2010

The Cost Effectiveness of 22 Approaches for Raising Student Achievement

Stuart S Yeh, *University of Minnesota-Twin Cities*

The Cost Effectiveness of 22 Approaches for Raising Student Achievement

*Stuart S. Yeh**

ABSTRACT

Review of cost-effectiveness studies suggests that rapid assessment is more cost effective with regard to student achievement than comprehensive school reform (CSR), cross-age tutoring, computer-assisted instruction, a longer school day, increases in teacher education, teacher experience or teacher salaries, summer school, more rigorous math classes, value-added teacher assessment, class size reduction, a 10% increase in per pupil expenditure, full-day kindergarten, Head Start (preschool), high-standards exit exams, National Board for Professional Teaching Standards (NBPTS) certification, higher teacher licensure test scores, high-quality preschool, an additional school year, voucher programs, or charter schools. Limitations of the study are discussed and the findings are interpreted with regard to studies of the effect of performance feedback on student motivation and perceptions of control over academic outcomes.

INTRODUCTION

Under the federal No Child Left Behind Act (NCLB), schools that are not making adequate yearly progress in raising student achievement in math and reading are subject to a progressive series of sanctions. As a consequence, policymakers have a pressing need for information about the most cost-effective approaches for raising math and reading achievement. The alternatives being considered are diverse, ranging from narrow achievement-focused interventions to complex reforms that aim to achieve multiple outcomes, in addition to increases in math and reading achievement.

Cost-effectiveness analysis provides standard techniques and a conceptual framework for comparing dissimilar interventions to inform policy decisions

Stuart S. Yeh is an Associate Professor at the University of Minnesota, Minneapolis, MN.
*The author is grateful to Arthur Reynolds for comments on an earlier draft of the manuscript.

(*see* Levin and McEwan, 2001). Typically, cost information is integrated with information from existing impact analyses to derive cost-effectiveness ratios. For example, Levin, Glass, and Meister (1987) evaluated the cost effectiveness of four dissimilar interventions—computer-assisted instruction, cross-age tutoring, class size reduction, and increases in instructional time—to assess the relative impact of each intervention per dollar.

Cost-effectiveness analysis is appropriate when policymakers are primarily concerned with the goal of reducing the number of students who perform poorly on measures of math and reading achievement. However, a concern that arises is that the outcomes that are not measured are not valued. There are two main strategies for addressing this issue. The first strategy is to substitute an approach such as cost-benefit analysis that is suited to handle comparisons involving multiple outcomes. However, comparisons across each of the major approaches for raising student achievement would only be feasible if cost-benefit analyses had been conducted for each approach. This information is not currently available. Therefore, an alternative strategy was pursued. This strategy aims to identify the most cost-effective approach for raising math and reading achievement, with the expectation that it will provide the information necessary to conserve and maximize social resources—including instructional time—that would become available for the pursuit of important goals (such as proficiency in art, music, and science) that are not measured by standardized tests of math and reading. In contrast, the adoption of an approach that is inefficient would reduce the resources available to address these other important objectives.

While previous reviews found a very small corpus of cost-effectiveness studies in education (Hummel-Rossi and Ashdown 2002; Levin 1991; Monk and King 1993), cost-effectiveness techniques have now been applied to evaluate the relative cost-effectiveness of 11 promising approaches for raising student achievement:

1. Voucher programs
2. Charter schools
3. 10% increase in per pupil expenditure
4. Increased educational accountability
5. Comprehensive school reform (CSR)
6. Class size reduction
7. High-quality preschool
8. NBPTS teacher certification
9. A proposal to use value-added teacher assessment to identify and replace the bottom quartile of novice teachers with new teachers
10. Raising teacher quality by requiring new candidates to meet a minimum SAT score of 1000
11. Rapid formative assessment, involving systems where student performance in math and reading is assessed 2–5 times per week

This review compares the cost-effectiveness of these approaches for raising student achievement, along with brief comparisons of 11 additional approaches for which cost-effectiveness information is available.

While the corpus of cost-effectiveness studies has expanded in recent years, it remains limited, and this review reflects those limitations. First, the dearth of studies prior to the author's recent publication of multiple cost-effectiveness studies means that the available corpus of studies is weighted toward comparisons conducted by the author.

A second limitation is more fundamental. The comparison of 22 approaches for raising student achievement is necessarily limited to standardized measures that are available in published reports. Almost all studies focus on math and reading achievement. While approaches such as class size reduction and CSR aim to improve achievement in subjects beyond math and reading, it is not feasible to compare the 22 approaches with regard to other subject areas because standardized measures of performance in those areas were not administered across the 22 approaches.

It is not accidental that most studies focus on the results of standardized tests in math and reading. It is well established that basic skills in math and reading predict educational attainment and earnings (Currie and Thomas 2001; Murnane, Willet, and Levy 1995; Neal and Johnson 1996; O'Neill 1990; Winship and Korenman 1999). Perhaps the most sophisticated study, by Winship and Korenman (1999), used a nationally representative longitudinal sample, accounted for the reciprocal effects of cognitive skills and schooling, and found that a one standard deviation difference in basic math and reading skills is ultimately associated with a 35.8% difference in annual earnings, controlling for a host of covariates including ability. Given the importance of math and reading achievement, it is useful to conduct a cost-effectiveness analysis that focuses on these outcomes so that resources devoted to math and reading achievement may be allocated efficiently, maximizing the availability of resources to address subjects beyond math and reading.

A third limitation of the present review is that published studies may differ with regard to the age, ethnicity, or level of poverty of the populations that were studied, or other factors influencing achievement. These differences make comparisons problematic. The results of the present review are necessarily limited to statements regarding the particular variations of the interventions that were studied, with the specific samples and under the specific conditions that were studied. Thus, the interpretation of the results of the present review should be tempered by an understanding of the inherent limitations and complexity of cost-effectiveness analysis. Ultimately, further research is needed to explore whether the conclusions presented in this review are stable across age, ethnicity,

level of poverty, and variations in the implementation of each intervention.

The purpose of the present review is to examine the existing corpus of educational cost-effectiveness studies, adjust all costs to the same base period (August 2006), critically analyze deficiencies in existing studies, synthesize what is currently known about the cost effectiveness of alternative approaches for raising student achievement, and permit readers to compare the cost effectiveness of every approach for which information is available, within the limits specified in the review. The Discussion section reviews important limitations of this study, then proposes an explanation of the results in terms of existing literature regarding the psychology of learning and achievement.

SEARCH PROTOCOL

Multiple searches were conducted using the Education Resources Information Clearinghouse (ERIC), Web of Science, Academic Search Premier, Education Fulltext, PolicyFile, and Econlit databases, as well as GoogleScholar, limited to the years 1999 through 2008. First, searches were conducted using the term *cost-effectiveness* in combination with *student achievement*. Second, successive searches were conducted using the term *cost-effectiveness* in combination with each individual term from the following list: *voucher program, charter school, economics of education, educational production, educational expenditure, educational accountability, comprehensive school reform, class size reduction, preschool, Perry, Abecedarian, formative assessment, progress monitoring, curriculum-based assessment, curriculum-based measurement, National Board, teacher evaluation,* and *teacher assessment*. Third, these searches were repeated using the term *cost per student* in lieu of *cost-effectiveness*.

A small number of older reports were located through citations in the set of more recent studies. A few benefit-cost studies provided the information necessary to calculate effectiveness-cost ratios and were included. Reports were excluded if:

a.) they omitted the information necessary to calculate effectiveness-cost ratios (annualized student achievement effect size, divided by annual cost per student),
b.) effect-size information was calculated from uncontrolled or weakly controlled research designs,
c.) cost information lacked detail or was collected for a single school,
d.) they did not involve a measure related to math or reading achievement with regard to American students in grades PK–12,
e.) they were limited to disabled or special education populations, or
f.) they merely summarized results from a previously published cost-effectiveness study.

Reports were limited to studies involving American populations because comparisons of effectiveness-cost ratios assume comparability of underlying population variances as well as comparability of cost estimates. The characteristics of American students and the American educational system are sufficiently different from their European and Asian counterparts to suggest that the population variances underlying the effect-size estimates may not be comparable across cultures. In addition, differences in purchasing power create difficulties in comparing costs internationally.

Only a limited number of reports included original cost-effectiveness analyses meeting the search criteria, reflecting the general lack of attention to cost-effectiveness methods in education and the need for additional studies. The current review focuses on the most recent cost-effectiveness studies, with particular attention to key methodological and analytical issues that have been overlooked by previous researchers. Some interventions and the cost to implement them can be described in simple terms without confusion, while other interventions raise thorny issues that require explication. Thus, less space is devoted to older studies and studies that require less explanation.

METHOD

All cost-effectiveness studies meeting the search protocol were included in the analysis. Each study was reviewed, with special attention to methodological and analytical weaknesses and errors in accounting for the full costs of each intervention, including social opportunity costs. All effect sizes were annualized. All costs were annualized and adjusted to the same base period (August, 2006) using the consumer price index (CPI). Effectiveness-cost ratios were calculated (annualized effect size divided by annual cost per student). Interventions were ranked according to effectiveness-cost ratios, using upper-bound estimates for all interventions other than rapid assessment (to provide conservative estimates of the difference in efficiency between rapid assessment and the alternatives).

RESULTS

The results for each intervention are presented in the order of cost effectiveness. Table 1 lists the annualized effect size, annual cost per student, and effectiveness-cost ratio (effect size divided by cost per student) for each intervention. All cost figures have been adjusted for inflation to the same base period (August, 2006), using the CPI. This adjustment is necessary to minimize the effect of inflation on the effectiveness-cost ratios in Table 1.

Table 1. Comparison of effect sizes, costs, and effectiveness-cost ratios for various interventions to raise student achievement

	Effect Size (SD)[a]		Cost[b]	Effectiveness-Cost Ratio[c]	
	Reading	*Math*		*Reading*	*Math*
Rapid Assessment (high estimates)	0.270[d]	0.392[e]	$9.45[f] $18.89[g]	0.028571	0.020752
Rapid Assessment (low estimates)	0.175[h]	0.324[i]	$9.45[f] $18.89[g]	0.018519	0.017152
Comprehensive school reform	0.510[j]	0.510[j]	$217.83	0.002341[k]	0.002341[k]
Cross-age tutoring	0.480	0.970	$555.61	0.000864	0.001746
Computer-assisted instruction	0.230	0.120	$238.12	0.000966	0.000504
Longer school day	0.070	0.030	$159.87	0.000438	0.000188
Teacher education	0.22	0.22	$702.62	0.000313	0.000313
Teacher experience	0.18	0.18	$702.62	0.000256	0.000256
Teacher salary	0.16	0.16	$702.62	0.000228	0.000228
Summer school	0.190	0.190	$1515.00	0.000125	0.000125
Rigorous math classes	--	0.200	$1,911.17	--	0.000105
Value-added teacher assessment	0.057	0.057	$624.72	0.000091	0.000091
Class Size Reduction	0.104[l]	0.090[l]	$1,379.28[m]	0.000075	0.000065
Nye et al. (2001)	0.120[n]	0.129[n]	$1,379.28[m]	0.000087	0.000094
Finn et al. (2001)					
10 percent increase in spending	0.083[o]	0.083[o]	$1118.83[p]	0.000075	0.000075
Full-day kindergarten	0.181	0.181	$2611.00	0.000069	0.000069
Head Start	0.324	0.165	$9,000.00	0.000036	0.000018
High standards exit exam	0.051[q]	-0.062[q]	$2025.97[r]	0.000025	-0.000031
NBPTS teacher certification	0.002	0.004	$326.53	0.000006	0.000012
Higher licensure test scores	0.004	0.015	$894.10	0.000004	0.000017
Perry Preschool	0.150[s]	0.155[s]	$12,147.03[t]	0.000012	0.000013
Abecedarian Preschool	0.150[u]	0.054[u]	$10,188.09[v]	0.000015	0.000005
Additional school year	0.15	0.15	$14,271.76	0.000011	0.000011
Voucher programs	0.032[w,x]	0.081[w,x]	$9,646.01[w,y]	0.000003	0.000008
Charter schools	0.009[w,z]	0.001[w,z]	$8,086.30[w,*]	0.000001	0.000000

Notes: See Appendix A

Rapid Assessment

The most cost-effective approach was rapid assessment, with four effectiveness-cost ratios ranging from 0.017152 to 0.028571 based on two randomized evaluations of the impact of the *Reading Assessment* program on reading achievement (Nunnery, Ross, and McDonald 2006; Ross, Nunnery, and Goldfeder 2004), one randomized and one national quasi-experimental evaluation of the impact of the *Math Assessment* program on math achievement (Ysseldyke and Bolt 2007; Ysseldyke and Tardrew 2007), and a cost-effectiveness analysis that

integrated the impact results with cost data (Yeh 2008).[1] It is useful to review the salient characteristics of this approach:

> *Reading Assessment* [is] 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, no date). 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).[2] 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 where students and teachers receive rapid, frequent feedback on student performance upon completion of every set of problems. (Yeh 2007, p. 417)

In reading, multiple choice quiz items test student knowledge and comprehension of each book's characters and plot (for fiction books) or declarative knowledge (for nonfiction books). Math problems range from simple arithmetic items suitable for first-grade students to calculus items for more advanced students. Students use a classroom device to scan bubble answer sheets, which are scored by the classroom computer—explaining how the process works while minimizing the burden on teachers. Rapid assessment is not intended to minimize the role of the teacher in providing comments about strategies for improvement. Instead, rapid assessment utilizes the advantages of technology to relieve teachers of the burden of providing low-level corrective feedback, facilitating the ability of teachers to focus on individual tutoring and to provide what Hattie and Timperly (2007) refer to as Level 2 and Level 3 feedback—comments by teachers regarding strategies for improvement, both specific process-oriented strategies and general self-regulatory strategies, and assessment of higher-order thinking skills. Since research literature indicates that tutoring is more effective than recitation, with an average effect size of 0.40 standard deviation (SD), placing teachers in the role of tutors may make them more effective teachers (Cohen, Kulik, and Kulik 1982).

Program activities do not displace standard reading and math activities. During designated periods of the day devoted to reading and math, 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. No additional time is allocated to reading or math instruction beyond standard 60 minute daily periods of reading and math instruction, nor is that time used in a way that is much different than standard reading and math learning activities. The primary difference is that books are selected according to each student's reading level, math problems are assigned according to each student's math level, and students and teachers are able to quickly diagnose areas where students are having difficulty (Yeh 2008).

A detailed analysis of the operation of the program, based on interviews with teachers, administrators, and students suggests that individualization of book selections and math assignments for each student's achievement level creates a task structure where task difficulty is individualized (Yeh 2006). In combination with individualization of performance expectations and rapid, test-based performance feedback, it appears that these characteristics promote a steady dose of positive, objective feedback to students, reinforcing student beliefs that they are competent learners who are able to master new skills and knowledge every day, and restructuring the experience of schooling to meet students' psychological need to feel competent, promoting engagement and commitment to work effort (Yeh 2006).

The cost analysis assumed the purchase of either *Reading* or *Math Assessment* software and a diagnostic assessment for each student receiving the rapid assessment intervention, plus mark scan devices for each math classroom. Initial costs were averaged over an enrollment of 500 students in 25 classrooms per building. Initial fees, teacher and administrator training, and the cost of the scanners were amortized over seven years, which was (arbitrarily) assumed to be the life of the program. For reading, the costs included access to 100,000 book quizzes for every student. For math, the costs included access to *Math Assessment* grade level libraries tagged to state standards for grades one through seven, as well as multiple subject area libraries for the secondary grades (pre-algebra, algebra 1, algebra 2, geometry, probability and statistics, pre-calculus, calculus, basic math, chemistry, physics). The cost analysis assumed that every classroom teacher and one administrator (for every 500 students) underwent a full-day training session for *Math Assessment* and a full-day training session for *Reading Assessment.* In addition, the cost analysis assumed a 50% teacher turnover rate during the seven-year implementation period, and assumed that

each new teacher received a full-day training session for *Math Assessment* and a full-day training session for *Reading Assessment*.

Sensitivity analyses suggested that purchasing a computer and printer for every classroom would not change the relative cost-effectiveness ranking of rapid assessment with regard to voucher programs, charter schools, a 10% increase in per pupil expenditure, increased educational accountability, CSR, class-size reduction, or high-quality preschool (Yeh 2007, 2008).

Teacher interviews and classroom observations suggested that the use of rapid assessment technology does not require skilled, highly trained teachers, nor does it supplant normal reading and math activities (Yeh 2006). The burden on teachers is minimal because assessments are self administered by students and scoring and reporting is handled by computer software. The burden is offset by the time saved due to the program's scoring and student progress monitoring features, which replace the time-consuming conventional tasks of grading math homework and assessing reading comprehension (Yeh 2006). The annual cost per student in 2006 dollars is $9.45 in reading and $18.89 in math, adjusted for the opportunity costs of teacher training time ($3.02 per student) and adjusted for the opportunity costs created by large upfront fixed costs (Yeh 2003).

Comprehensive School Reform (CSR)

The second most cost-effective approach was comprehensive school reform (CSR), based on a meta-analysis of effect sizes for 29 school reform models by Borman, Hewes, Overman, and Brown (2003), cost data from a range of sources (Herman 1999; King 1994; Odden 2000), and a cost-effectiveness analysis that integrated the impact and cost data (Yeh 2008). Of the group of school-reform models categorized by Borman and others (2003) as having the *Strongest* or *Highly Promising Evidence of Effectiveness*, the model with the highest effect size, and highest effectiveness-cost ratio, was *Expeditionary Learning Outward Bound* (effect size = 0.51 SD, effectiveness-cost ratio = 0.002341, *see* Yeh 2008). None of the other 28 reform models had an effectiveness-cost ratio larger than the lowest ratio for rapid assessment (0.017152, *see* Yeh 2008). The results suggest that rapid assessment is roughly an order of magnitude more effective per dollar compared to the most cost-effective CSR model. However, the true advantage for rapid assessment is likely to be substantially larger than indicated by the ratios in Table 1. Given Borman and other's (2003) conclusion that effect-size estimates for reform models other than *Success for All,* the *School Development Model,* and *Direct Instruction* are unreliable, the true upper bound for the school reform effectiveness-cost ratios is likely to be closer to the maximum ratio calculated for the three models for which there is reliable evidence (Yeh 2008, p. 21):

> Rapid assessment is a minimum of 59 times as cost effective as *Direct Instruction*, the most cost-effective of the three models, suggesting that the true advantage of rapid assessment may be 59 times as large as the cost-effectiveness of CSR. Regardless, the cost-effectiveness analysis suggests that CSR is not an efficient approach for improving student achievement. This lack of efficiency may explain previous research findings suggesting that enthusiasm for CSR wanes over time and leads to teacher burnout, conflict, disengagement, and exhaustion (Little and Bartlett 2002).

The advantage of rapid assessment is even greater if the cost effectiveness of CSR is based on Borman and Hewes's (2002) alternative estimate that *Success for All* produces "sustained" effect sizes of 0.29 SD in reading and 0.11 in math (over a 3.84 year period of intervention) at a total cost per student of $3,689.04—suggesting small effectiveness-cost ratios of 0.000079 in reading and 0.000030 in math after adjusting for inflation.

Cross-Age Tutoring

Levin, Glass, and Meister (1987) estimated that cross-age tutoring (tutoring of younger students by older peers) would improve student achievement by 0.97 SD in math and 0.48 SD in reading, at an annual cost of $555.61 per student (in 2006 dollars), suggesting effectiveness-cost ratios of 0.001746 in math and 0.000864 in reading.

Computer-Assisted Instruction

Computer-assisted instruction is estimated to improve student achievement by 0.12 SD in math and 0.23 SD in reading, at an annual cost of $238.12 per student (in 2006 dollars), suggesting an effectiveness-cost ratio of 0.000504 in math and 0.000966 in reading (Levin, Glass, and Meister 1987).

Longer School Day

Lengthening the school day by 60 minutes is estimated to improve student achievement by 0.03 SD in math and 0.07 SD in reading, at an annual cost of $159.87 per student (in 2006 dollars), suggesting an effectiveness-cost ratio of 0.000188 in math and 0.000438 in reading (Levin, Glass, and Meister 1987).

Teacher Education

Greenwald, Hedges, and Laine (1996), based on meta-analytic results, estimated that substitution of teachers with a masters degree for teachers with a bachelors degree is associated with an increase in student achievement of 0.22 SD, at a cost per pupil of $702.62 (in 2006 dollars), resulting in an effectiveness-cost ratio of 0.000313.

Teacher Experience

Greenwald, Hedges, and Laine (1996), based on meta-analytic results, estimated that teachers with more experience are associated with an increase in student achievement of 0.18 SD, at a cost per pupil of $702.62 (in 2006 dollars), resulting in an effectiveness-cost ratio of 0.000256.

Teacher Salary

Greenwald, Hedges, and Laine (1996), based on meta-analytic results, estimated that increases in teacher salaries are associated with an increase in student achievement of 0.16 SD, at a cost per pupil of $702.62 (in 2006 dollars), resulting in an effectiveness-cost ratio of 0.000228.

Summer School

Belfield (2006a) drew upon meta-analytic results of summer school programs estimating a median effect size of 0.19 SD (Cooper et al. 2000), and per pupil costs estimated at a total of $1,515 (Borman and Dowling 2006), suggesting an effectiveness-cost ratio of 0.000125.

More Rigorous Math Courses

Students who take a couple of advanced math courses score approximately 0.20 SD higher on math tests (Mayer and Peterson 1999). Assuming that student achievement can be boosted at no cost, allowing more students to enter and complete advanced math courses, and assuming that it is possible to elicit the required increase in the supply of advanced math teachers through a targeted increase of 33% in the salaries of these teachers, the annual cost would be $1,911.17 per student, in 2006 dollars (Mayer and Peterson 1999). Both of these assumptions are probably unrealistic, suggesting that the effectiveness-cost ratio of 0.000105 is an upper-bound estimate.

Value-Added Teacher Assessment

Yeh and Ritter (2009) examined the effects of Gordon, Kane, and Staiger's (2006) proposal to improve the quality of the teaching force by identifying and replacing the bottom quartile of novice teachers, equal to 68,299 of the estimated 3,161,000 elementary and secondary public school teachers in 2006 (*see* National Center for Education Statistics [NCES] 2005c), using value-added assessments of teacher performance. Their proposal is one of the featured proposals by the Hamilton Project at the Brookings Institution and, thus, has attracted attention as perhaps the most promising strategy for implementing a value-added teacher

assessment system. The significance of Gordon, Kane, and Staiger's (2006) proposal is that it converts the general strategy of value-added teacher assessment into a concrete proposal that can be evaluated. It is currently the only proposal of this type that adequately addresses the fact that most teachers are tenured and, therefore, cannot be replaced without adequate cause. However, Gordon, Kane, and Staiger's (2006) data imply a small effect size of only 0.057 SD.

Gordon, Kane, and Staiger (2006) propose to replace bottom quartile rejected teachers with individuals who would be hired through alternative teacher certification programs. These programs primarily serve individuals who have accumulated substantial work experience in occupations other than teaching but wish to change occupations and enter teaching. Alternative certification typically allows these individuals to enter teaching relatively quickly, under the supervision of mentors and while electing courses in curriculum and instructional methods.

Expansion of alternative certification programs to accommodate the additional demand for new teachers created by Gordon, Kane, and Staiger's (2006) proposal would generate substantial costs. The total annual cost of implementing the proposal includes:

- costs to the chain of sending employers and replacement employees triggered by the proposal;
- costs to the hiring school district, the new teacher, the terminated teacher, and other costs to society;
- adjustments for the gain in output (a benefit) of the terminated teacher once he or she has been retrained and has started a new job;
- psychic gains to the new teacher and psychic gains to the individual who voluntarily replaces the replacement teacher;
- the cost of alternative certification;
- the cost to raise salaries for all teachers that is necessary to recruit replacements for the bottom quartile of novice teachers; and
- the cost of the assessments (Yeh and Ritter 2009).

The total annual cost per student is $624.72 (Yeh and Ritter 2009). If the effect size of Gordon, Kane, and Staiger's (2006) proposal is 0.057 SD per year, the effectiveness-cost ratio is 0.000091.

Class Size Reduction

Yeh (2009a) integrated Reichardt's (2000) cost analysis with estimates of student achievement effect sizes from the Tennessee STAR experiment by Finn and others (2001) and Nye, Hedges, and Konstantopoulos (2001) to estimate the cost effectiveness of class size reduction. Perhaps the most optimistic estimate, by Finn and others (2001), employed hierarchical linear modeling, controlled for an array of covariates, and isolated the impact of class size reduction according

to duration of exposure. By the end of second grade, the achievement of students randomly assigned to classrooms of 13–17 students increased relative to students assigned to classrooms with 22–26 students, with annualized effect sizes equal to 0.120 SD in reading and 0.129 SD in math. The annual inflation-adjusted cost per student to reduce class size from 24 to a ceiling of 17 students per class is $1,379.28 (Reichardt 2000), resulting in effectiveness-cost ratios ranging from 0.000065 to 0.000094 (Yeh 2009a). Other estimates of the cost-effectiveness of class size reduction are limited by relatively unsophisticated cost calculations.

Most alternative estimates are smaller than Yeh's (2009a) estimates. Aos, Miller, and Mayfield (2007) analyzed 38 recent, methodologically rigorous evaluations of class size reduction and estimated that each one-student reduction in class size increases student achievement by 0.019 SD in grades K–2, and 0.007 SD in grades 3–6, or an average of 0.013 SD. This study found that each one-student reduction in class size would cost $217 per student per year, suggesting an effectiveness-cost ratio of 0.000060. Krueger (2003) drew upon existing estimates that class size reduction improved student achievement by 0.20 SD in Tennessee, at a cost of $4,428.37 in 2006 dollars, suggesting an effectiveness-cost ratio of 0.000045. Borman and Hewes (2002) estimated effectiveness-cost ratios equivalent to 0.000038 in reading and 0.000052 in math after adjusting for inflation, based on effect-size estimates from Krueger and Whitmore's (2001) analysis of the Tennessee STAR data, and "total" cost information from Levin, Glass, and Meister (1987). In addition, Borman and Hewes (2002) reported a second set of effectiveness-cost ratios, accepting Levin, Glass, and Meister's (1987) assumption that class size reduction improves student achievement in subjects other than math and reading, as well as the assumption that the cost of class size reduction is divisible across those subjects. Mayer and Peterson (1999) estimated that class size reduction of one-third (from 24 to 16 students) would improve student achievement by 0.20 SD over two years, or 0.10 SD per year, at an annual cost of $1,551.28 per student, in 2006 dollars, suggesting an effectiveness-cost ratio of 0.000064. Greenwald, Hedges, and Laine (1996), based on meta-analytic results, estimated that decreases in pupil-teacher ratios are associated with an increase in student achievement of 0.04 SD at a cost per pupil of $702.62, in 2006 dollars, resulting in an effectiveness-cost ratio of 0.000057.

Three estimates of the cost effectiveness of class size reduction are somewhat larger than Yeh's (2009a) estimates, primarily due to cost estimates that are far below the estimates by Aos, Miller, and Mayfield (2007), Krueger (2003), Reichardt (2000), and Mayer and Peterson (1999). Grissmer (2002), based on the Tennessee STAR data, estimated that reducing class size from 24 to 16 students would improve student achievement for each cohort of students by 0.30 SD over four years, or 0.075 SD per year, at an annual cost of $531.49 per student, in

2006 dollars, suggesting an effectiveness-cost ratio of 0.000141. Grissmer (2002) elected to shrink the $531.49 per-pupil cost estimate by a factor of 4/13, arguing that the intervention is implemented in grades K–3 and, thus, the annual per-pupil cost should be deflated to reflect implementation in four of 13 grade levels. However, the relevant metric is the annual cost per pupil in the grades where the intervention is delivered, not the cost spread over every pupil from K–12. Thus, Grissmer's (2002) shrunken-cost estimate is analytically inconsistent with his impact estimate and is underestimated by a factor of 4/13. Grissmer (2002) also estimated the cost effectiveness of adding teacher aides to regular-sized classrooms, which was nearly three times lower than his overestimated figure for the cost effectiveness of class size reduction.

Grissmer, Flanagan, Kawata, and Williamson (2000), based on the Tennessee STAR data, estimated that decreases in pupil-teacher ratios are associated with an increase in student achievement of 0.10 SD at a cost per pupil of $281.05, in 2006 dollars, resulting in an effectiveness-cost ratio of 0.000356.

Levin, Glass, and Meister (1987) estimated that class size reduction from 25 to 20 students would improve student achievement by 0.09 SD in math and 0.05 SD in reading, at an annual cost of $736.45 per student, in 2006 dollars, suggesting effectiveness-cost ratios of 0.000122 in math and 0.000068 in reading. Levin, Glass, and Meister (1987) elected to reduce their cost estimate by an additional two-thirds under the assumption that class size reduction improves student achievement in subjects other than math and reading.

Ten Percent Increase in Spending

Yeh (2007) estimated a maximum effectiveness-cost ratio of 0.000075 for a 10% increase in existing patterns of expenditure per pupil (equal to $1,118.83), based on Greenwald, Hedges, and Laine's (1996) relatively optimistic meta-analytic estimate that this would increase student achievement in math and reading by 0.083 SD per year. In contrast, Hanushek (1986, 1989, and 1997) estimated that the relationship between educational spending and student achievement is much weaker. Yeh's (2007) results suggest that the achievement gains per dollar from rapid assessment are more than two orders of magnitude larger than the gains that accrue through a 10% increase in existing patterns of educational expenditures.

The results are consistent with alternative estimates. Grissmer (2002) estimated that a comparable increase in per pupil expenditure of $1,207.94 (in 2006 dollars) would increase student achievement by 0.042 to 0.098 SD, or an average of 0.07 SD, suggesting an effectiveness-cost ratio of 0.000058. Wenglinsky (1997) argued that the results of previous meta-analyses (Greenwald, Hedges, and Laine 1996;

Hanushek 1989; Hedges, Laine, and Greenwald 1994) were inconclusive because they relied on studies that did not involve nationally representative populations, and failed to properly model the paths by which particular types of expenditure may influence student achievement. Using path modeling with National Assessment of Educational Progess (NAEP) data, Wenglinsky (1997) estimated that a $1,476.47 increase in instructional expenditures (in 2006 dollars) would reduce pupil-teacher ratios, increasing math achievement by 0.04 SD, suggesting an effectiveness-cost ratio of 0.000027.

Full-Day Kindergarten

Aos, Miller, and Mayfield (2007) analyzed 23 methodologically rigorous evaluations of full-day kindergarten compared to half-day kindergarten, and estimated that student achievement would increase by an average of 0.181 SD. However, the gains disappeared by the end of first grade. Aos, Miller, and Mayfield (2007) calculated that the incremental cost of full-day kindergarten compared to half-day kindergarten is $2,611 per student per year. If achievement increases by 0.181 SD, the effectiveness-cost ratio is 0.000069.

Head Start

Ludwig and Phillips (2008) drew upon results from the first randomized study of the Head Start preschool program, finding an average effect size of 0.324 SD regarding letter-naming and 0.165 SD regarding math. Ludwig and Phillips (2008) estimated a cost of $9,000 per child, suggesting corresponding effectiveness-cost ratios equal to 0.000036 and 0.000018.

High-Standards Exit Exam

Increased educational accountability may be defined as the implementation of high standards exit exams. A majority of states have now implemented high school exit exams that students must pass to graduate from high school, affecting 70% of students nationwide (Gayler et al. 2004) and, with few exceptions, requiring mastery of material at the high school level (Center on Education Policy 2005).

To calculate the cost effectiveness of increased educational accountability, Yeh (2007) drew upon Hanushek and Raymond's (2005) longitudinal panel study with controls for fixed-state effects as well as measures of parental education, school spending, and race. This study found that increased accountability is associated with an impact of 0.2 SD over four years, or 0.05 SD per year. Based on a study by the General Accounting Office (2003), the annual cost of administering accountability-related assessments was conservatively estimated to be $12.73 per student. However, unlike previous researchers (*see* Hoxby,

2002) who focused on the nominal cost of the assessments, Yeh (2007) included the cost to society of the incremental increase in the number of students who are denied diplomas solely because exit exams raise the graduation bar. To calculate these costs, Yeh (2007) drew upon Warren, Jenkins, and Kulick's (2006) estimate of the incremental increase in the fraction of students who were denied diplomas as a consequence of 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 and controlling for state and year fixed effects, high school completion rates were 2.1 percentage points lower in states with this type of exit exam, implying that an extra 89,908 students would be denied diplomas annually (0.021 X national public student ninth-grade enrollment of 4,281,345 in academic year 2004–2005 [U.S. Department of Education (USDE) 2005]).

To calculate the economic value of the lost diplomas, Yeh (2007) drew upon Flores-Lagunes and Light's (2004) estimates of the "sheepskin" (signaling) effect, as well as Census Bureau estimates of lifetime earnings by level of education (Day and Newburger 2002), to calculate 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. Yeh (2007) estimated that the cost per pupil served is $185.01.

This figure is underestimated. After adjustments to reflect Flores-Lagunes and Light's (2007) latest estimates, and basing the per pupil cost on national public student ninth-grade enrollment of 4,281,345 in academic year 2004–2005 (USDE 2005), the cost per pupil served is $2,013.24 to raise the graduation bar. This figure is considerably higher than Yeh's (2007) original estimate, based on total PK–12 enrollment of 48,359,697.

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 ($2,013.24) or $2,025.97. Using the adjusted 0.05 SD effect-size estimate derived from Hanushek and Raymond (2005), the effectiveness-cost ratio is 0.000025, implying that the achievement gains per dollar through rapid assessment are 686 times larger than the gains through accountability testing.

A more recent analysis suggests that this estimate of the cost-effectiveness of educational accountability is overstated. Grodsky, Warren, and Kalogrides (2008) analyzed longitudinal NAEP data with controls for state-fixed effects as well as

gender, race, parental education, grade level, a measure of "home resources," age, and difficulty of each state's exit exam. This study, which provides the best available evidence for the impact of accountability policies, found no significant effect of exit exams on student achievement. While the results are not statistically significant, they imply mixed effect sizes of 0.051 SD in reading and negative 0.062 SD in math, in comparison with Hanushek and Raymond's (2005) positive 0.05 SD effect size estimate. Using the 0.051 SD reading achievement effect-size estimate from Grodsky, Warren, and Kalogrides (2008), the effectiveness-cost ratio is 0.000025 (and since Grodsky, Warren, and Kalogride's [2008] impact estimate with regard to math is negative, the effectiveness-cost ratio is a negative 0.000031 with regard to math achievement).

Alternative estimates of the cost and cost effectiveness of increased educational accountability are flawed. Mayer and Peterson (1999) estimated that external examinations during the senior year of high school increase eighth-grade math scores by 0.2 SD, at an annual cost of $248.20 per student (in 2006 dollars), suggesting an effectiveness-cost ratio of 0.000806. However, the effect size estimate was based on studies that failed to control for fixed effects, while the cost estimate ignored the cost to society of the lost economic output due to exit examinations that deny diplomas to more students. Similarly, Hoxby (2002) ignored the costs of denying diplomas to students. She estimated that the annual cost per student of test-based accountability systems ranges from a low of $2.05 to a high of $38.90 across 25 states, in 2006 dollars, and concluded that test-based accountability is highly cost effective. However, testing without consequences has little effect on student achievement (Hanushek and Raymond 2005). Thus, the most defensible approach is to combine effect size estimates from studies of exit exams and other consequential accountability systems with cost estimates that include the full costs of consequential accountability, including the economic costs of raising the high school exit bar.

NBPTS Teacher Certification

Hopes that teacher certification would reliably distinguish effective and ineffective teachers have foundered as evidence has accumulated that certification status is a weak predictor of teacher performance (Hanushek et al. 2005; Hanushek, Kain, and Rivkin 1999). Perhaps the most sophisticated and rigorous certification system is the process developed by the National Board for Professional Teaching Standards (NBPTS) in collaboration with the Educational Testing Service (ETS). This system is widely considered to be the standard for teacher certification systems, involving an examination on six 30-minute exercises, specific to the candidate's chosen certificate area, at one of 300 NBPTS computer-based testing

centers across the U.S., and evaluation of portfolio evidence, including video recordings of teacher-student interactions. The NBPTS certification system offers perhaps the strongest test of the hypothesis that certification can distinguish between more and less effective teachers.

Seven large-scale studies have now been conducted of NBPTS certification (Cavalluzzo 2004; Clotfelter, Ladd, and Vigdor 2006, 2007; Goldhaber and Anthony 2007; Harris and Sass 2007; Ladd, Sass, and Harris 2007; Sanders, Ashton, and Wright 2005). These studies offer the necessary power to detect effects, if they exist, and controlled for student or school fixed effects or used hierarchical linear modeling (HLM). However, the average estimate of the signaling effect of NBPTS certification, in comparison with teachers who were never certified, was only 0.002 SD in reading and 0.004 SD in math (Yeh in press-a).

Rice and Hall (2008) estimated that the full societal cost of NBPTS certification, including opportunity costs, averaged $25,665.37 (adjusted to 2006 dollars) per participant, per year. However, while Rice and Hall (2008) assumed that the certification process averages one year, it actually takes two years (Goldhaber and Anthony 2005; Sanders, Ashton, and Wright 2005). After adjusting for this difference, the total cost per participant is $51,330.74. Amortized over an average teaching career of 7.86 years as an NBPTS-certified teacher (after a minimum three years before becoming eligible to apply for NBPTS certification and a certification process averaging two years [Goldhaber and Anthony 2005; Sanders, Ashton, and Wright 2005]), and averaged over 20 students per classroom, the annual cost of NBPTS certification per student is $326.53, resulting in effectiveness-cost ratios of 0.000006 in reading and 0.000012 in math.

Licensure Test Scores

The use of licensure test scores to distinguish between strong and weak teachers is a less expensive approach to raising teacher quality, compared to NBPTS certification. A few early studies (Ehrenberg and Brewer 1995; Ferguson 1991; Ferguson and Ladd 1996; Strauss and Sawyer 1986) suggested that teacher performance on general tests of math or reading might be useful indicators of teacher effectiveness. Until recently, however, no study of teacher performance on tests of math and reading controlled for nonrandom teacher sorting or student fixed effects. These controls are important because recent studies demonstrate that positive correlations between the strength of teacher qualifications and student achievement observed in cross-sectional data are driven largely by sorting of teachers and students across schools and, to a lesser extent, within schools (Clotfelter, Ladd, and Vigdor 2006, 2007; Goldhaber 2007). Once these controls

are implemented, the observed relationship between teacher performance on licensure tests and student achievement is greatly attenuated.

Clotfelter, Ladd, and Vigdor (2007) provide the best available econometric estimate of the effect of high-ability individuals on student achievement, where ability is measured through reading, writing, and math test scores. Clotfelter, Ladd, and Vigdor (2007) employed statewide longitudinal data covering all North Carolina students in grades 3, 4, and 5 in years 1995–2004, matched student test scores to specific teachers and, by controlling for student fixed effects, controlled for observable as well as unobservable characteristics of students that may influence achievement. However, the effect of teacher ability, measured by licensure exams, was very small. Teachers who scored 1 SD above the average on teacher licensure exams boosted math achievement by 0.015 SD and reading achievement by 0.004 SD, relative to the average teacher (Clotfelter, Ladd, and Vigdor 2007, Table 7).

The cost-effectiveness of this approach may be estimated by calculating the cost of setting a minimum SAT® score of 1,000, augmented with analyses to determine the sensitivity of the calculations to the range of plausible relationships between licensure and SAT® test scores (Yeh 2009b). Based on econometric analyses, Manski (1987) estimated that setting a minimum SAT® score of 1,000 would result in an average SAT® score of 1,130, which is one SD above the average score of 905 achieved by a nationally representative sample of teachers (Angrist and Guryan 2004; College Board 2006). Of the 111,700 members of the National Education Longitudinal Study of 1988 (NELS:88) cohort who were college educated and working in the year 2000, 0.8% (894 individuals) chose the teaching profession. Assuming that the wage elasticity of the supply of teachers is three—at the upper end of estimates by Manski (1987), and Ballou & Podgursky (1995)—and assuming a very high probability (.88) that a job applicant receives an offer, wages must increase by 44.65% in order to increase the annual supply of high ability teacher applicants by an amount sufficient to replace the 54.1% of each cohort of new teachers who may be expected to score below 1,000 on the SAT® (.541 X 894 = .88 X .459 X 44.65 X .03 X 894). This figure is underestimated to the extent that the true wage elasticity is lower than three, or the probability that a job applicant receives an offer is less than 88%, suggesting that a 44.65% increase in salary, equal to $17,882 per year, is a lower-bound estimate of the wage premium needed to attract sufficient teacher applicants scoring a minimum of 1,000 on the SAT.® The annual cost to society is equivalent to $894.10 per student, assuming 20 students per classroom, suggesting an effectiveness-cost ratio of 0.000017 in math and 0.000004 in reading.

High-Quality Preschool

Roughly similar in cost effectiveness was high-quality preschool, based on impact estimates from Schweinhart, Barnes, and Weikart (1993) and Ramey and others (2000), cost estimates from Barnett (1992) and Barnett and Masse (2007), and a cost-effectiveness analysis that integrated the impact and cost data, resulting in four effectiveness-cost ratios ranging from 0.000005 to 0.000015 (Yeh 2008). The reported effect sizes for children participating in Perry preschool were 0.150 SD in reading and 0.155 SD in math (annualized over a two-year implementation period) at the end of second grade (Schweinhart, Barnes, and Weikart 1993), but the annualized cost was $12,147.03 (Barnett 1992), resulting in small effectiveness-cost ratios of 0.000012 in reading and 0.000013 in math (Yeh 2008). Results for participants in Abecedarian preschool were roughly comparable to the results for participants in Perry preschool. Annualized effect sizes for third-grade children were 0.150 in reading and 0.054 in math over a five-year implementation period (Ramey et al. 2000), at an annual cost of $10,188.09 (Barnett and Masse 2007) in 2006 dollars, adjusted for the value of formal and informal daycare services provided to the control group, resulting in effectiveness-cost ratios of 0.000015 in reading and 0.000005 in math (Yeh 2008). These results suggest that rapid assessment is approximately three orders of magnitude more cost effective than high-quality preschool with regard to student achievement.[3] However, some studies of high-quality preschool detect reductions in crime (although studies of the Abecedarian program did not), and cost-effectiveness analyses that focus only on gains in student achievement ignore the benefits of reductions in crime (to the extent that those benefits are not the result of gains in student achievement).[4]

After adjusting for inflation, Borman and Hewes' (2002) effectiveness-cost ratios for the Perry and Abecedarian programs are comparable, ranging from 0.000008 to 0.000027, based on effect-size estimates from Schweinhart, Barnes, and Weikart (1993) and Ramey and others (2000), and cost information from Barnett (1992) and Developmental Center for Handicapped Persons (1987). In addition, Borman and Hewes (2002) also reported a second set of effectiveness-cost ratios, under the assumption that the cost of preschool is divisible into the portions responsible for increased student achievement, versus reductions in crime and other social benefits. As a practical matter, however, society incurs the full cost of preschool, and cannot choose to incur the partial cost that is assumed in the second set of ratios. The full cost is relevant if policymakers are strictly concerned with selection of the most cost-effective approach for raising student achievement with a limited amount of dollars.

Additional School Year

An additional year of schooling is estimated to improve student achievement by 0.10 to 0.20 SD, or an average of 0.15 SD, at an annual cost ranging from $6,205.11 to $22,338.41 (an average of $14,271.76) per student, in 2006 dollars, suggesting an effectiveness-cost ratio of 0.000011 (Mayer and Peterson 1999).

Voucher Programs

The cost effectiveness of voucher programs was estimated based on effect sizes averaged across three random assignment or lottery-based evaluations (Greene, Peterson, and Du 1999; Howell et al. 2002; Rouse 1998) and a cost-effectiveness analysis that integrated the impact results with cost estimates (Yeh 2007). While previous cost estimates focused on the nominal costs of the vouchers themselves, or the costs of voucher record keeping, monitoring, information, and adjudication services, Yeh (2007) pointed out that 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 and others (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) (NCES 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. (p. 425)

Yeh (2007) performed similar calculations suggesting that the same issue arises in each of the major voucher programs, and pointed out that this would not have allowed any 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 (9 to 1; *see* NCES 2005a). If public school students enroll in private schools, the private school classrooms fill up nine 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).

The issue remains even if voucher programs are adopted nationwide and all schools are privatized. The assumption underlying voucher programs is that competition improves student achievement. However, competitive pressure only exists if a significant proportion of students transfers from low-performing to higher-performing schools every year. The logic underlying voucher programs requires that they continue to draw students at random from the vast majority of lower-performing schools and funnel them into the top 10% of all schools, creating a need to build new schools and hire new teachers and administrators without compensating cost reductions in the sending schools. If at some point in the future this competitive pressure has raised student achievement to a satisfactory level, it may be reasonable to eliminate voucher programs and their associated costs, but the situation is then identical to the current situation where there is little or no competitive pressure (in areas where voucher programs are lacking), with presumably the same adverse effects on student achievement. The only way voucher programs can remain effective is if they continue to divert students into the top 10% of all schools.

Yeh (2007) have suggested that 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 students with disabilities and limited English proficiency (Belfield 2006b; Levin 1998; Levin and Driver 1997). While comprehensive cost data are not available for private schools, such data are available for charter schools (Nelson, Muir, and 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, Muir, and Drown 2003). For example, in Texas and Minnesota, charter school teacher salaries are over $13,760 lower, and in Pennsylvania more than $17,513 lower, in inflation-adjusted dollars, than comparable host district teacher salaries (Nelson, Muir, and Drown 2003). Yeh (2007) combined charter school cost information 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.

This figure may be checked against the per pupil private school expenditure figure calculated from the Digest of Education Statistics. Expenditures by private elementary and secondary schools totaled $39,300,000,000 in 2003–04 current dollars (USDE 2007, Table 26). Dividing by 2003 total private elementary and secondary enrollment of 6,099,221 (USDE 2007, Table 55), suggests per pupil expenditure of $6443.45, or $7,094.05 in 2006 dollars.

This figure ($7,094.05) is nearly identical to the figure Yeh (2007) estimated based on charter school expenditures ($7,098.88), before adding services that are provided by public schools but not private schools: food service ($323.00), special education ($1,082.55), transportation ($2,005.57), and compensatory, vocational, bilingual, and career and technology education ($793.83), plus, services that would be required by a voucher program: voucher record keeping and monitoring ($68.19), information services ($50.81), and adjudication services ($49.82).

Although Yeh (2007) assumed that private school expenditures are lower than expenditures by their public school counterparts, Levin (1998) obtained site-based expenditure and revenue data for Milwaukee public schools and their voucher school counterparts and concluded that voucher schools in Milwaukee are receiving at least as much revenue per student as the public schools. This suggests that Yeh's (2007) cost estimates for voucher programs are, if anything, underestimated.

Levin (1998) estimated annual core per pupil costs of $4,581.15, plus $1,980.89 to provide the full range of services provided by the public schools, as well as the record-keeping, monitoring, transportation, information, and adjudication services that would be required by a voucher program, or a total of $6562.04, in 2006 dollars. Levin (1998) finds Rouse's (1998) estimates of the effectiveness of the Milwaukee voucher program to be most valid. She found an effect size of 0.09 SD in math and zero in reading. Dividing 0.09 SD by Levin's cost figure of $6,562.04 results in an effectiveness-cost ratio of 0.000014—somewhat higher than Yeh's (2007) maximum ratio of 0.000008.

Mayer and Peterson (1999) estimated that voucher programs would improve student achievement by 0.20 SD at no cost. However, the assertion that voucher programs have no cost is untenable.

Charter Schools

Yeh (2007) integrated cost data with impact results from existing evaluations of charter schools to estimate cost effectiveness. While charter schools are

heterogeneous, the available evidence fails to demonstrate that any particular type of charter school is especially effective. Thus, it is appropriate to ask whether charter school programs are, in general, an efficient use of public resources. Since no large-scale random assignment studies have been completed, Yeh (2007) averaged effect sizes drawn from three large-scale impact evaluation studies that used panel data from Florida, North Carolina, and Texas and applied regression discontinuity methods to control for selection bias, maturation, and history effects (Bifulco and Ladd 2006; Hanushek et al. 2005; Sass 2006). The average effect size across these key studies of charter schools was a negative 0.088 SD in reading and a negative 0.116 SD in math. The effect of competition from charter schools on the performance of traditional public schools was inconsequential (Bifulco and Ladd 2006; Sass 2006). However, in all three studies, negative effects were largest in the first year of charter school operation and diminished over time. The average effect size for charter schools after five years in operation was 0.009 SD in reading and 0.001 SD in math.

Yeh's (2007) cost-effectiveness analysis drew upon much of the same cost data and logic used in the analysis of 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, including the public school cost savings attributable to students transferring to charter schools.

As illustrated by the calculations regarding voucher programs, only a fraction of the costs incurred by charter schools when students transfer from regular schools to charter schools 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).

DISCUSSION

The results suggest that rapid assessment is significantly more cost effectiveness than any of the 21 alternative approaches for raising student achievement listed in Table 1. However, this conclusion must be tempered by limitations in the available data. First, the comparison was limited to measures of math and reading achievement. While these measures are strongly related to future earnings (Currie and Thomas 2001; Murnane, Willet, and Levy 1995; Neal and Johnson

1996; O'Neill 1990; Winship and Korenman 1999), suggesting that achievement in these areas is important, the results should not be interpreted to mean that all spending on education should be directed toward rapid assessment. Instead, the results suggest that in schools where student achievement in math and reading is low, it may be cost effective to reallocate a portion of funding to rapid assessment. For example, the results suggest that a school system that is currently using voucher programs that cost an average of $9,646.01 per student annually could reallocate $28.34 of this to rapid assessment programs, improve achievement in math and reading, and generate savings of $9,617.67 per student that could be directed toward other important educational objectives. Thus, the application of the results of the current study would permit the vast majority of each school's budget, including the portion of funding that is saved as a consequence of rapid assessment's efficiency, to be directed toward important educational objectives that are not measured by standardized assessments in math and reading. It may be the case that funding should be directed toward some of the 21 alternatives analyzed here, if future research indicates that those alternatives are cost-effective approaches for addressing objectives other than math and reading achievement. The limitation of the current study is that it does not provide guidance about where to allocate funding with regard to those other objectives.

Second, the results of the present analysis are conditional on future studies that are needed to determine if the present results generalize across the full range of student and school characteristics. It is important to note that, while the rapid assessment approach appears to work at all grade levels in math, effects tend to fade out in the upper grades in reading. More research is needed to explore why this happens. It may be the case that future studies indicate that some of the 21 alternatives that I analyzed are more cost-effective than rapid assessment in raising student achievement in reading, especially for older children. Interestingly, however, the only study of rapid assessment that analyzed the influence of ability level found that rapid assessment was equally effective for gifted students (effect size = 0.45 SD) and nongifted students (effect size = 0.47 SD) with regard to math achievement (Ysseldyke et al. 2004).

Third, the results of the present analysis are conditional on the specific variations of each treatment that were analyzed, and the conditions under which each treatment was studied. It is possible that other variations of rapid assessment will prove to be less cost effective, either because they are less effective or more costly, or that the variation of rapid assessment that is the focus of the present analysis will prove to be less cost-effective under conditions that differ from the conditions of the studies reported here.

With these caveats, the results suggest that the variation of rapid assessment that is the focus of the present analysis, used with the population of students and

under the conditions characterized by the studies reported here, is approximately one magnitude (10 times) more cost effective than CSR, cross-age tutoring, or computer-assisted instruction; two magnitudes more cost-effective than a longer school day, increases in teacher education, teacher experience or teacher salaries, summer school, more rigorous math classes, value-added teacher assessment, class size reduction, a 10% increase in per pupil expenditure, or full-day kindergarten; three magnitudes more cost effective than Head Start, high-standards exit exams, NBPTS teacher certification, higher teacher licensure test scores, high-quality preschool, an additional school year, or voucher programs; and four magnitudes more cost-effective than charter schools.

A magnitude of gain implies that, for every dollar invested in rapid assessment rather than the alternative, a given increase in student achievement could be purchased for 10 times as many students. Two magnitudes imply a 100-fold gain, three magnitudes imply a 1,000-fold gain, and four magnitudes imply a 10,000-fold gain. The size of the differences in cost-effectiveness suggests that the results are robust even if future research indicates that the effect sizes or costs are incorrectly estimated by factors of 10 or more (Figure 1).

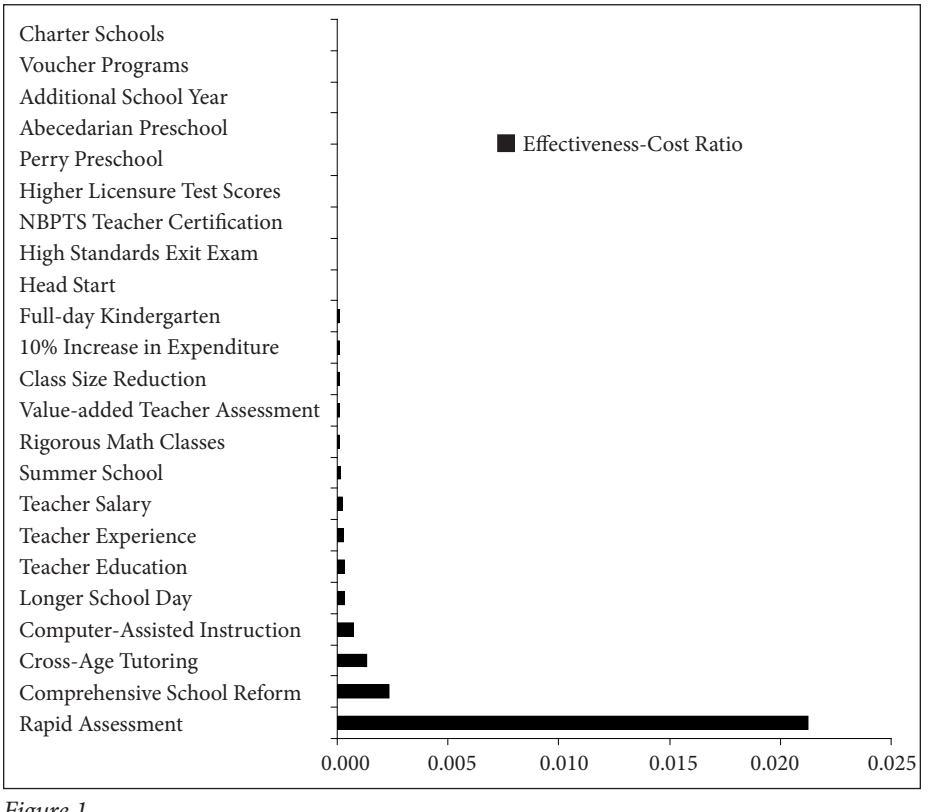

Figure 1.

If rapid assessment is indeed as cost-effective as it appears to be, what might explain the results? A review of research regarding the development of children's feelings of competence, the phenomenon of learned helplessness, and factors influencing persistence and intrinsic motivation, suggests that the individualization of task difficulty and performance feedback that are associated with rapid assessment may be critical in developing students' perceived control over academic outcomes and, thus, student engagement and achievement (Yeh in press-b). A student's perceived control over his or her academic performance is strongly predictive of academic achievement (Brookover et al. 1979; Brookover et al. 1978; Coleman et al. 1966; Crandall, Katkovky, and Crandall 1965; Kalechstein and Nowicki 1997; Keith, Pottebaum, and Eberhart 1986; Skinner, Wellborn, and Connell 1990; Teddlie and Stringfield 1993). There is a feedback loop between performance and control beliefs, with high performance leading to subsequent perceptions of control, so that early achievement strongly influences later achievement, and does so primarily by increasing students' sense of personal control (Musher, Eizenman, Nesselroade, and Schmitz 2002; Ross and Broh 2000; Skinner, Wellborn, and Connell 1990). Thus, when children believe that they can exert control over success in school, they perform better on cognitive tasks. And, when children succeed in school, they are more likely to view school performance as a controllable outcome (Skinner, Wellborn, and Connell 1990, p. 22).

Performance feedback plays a key role in this explanation. Feedback increases student motivation (Robinson, DePascale, and Roberts 1989; Ryan, Mims, and Koestner 1983). Feedback also increases achievement. 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 were not frequently tested (Bangert-Drowns et al. 1991; Fuchs and Fuchs 1986; Kluger and DeNisi 1996). A meta-analysis of 21 experimental studies that focused on studies involving testing found that students who were tested two to five times per week outperformed students who were not frequently tested, with an average effect size of 0.7 SD (Fuchs and Fuchs 1986). 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 and Fuchs 1986). A second meta-analysis of 40 feedback studies (Bangert-Drowns et al. 1991) that included nontesting feedback (such as 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 that included studies involving nontesting feedback, as well as studies involving testing feedback, found that praise or criticism attenuated the effectiveness of feedback (Kluger and DeNisi 1996). Emotionally neutral (i.e. testing) feedback which is void of praise or criticism "is likely to yield impressive gains in performance, possibly exceeding 1 SD"—much higher than the average effect size of 0.4 SD when all types of feedback studies were grouped together (Kluger and DeNisi 1996, p. 278). 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 and Timperley 2007). These results suggest that effective feedback systems involve frequent testing (two to five times per week) with nonjudgmental feedback presented immediately after each test. Under these conditions, the three meta-analyses of feedback interventions (Bangert-Drowns et al. 1991; Fuchs and Fuchs 1986; Kluger and DeNisi 1996) suggest that the effect size for testing feedback is no lower than 0.7 SD.

Individualization of task difficulty and rapid performance feedback are integral features of the particular type of rapid assessment program that is the focus of the current literature review. While other types of rapid assessment systems do not require schools to shelve books according to reading level, and do not provide software that prints individualized sets of math problems, they provide teachers with the information needed to individualize task difficulty. Rapid assessment means that teachers can quickly observe the types of items that cause difficulty for particular students and can make appropriate adjustments in task difficulty. To the extent that rapid assessment technology provides the information needed by teachers in order to individualize instruction, this technology is likely to promote actions by teachers necessary to achieve individualization.

The importance of individualization becomes clear when the current structure of schooling is analyzed with respect to the literature regarding control and achievement (Yeh in press-b). The existing structure of schooling, where students are grouped into age cohorts and constantly compared with their same-age classmates, inadvertently undermines the perceived control of below-average students over academic outcomes (Yeh in press-b). These students are constantly presented with tasks geared toward the average learner in their age cohort, and their performances are constantly compared to the average learner in this cohort (Yeh in press-b). The experience of schooling for these students is that they constantly receive test and grade reports that categorize them as below-average students and reinforce poor self images. Demoralized, they become passive and disengaged. They reduce their effort, further reducing their performance (Yeh in press-b). In contrast, the current structure of schooling reinforces the beliefs of above-average students that they are competent, because they regularly receive

test and grade reports that reinforce this categorization. But while their efforts are not sabotaged, this structure fails to adjust the level of task challenges to their level of competence, commonly resulting in boredom.

A review of existing research literature suggests that both syndromes can be addressed through interventions such as rapid assessment that individualize task difficulty on a daily basis, ensures that each student experiences more successes than failures, and provides performance feedback, autonomy in task execution, and an accelerating standard of performance (Yeh in press-b). Rapid assessment, unlike the alternatives to which it is being compared, provides a simple process by which students select and read books and complete a selection of math problems tailored to each student's current ability level (Yeh in press-b). This individualization of task difficulty on a daily basis serves to adjust reading and math activities so that they are challenging but not beyond the current ability of each individual learner. If individualization of task difficulty and rapid performance feedback foster student feelings of competence and control, and if feelings of competence and control are central to engagement and therefore achievement, there is reason to think that the implementation of rapid assessment systems intervenes in a way that changes the experience of schooling for underachieving students.

This hypothesis is supported by interviews with teachers and administrators and classroom observations in a Texas school district that implemented rapid assessment programs (Yeh 2006). According to the teachers, the use of the rapid assessment system promoted student feelings of competence and control, as well as engagement and achievement (Yeh 2006). Below-average students could work side-by-side with their more able peers, successfully complete a book or set of math problems, and feel the same sense of accomplishment as their peers, because task difficulty and performance feedback were individualized for each learner. Thus, to the extent that rapid assessment promotes task individualization, thereby ensuring that most students are successful and feel successful most of the time, and to the extent that rapid assessment accomplishes this through investments in software that can be spread over several hundred students per school and amortized over seven or more years, rapid assessment may address a critical need at relatively low cost. This could explain why it is much more cost-effective than alternative approaches for raising student achievement.

While the results of the current review of cost-effectiveness literature suggest the advantages of rapid assessment, the use of cost-effectiveness analysis to inform policy decisions is hindered by the persistent criticism that this type of analysis assumes that math and reading are the *only* important educational objectives and, furthermore, that standardized tests are the best way of measuring progress toward those objectives. However, it may be more accurate to say that cost-

effectiveness analysis assumes that improvements in math and reading skills are two important educational objectives, that standardized tests are one important way to measure progress toward those objectives, and the implementation of the most efficient approaches for raising math and reading achievement would conserve social resources that would then become available to pursue additional educational goals beyond the important goals of raising student achievement in math and reading. Interviews with teachers and classroom observations in a district that implemented rapid assessment systems district-wide suggest that these systems permitted teachers to efficiently teach math and reading skills, freeing up time that would otherwise have to be devoted to test preparation for the state-mandated test, thereby avoiding the basic skills-dominated curriculum that critics fear (Yeh 2006). In contrast, the persistent reliance on highly inefficient approaches for raising math and reading achievement in districts that do not use rapid assessment systems forces teachers and students to devote inordinate amounts of time and energy to basic skills, narrowing the curriculum and subverting the broader educational goals that are favored by critics of standardized testing (Yeh 2006).

If rapid assessment is more cost effective than the alternatives examined here, why has it not been more widely adopted? Yeh (2007) listed common objections. The most substantive concerns are that rapid assessment does not replace good instruction, it can be misused, and it does not foster critical thinking (Biggers 2001). However, rapid assessment is neither designed to supplant regular instruction nor to teach critical thinking skills, and is better understood as a tool designed to support teachers, providing useful information about student progress so that teachers may be more effective in teaching basic skills. 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.

Rapid assessment is unlikely to be a panacea. As with any other intervention, it can be implemented poorly, with poor results. Overall, however, the available research evidence suggests that rapid assessment may address a fundamental need to structure the experience of schooling so that both low- and high-ability students feel challenged but not overwhelmed, improving engagement and achievement in a cost-effective way.

Endnotes

1. *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.
2. *Ibid.*
3. Note that the math and reading achievement measures for the Perry and Abecedarian programs were administered two to three years after the preschool treatment ended. Since the effect of a treatment may be expected to decay over time, the delay in measurement may disadvantage high-quality preschool, compared to rapid assessment and the other interventions where achievement was measured immediately after the treatments ended.
4. If the effects of preschool on earnings and reductions in crime are mediated entirely by increased student achievement, it is appropriate to infer that any intervention that improves student achievement by an equivalent amount would have similar effects on earnings and crime, and a cost-effectiveness analysis that ranks each intervention along the dimension of student achievement accurately predicts rankings along the dimensions of earnings and reductions in crime. If, however, the effects of preschool on earnings and reductions in crime involve a different path, it would not be appropriate to infer that rankings along the dimension of student achievement accurately predict rankings along the dimensions of earnings and reductions in crime. The best available data suggest that a significant portion of the effects of preschool is channeled through increased student achievement but other mediators are also important (Reynolds, Ou, and Topitzes 2004).

APPENDIX A: NOTES TO TABLE 1

[a] Annualized effect size.
[b] Annual cost per student, adjusted for inflation to 2006 dollars.
[c] Effect size in SD units divided by annual cost per student.
[d] Ross et al. (2004)
[e] Ysseldyke and Tardrew (2007).
[f] Reading.
[g] Math.
[h] Nunnery et al. (2006).
[i] Ysseldyke and Bolt (2007).
[j] (Yeh 2008). Of the subgroup of comprehensive school reform models categorized by Borman, Hewes, Overman, and Brown (2003) as having the *Strongest* or *Highly Promising Evidence of Effectiveness*, the model with the highest effect size was *Expeditionary Learning Outward Bound* (effect size = 0.51 SD).
[k] (Yeh 2008). Of the subgroup of comprehensive school reform models categorized by Borman, Hewes, Overman, and Brown (2003) as having the *Strongest* or *Highly Promising Evidence of Effectiveness*, the model with the highest effectiveness-cost ratio was *Expeditionary Learning Outward Bound* (effectiveness-cost ratio = 0.002341).
[l] Nye, Hedges, and Konstantopoulos (2001). Average of effect sizes in grades 1, 2, and 3.
[m] Reichardt (2000). Annual cost per student of reducing class size from 24 to a ceiling of 17 students per class, adjusted for inflation using the September, 1997 price index (161.2) and the August, 2006, price index (203.9).
[n] Finn, Gerber, Achilles, and Boyd-Zaharias (2001). In grade 2, the achievement advantage for students who participated in small classes for 1, 2, and 3 years was 0.12 SD, 0.24 SD, and 0.36 SD respectively in reading, or an average of 0.12 SD per year, and 0.16 SD, 0.24 SD, and 0.32 SD respectively in math, or an average of 0.129 SD per year.
[o] Greenwald, Hedges, and Laine's (1996) meta-analytic estimate that a 10% increase in per-pupil expenditure would increase student achievement in math and reading by 0.083 SD annually.
[p] Total expenditure per pupil in constant 2004–2005 dollars equals $10,464 (NCES 2005b). A 10% increase equals $1,046.40, or $1,118.83 in 2006 dollars, using the January 2005 price index (190.7) and the August 2006 price index (203.9).
[q] Grodsky, Warren, and Kalogrides (2008).
[r] Including the cost to society of the incremental increase in the number of students who are denied diplomas solely because exit exams raise the graduation bar.
[s] Schweinhart, Barnes, and Weikart (1993), Table 13, annualized over two-year period.
[t] Barnett (1992), adjusted for inflation using the January, 1985, price index (105.5) and the August, 2006, price index (203.9).
[u] Ramey et al. (2000), Figure 3, annualized over five-year period.
[v] Barnett and Masse (2007). Cost in a public school setting, minus the value of formal and informal childcare services provided to the control group, and adjusted for inflation using the January, 2002, price index (177.1) and the August, 2006, price index (203.9). Note that since preschool costs are incurred in years prior to the K-6 years when rapid assessment is typically implemented, preschool costs are underestimated relative to the cost estimate for rapid assessment (costs for rapid assessment should be discounted to the time period when preschool costs are incurred).
[w] Yeh (2007).
[x] Average effect size across three randomized and lottery-based studies.
[y] 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. Private school costs were reduced by public school savings of 17.38%.
[z] Average effect size across 3 large-scale studies controlling for student fixed effects, after five-year shakedown period.

[*]To the extent that charter schools are not simply converted public schools, but instead are newly created schools that require new facilities, administrators, and 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. Charter school costs were reduced by regular school savings of 17.38%.

References

Angrist, J. D., and J. Guryan. 2004. *Does teacher testing raise teacher quality? Evidence from state certification requirements*. no city: National Bureau of Economic Research.

Aos, S., M. Miller, and J. Mayfield. 2007. *Benefits and costs of K–12 educational policies: Evidence-based effects of class size reductions and full-day kindergarten*. (Document 07-03-2201). Olympia: Washington State Institute for Public Policy.

Ballou, D., and M. Podgursky. 1995. "Recruiting smarter teachers." *The Journal of Human Resources* 30: 326-338.

Bangert-Drowns, R. L., C. C. Kulik, J. A. Kulik, and M. Morgan. 1991. "The instructional effect of feedback in test-like events." *Review of Educational Research* 61: 213-238.

Barnett, W. S. 1992. "Benefits of compensatory preschool education." *The Journal of Human Resources* 27: 279-312.

Barnett, W. S., and L. N. Masse. 2007. "Comparative benefit–cost analysis of the Abecedarian program and its policy implications." *Economics of Education Review* 26: 113-125.

Belfield, C. 2006a. *The use of cost-benefit analysis in guiding investments in human capital in elementary and secondary school*. Flushing, NY: City University of New York.

Belfield, C. R. 2006b. *The evidence on education vouchers: An application to the Cleveland Scholarship and Tutoring Program* Flushing, NY: Queens College, City University of New York, and National Center for the Study of Privatization in Education, Teachers College.

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

Biggers, D. 2001. "The argument against Accelerated Reader." *Journal of Adolescent & Adult Literacy* 45: 72-75.

Borman, G. D., and N. M. Dowling. 2006. "Longitudinal achievement effects of multiyear summer school: Evidence from the Teach Baltimore randomized field trial." *Educational Evaluation and Policy Analysis* 28: 25-48.

Borman, G. D., and G. M. Hewes. 2002. "The long-term effects and cost-effectiveness of Success for All." *Educational Evaluation and Policy Analysis* 24: 243–266.

Borman, G. D., G. M. Hewes, L. T. Overman, and S. Brown. 2003. "Comprehensive school reform and achievement: A meta-analysis." *Review of Educational Research* 73: 125-230.

Brookover, W. B., C. H. Beady, P. K. Flood, J. G. Schweitzer, and J. M. Wisenbaker. 1979. *Schools, social systems and student achievement: Schools can make a difference*. New York: Praeger.

Brookover, W. B., J. G. Schweitzer, J. M. Schneider, C. H. Beady, P. K. Flood, and J. M. Wisenbaker. 1978. "Elementary school social climate and school achievement." *American Educational Research Journal* 15: 301-318.

Cavalluzzo, L. "Is National Board certification an effective signal of teacher quality?" The CNA Corporation. http://www.cna.org/documents/CavaluzzoStudy.pdf (Accessed October 29, 2006).

Center on Education Policy. 2005. *High school exit exams: Basic features*. (Exit Exams Policy Brief 1). Washington, D.C.: Center on Education Policy.

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

———. 2007. *How and why do teacher credentials matter for student achievement?* (Working Paper 2). no city: National Center for Analysis of Longitudinal Data in Education Research.

Cohen, P. A., J. A. Kulik, and C.-L. C. Kulik. 1982. "Educational outcomes of tutoring: A meta-analysis of findings." *American Educational Research Journal* 19: 237-248.

Coleman, J. S., E. Campbell, C. Hobson, J. McPartland, A. Mood, R. Weinfeld, and R. York. 1966. *Equality of educational opportunity*. Washington, D. C.: Government Printing Office.

College Board. "SAT percentile ranks: Critical reading, mathematics, and writing." http://www.collegeboard.com/prod_downloads/highered/ra/sat/SATPercentileRanks.pdf (Accessed February 17, 2007).

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

Crandall, V., W. Katkovky, and U. Crandall. 1965. "Children's beliefs in their own control of reinforcements in intellectual academic achievement situations." *Child Development* 36: 91-109.

Currie, J., and D. Thomas. 2001. "Early test scores, socioeconomic status, school quality and future outcomes." *Research in Labor Economics* 20: 103-132.

Day, J. C., and E. C. Newburger. 2002. *The big payoff: Educational attainment and synthetic estimates of work-life earnings*. (Report P23-210). Washington, D.C.: U.S. Census Bureau.

Developmental Center for Handicapped Persons. 1987. *Early Intervention Research Institute, final report*. Logan, UT: Utah State University. (ERIC Document Reproduction Service No. ED293242).

Ehrenberg, R. G., and D. J. Brewer. 1995. "Did teacher verbal ability and race matter in the 1960s? Coleman revisited." *Economics of Education Review* 14: 1-21.

Ferguson, R. F. 1991. "Paying for public education: New evidence on how and why money matters." *Harvard Journal on Legislation* 28: 465-498.

Ferguson, R. F., and H. F. Ladd. 1996. How and why money matters: An analysis of Alabama schools. In H. F. Ladd, ed.*Holding schools accountable: Performance-based reform in education*. Washington, D. C.: The Brookings Institution.

Finn, J. D., S. B. Gerber, C. M. Achilles, and J. Boyd-Zaharias. 2001. "The enduring effects of small classes." *Teachers College Record* 103: 145-183.

Flores-Lagunes, A., and A. Light. 2004. *Identifying sheepskin effects in the returns to education*. Tucson, AZ: Department of Economics, University of Arizona.

———. 2007. *Interpreting sheepskin effects in the returns to education*. Tucson, AZ and Columbus, OH: University of Arizona and Ohio State University.

Fuchs, L. S., and D. Fuchs. 1986. "Effects of systematic formative evaluation: A meta-analysis." *Exceptional Children* 53: 199-208.

Gayler, K., N. Chudowsky, M. Hamilton, N. Kober, and M. Yeager. 2004. *State high school exit exams: A maturing reform*. Washington, D.C.: Center on Education Policy.

Goldhaber, D. 2007. "Everyone's doing it, but what does teacher testing tell us about teacher effectiveness?" *Journal of Human Resources* 42: 765-794.

Goldhaber, D., and E. Anthony. "Can teacher quality be effectively assessed? National Board certification as a signal of effective teaching." http://www.urban.org/UploadedPDF/411271_teacher_quality.pdf (Accessed October 29, 2006).

———. 2007. "Can teacher quality be effectively assessed? National Board certification as a signal of effective teaching." *Review of Economics and Statistics* 89: 134-150.

Gordon, R., T. J. Kane, and D. O. Staiger. 2006. *Identifying effective teachers using performance on the job (Discussion Paper 2006-01)*. Washington, D.C.: The Brookings Institution.

Greene, J. P., P. E. Peterson, and J. Du. 1999. "Effectiveness of school choice: The Milwaukee experiment." *Education and Urban Society* 31: 190-213.

Greenwald, R., L. V. Hedges, and R. D. Laine. 1996. "The effect of school resources on student achievement." *Review of Educational Research* 66: 361-396.

Grissmer, D. 2002. Cost-effectiveness and cost-benefit analysis: The effect of targeting interventions. In H. M. Levin and P. J. McEwan, ed.*Cost-effectiveness and educational policy*. Larchmont, NY: Eye on Education.

Grissmer, D. W., A. Flanagan, J. Kawata, and S. Williamson. 2000. *Improving student achievement: What state NAEP test scores tell us*. Santa Monica, CA: RAND Corporation.

Grodsky, E., J. R. Warren, and D. Kalogrides. 2008. "State high school exit examinations and NAEP long-term trends in reading and mathematics, 1971-2004." *Educational Policy* (Accessed February 14, 2009).

Hanushek, E. A. 1986. "The economics of schooling: Production and efficiency in public schools." *Journal of Economic Literature* 24: 1141-1171.

———. 1989. "The impact of differential expenditures on school performance." *Educational Researcher* 18: 45-50.

———. 1997. "Assessing the effects of school resources on student performance: An update." *Educational Evaluation and Policy Analysis* 19: 141-164.

Hanushek, E. A., J. F. Kain, D. M. O'Brien, and S. G. Rivkin. 2005. *The market for teacher quality*. (Working Paper 11154). Cambridge, MA: National Bureau of Economic Research.

Hanushek, E. A., J. F. Kain, and S. G. Rivkin. 1999. *Do higher salaries buy better teachers? National Bureau of Economic Research Working Paper No. 7082*.

Hanushek, E. A., J. F. Kain, S. G. Rivkin, and G. F. Branch. 2005. *Charter school quality and parental decision making with school choice (NBER Working Paper 11252)*. Cambridge, MA: National Bureau of Economic Research.

Hanushek, E. A., and M. E. Raymond. 2005. "Does school accountability lead to improved student performance?" *Journal of Policy Analysis and Management* 24: 297-327.

Harris, D. N., and T. R. Sass. 2007. *The effects of NBPTS-certified teachers on student achievement*. Madison, WI and Tallahassee, FL: University of Wisconsin and Florida State University.

Hattie, J., and H. Timperley. 2007. "The power of feedback." *Review of Educational Research* 77: 81-112.

Hedges, L. V., R. D. Laine, and R. Greenwald. 1994. "Money does matter somewhere: A reply to Hanushek." *Educational Researcher* 23: 9-10.

Herman, R., Aladjem, D., McMahon, P., Masem, E., Mulligan, I., O'Malley, A. S., Quinones., S., Reeve, A., Woodruff, D. 1999. *An educator's guide to schoolwide reform*. Arlington, VA: American Institutes for Research.

Howell, W. G., P. J. Wolf, D. E. Campbell, and P. E. Peterson. 2002. "School vouchers and academic performance: Results from three randomized field trials." *Journal of Policy Analysis and Management* 21: 191-217.

Hoxby, C. M. 2002. *The cost of accountability*. (Working Paper 8855). Cambridge, MA: National Bureau of Economic Research.

Hummel-Rossi, B., and J. Ashdown. 2002. "The state of cost-benefit and cost-effectiveness analyses in education." *Review of Educational Research* 72: 1-30.

Kalechstein, A. D., and S. Nowicki, Jr. 1997. "A meta-analytic examination of the relationship between control expectancies and academic achievement: An 11-yr follow-up to *Findley* and *Cooper*." *Genetic, Social, & General Psychology Monographs* 123: 27-56.

Keith, T. Z., S. M. Pottebaum, and S. Eberhart. 1986. "Effects of self concept and locus of control on academic achievement: A large sample path analysis." *Journal of Psychoeducational Assessment* 4: 61-72.

King, J. R. 1994. "Meeting the educational needs of at-risk students: A cost analysis of three models." *Educational Evaluation and Policy Analysis* 16: 1-19.

Kluger, A. N., and A. DeNisi. 1996. "The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory." *Psychological Bulletin* 119: 254-284.

Krueger, A. B. 2003. "Economic considerations and class size." *The Economic Journal* 113: F34-F63.

Krueger, A. B., and D. M. Whitmore. 2001. *Would smaller classes help close the Black-White achievement gap?* (Working Paper 451). Princeton, NJ: Princeton University and NBER.

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

Levin, H. M. 1991. Cost-effectiveness at quarter century. In M. W. McLaughlin and D. C. Phillips, ed.*Evaluation and education at quarter century*. Chicago: University of Chicago Press.

———. 1998. "Educational vouchers: Effectiveness, choice, and costs." *Journal of Policy Analysis and Management* 17: 373-392.

Levin, H. M., and C. E. Driver. 1997. "Costs of an educational vouchers system." *Education Economics* 5: 265-283.

Levin, H. M., G. V. Glass, and G. Meister. 1987. "A cost-effectiveness analysis of computer-assisted instruction." *Evaluation Review* 11: 50-72.

Levin, H. M., and P. J. McEwan. 2001. *Cost effectiveness analysis: Methods and applications*. 2nd ed. Thousand Oaks, CA: Sage.

Little, J., and L. Bartlett. 2002. "Career and commitment in the context of comprehensive school reform." *Teachers and teaching: Theory and practice* 8: 345-354.

Ludwig, J., and D. A. Phillips. 2008. "Long-term effects of Head Start on low-income children." *Annals of the New York Academy of Sciences* 1136: 257–268.

Manski, C. F. 1987. Academic ability, earnings, and the decision to become a teacher: Evidence from the National Longitudinal Study of the high school class of 1972. In D. Wise, ed.*Public sector payrolls*. Chicago: University of Chicago Press.

Mayer, S. E., and P. E. Peterson. 1999. The costs and benefits of school reform. In S. E. Mayer and P. E. Peterson, ed.*Earning and learning: How schools matter*. Washington, DC: Brookings Institution Press.

Monk, D. H., and J. A. King. 1993. Cost analysis as a tool for education reform. In S. L. Jacobson and R. Berne, ed.*Reforming education: The emerging systemic approach*. Thousand Oaks, CA: Corwin Press.

Murnane, R., J. Willet, and F. Levy. 1995. "The growing importance of cognitive skills in wage determination." *Review of Economics and Statistics* 77: 251-266.

Musher Eizenman, D. R., J. R. Nesselroade, and B. Schmitz. 2002. "Perceived control and academic performance: A comparison of high and low performing children on within person change patterns." *International Journal of Behavioral Development* 26: 540-547.

National Center for Education Statistics. "Characteristics of the 100 largest public elementary and secondary school districts in the United States: 1999-2000." http://nces.ed.gov/pubs2001/100_largest/tableAA4.asp (Accessed July 21, 2006).

———. "Characteristics of the 100 largest public elementary and secondary school districts in the United States: 2000-2001." http://nces.ed.gov/pubs2002/100_largest/table_02_1.asp (Accessed July 21, 2006).

———. "The condition of education." U.S. Department of Education. http://nces.ed.gov/programs/digest/d05/tables/dt05_057.asp and http://nces.ed.gov/programs/digest/d05/tables/dt05_033.asp (Accessed March 13, 2007).

———. "Digest of education statistics, Table 162. Total and current expenditure per pupil in public elementary and secondary schools: Selected years, 1919-20 through 2002-03." U.S. Department of Education. http://nces.ed.gov/programs/digest/d05/tables/dt05_162.asp (Accessed September 8, 2006).

———. "Digest of education statistics: 2005." U.S. Department of Education. http://nces.ed.gov/programs/digest/d05/tables/dt05_004.asp (Accessed December 1, 2006).

Neal, D., and W. R. Johnson. 1996. "The role of pre-market factors in black-white wage differences." *Journal of Political Economy* 104: 869-895.

Nelson, F. H. "How much thirty thousand charter schools cost." American Federation of Teachers. http://www.aft.org/topics/charters/downloads/howmuch.pdf (Accessed July 27, 2006).

Nelson, F. H., E. Muir, and R. Drown. 2003. *Paying for the vision: Charter school revenue and expenditures. National charter school finance study*. Washington, D.C.: American Federation of Teachers.

Nunnery, J. A., S. M. Ross, and A. McDonald. 2006. "A randomized experimental evaluation of the impact of Accelerated Reader/Reading Renaissance implementation on reading achievement in grades 3 to 6." *Journal of Education for Students Placed At Risk* 11: 1-18.

Nye, B., L. V. Hedges, and S. Konstantopoulos. 2001. "Are effects of small classes cumulative? Evidence from a Tennessee experiment." *Journal of Educational Research* 94: 336-345.

O'Neill, J. 1990. "The role of human capital in earnings differences between black and white men." *Journal of Economic Perspectives* 4: 25-46.

Odden, A. 2000. "The costs of sustaining educational change through comprehensive school reform." *Phi Delta Kappan* 81: 433-438.

Ramey, C. T., F. A. Campbell, M. Burchinal, M. L. Skinner, D. M. Gardner, and S. L. Ramey. 2000. "Persistent effects of early childhood education on high-risk children and their mothers." *Applied Developmental Science* 4: 2-14.

Reichardt, R. E. 2000. The cost of class size reduction: Advice for policymakers, RAND Graduate School of Policy Studies, Santa Monica, CA.

Renaissance Learning. no date. *STAR Reading: Understanding reliability and validity*. Wisconsin Rapids, WI: Renaissance Learning.

Reynolds, A. J., S. R. Ou, and J. W. Topitzes. 2004. "Paths of effects of early childhood intervention on educational attainment and delinquency: A confirmatory analysis of the Chicago Child-Parent Centers." *Child Development* 75: 1299-1328.

Rice, J. K., and L. J. Hall. 2008. "National Board certification for teachers: What does it cost and how does it compare?" *Education Finance and Policy* 3: 339-373.

Robinson, S. L., C. DePascale, and F. C. Roberts. 1989. "Computer delivered feedback in group based instruction: Effects for learning disabled students in mathematics." *Learning Disabilities Focus* 5: 28-35.

Ross, C. E., and B. A. Broh. 2000. "The roles of self esteem and the sense of personal control in the academic achievement process." *Sociology of Education* 73: 270-284.

Ross, S. M., J. Nunnery, and E. Goldfeder. 2004. *A randomized experiment on the effects of Accelerated Reader/Reading Renaissance in an urban school district: Final evaluation report*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.

Rouse, C. E. 1998. "Private school vouchers and student achievement: An evaluation of the Milwaukee Parental Choice Program." *The Quarterly Journal of Economics* 113: 553-602.

Ryan, R. M., V. Mims, and R. Koestner. 1983. "The relationship of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory." *Journal of Personality and Social Psychology* 45: 736-750.

Sanders, W. L., J. J. Ashton, and S. P. Wright. 2005. *Comparison of the effects of NBPTS certified teachers with other teachers on the rate of student academic progress*. SAS Institute, Inc.

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

Schweinhart, L. J., H. V. Barnes, and D. P. Weikart. 1993. *Significant benefits: The High/Scope Perry Preschool Study through age 27*. Ypsilanti, MI: High/Scope Press.

Skinner, E. A., J. G. Wellborn, and J. P. Connell. 1990. "What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school." *Journal of Educational Psychology* 82: 22-32.

Strauss, R. P., and E. A. Sawyer. 1986. "Some new evidence on teacher and student competencies." *Economics of Education Review* 5: 41-48.

Teddlie, C., and S. Stringfield. 1993. *Schools do make a difference: Lessons learned from a 10 year study of school effects*. New York: Teachers College Press.

U.S. Department of Education. "Common core of data: State nonfiscal survey." National Center for Education Statistics. http://nces.ed.gov/ccd/bat/ (Accessed August 12, 2007).

———. 2007. *Digest of education statistics 2006*. (NCES 2007-017). Washington, D.C.: National Center for Education Statistics.

U.S. General Accounting Office. 2001. *School vouchers: Publicly funded programs in Cleveland and Milwaukee*. (Report Number GAO-01-914). Washington, D.C.: U.S. General Accounting Office.

United States General Accounting Office. 2003. *Title I: Characteristics of tests will influence expenses; Information sharing may help states realize efficiencies*. (GAO-03-389). Washington, D.C.: Author.

Warren, J. R., K. N. Jenkins, and R. B. Kulick. 2006. "High school exit examinations and state-level completion and GED rates, 1975-2002." *Educational Evaluation and Policy Analysis* 28: 131-152.

Wenglinsky, H. 1997. "The effect of school district spending on academic achievement." *Sociology of Education* 70: 221-237.

Winship, C., and S. D. Korenman. 1999. Economic success and the evolution of schooling and mental ability. In S. E. Mayer and P. E. Peterson, ed.*Earning and learning: How schools matter*. Washington, D. C.: Brookings Institution Press.

Yeh, S. S. 2006. "High stakes testing: Can rapid assessment reduce the pressure?" *Teachers College Record* 108: 621-661.

———. 2007. "The cost-effectiveness of five policies for improving student achievement." *American Journal of Evaluation* 28: 416-136.

———. 2008. "The cost-effectiveness of comprehensive school reform and rapid assessment." *Education Policy Analysis Archives* 16. http://epaa.asu.edu/epaa/v16n13/ (Accessed November 8, 2008).

———. 2009a. "Class size reduction or rapid formative assessment? A comparison of cost-effectiveness." *Educational Research Review* 4: 7-15.

———. 2009b. "The cost-effectiveness of raising teacher quality." *Educational Research Review* 4: 220-232.

———. in press-a. "The cost-effectiveness of NBPTS teacher certification." *Evaluation Review*.

———. in press-b. "Understanding and addressing the achievement gap through individualized instruction and formative assessment." *Assessment in Education*.

Yeh, S. S., and J. Ritter. 2009. "The cost-effectiveness of replacing the bottom quartile of novice teachers through value-added teacher assessment." *Journal of Education Finance* 34: 426-451.

Ysseldyke, J., and D. M. Bolt. 2007. "Effect of technology-enhanced continuous progress monitoring on math achievement." *School Psychology Review* 36: 453-467.

Ysseldyke, J., and S. Tardrew. 2007. "Use of a progress-monitoring system to enable teachers to differentiate math instruction." *Journal of Applied School Psychology* 24: 1-28.

Ysseldyke, J., S. Tardrew, J. Betts, T. Thill, and E. Hannigan. 2004. "Use of an instructional management system to enhance math instruction of gifted and talented students." *Journal for the Education of the Gifted* 27: 293-310.