additional computers and printers to implement rapid assessment. This assumption is based on research implying that virtually all classrooms will have access to at least one online computer by the year 2006 (Parsad & Jones, 2005) plus the ability of all online computers to print to a high capacity printer in a school's media center (Reilly, n.d.); this was verified through interviews with teachers and through observations of classrooms in which rapid assessment was used. However, if each classroom requires new equipment, an entire system, including a computer, monitor, keyboard, mouse, software, service plan and printer, may be purchased from Dell for $1,015. Assuming that a complete system is purchased for every classroom of 20 students, and the cost is amortized over a 7-year period, the annual cost per student is $7.25, raising the total cost of rapid assessment to $29.52 per student. The effectiveness-cost ratio falls to 0.01080623, but rapid assessment remains 146 times as cost-effective as a 10% increase in per-pupil expenditure, 1,828 times as cost-effective as voucher programs, 17,429 times as cost-effective as charter schools, and 43 times as cost-effective as increased educational accountability.

Sensitivity Analysis II

A second assumption, based on interviews with teachers and observations of classrooms in which rapid assessment is used, is that the rapid assessment software saves more teacher time (primarily time that would otherwise be spent grading math homework and assessing reading comprehension) than is consumed in noninstructional tasks, such as supervising student use of the computer, scanner, and printer. However, if teachers do not save time and, instead, lose 15 min per day (or 1.25 hr per week), the annual cost is $81.06 per student, assuming 20 students per teacher and an annual teacher salary of $51,880, adjusted for inflation using the January 2000 price deflator (168.8) and the August 2006 price deflator (203.9; U.S. Department of Education, 2005c). The total cost for rapid assessment increases to $103.33 per student, reducing the effectiveness-cost ratio to 0.00308720, but rapid assessment remains 42 times as cost-effective as a 10% increase in per-pupil expenditure, 522 times as cost-effective as voucher programs, 17,429 times as cost-effective as charter schools, and 43 times as cost-effective as increased educational accountability.
Conclusion

Comparisons of effect sizes suggest that rapid assessment is potentially a more promising approach for improving student achievement than increases in the preexisting pattern of spending, voucher programs, charter schools, or testing-based accountability. The research findings presented above suggest that rapid assessment is 4 times as effective as a 10% increase in per-pupil expenditure, 6 times as effective as vouchers, 64 times as effective as charter schools, and 6 times as effective as test-based accountability. The cost-effectiveness calculations suggest that achievement gains per dollar from rapid assessment are 193 times the gains from vouchers, 23,166 times the gains from charter schools, and 57 times the gains from increased accountability.

In concrete terms, the research findings suggest that improving student achievement by 0.319 SD (or approximately 3 months of learning) for every student would take less than 1 year and $1.1 billion, using rapid assessment. To achieve the same effect size through increased spending would take 4 years and $208.7 billion. Voucher programs would take 6 years and $2.6 trillion to achieve the same effect. Charter schools would take 64 years and $25.1 trillion, and increased accountability would take 6 years and $61.3 billion. These calculations suggest the magnitude of the differences in effectiveness and cost.

If rapid assessment is more effective than the alternatives examined here, an obvious question is why it has not been more widely adopted. Although teachers in 65,000 schools have adopted one or more of the rapid assessment programs (Northwest Regional Educational Laboratory, 2006), only a minority of those schools have adopted the programs school-wide. Reading specialists are concerned that Reading Assessment does not include a measure of oral reading comprehension or teacher observations of reading behavior, that the Flesch-Kincaid reading index that is used to establish the readability level of trade books does not account for the reader’s level of interest in particular topics, and that reading comprehension is assessed through a multiple-choice test that does not allow for written responses (Biggers, 2001). Teachers are also concerned that student progress is measured through a point system that rewards students for reading long books with high levels of difficulty (determined by the Flesch-Kincaid readability index) without accounting for student effort, and company literature refers to the use of pizza lunches, skating parties, and recognition buttons to reward students who make rapid progress, posing the threat of extinction once these rewards are withdrawn and the possibility that the internal motivation to read is reduced (Biggers, 2001). Teacher praise and constructive feedback may be more effective (Biggers, 2001). In addition, specialists express concern that Reading Assessment is not an instructional program, it does not mention the teacher’s role in providing direct instruction in reading strategies, it is not a balanced reading program, some schools rely too heavily on it, and initial costs are substantial (Biggers, 2001). Specialists cite research suggesting that students achieve at higher rates when free reading time is combined with direct instruction in reading activities and reading-extension activities (Biggers, 2001). Specialists as well as researchers question the early research studies evaluating the effectiveness of Reading Assessment (Biggers, 2001; Krashen, 2003, 2005). Similar concerns have been expressed by math teachers about the Math Assessment program: It does not replace good instruction, can be misused, and does not foster critical thinking (WikEd, 2006), although research suggests that the program can be effective with gifted students, with an effect size of 0.45 SD compared with gifted students who do not receive Math Assessment (Ysseldyke, Tardrew, Betts, Thill, & Hannigan, 2004). Perhaps most importantly, given that neither program is a complete curriculum, they are generally not approved for purchase with state funds (see, e.g., California State Board of Education, 1999).

1. As noted by a revision of the text.
2. The author has written student achievement. In revised the text.
3. Reading Assessment: appearance that the author has written student achievement. In the text.
4. The author has written student achievement. In the text.
5. Average baseline is the Iowa Tests of Basic S NCE points; thus, 0.6 IOWA_FAQ.pdf. Through the text.
6. As noted by a revised in the text.

Amrein, A. L., & Berlin
Analysis Archives, It
Amrein, A. L., & Berlin
An analysis of NAEF school graduation ef
0211-126-EPRU.pdf
Resolving these concerns probably depends on a better understanding of the strengths and limitations of rapid assessment. Rapid assessment is neither designed to supplant regular instruction nor to teach critical thinking skills. Instead, it is better understood as a tool designed to support teachers, providing useful information about student progress so that teachers may be more effective teachers of basic skills. However, to the extent that this reduces the amount of time required to teach those skills and improves student preparation for learning critical thinking skills, rapid assessment may also help teachers who aim to spend more time on critical thinking activities.

Finally, it is likely that implementation matters. If teachers are inflexible about allowing students to read books outside their established reading levels, students may become frustrated. If teachers do not monitor students, it is possible that some of them will share their answers to the reading and math quizzes with other students. Students may choose to read books full of illustrations to rack up points. However, if rapid assessment is understood as a tool with specific strengths and limitations, and if it is implemented with care by teachers and principals, the available evidence suggests that it can be more effective in raising student achievement in math and reading than a 10% increase in existing patterns of educational spending, voucher programs, charter schools, or increased educational accountability.

Notes

1. As noted by a reviewer, comparison of effect sizes across studies assumes consistency in underlying population variances. However, it seems unlikely that the magnitude of the differences in cost-effectiveness that are reported in this study are entirely attributable to underlying differences in population variances.

2. The author has written multiple manuscripts comparing rapid assessment with various alternatives for raising student achievement. In an effort to maintain consistency, the descriptions, reviews of literature, and cost estimates regarding rapid assessment are essentially identical in each manuscript.

3. Reading Assessment, Math Assessment, and the Rapid Assessment Corporation are pseudonyms, to avoid the appearance that the author endorses the assessment software. The author is neither affiliated with nor has received any funding from the vendor.

4. R. Bangert-Drowns (personal communication, June 7, 2006) estimated that the average duration of the 40 studies in his meta-analysis (Bangert-Drowns et al., 1991) was 1.5 to 2 weeks. D. Fuchs (personal communication, June 8, 2006) estimated that the average duration of the 21 studies in his meta-analysis (Fuchs & Fuchs, 1986) was 10 to 14 weeks. A. Kluger (personal communication, June 13, 2006) calculated that the average duration of the 131 studies in his meta-analysis (Kluger & DeNinì, 1996) was 17.8 days. Using the midpoints of each range, the weighted average duration of the feedback interventions in the three meta-analyses was 23.9 days (3.4 weeks).

5. Average baseline scores of all students in the three studies ranged from the 19th to the 30th national percentile on the Iowa Tests of Basic Skills (ITBS; Howell et al., 2002). In this range on the ITBS, each NPR point is equal to 0.636 NCE points; thus, 0.9 NPR points equal 0.572 NCE points (see http://www.usoe.k12.ut.us/eval/DOCUMENTS/IOWA_FREE.pdf). Throughout this review, when gains were reported using the NCE metric, effect sizes were calculated by dividing the NCE gain score by the NCE standard deviation (21.6).

6. As noted by a reviewer, these studies typically examine a small fraction of charter school students—those who enter in fifth grade and fill a small number of slots created by exiting students. This may ignore heterogeneity by grade level and by years spent in the charter school. Second, and more importantly, students who switch into charter schools at late ages may be different from students who do not switch in ways that cannot be controlled using student-fixed effects.

References


Because of its researchers, I evaluate hypotheses explanatory power (Coryn, 2007a). The enterprise since b