



Education Policy Analysis Archives/Archivos
Analíticos de Políticas Educativas

ISSN: 1068-2341

epaa@alperin.ca

Arizona State University
Estados Unidos

Yeh, Stuart S.

The Cost-Effectiveness of Comprehensive School Reform and Rapid Assessment

Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas, vol. 16, 2008, pp. 1-32

Arizona State University

Arizona, Estados Unidos

Disponible en: <http://www.redalyc.org/articulo.oa?id=275020545012>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

EDUCATION POLICY ANALYSIS ARCHIVES

English Editor: Sherman Dorn
College of Education
University of South Florida

Spanish Editor: Gustavo Fischman
Mary Lou Fulton College of Education
Arizona State University

Volume 16 Number 13

June 30, 2008

ISSN 1068-2341

The Cost-Effectiveness of Comprehensive School Reform and Rapid Assessment

Stuart S. Yeh
University of Minnesota

Citation: Yeh, S. S. (2008). The cost-effectiveness of comprehensive school reform and rapid assessment. *Education Policy Analysis Archives*, 16(13). Retrieved [date] from <http://epaa.asu.edu/epaa/v16n13/>.

Abstract

Analysis of the cost-effectiveness of 29 Comprehensive School Reform (CSR) models suggests that all 29 models are less cost-effective than an alternative approach for raising student achievement, involving rapid assessment systems that test students 2 to 5 times per week in math and reading and provide rapid feedback of the results to students and teachers. Results suggest that reading and math achievement could increase approximately one order of magnitude greater for every dollar invested in rapid assessment rather than CSR. The results also suggest that reading and math achievement could increase two orders of magnitude for every dollar invested in rapid assessment rather than class size reduction and three orders of magnitude for every dollar invested in rapid assessment rather than high quality preschool.

Keywords: comprehensive school reform; formative evaluation; cost effectiveness; reading; math.



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives (Archivos Analíticos de Políticas Educativas)**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-nd/2.5/>. All other uses must be approved by the author(s) or **EPAA/AAPE**. **EPAA/AAPE** is published jointly by the Colleges of Education at Arizona State University and the University of South Florida. Articles are indexed by the Directory of Open Access Journals, H.W. Wilson & Co., and SCOPUS.

La relación costo-eficacia de reformas escolares comprensivas y evaluaciones rápidas

Abstract

Análisis de la relación costo-eficacia, de 29 modelos de reformas escolares comprensivas (REC) sugiere que todos esos modelos son menos rentables que un enfoque alternativo para aumentar el rendimiento de los estudiantes, que involucra sistemas de evaluación rápidos que ponen a prueba los estudiantes de 2 a 5 veces por semana en las áreas de matemáticas y lectura, proporcionando información de los resultados a los estudiantes y profesores de manera rápida. Los resultados sugieren que por cada dólar invertido se podría aumentar el aprendizaje de la lectura y matemáticas en un orden de magnitud mayor usando el sistema de evaluación rápida en lugar de la REC. Los resultados también sugieren que el aprendizaje de la lectura y matemáticas podría aumentar dos órdenes de magnitud por cada dólar invertido usando evaluación rápida en lugar de reducir el tamaño de las clases y tres órdenes de magnitud por cada dólar invertido usando evaluación rápida en vez de invertir en educación preescolar de calidad.

Palabras claves: reformas escolares comprensivas; evaluación formativa; costo-eficacia; lectura; matemáticas.

Introduction

Comprehensive school reform (CSR) may be defined as externally developed school improvement programs known as “whole-school” or “comprehensive” reforms emphasizing a coherent vision of education, a challenging curriculum, and high expectations for academic achievement. CSR is often implemented at the elementary school level, although several models target middle or high schools. CSR typically involves intensive staff development, increased attention to instruction and the needs of individual students, and parent involvement. These programs originated in 1991, when President George H. W. Bush announced the creation of a private-sector organization called the New American Schools Development Corporation (NAS), which was intended to support the creation of “break the mold” models of American schools that would enable all students to achieve world-class standards in core academic subjects (Kearns & Anderson, 1996). NAS solicited and received nearly 700 proposals in February, 1992. Eleven were chosen for a three-year program of development and testing. Subsequently, NAS dropped four models but provided more than \$150 million over the past decade to develop and “scale up” implementation of the remaining seven models.

In 1997, the U.S. Congress spurred the development of CSR by passing legislation to fund the Comprehensive School Reform Demonstration (CSR/D) Program, which provided \$50,000 per year for three years to qualifying schools. In 2001, the reauthorization of Title I limited CSR funding to “scientifically based” whole-school reform models, increasing pressure on CSR developers to show that the models improved student achievement (U.S. Department of Education, 2002b). Congressional appropriations for CSR totaled \$1.9 billion from 1998 to 2006 (U.S. Department of Education, 2004, 2006), in addition to over \$150 million provided by NAS (Borman, Hewes, Overman, & Brown, 2004). Thus, funding for CSR has totaled well over \$2 billion. CSR has expanded to include over 800 different reform models and has been implemented in 5,160 schools nationwide (Rowan, Barnes, & Camburn, 2004). It is estimated that somewhere between 10% and

20% of all elementary schools in the United States have adopted an external model of CSR or are working with their own locally developed model (Rowan et al., 2004). Forty-five percent of all Title I schools, and 80% of the highest-poverty schools, operate a schoolwide program (Heid & Webber, 1999).

The theory of action underlying CSR is that comprehensive changes are needed in multiple areas including staff attitudes and school organization, as well as curriculum and instruction, for these changes to be effective in improving student achievement (Borman et al., 2004). It is believed that the impact of isolated changes in any individual area is likely to be undermined by dysfunction in other areas. For example, changes in attitudes and organizational behaviors may be ineffective without improvements in curriculum and instruction, and improvements in curriculum and instruction can easily be undermined by dysfunctional attitudes and organizational behaviors. Thus, the theory of action underlying CSR is that all of these areas must be simultaneously addressed to improve student achievement.

The attraction of CSR may be explained by the plausibility of this grand model. Furthermore, school leaders and policymakers tend to emphasize dramatic reform efforts rather than narrowly-tailored interventions because dramatic reforms are highly visible symbols of a commitment to change (Tyack & Cuban, 1995). However, enormous resources have been devoted to CSR, yet the results have been meager. There is a need to re-examine the cost and effectiveness of CSR, and to compare its cost-effectiveness with promising alternatives.

While the notion that public schools need to be overhauled from top to bottom may be attractive, the history of school reform suggests that incremental reforms are more likely to persist than dramatic reforms because incremental reforms are more easily integrated into existing structures and routines (Tyack & Cuban, 1995). Hattie and Timperley (2007) systematically reviewed the available meta-analytic evidence regarding a broad range of interventions to improve student achievement. They identified performance feedback, which may be considered an incremental reform, as having one of the largest effect sizes (0.79 standard deviations, or *SD*). Thus, the purpose of this paper is to compare CSR with rapid assessment systems that provide feedback to students and teachers regarding student performance in math and reading.

CSR Impacts

Evidence regarding the effects of CSR has emerged very slowly. Early schoolwide reforms failed to produce compelling evidence of improved student achievement (Wong, 2001; Wong & Meyer, 1998). As a result, reviews of the research literature were limited to practitioner-oriented summaries of the general attributes of the CSR models, the level of support provided by developers, the costs associated with implementing the models, and narrative appraisals of the research supporting each CSR design (see Herman et al., 1999; Northwest Regional Educational Laboratory, 1998; Northwest Regional Educational Laboratory, 2005; Slavin & Fashola, 1998; Traub, 1999; Wang, Haertel, & Walberg, 1997). These reviews did not provide quantitative meta-analyses of the overall effects of CSR nor the effects of the various CSR models.

More recently, however, Borman, Hewes, Overman, and Brown (2003) synthesized “all known research on the achievement effects of the most widely implemented, externally developed school improvement programs known as ‘whole-school’ or ‘comprehensive’ reforms” and conducted the first meta-analysis of CSR effects (p. 126). This exhaustive review of 232 studies of achievement, regarding 29 of the most widely implemented CSR models, found that several models produced large effect sizes. However, the number and methodological quality of the research studies regarding those models was not sufficient to draw firm conclusions (Borman et al., 2003).

Three models met the highest standard of evidence and “are the only CSR models to have clearly established, across varying contexts and varying study designs, that their effects are relatively robust and that the models, in general, can be expected to improve test scores” (Borman et al., 2003, p. 168). The best estimate of effect size is derived by examining studies involving comparison groups.¹ These effect sizes were small: *Direct Instruction* (effect size = 0.15 *SD*), the *School Development Program* (effect size = 0.05 *SD*), and *Success for All* (effect size = 0.18 *SD*). Acknowledging disappointing results, researchers conceded that “CSR.... is not the panacea for closing the achievement gap and decreasing high school drop out rates for poor and historically underserved students of color” (Ross & Gil, 2004, p. 170).

Unexplained Variation in Results

Borman et al. (2003) noted wide variation in achievement that was not explained by the CSR models. The results suggest that CSR models fail to account for important factors that influence student achievement:

The heterogeneity of the CSR effect and the fact that few of the general reform components helped explain that variability suggest that the differences in the effectiveness of CSR are largely due to unmeasured program-specific and school-specific differences in implementation. (Borman et al., 2003, p. 166)

Furthermore, several factors previously presumed to influence effect sizes were not significant:

Our regression analysis suggested that whether a CSR model, in general, requires the following components explains very little in terms of the achievement outcomes the school can expect: (a) ongoing staff professional development, (b) measurable goals and benchmarks for student learning; (c) a faculty vote to increase the likelihood of model acceptance and buy-in; and (d) the use of specific and innovative curricular materials and instructional practices designed to improve teaching and student learning. (Borman et al., 2003, p. 166)

The insignificant impact produced by explicitly requiring ongoing staff development, educational standards, faculty buy-in, and innovative curriculum and instruction suggests that incorporating these four components may not significantly improve outcomes. What explains the lack of effects? One possibility is that among schools where the requirements are in place implementation of these components may be poor. While this may be the case, over \$2 billion have been invested in implementation of CSR since 1991 (Borman et al., 2004; U.S. Department of Education, 2004, 2006). If this massive investment has failed to ensure strong implementation, it seems unlikely that further improvements will be easily achieved.

A second possibility is that, among schools that are not subject to the requirements, teachers are already highly committed, are participating in professional development activities, and are implementing educational standards and innovative curriculum and instruction. This would explain why there is little difference in outcomes between schools that do and do not require these components. Essentially, these components may already be in place in most schools. For example,

¹ Effect sizes calculated from studies using comparison groups are comparable to the effect sizes presented (below) for studies of rapid assessment feedback interventions, all of which involve effect sizes calculated from studies using comparison groups. Borman et al. (2003) also calculated and reported effect sizes for a larger group of studies, including studies where no comparison group was used and effect sizes were calculated from pretest to posttest for the treatment group only. However, this method fails to address threats to internal validity, including maturation and history.

teachers and principals generally feel tremendous pressure to raise student achievement (Pedulla et al., 2003) and arguably must be highly committed to pursue teaching as a career, given the low pay and stressful working conditions. Thus, levels of commitment may not vary significantly between the two groups of schools. Furthermore, teachers routinely participate in professional development activities. For example, 98.3% of all teachers in the nationally representative *Schools and Staffing Survey* indicated that they had participated in some form of teacher professional development within the 12 months prior to the survey and 72.6% participated in “regularly scheduled collaboration with other teachers” (Choy, Chen, Bugarin, & Broughman, 2006, Table 16, p. 49). Thus, participation in professional development activities may not vary significantly between the two groups of schools.

Similarly, the proportion of schools that implement educational standards is unlikely to differ between the two groups because all schools came under strong pressure to implement standards starting with the passage of *Goals 2000* in 1994 (U.S. Department of Education, 1998) and subsequently reinforced by the requirements of the federal *No Child Left Behind Act of 2001* (U.S. Department of Education, 2002a). A study involving a nationally-representative sample of public elementary and secondary schools found that 72% of all principals reported using content standards to a great extent to guide curriculum and instruction in reading and mathematics (Heid & Webber, 1999). Furthermore, “there were no significant differences in the use of content standards between Title I and non-Title I schools or between schoolwide programs and targeted assistance schools or between highest-poverty and low-poverty schools” (Heid & Webber, 1999). A nationally representative survey of teachers in Title I schools found that 79% reported teaching to standards in reading and 66% reported teaching to standards in math (U.S. Department of Education, 2002c).

Finally, the rate at which schools implement innovative curricula may not differ across the two groups of schools because the pressure to raise test scores as a result of the *No Child Left Behind Act* exerts tremendous pressure on teachers and principals to seek innovative curricula. Many schools are rapidly adopting innovative curricula, regardless of any *requirement* to do so. In addition to the 1800 schools across the United States that have adopted *Success for All* (Borman et al., 2004), numerous schools have adopted other innovative curricula. By 2003, over 2.8 million elementary school students used the National Science Foundation (NSF)-funded *Everyday Mathematics Program* (University of Chicago School Mathematics Project, 2003). Students in 5,000 middle schools used NSF-funded math programs (Clayton, 2000), including students in over 2,200 school districts that have adopted the *Connected Math Program* (National Science Foundation, undated), and at least 500 high schools used the NSF-funded *Core-Plus Program* (Clayton, 2000).

A third possibility is that staff development, educational standards, faculty buy-in, and innovative curriculum and instruction—as currently designed and implemented—are simply inadequate for the purpose of improving student achievement. This is perhaps the most straightforward interpretation of Borman et al.’s (2003) results. Variation in these components does not explain variation in outcomes, and the addition of these components is unlikely to improve outcomes. The addition of components designed to involve parents may even have a negative impact: “The one reform attribute that was a statistically significant predictor of effect size suggested that CSR models that require the active involvement of parents and the local community in school governance and improvement activities tend to achieve worse outcomes than models that do not require these activities” (p. 95). This finding is consistent with a previous review of 41 evaluation studies of parental involvement, which found “little empirical support for the widespread claim that parent involvement programs are an effective means of improving student achievement” (Mattingly, Prislín, McKenzie, Rodriguez, & Kayzar, 2002, p. 549).

Another puzzling result is inconsistent with the basic CSR assumption that broad, whole school changes are necessary. The effect sizes for *Success for All* ($d=0.18$) and *Direct Instruction* ($d=0.15$), both of which focus primarily on changing a school’s instructional practices, exceed the

average effect size for all CSR models ($d=0.12$), including models where changes in instruction are only one component of much broader school reforms. This result undermines the argument that broad reforms are more effective than reforms that focus narrowly on changes in instruction and suggests instead that narrowly focused models may be somewhat more effective. At the same time, the small effect sizes for the narrowest models (*Success for All* and *Direct Instruction*) suggest that narrowing the CSR approach to focus on instruction is unlikely to result in sharp improvement in student achievement.

In sum, Borman et al.'s (2003) results cast doubt on the thesis that the meager gains from CSR can be significantly improved by fine-tuning the models. Neither adding nor subtracting components promises to substantially change CSR outcomes. The unavoidable implication is that as currently implemented, CSR models are a flawed strategy for improving student achievement. The small, uneven effects suggest that we cannot reliably expect large improvements from implementing any of the CSR models. The meager returns to the \$2 billion already invested in CSR implementation efforts suggest that any further gains in student achievement will not be easy and will rely on future breakthroughs in overcoming the barriers inherent in implementing large, complex whole-school reforms. For these reasons, it would be unrealistic to expect that improved implementation of existing CSR models will significantly improve student achievement. Instead, these results suggest a need to examine the cost-effectiveness of CSR compared to alternatives such as rapid assessment.

Cost-Effectiveness of CSR

For the purpose of the cost-effectiveness analysis, I drew upon Borman et al.'s (2003) impact estimates for each of the 29 CSR models included in the meta-analysis. Importantly, these estimates were obtained over varying periods of time and thus are not directly comparable across the 29 CSR models. Borman et al. reported information regarding the average duration of implementation for each model, but this information does not correspond to the duration of the constituent research studies in the meta-analysis and cannot be used to annualize the reported effect sizes. Since some of the constituent research studies were conducted over multi-year periods, the reported effect sizes are only upper bound estimates of the annualized effect sizes.

Special attention is given to the three CSR models for which there is reliable impact evidence: *Success for All*, the *School Development Model*, and *Direct Instruction*. Cost information was drawn from three sources (Herman et al., 1999; King, 1994; Odden, 2000). King (1994) conducted a detailed, refereed cost analysis of three CSR models including *Success for All* and the *School Development Model*. A second cost analysis of the same three models lacks detail, was not refereed, and was excluded (Barnett, 1996). Cost information for *Direct Instruction* as well as several other CSR models was adapted from Herman et al. (1999). For comparison purposes, Odden's (2000) lower bound estimate of the costs of the major CSR models was also provided. All costs were adjusted for inflation to the same period (August, 2006) using the consumer price index (CPI).

Ideally, the Elementary/Secondary Price Index would be used to adjust the costs of educational inputs for inflation. However, this index is not available after 1994. The use of the CPI significantly underestimates inflation in the costs of educational inputs but the distortion is less than the distortion when using traditional measures such as the gross domestic product (GDP) price deflator. For example, the costs of elementary and secondary educational inputs increased by 224 percent between 1974 and 1994, while the CPI and the GDP price deflator increased 190 and 173 percent, respectively (U.S. Department of Education, 1997, Table 38). Thus, the CPI is the best available measure of inflation in the costs of educational inputs.

CSR involves significant changes in the way schools are organized and operated, typically including wholesale changes in staffing patterns, curriculum and instruction. Thus, CSR typically requires large increases in staffing and training costs. These changes are reflected in King's (1994) estimates of the annual per-pupil costs of the additional staffing and training required for implementing *Success for All* and the *School Development Model* during the initial year of model implementation. Differences in costs are related to differences in stated objectives: *Success for All* emphasizes changes in instructional practices, including substantial amounts of individual tutoring, while the *School Development Model* aims to foster children's social and emotional development and improvements in school climate. King's figures were adjusted downward to estimate average annual ongoing costs by amortizing initial fixed fees and high initial training costs over the expected life of the programs. In general, unless the developer indicated otherwise, annual ongoing training costs were assumed to be half the year 1 costs. This assumption reflects the need to provide ongoing training of existing staff and to train new staff who are hired during the life of the program, but it may underestimate the true ongoing costs. The expected six year life of the CSR programs was based on data suggesting that nearly one-third of a sample of 395 urban, disadvantaged, low-achieving elementary and middle schools dropped or switched CSR models within a 2-year period (Taylor, 2006).² Thus, initial fixed fees and high initial training costs were amortized over a 6-year period.

King's (1994) data were adjusted by monetizing the value of the additional teacher and principal time that King estimated would be necessary to implement the CSR models, based on average teacher and principal salaries. At an average teacher salary of \$51,880 per year and average of 184 work days per year, teacher time is valued at \$1.76 per hour, per student, assuming a class size of 20 students. Similarly, principal time was valued at \$0.09 per student per hour, assuming an average principal salary of \$75,857, a contract lasting 42 weeks per year, and 500 students per building. Following King, the opportunity costs of increased parent and student time that would be required to implement the CSR models were not monetized because these costs are difficult to value. Therefore, the cost estimates presented here are underestimated by the value of that time.

Herman et al. (1999) provide the best available cost estimates for *Direct Instruction*, based on cost data from a sample of 4 sites using the model as well as cost data from the developer. (Cost data are also available from a more recent source but are based solely on information from each developer's website and do not include release time for peer coaches and staff training, conferences, curricular materials for students, or travel to model schools; American Institutes for Research, 2006).³ The developer estimated that all 25 teachers (in a school of 500 students) would require release time of 9.5 days in the first year and 4.5 days in each following year. I amortized the extra 5 days of release time in the first year over the expected 6-year life of the CSR program. *Direct Instruction* emphasizes scripted pre-planned curricula and, thus, the costs for curriculum materials are proportionately larger compared to *Success for All* or the *School Development Model*. The costs for all three models are summarized in Table 1.

² In a parallel fashion, in every biennium, the national election day in November sees one third of U.S. Senate seats up for election, as their fixed terms are six years long.

³ Sikorski (1990) also conducted a cost-effectiveness analysis of Direct Instruction. However, she calculated the total rather than the incremental increase in costs associated with the intervention. Therefore, her cost figures are not consistent with the incremental cost method used throughout the current paper. In addition, Sikorski only analyzed cost-effectiveness with regard to reading instruction.

Table 1

Annual costs per student to implement Success For All,^a the School Development Model,^a and Direct Instruction

| Component | Success for All ^a | | School Development ^a | | Direct Instruction ^b | |
|-----------------------------|------------------------------|---------------------|---------------------------------|---------------------|---------------------------------|--------------------|
| | Low | High | Low | High | Low | High |
| Extra Staff | \$540.22 | \$1,468.30 | \$138.52 | \$415.56 | \$124.10 | \$372.31 |
| Initial Fee | 3.46 ^c | 3.46 ^c | N/A | N/A | N/A | N/A |
| Training | 94.64 ^d | 176.15 ^d | 146.28 ^d | 355.02 ^d | 161.33 | 161.33 |
| Materials | N/A ^e | N/A ^e | N/A ^e | N/A ^e | 155.13 | 155.13 |
| Teacher time ^f | 162.80 ^g | 472.12 ^g | 97.68 ^g | 586.08 ^g | 75.09 ^h | 75.09 ^h |
| Principal time ⁱ | 3.78 | 7.56 | 3.78 | 11.34 | 0.72 | 0.72 |
| Total | \$804.90 | \$2,127.59 | \$386.26 | \$1,368.00 | \$516.37 | \$764.58 |

All per-pupil estimates based on a school of 500 students. N/A = not applicable or not available.

^aBased on King (1994), Tables 2 and 4. Adjusted for inflation using the March, 1994 price deflator (147.2) and the August, 2006 price deflator (203.9).

^bBased on Herman, R., Aladjem, D., McMahon, P., Masem, E., Mulligan, I., O'Malley, A. S., Quinones, S., Reeve, A., Woodruff, D. (1999). Adjusted for inflation using the January, 1999, price deflator (164.3) and the August, 2006 price deflator (203.9).

^cAmortized over the expected six year life of the CSR program.

^dAssumes that annual training costs in years 2 through 6 are half the costs in year 1, and extra initial costs are amortized over the six year life of the CSR program.

^eKing (1994) focused solely on personnel costs, asserting that nonpersonnel costs are small. Thus, the total costs of *Success for All* and the *School Development Model* are underestimated by the amount of nonpersonnel costs.

^fAssuming annual teacher salary of \$51,880 (U.S. Department of Education, 2005) and an average of 184 work days (37 weeks) per year, the cost of teacher time is \$1.76 per student per hour, adjusted for inflation.

^gKing's (1994) "partial" staffing was assumed to apply to half of the teaching staff.

^hDirect Instruction requires 5 days of teacher release time in year 1 in addition to the annual requirement of 4.5 days; the 5 days were amortized over the expected 6 year life of the CSR program.

ⁱAssuming annual principal salary of \$75,857 (National Center for Education Statistics, 2007) and a contract lasting 42 weeks per year, the cost of principal time is \$0.09 per student per hour, adjusted for inflation.

King's (1994) cost analysis and Borman et al.'s (2003) meta-analysis of effect sizes permit calculations of lower and upper bound effectiveness-cost ratios (effect size divided by the annual cost per student) for *Success for All* (effect size = 0.18 *SD*) and the *School Development Model* (effect size = 0.05 *SD*). Effectiveness-cost ratios for *Direct Instruction* (effect size = 0.15 *SD*) may be calculated using Herman et al.'s (1999) cost figures and Borman et al.'s meta-analysis. The upper bound effectiveness-cost ratios for these three models are, respectively, 0.000224, 0.000129, and 0.000290. (The policy implications of this paper are based on the large differences in effectiveness-cost ratios between CSR and rapid assessment, not the relatively small differences in effectiveness-cost ratios across individual CSR models.)

In addition to the results above regarding the CSR models that Borman et al. (2003) categorized as having the *Strongest Evidence of Effectiveness*, it is useful to establish a tentative upper bound on the cost-effectiveness of the CSR models that Borman et al. (2003) categorized as

having *Highly Promising Evidence of Effectiveness*, based on studies with comparison groups.⁴ The results of this analysis are tentative for three reasons. First, Borman et al. indicated that the number and methodological quality of the research studies regarding those models was not sufficient to draw firm conclusions. In addition, the cost data may not be reliable. Finally, Borman et al.'s effect size estimates were not annualized and therefore are not directly comparable. As the number and methodological quality of CSR research studies increases and annualized effect size estimates become available, it is probable that the effect size estimates for many of these models will decrease and the estimated costs will increase. Thus, the purpose of this analysis is limited to establishing an upper bound estimate of the cost effectiveness of CSR in relation to the cost-effectiveness of rapid assessment, rather than establishing the relative cost-effectiveness among the various CSR models.

The CSR model with the highest effect size in the group categorized as having *Highly Promising Evidence of Effectiveness* is *Expeditionary Learning Outward Bound* (effect size = 0.51 *SD*) (Borman et al., 2003). Based on a sample of 3 sites and information from the developer, the annual inflation-adjusted cost in years 1 and 2 is \$100,522.82, including professional development, teacher release time, and materials, but excluding travel, stipends, and expedition costs (Herman et al., 1999). Annual costs decline 20% in year 3, 36% in year 4, and 48.8% in year 5 (Herman et al.,

⁴ Separate cost-effectiveness analyses suggest that *Reading Assessment* and *Math Assessment* are more cost-effective than the remaining CSR models in the Borman et al. meta-analysis, including those that only offered *Promising Evidence of Effectiveness* and those that required the *Greatest Need for Additional Research*. The model with the highest effect size reported by Borman et al. (2003) is *Integrated Thematic Instruction (ITI)*. The effect size of 0.92 *SD* is based on a single matched group quasi-experimental study that compared students in 1 school that used ITI with 1 school that did not. This unpublished doctoral dissertation compared 19 ITI students with 45 non-ITI students over a two-year period, starting in 3rd grade, with regard to reading achievement, implying an annualized effect size of 0.46 *SD*. Average annual professional development costs over the first 3 years are \$73,000, excluding the value of teacher time for required training workshops (American Institutes for Research, 2006). If 25 teachers participate in a one-week training workshop and their time is valued at \$282 per day, the annual cost is \$35,250. Materials, including a library of professional books that each school is required to purchase, are estimated to cost approximately \$5,000.

Assuming that annual training costs in years 4, 5, and 6 are half the annual costs in years 1, 2, and 3, and if the cost of materials and high initial training costs are amortized over the average 6 year life of a CSR program, the total per pupil cost is \$164.04 per year, or \$116.66 less than the lowest cost estimated by Odden (2000) for all of the CSR models in his cost analysis, after adjusting for inflation (\$280.70). The effectiveness-cost ratio for *Integrated Thematic Instruction* is 0.002804. The CSR model with the second-highest effect size in Borman et al.'s (2003) meta-analysis is the *Paideia* model ($d = 0.57$ *SD*). Based on a sample of 3 sites and information from the developer, the first year cost, adjusted for inflation using the January, 1999, price deflator (164.3) and the August, 2006, price deflator (203.9), is \$181,189.29, including the salary of a facilitator, professional development, teacher release time, and materials (Herman et al., 1999). The annual inflation-adjusted cost in subsequent years includes the cost of a school facilitator (\$62,051.13), teacher release time (\$28,200 for 4 days for 25 teachers), and assessments (\$43,435.79) (Herman et al., 1999). There are additional costs in year 2 for the implementation of coaching (\$55,846.01) (Herman et al., 1999). Amortizing high initial costs in years 1 and 2 over the expected 6-year life of a CSR program, the annual cost per student in a school with 500 students is \$301.82, adjusted for inflation. The effectiveness-cost ratio is 0.001889.

Effect sizes for the remaining CSR models in Borman et al.'s (2003) meta-analysis are smaller than the 0.46 *SD* effect size for *Integrated Thematic Instruction*, and cost data are either not available beyond the initial year of implementation, or suggest annual costs equal to or greater than the cost of *Integrated Thematic Instruction*, implying maximum effectiveness-cost ratios no larger than 0.002804, the ratio for *Integrated Thematic Instruction*. This ratio is smaller than the smallest ratios for *Reading* and *Math Assessment*.

1999) and year 6 costs are assumed to equal year 5 costs. Travel, stipends and expedition costs total \$1,365.12 per teacher per year, or \$34,128 for 25 teachers (Herman et al., 1999). Amortizing high initial costs over the expected 6 year life of a CSR program, the annual cost per student in a school with 500 students is \$217.83, adjusted for inflation. The effectiveness-cost ratio is 0.002341.

The CSR model with the second-highest effect size in this group is *Roots and Wings* (effect size = 0.35 *SD*) (Borman et al., 2003). *Roots and Wings* incorporates the elements of *Success for All* but includes additional components as well (*MathWings* and *WorldLab*). While a detailed cost analysis is not available for *Roots and Wings*, the estimate derived above from King's (1994) data regarding *Success for All* suggests that a lower-bound estimate for the cost of *Roots and Wings* is a minimum of \$804.90 per student per year. Thus, the effectiveness-cost ratio for *Roots and Wings* has a maximum of 0.000435.

The effect size for the remaining CSR model in this group, *Modern Red Schoolhouse*, is 0.17 *SD* (Borman et al., 2003). Information from the developer suggests annual costs for materials ranging from \$10,341 to \$124,102, costs for teacher release time (13 days for each member of a teaching staff of 25 teachers) equal to \$4,581.79, plus other costs averaging \$86,872, adjusted for inflation (Herman et al., 1999). While the effect size is one-third the size of the effect for *Expeditionary Learning Outward Bound*, the costs are not substantially different, implying a maximum effectiveness-cost ratio considerably smaller than 0.002341, the ratio for *Expeditionary Learning Outward Bound*.

Table 2 summarizes the effectiveness-cost ratios for the six models described above. The highest effectiveness-cost ratio is 0.002341, for *Expeditionary Learning Outward Bound*. Therefore, the upper bound effectiveness-cost ratio for the six CSR models with the *Strongest* or *Highly Promising* evidence of effectiveness, according to Borman et al. (2003), is 0.002341. The lowest effectiveness-cost ratio is 0.000036, for King's (1994) upper bound cost estimate and Borman et al.'s 0.05 *SD* effect size estimate for the *School Development Model*.

Table 2

Effectiveness-cost ratios for CSR models with the strongest or highly promising evidence of effectiveness

| Program | Effect Size (<i>SD</i>) ^a | Cost ^b | Effectiveness-Cost Ratio ^c |
|---|--|------------------------|---------------------------------------|
| <i>Success for All</i> | 0.18 | \$ 804.90 ^d | 0.000224 |
| | 0.18 | 2,127.59 ^d | 0.000085 |
| <i>School Development Model</i> | 0.05 | 386.26 ^d | 0.000129 |
| | 0.05 | 1,368.00 ^d | 0.000036 |
| <i>Direct Instruction</i> | 0.15 | 516.37 ^d | 0.000290 |
| | 0.15 | 764.58 ^d | 0.000196 |
| <i>Expeditionary Learning Outward Bound</i> | 0.51 | 217.83 | 0.002341 |
| <i>Roots and Wings</i> | 0.35 | 804.90 | 0.000435 |
| <i>Modern Red Schoolhouse</i> | 0.17 | see text | < 0.002341 |

All cost figures adjusted for inflation.

^aFrom Borman et al. (2003). Note that Borman did not annualize his effect size estimates, i.e., the effect sizes in individual studies of CSR may have been achieved over multi-year periods.

^bAnnual cost per student in dollars.

^cEffect size in *SD* units divided by annual cost per student in dollars.

^dFrom Table 1.

Perhaps the most widely-recognized cost analysis of CSR was conducted by Odden (2000). Based on interviews with CSR developers, Odden estimated the additional ongoing staffing and training costs that would be needed to implement the major CSR models, beyond “core” costs equal to \$1.36 million (adjusted for inflation), that each school would incur if it provided 1 teacher for every 25 students and 1 principal for every 500 students. Odden (2000) estimated that the additional annual cost per student for the major CSR models ranged between \$290.01 and \$900.57, adjusted for inflation.⁵ These models apparently included *Roots and Wings*, which is a more elaborate version of *Success for All*, as well as the *Comer School Development Model*, *ATLAS*, *Expeditionary Learning-Outward Bound*, *Modern Red Schoolhouse*, *Co-NECT*, and others. While Odden did not provide precise estimates linked to each model, his lower-bound cost figure provides a useful benchmark for comparing the cost data from Herman et al. (1999), which in many cases appear to underestimate the true costs of individual CSR models.

Rapid Assessment Impacts

While CSR models have multiple goals, a primary goal is to improve student achievement in math and reading. An alternative to CSR that has largely been overlooked is to provide performance feedback through rapid formative assessments of math and reading performance two to five times weekly. Rapid assessment systems may be defined as systems that provide testing feedback to students and teachers regarding student performance in math and reading two to five times weekly. Positive effects of feedback on student engagement and achievement have been demonstrated in numerous studies dating back to the 1960s. For example, Smith, Brethower, and Cabot (1969) found that having students chart their progress significantly improved motivation and output.

In a second study, Robinson, DePascale, & Roberts (1989) randomly assigned 5th- and 6th-grade students to two groups. Both groups of students worked on identical sets of math problems in the same classroom at the same time with the same teacher. In the first session, neither group received feedback. In the second session, Group 1 received feedback, while Group 2 did not. In the third session, both groups received feedback. In the fourth session, neither group received feedback. The results showed that whenever a group received feedback, students in that group completed more problems with greater accuracy, compared to the baseline condition. Whenever feedback was withdrawn, the completion and accuracy rates dropped. The design of this study virtually rules out any explanation other than the conclusion that feedback caused improved student engagement and achievement. It is difficult to attribute the results of this experiment to individual differences in student characteristics, teacher characteristics, classrooms, or schools. The research design controlled for those differences.

⁵ Adjusted for inflation using the June 1997 price deflator (160.3) and the August 2006 price deflator (203.9). (See U.S. Department of Labor, 2006, for price deflators).

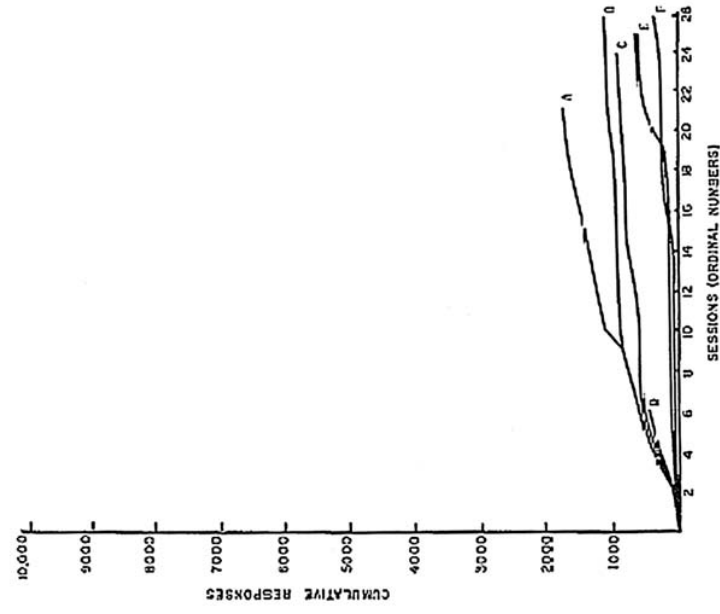


Figure 1. Language arts task output without performance feedback. Adapted from *Journal of Experimental Child Psychology*, (8)1, Smith, D. E. P., Brethower, D., & Cabot, R., Increasing Task Behavior in a Language Arts Program by Providing Reinforcement, p. 48, Copyright 1969, with permission from Elsevier.

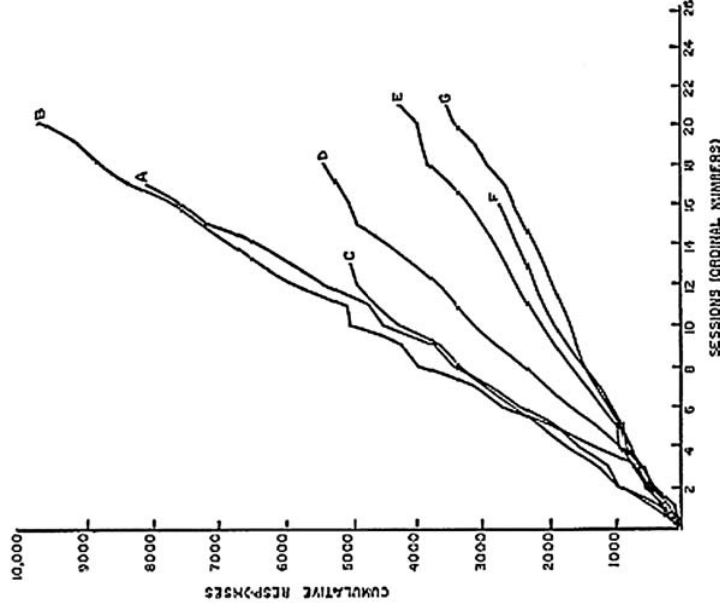


Figure 2. Language arts task output with performance feedback. Adapted from *Journal of Experimental Child Psychology*, (8)1, Smith, D. E. P., Brethower, D., & Cabot, R., Increasing Task Behavior in a Language Arts Program by Providing Reinforcement, p. 58, Copyright 1969, with permission from Elsevier.

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, Kulik, Kulik, & Morgan, 1991; Fuchs & Fuchs, 1986; Kluger & 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 standard deviations (*SD*) (Fuchs & Fuchs, 1986), equivalent to raising the achievement of an average nation such as the United States to the level of the top five nations (Black & Wiliam, 1998). When teachers were required to follow rules about using the assessment information to change instruction for students, the average effect size exceeded 0.9 *SD*, and when students were reinforced with material tokens, in addition to the frequent testing, the average effect size increased even further, exceeding 1.1 *SD* (Fuchs & Fuchs, 1986).

A second meta-analysis of 40 feedback studies (Bangert-Drowns et al., 1991) that included studies involving 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 & DeNisi, 1996). Emotionally neutral (i.e., testing) feedback that 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 lumped together (Kluger & 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 & Timperley, 2007).

These results suggest the nature of effective feedback systems: nonjudgmental, involving frequent testing (two to five times per week), presented immediately after a test. Under these conditions, the three meta-analyses of feedback interventions suggest that the effect size for testing feedback is no lower than 0.7 *SD* (Bangert-Drowns et al., 1991; Fuchs & Fuchs, 1986; Kluger & DeNisi, 1996). However, the meta-analyses generally involved short implementations of rapid assessment (the average duration across all studies in the three meta-analyses was only 3.4 weeks), often with students in special education who may not be representative of the general student population, and the effectiveness of rapid assessment in large-scale field trials may differ.⁶ To avoid these difficulties in generalizing, it is useful to examine the best controlled field trials of two widely implemented variants of rapid assessment whose characteristics match the previously cited characteristics of effective feedback systems. *Reading Assessment* and *Math Assessment* provide immediate testing feedback in reading and math to each student, two to five times per week and have been implemented in classrooms in over 65,000 schools (Northwest Regional Educational Laboratory, 2006), including statewide implementation of *Reading Assessment* in Idaho

⁶ R. Bangert-Drowns (personal communication, June 7, 2006) estimated that the average duration of the 40 studies in his meta-analysis (Bangert-Drowns et al., 1991) was 1.5 to 2 weeks. D. Fuchs (personal communication, June 8, 2006) estimated that the average duration of the 21 studies in his meta-analysis (Fuchs & Fuchs, 1986) was 10–14 weeks. A. Kluger (personal communication, June 13, 2006) calculated that the average duration of the 131 studies in his meta-analysis (Kluger & DeNisi, 1996) was 17.8 days. Using the midpoints of each range, the weighted average duration of the feedback interventions in the three meta-analyses was 23.9 days, or 3.4 weeks. In contrast, G. Borman (personal communication, July 7, 2006) stated that the modal duration of the studies in his meta-analysis of CSR effects (Borman et al., 2003) was one school year (39 weeks).

(Renaissance Learning, 2002).⁷ (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.)

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, n.d.). This selection process helps students to avoid the frustrating experience of choosing a book that is too difficult. After finishing a book, the student completes a computer-based quiz, unique to the book, that is intended to monitor basic reading comprehension (*Rapid Assessment Corporation* has created more than 100,000 quizzes). Similarly, *Math Assessment* is a popular program that provides individualized, printed sets of math problems, a system of assessing student performance on those problems, and a scoring system where students and teachers receive rapid, frequent feedback on student performance upon completion of every set of problems.

Two randomized experiments evaluated the effectiveness of the *Reading Assessment* program (Nunnery, Ross, & McDonald, 2006; Ross, Nunnery, & Goldfeder, 2004). The first experiment involving 1,665 Memphis students (a district where 71 percent of all students are eligible for free/reduced price lunch) found an average effect size of 0.270 *SD* per grade in grades K through 6 on the *STAR Reading* test, over a 9-month school year (Ross et al., 2004). Using HLM, the second experiment involving 978 students (89.9% African American and 83% eligible for free/reduced price lunch) found an average effect size of 0.175 *SD* per grade in grades 3 through 6 on the *STAR Reading* test and the *STAR Early Literacy* test over a 9-month school year (Nunnery et al., 2006). These two estimates suggest upper and lower bound figures for the effect size of *Reading Assessment* with regard to a highly disadvantaged population of students.

The only randomized study of *Math Assessment*, which involved 1,880 students in grades 2 through 8 in 80 classrooms and 7 states, found an effect size of 0.324 *SD* over a 7-month period on the *STAR Math* test, after controlling for treatment integrity (Ysseldyke & Bolt, 2007). The only national, refereed quasi-experimental evaluation of *Math Assessment*, involving 2,202 students in grades 3 through 10 in 125 classrooms and 24 states, found that students in the treatment group gained an average of 0.392 *SD* per grade over one semester (18 weeks) on the *STAR Math* test, compared to students not receiving *Math Assessment* (at pretest the scores of treatment and comparison students were not significantly different) (Ysseldyke & Tardrew, 2007). These two estimates suggest upper and lower bound figures for the effect size of *Math Assessment*.

In studies involving the effects of frequent testing, a question that arises is whether gains that are attributed to the treatment might be an artifact resulting from alignment of the formative tests with the criterion measures. With regard to *Reading Assessment*, the frequent book quizzes aim to assess each student's reading comprehension with regard to individual books that are individually selected by each student from a large library. Thus, the content of the quizzes is aligned with the content of individual books rather than the criterion *STAR Reading* assessment. With regard to *Math Assessment*, students are assigned individualized math problem sets that are tailored and

⁷ Although rapid assessment has been implemented in classrooms in 65,000 schools, it has not been implemented in every classroom in those schools. The results of the only statewide implementation of *Reading Assessment* are difficult to interpret because the evaluation did not include a control group (Renaissance Learning, 2002).

aligned with state standards for mathematics instruction in grades 1 through 7, as well as standards for basic math, pre-algebra, algebra 1, algebra 2, geometry, probability and statistics, pre-calculus, and calculus in the secondary grades. The content of the *STAR Math* criterion test is aligned with national standards but is also aligned with the content of the math problem sets (which are rapidly scored and provide rapid performance feedback) to the extent that the state and national standards overlap.

Costs of Rapid Assessment

Tables 3 and 4 list the cash costs associated with the implementation of Reading Assessment and Math Assessment.⁸ In addition to the core software programs (either Reading Assessment or Math Assessment), it is assumed that a diagnostic assessment (either STAR Reading or STAR Math) is purchased and implemented for each student receiving the rapid assessment intervention. In addition, mark scan devices are purchased for each math classroom. For the purpose of calculating the costs per student, it is assumed that initial one-time costs are averaged over an enrollment of 500 students in 25 classrooms per building. Initial fees, teacher and administrator training, and the cost of the scanners are amortized over 7 years, which is (arbitrarily) assumed to be the life of the program (schools that choose to continue using the programs for a longer period of time would effectively reduce the annual cost). For reading, the costs include access to 100,000 book quizzes for every student. For math, the costs include access to Math Assessment grade level libraries tagged to state standards for grades 1 through 7 and 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 assessment programs are simple to implement; thus, an administrator could instruct each teacher regarding the use of the software. However, the Rapid Assessment Corporation offers full day training sessions costing \$149 per teacher, and the cost analysis assumes that every classroom teacher and one administrator for every 500 students completes a full-day training session for Math Assessment and a full-day training session for Reading Assessment. In addition, the cost analysis assumes a 50% teacher turnover rate during the 7-year implementation period and assumes that each new teacher receives a full-day training session for *Math Assessment* and a full-day training session for *Reading Assessment*.

Implementation requires that each classroom of students has access to one computer and one printer (math problems are printable so that students can work individually without using a computer). Based on a nationally-representative survey, 93% of all instructional classrooms were online by 2003, implying that students in those classrooms had access to at least one classroom computer, and a linear extrapolation of recent trends in online access suggests that 100 percent of classrooms had access to a classroom computer by 2006 (Parsad & Jones, 2005). In addition, researchers note that available computer resources are frequently underutilized (Cuban, 2001). While most classrooms that have a computer also have a printer, the printer may be cheap and unreliable. However, since internet-connected computers can be linked through a local area network (LAN) to

⁸ Costs in Tables 3 and 4 reflect the operating experience of schools in a typical district, audited by the researcher through classroom observation of operating procedures as well as teacher and administrator interviews in 8 schools (spanning elementary, middle, and high school levels). While a school of 500 students taking 2 to 5 assessments per week in math and reading suggests that 2,000 to 5,000 assessments are processed weekly, the burden on teachers is minimal because students scan their own bubble sheets, the software scores each assessment, and summary reports are available to teachers and administrators electronically.

print from any printer in the same building, it is feasible to utilize high-capacity printers in the school's media center (Reilly, n.d.). It is also possible to print to a Xerox machine in the same building if the Xerox is equipped with a LAN card (Reilly, n.d.).

Table 3
Cash costs to implement Reading Assessment

| Item | Fixed Cost (per school) | Annual Variable Cost (per student) | Annual Variable Cost (500 students) | Total Annual Cost (per student) ^a |
|---|----------------------------|--|---|--|
| Reading Assessment | \$1,499.00 | \$4.00 | \$2,000.00 | \$4.43 |
| STAR Reading | 1,499.00 | 0.39 | 195.00 | 0.82 |
| Reading Assessment Training ^b | 5736.50 | — | — | 1.64 |
| Total | \$8,734.50 | 4.39 | \$2,195.00 | \$6.89 |
| After 10% Discount | \$7,861.05 | 3.95 | \$1,975.50 | \$6.20 |

Source: Rapid Assessment Corporation, August 8, 2006

^aAssuming fixed costs are spread over 500 students and averaged over a 7 year implementation period.

^b\$149/full day training X (37.5 teachers + 1 administrator).

Table 4
Cash costs to implement Math Assessment

| Item | Fixed Cost (per school) | Annual Variable Cost (per student) | Annual Variable Cost (500 students) | Total Annual Cost (per student) ^a |
|--|----------------------------|--|---|--|
| Math Assessment | \$1,499.00 | \$4.00 | \$2,000.00 | \$4.43 |
| STAR Math | 1,499.00 | 0.39 | 195.00 | 0.82 |
| AccelScan scanners and cards | 7,875.00 ^c | 8.10 ^d | 4,050.00 | 10.35 |
| Math Assessment Training ^b | 5,736.50 | — | — | 1.64 |
| TOTAL | \$16,609.50 | \$12.49 | \$6,245.00 | \$17.24 |
| After 10% Discount | \$14,948.55 | \$11.24 | \$5,620.50 | \$15.52 |

Source: Rapid Assessment Corporation, August 8, 2006

^aAssuming fixed costs are spread over 500 students and averaged over a 7 year implementation period.

^b\$149/full day training X (37.5 teachers + 1 administrator).

^cAssuming 25 classrooms per school X \$315 per scanner.

^d180 instructional days per 9 month year X 1 mark card @ \$.045.

For these reasons as well as the author's observations of program operations, the cost analysis assumes that it would be feasible to implement the rapid assessment programs without purchasing additional computers or printers. Thus, the cash costs of rapid assessment are primarily the costs listed in Tables 3 and 4. 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.⁹

Since the use of rapid assessment technology does not supplant normal reading and math activities, the opportunity costs of using rapid assessment involve primarily the time required by teachers to monitor students. Through interviews with teachers and administrators and classroom observations, as well as review of program documents, the researcher verified that 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. Thus, to the extent that *Reading* and *Math Assessment* activities do not displace the reading and math activities that may be expected in the absence of rapid assessment, and given that the assessments are self-administered by students and scoring and reporting is handled by computer software, the opportunity costs of implementing the program primarily involve the time required by teachers to ensure that students select and read appropriate books, take the comprehension quiz without assistance from other students, and complete and scan their answers to assigned math problems.

According to teachers who were interviewed, the time saved 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, more than offset the opportunity costs of helping students to select books and monitoring student use of the classroom computer (Yeh, 2006). Thus, the annual cost of rapid assessment, including the opportunity costs of operating the program, remains a total of \$9.45 per student in reading and \$18.89 in math.

Comparisons of Cost-Effectiveness

In principle, the effectiveness-cost ratios for *Reading* and *Math Assessment* may be compared to the corresponding ratios for CSR in order to assess relative cost-effectiveness. However, ratios of cost-effectiveness ratios are sensitive to small changes in denominators and will have relatively large standard errors. With this caveat, it is useful to offer tentative conclusions about the relative cost-effectiveness of *Reading* and *Math Assessment* compared to CSR and various alternative interventions. The purpose is provide policymakers with general guidance about the gross magnitude of differences in cost-effectiveness rather than claims about the precise size of those differences. A conservative approach is to examine the lowest effectiveness-cost ratios for *Reading*

⁹ Large fixed costs (\$7,861.05 in reading and \$14,948.55 in math) incurred at start-up create opportunity costs equal to the income that would otherwise be earned if this amount were instead expended in 7 equal annual installments (the arbitrary lifetime of the program) and the remaining funds were invested in an interest-bearing account. Assuming a real interest rate of 3 percent and a discount rate of 3.5 percent, the foregone income is \$819.55 in reading and \$1,229.04 in math, or \$0.23 per student per year in reading and \$0.35 per student per year in math, in a school with 500 students amortized over the 7-year lifetime of the program.

and *Math Assessment* in relation to the highest effectiveness-cost ratios for CSR as well as two popular alternatives for raising student achievement: class size reduction and high quality preschool.

Table 5
Comparison of effect size, cost, and effectiveness-cost ratios

| Program | Effect Size (<i>SD</i>) ^a | | Cost ^c | Effectiveness-Cost Ratio ^b | |
|--------------------------|--|--------------------|--|---------------------------------------|----------|
| | Reading | Math | | Reading | Math |
| Rapid Assessment | | | | | |
| (high estimates) | 0.270 ^d | 0.392 ^e | \$ 9.45 ^f 18.89 ^g | 0.028571 | 0.020752 |
| (low estimates) | 0.175 ^h | 0.324 ⁱ | 9.45 ^f 18.89 ^g | 0.018519 | 0.017152 |
| Class Size Reduction | | | | | |
| Nye et al. (2001) | 0.104 ^j | 0.090 ^j | 1,379.28 ^k | 0.000075 | 0.000065 |
| Finn et al. (2001) | 0.120 ^l | 0.129 ^l | 1,379.28 ^k | 0.000087 | 0.000094 |
| Perry Preschool | 0.150 ^m | 0.155 ^m | 12,147.03 ⁿ | 0.000012 | 0.000013 |
| Abecedarian Preschool | 0.150 ^o | 0.054 ^o | \$ 10,188.09 ^p | 0.000015 | 0.000005 |

^aAnnualized effect size.

^bEffect size in *SD* units divided by annual cost per student.

^cAnnual cost per student, adjusted for inflation.

^dFrom Ross et al. (2004)

^eFrom Nunnery et al. (2006)

^fReading.

^gMath.

^hFrom Ysseldyke and Tardrew (2007).

ⁱFrom Ysseldyke and Bolt (2007).

^jNye, Hedges, and Konstantopoulos (2001). Average of effect sizes in grades 1, 2, and 3.

^kFrom 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 deflator (161.2) and the August, 2006, price deflator (203.9).

^lFinn, 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.

^mFrom Schweinhart, Barnes, & Weikart (1993), Table 13, annualized over 2 year period.

ⁿFrom Barnett (1992), adjusted for inflation using the January, 1985, price deflator (105.5) and the August, 2006, price deflator (203.9).

^oFrom Ramey et al. (2000), Figure 3, annualized over 5 year period.

^pFrom 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 deflator (177.1) and the August, 2006, price deflator (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).

The lowest effectiveness-cost ratio for *Reading Assessment* is 8 times larger than the highest effectiveness-cost ratio for the most cost-effective CSR model in Borman et al.'s (2003) grouping of

models demonstrating *Strongest Evidence* or *Highly Promising Evidence* of effectiveness (Tables 2 and 5). The lowest effectiveness-cost ratio for *Math Assessment* is 7 times larger than the highest effectiveness-cost ratio for the most cost-effective CSR model in this group. The results suggest that *Reading* and *Math Assessment* are roughly an order of magnitude more effective per dollar compared to the most cost-effective CSR model in this group.¹⁰ For comparison purposes, Table 5 lists annualized effect sizes, the annual cost per pupil, and effectiveness-cost ratios for class size reduction, Perry preschool, and Abecedarian preschool.

Perhaps the most optimistic assessment of class size reduction employed hierarchical linear modeling, controlled for an array of covariates, and isolated the impact of class size reduction according to duration of exposure (Finn, Gerber, Achilles, & Boyd-Zaharias, 2001). After 1, 2, and 3 years of exposure to class size reduction, the achievement of students randomly assigned to classrooms of 13–17 students was higher than the achievement of students randomly assigned to classrooms with 22–26 students, with annualized effect sizes equal to 0.120 *SD* in reading and 0.129 *SD* in math per year at the end of 2nd grade. 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 of 0.000087 in reading and 0.000094 in math. The lowest effectiveness-cost ratio for *Reading Assessment* is 213 times the highest effectiveness-cost ratio for class size reduction. The lowest effectiveness-cost ratio for *Math Assessment* is 182 times the highest effectiveness-cost ratio for class size reduction. These results suggest that *Reading* and *Math Assessment* are approximately two orders of magnitude more cost-effective than class size reduction.

The advantage of rapid assessment compared to preschool is even stronger than the advantage compared to class size reduction. 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 2nd grade (Schweinhart, Barnes, & Weikart, 1993), but the annualized cost was \$12,147.03, resulting in small effectiveness-cost ratios of 0.000012 in reading and 0.000013 in math. Furthermore, it is not clear that the children participating in the treatment outperformed children in the control group with regard to student achievement. After correcting for family-wise error, only two of a total of 24 tests reached statistical significance, suggesting that the overwhelming majority of statistical tests indicated that there was no significant difference in the achievement of children who participated in Perry preschool compared to children who did not participate. Results for participants in Abecedarian preschool were roughly comparable to the results for participants in Perry preschool. Annualized effect sizes for 3rd grade children were 0.150 in reading and 0.054 in math over a 5-year implementation period, at an annual cost of \$10,188.09, adjusted for inflation and also 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.

The lowest effectiveness-cost ratio for *Reading Assessment* is 1,235 times the highest effectiveness-cost ratio for high quality preschool. The lowest effectiveness-cost ratio for *Math Assessment* is 1,319 times the highest effectiveness-cost ratio for high quality preschool. These results suggest that *Reading* and *Math Assessment* are approximately three orders of magnitude more cost-effective than high quality preschool. As noted above, ratios of cost-effectiveness ratios are sensitive to small changes in denominators and will have relatively large standard errors. The analysis does suggest, however, that *Reading* and *Math Assessment* are significantly more cost-effective than CSR, class size reduction, or high quality preschool. The difference in cost-effectiveness is approximately one order of magnitude compared to CSR, two orders of magnitude compared to class size reduction, and three orders of magnitude compared to high quality preschool.

¹⁰ In effect, this comparison is between the point estimates that most underestimate the relative advantage of rapid assessment.

Sensitivity Analysis I

A key assumption is that schools do not need to purchase additional computers and printers in order to implement rapid assessment. This assumption is based on research implying that virtually all classrooms had access to at least one online computer by 2006 (Parsad & Jones, 2005), plus the ability of all online computers to print from a high capacity printer in a school's media center (Reilly, n.d.), and was verified through interviews with teachers and observations of classrooms where rapid assessment was used. However, if each classroom requires new equipment, an entire system, including a computer, monitor, keyboard, mouse, software, service plan and a high capacity laser printer, may be purchased from Dell for \$1,015. Assuming that a complete system is purchased for every classroom of 20 students and is amortized over a 7-year period, the annual cost per student is \$7.25. Splitting this cost between reading and math raises the total cost of *Reading Assessment* from \$9.45 to \$13.08 per student. Using the low effect size estimate of 0.175 *SD*, the effectiveness-cost ratio falls to 0.013379 but *Reading Assessment* remains a minimum of 6 times as cost-effective as CSR, 154 times as cost-effective as class size reduction, and 892 times as cost-effective as high quality preschool. The total cost of *Math Assessment* rises from \$18.89 to \$22.52. Using the 0.324 *SD* effect size estimate, the effectiveness-cost ratio falls to 0.014387 but *Math Assessment* remains a minimum of 6 times as cost-effective as CSR, 153 times as cost-effective as class size reduction, and 1107 times as cost-effective as high quality preschool.

Sensitivity Analysis II

A second assumption, based on interviews with teachers and observations of classrooms where rapid assessment is used, is that the rapid assessment software saves more teacher time (primarily time that would otherwise be spent grading math homework and assessing reading comprehension) than is consumed in noninstructional tasks such as supervising student use of the computer, scanner and printer. However, if teachers do not save time and instead lose 15 minutes per day (or 1.25 hours per week), the annual cost is \$81.06 per student, assuming 20 students per teacher and an annual inflation-adjusted teacher salary of \$51,880 (U.S. Department of Education, 2005). Splitting this cost between reading and math raises the total cost of *Reading Assessment* from \$9.45 to \$49.98 per student. Using the 0.175 *SD* effect size estimate, the effectiveness-cost ratio falls to 0.003501 but *Reading Assessment* remains a minimum of 1.5 times as cost-effective as CSR, 40 times as cost-effective as class size reduction, and 233 times as cost-effective as high quality preschool. The total cost of *Math Assessment* rises from \$18.89 to \$59.42. Using the 0.324 *SD* effect size estimate, the effectiveness-cost ratio falls to 0.005453, but *Math Assessment* remains a minimum of 2.3 times as cost-effective as CSR, 58 times as cost-effective as class size reduction, and 419 times as cost-effective as high quality preschool.

Discussion

The results indicate that rapid assessment represents a much more cost-effective approach than CSR, class size reduction, or high quality preschool. The true advantage for rapid assessment is likely to be substantially larger than indicated by the point-estimate ratios in Tables 2 and 5. Given the unreliability of CSR effect size and cost estimates for models other than *Success for All*, the *School Development Model*, and *Direct Instruction*, the true upper bound for the CSR effectiveness-cost ratios is likely to be closer to the maximum ratio calculated for the three models for which there

is reliable evidence. 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 & Bartlett, 2002).

This inefficiency may also indicate fundamental problems with CSR. A basic assumption underlying CSR is that comprehensive, whole school reforms that “break the mold” to create radically different schools are necessary to achieve significant gains in student achievement. However, of the models for which there is strong evidence of effectiveness, the most cost-effective (*Direct Instruction*) is the most traditional, prescribing conventional teaching methods and reinforcing what Tyack and Cuban (1995) call the fundamental grammar of schooling. This suggests that the basic assumption underlying CSR may not be correct.

A second assumption underlying CSR is that improvements in school culture and improved teaching are adequate to improve student achievement. CSR is not designed to directly influence student engagement, although several models stress the importance of establishing respect and trust and a positive school culture may be expected to foster positive attitudes. However, without deeper insight into factors that build intrinsic interest in academic achievement, it may be unrealistic to expect that CSR will significantly improve student engagement in academic work. Instead, CSR is typically designed to transform the school structure that supports teachers, providing a positive, collegial environment where teachers work together to solve instructional problems. In some cases, such as with *Direct Instruction* and *Success for All*, explicit guidance is provided regarding instructional strategies. In other cases, explicit guidance is provided regarding a focus on academics, such as with the *Modern Red Schoolhouse*. While a positive environment and good teaching may indirectly serve to engage students, none of the CSR models offer deep insight into factors that build intrinsic interest in academic achievement, and none of the models specify how schools may be structured to foster student engagement.

CSR is not designed to address low student engagement, yet lack of engagement is epidemic in the public schools. Data from the Education Longitudinal Survey (Ingels et al., 2005), which is a nationally-representative survey of 10th-graders, indicated that only 24% liked school a great deal—65% percent reported that they liked school “somewhat,” and 12% said they “did not like it at all”—suggesting that by their 10th grade year, the vast majority of students were, at best, lukewarm about school. An even larger majority—81%—indicated that “the teaching is good”—suggesting that poor teaching is not an adequate explanation for low achievement and improved teaching is unlikely to produce dramatic improvements in student achievement. To the extent that the basic reason for low student achievement is low engagement, current CSR models are not adequate to address this challenge.

What explains the dramatic differences in the cost-effectiveness of rapid assessment in comparison with CSR and class size reduction? If the basic reason for low student achievement is low engagement in academic work, and if performance feedback serves to engage students, then an intervention such as rapid assessment may be more precise and effective than broad interventions such as CSR or class size reduction.

Existing research suggests that performance feedback engages students by reinforcing student self-efficacy. A student’s perceived control over his or her academic performance is strongly predictive of academic achievement (Brookover, Beady, Flood, Schweitzer, & Wisenbaker, 1979; Brookover et al., 1978; Coleman et al., 1966; Crandall, Katkovky, & Crandall, 1965; Kalechstein & Nowicki, 1997; Keith, Pottebaum, & Eberhart, 1986; Skinner, Wellborn, & Connell, 1990; Teddlie &

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, & Schmitz, 2002; Ross & Broh, 2000; Skinner et al., 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 et al., 1990). To the extent that rapid assessment enables teachers to provide individualized instruction, keeping each student in his or her own zone of proximal development, students are more likely to be successful and feel that they can control their performance. Feelings of control reinforce effort, which improves achievement, which reinforces feelings of control and engagement.

CSR focuses on creating a school environment that supports teachers and effective teaching, yet it neglects to provide a system of rapid assessment and individualized curricula where every book that is read and every set of math problems that is accurately completed is quickly acknowledged through objective feedback to students. Thus, students in CSR schools typically do not have the type of feedback system that research suggests is effective in promoting engagement (Robinson et al., 1989). This lack of feedback may explain why CSR is far less effective than rapid assessment. A similar analysis suggests why class size reduction is poorly suited to the task of engaging students in academic work: reducing class size is, at best, a weak strategy to improve performance feedback.

Rapid assessment could easily be integrated into CSR. The research findings presented here suggest that this may be an effective way of improving CSR outcomes. On the other hand, rapid assessment could also be implemented as a stand-alone intervention and might improve outcomes a minimum of 7 times as efficiently as CSR. It seems likely that stand-alone implementation could be accomplished nationwide much more quickly than complex whole-school reforms such as CSR, and at a fraction of the cost.

Similarly, high quality preschool is an enormously complex, expensive intervention, requiring highly skilled early childhood educators. In contrast, rapid assessment is relatively simple to implement and could be implemented in existing classrooms, with existing teachers. The results of the studies in Memphis suggest that it can be effective with economically disadvantaged, minority students, while the results of the cost-effectiveness analysis comparing rapid assessment and high quality preschool suggest that rapid assessment is dramatically more cost-effective. Funding for rapid assessment is likely to be a more productive use of scarce resources, compared to CSR, class size reduction, or high quality preschool.

References

- American Institutes for Research. (2006). *CSRQ Center report on elementary school comprehensive school reform models*. Washington, DC: American Institutes for Research.
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213–238.
- Barnett, W. S. (1992). Benefits of compensatory preschool education. *The Journal of Human Resources*, 27(2), 279–312.

- Barnett, W. S. (1996). Economics of school reform: Three promising models. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 299–326). Washington, DC: The Brookings Institution.
- Barnett, W. S., & Masse, L. N. (2007). Comparative benefit–cost analysis of the Abecedarian program and its policy implications. *Economics of Education Review*, 26, 113–125.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.
- Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125–230.
- Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2004). Comprehensive school reform and achievement: A meta-analysis. In C. T. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 53–108). Washington, DC: The National Clearinghouse for Comprehensive School Reform.
- Brookover, W. B., Beady, C. H., Flood, P. K., Schweitzer, J. G., & Wisenbaker, J. M. (1979). *Schools, social systems and student achievement: Schools can make a difference*. New York: Praeger.
- Brookover, W. B., Schweitzer, J. G., Schneider, J. M., Beady, C. H., Flood, P. K., & Wisenbaker, J. M. (1978). Elementary school social climate and school achievement. *American Educational Research Journal*, 15, 301–318.
- Choy, S. P., Chen, X., Bugarin, R., & Broughman, S. P. (2006). *Teacher professional development in 1999–2000: What teachers, principals, and district staff report*. Retrieved July 5, 2006
- Clayton, M. (2000). How a new math program rose to the top, *Christian Science Monitor* (May 23 ed.).
- Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, R., et al. (1966). *Equality of educational opportunity*. Washington, D. C.: Government Printing Office.
- Crandall, V., Katkovky, W., & Crandall, U. (1965). Children's beliefs in their own control of reinforcements in intellectual academic achievement situations. *Child Development*, 36, 91–109.
- Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.
- Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (2001). The enduring effects of small classes. *Teachers College Record*, 103(2), 145–183.
- Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53(3), 199–208.

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Heid, C., & Webber, A. (1999). *School-level implementation of standards-based reform: Findings from the Public School Survey on Education Reform*. Retrieved October 1, 2006, from <http://www.ed.gov/pubs/SchoolSurvey/title.html>.
- 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.
- Ingels, S. J., Burns, L. J., Charleston, S., Chen, X., Cataldi, E. F., & Owings, J. A. (2005). *A Profile of the American high school sophomore in 2002: Initial results from the base year of the Education Longitudinal Study of 2002 (NCES 2005-338)*. Retrieved August 22, 2006, from <http://nces.ed.gov/pubs2005/2005338.pdf>.
- Kalechstein, A. D., & Nowicki, S., 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(1), 27–56.
- Kearns, D., & Anderson, J. (1996). Sharing the vision: Creating New American Schools. In S. Stringfield, S. Ross & L. Smith (Eds.), *Bold plans for school restructuring* (pp. 9–23). Mahwah, NJ: Erlbaum.
- Keith, T. Z., Pottebaum, S. M., & Eberhart, S. (1986). Effects of self concept and locus of control on academic achievement: A large sample path analysis. *Journal of Psychoeducational Assessment*, 4(1), 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), 1–19.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.
- Little, J., & Bartlett, L. (2002). Career and commitment in the context of comprehensive school reform. *Teachers and teaching: Theory and practice*, 8(3-4), 345–354.
- Mattingly, D., Prislun, R., McKenzie, T., Rodriguez, J., & Kayzar, B. (2002). Evaluating evaluations: The case of parent involvement programs. *Review of Educational Research*, 72(4), 549–576.
- Musher Eizenman, D. R., Nesselroade, J. R., & Schmitz, B. (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(6), 540–547.
- National Center for Education Statistics. (2007). *The condition of education: Table 34-2. Number and percentage distribution of school principals, by school level, school type, and selected professional characteristics: School years 1993–94, 1999–2000, and 2003–04*. Retrieved

- September 9, 2007, from
<http://nces.ed.gov/programs/coe/2007/section4/table.asp?tableID=726>.
- National Science Foundation. (undated). *Education: Lessons about learning*. Retrieved September 6, 2006, from <http://www.nsf.gov/about/history/nsf0050/pdf/education.pdf>.
- Northwest Regional Educational Laboratory. (1998). *Catalog of school reform models: First edition*. Portland, OR: Author. (ERIC Document Reproduction Service No. ED 439 510.)
- Northwest Regional Educational Laboratory. (2005). *Catalog of school reform models*. Portland, OR: Author. Retrieved July 5, 2006, from <http://www.nwrel.org/scpd/catalog/index.html>.
- Northwest Regional Educational Laboratory. (2006, October 4). *The catalog of school reform models*. Portland, OR: Author. Retrieved March 5, 2007, from <http://www.nwrel.org/scpd/catalog/ModelDetails.asp?ModelID=48>.
- Nunnery, J. A., Ross, S. M., & McDonald, A. (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), 1–18.
- Nye, B., Hedges, L. V., & Konstantopoulos, S. (2001). Are effects of small classes cumulative? Evidence from a Tennessee experiment. *Journal of Educational Research*, 94(6), 336–345.
- Odden, A. (2000). The costs of sustaining educational change through comprehensive school reform. *Phi Delta Kappan*, 81(6), 433–438.
- Parsad, B., & Jones, J. (2005). *Internet access in U.S. public schools and classrooms: 1994–2003 (NCES 2005-015)*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy, Lynch School of Education, Boston College.
- Ramey, C. T., Campbell, F. A., Burchinal, M., Skinner, M. L., Gardner, D. M., & Ramey, S. L. (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.
- Reilly, R. (2002, October). There is a printer and a Xerox machine in your classroom that you can't see! *Teachers.net Gazette*, 3(10). Retrieved August 5, 2006, from <http://www.teachers.net/gazette/OCT02/reilly.html>.

- Renaissance Learning. (2002). *Results from a three-year state-wide implementation of Reading Renaissance in Idaho*. Madison, WI: Renaissance Learning.
- Renaissance Learning. (n.d.). *STAR Reading: Understanding reliability and validity*. Wisconsin Rapids, WI: Renaissance Learning.
- Robinson, S. L., DePascale, C., & Roberts, F. C. (1989). Computer delivered feedback in group based instruction: Effects for learning disabled students in mathematics. *Learning Disabilities Focus*, 5(1), 28–35.
- Ross, C. E., & Broh, B. A. (2000). The roles of self esteem and the sense of personal control in the academic achievement process. *Sociology of Education*, 73(4), 270–284.
- Ross, S. M., & Gil, L. (2004). The past and future of comprehensive school reform: Perspectives from a researcher and practitioner. In C. T. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 151–174). Washington, DC: National Clearinghouse for Comprehensive School Reform.
- Ross, S. M., Nunnery, J., & Goldfeder, E. (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, University of Memphis.
- Rowan, B., Barnes, C., & Camburn, E. (2004). Benefiting from comprehensive school reform: A review of research on CSR implementation. In C. T. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 1–52). Washington, DC: National Clearinghouse for Comprehensive School Reform.
- Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). *Significant benefits: The High/Scope Perry Preschool Study through age 27*. Ypsilanti, MI: High/Scope Press.
- Sikorski, M. F. (1990). *Cost-effectiveness analysis of mastery learning and direct instruction in reading*. Unpublished doctoral dissertation, Northern Illinois University, Dekalb, IL.
- Skinner, E. A., Wellborn, J. G., & Connell, J. P. (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(1), 22–32.
- Slavin, R. E., & Fashola, O. S. (1998). *Show me the evidence!* Thousand Oaks, CA: Corwin Press.
- Smith, D. E. P., Brethower, D., & Cabot, R. (1969). Increasing task behavior in a language arts program by providing reinforcement. *Journal of Experimental Child Psychology*, 8(1), 45–62.
- Taylor, J. E. (2006). The struggle to survive: Examining the sustainability of schools' comprehensive school reform efforts. *Journal of Education for Students Placed At Risk*, 11(3&4), 331–352.
- Teddlie, C., & Stringfield, S. (1993). *Schools do make a difference: Lessons learned from a 10 year study of school effects*. New York: Teachers College Press.

- Traub, J. (1999). *Better by design? A consumer's guide to schoolwide reform*. Washington, DC: Thomas B. Fordham Foundation.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
- U.S. Department of Education. (1997). *Digest of education statistics* (NCES 98-015). Washington, DC: National Center for Education Statistics.
- U.S. Department of Education. (1998, April 30). *Goals 2000: Reforming education to improve student achievement*. Washington, DC: Author. Retrieved July 6, 2006, from <http://www.ed.gov/PDFDocs/g2kfinal.pdf>.
- U.S. Department of Education. (2002a). *Public Law 107-110, 107th Congress: An Act to close the achievement gap with accountability, flexibility, and choice, so that no child is left behind*. Washington, DC: Author. Retrieved October 1, 2007, from <http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf>.
- U.S. Department of Education. (2002b). *Scientifically based research and the Comprehensive School Reform (CSR) Program*. Washington, DC: Author. Retrieved June 30, 2006, from <http://www.ed.gov/programs/compreform/legislation.html>.
- U.S. Department of Education. (2002c). *Teacher professional development in Title I schools: Recent evidence from the National Longitudinal Survey of Schools*. Washington, DC: Author. Retrieved October 1, 2006, from http://www.ed.gov/offices/OUS/PES/professional_develop03.pdf.
- U.S. Department of Education. (2004). *Implementation and early outcomes of the Comprehensive School Reform Demonstration (CSR/D) Program* (No. 2004-15). Jessup, MD: Policy and Program Studies Service, U.S. Department of Education.
- U.S. Department of Education. (2005). *Digest of education statistics*. Washington, DC: Author. Retrieved December 1, 2006, from http://nces.ed.gov/programs/digest/d05/tables/dt05_075.asp.
- U.S. Department of Education. (2006, March 6). *Comprehensive school reform program: Funding status*. Washington, DC: Author. Retrieved July 6, 2006, from <http://www.ed.gov/programs/compreform/funding.html>.
- U.S. Department of Labor. (2006, September 15). *Consumer price index*. Washington, DC: Author. Retrieved October 5, 2006, from <ftp://ftp.bls.gov/pub/special.requests/cpi/cpi.txt>.
- University of Chicago School Mathematics Project. (2003). *Welcome to the UCSMP Everyday Mathematics Center*. Chicago: Author. Retrieved September 6, 2006, from <http://everydaymath.uchicago.edu/>.

- Wang, M. C., Haertel, G. D., & Walberg, H. (1997). *What do we know? Widely implemented school improvement programs*. Philadelphia: Temple University Center for Research in Human Development and Education.
- Wong, K., & Meyer, S. (2001). Title I schoolwide programs as an alternative to categorical practices: An organizational analysis of surveys from the Prospects study. In G. D. Borman, S. C. Stringfield & R. E. Slavin (Eds.), *Title I: Compensatory education at the crossroads* (pp. 195–234). Mahwah, NJ: Erlbaum.
- Wong, K. K., & Meyer, S. J. (1998). Title I schoolwide programs: A synthesis of findings from recent evaluation. *Educational Evaluation and Policy Analysis*, 20(2), 115–136.
- Yeh, S. S. (2006). High stakes testing: Can rapid assessment reduce the pressure? *Teachers College Record*, 108(4), 621–661.
- Yeh, S. S. (2007). The cost-effectiveness of five policies for improving student achievement. *American Journal of Evaluation*, 28(4), 416–436.
- Ysseldyke, J., & Bolt, D. M. (2007). Effect of technology-enhanced continuous progress monitoring on math achievement. *School Psychology Review*, 36(3), 453–467.
- Ysseldyke, J., & Tardrew, S. (2007). Use of a progress-monitoring system to enable teachers to differentiate math instruction. *Journal of Applied School Psychology*, 24(1), 1–28.

About the Author

Stuart S. Yeh

University of Minnesota

Email: yehxx008@umn.edu

Stuart Yeh is Associate Professor and Coordinator of the Evaluation Studies Program in the Department of Educational Policy and Administration at the University of Minnesota. He recently completed cost effectiveness analyses of 10 approaches for raising student achievement, including voucher programs, charter schools, a 10 percent increase in existing patterns of educational expenditures, increased educational accountability, comprehensive school reform, teacher certification by the National Board for Professional Teaching Standards, a proposal to replace the bottom quartile of novice teachers with new teachers, class size reduction, raising teacher quality by requiring new candidates to meet a minimum SAT score of 1000, and rapid assessment, which is an intervention that provides rapid performance feedback to students and teachers.

EDUCATION POLICY ANALYSIS ARCHIVES <http://epaa.asu.edu>

Editor: Sherman Dorn, University of South Florida

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Sherman Dorn, epaa-editor@shermamdorn.com.

Editorial Board

| | |
|-----------------------|------------------------|
| Noga Admon | Jessica Allen |
| Cheryl Aman | Michael W. Apple |
| David C. Berliner | Damian Betebenner |
| Robert Bickel | Robert Bifulco |
| Anne Black | Henry Braun |
| Nick Burbules | Marisa Cannata |
| Casey Cobb | Arnold Danzig |
| Linda Darling-Hammond | Chad d'Entremont |
| John Diamond | Amy Garrett Dikkers |
| Tara Donohue | Gunapala Edirisooriya |
| Camille Farrington | Gustavo Fischman |
| Chris Frey | Richard Garlikov |
| Misty Ginicola | Gene V Glass |
| Harvey Goldstein | Jake Gross |
| Hee Kyung Hong | Aimee Howley |
| Craig B. Howley | William Hunter |
| Jaekyung Lee | Benjamin Levin |
| Jennifer Lloyd | Sarah Lubienski |
| Les McLean | Roslyn Arlin Mickelson |
| Heinrich Mintrop | Shereeza Mohammed |
| Michele Moses | Sharon L. Nichols |
| Sean Reardon | A.G. Rud |
| Ben Superfine | Cally Waite |
| John Weathers | Kevin Welner |
| Ed Wiley | Terrence G. Wiley |
| Kyo Yamashiro | Stuart Yeh |

EDUCATION POLICY ANALYSIS ARCHIVES <http://epaa.asu.edu>

**New Scholar Board
English Language Articles
2007–2009**

| | |
|---------------------------|------------------------|
| Wendy Chi | Corinna Crane |
| Jenny DeMonte | Craig Esposito |
| Timothy Ford | Samara Foster |
| Melissa L. Freeman | Kimberly Howard |
| Nils Kauffman | Felicia Sanders |
| Kenzo Sung | Tina Trujillo |
| Larisa Warhol | |

Archivos Analíticos de Políticas Educativas <http://epaa.asu.edu>

Editor

Gustavo E. Fischman

Arizona State University

Asistentes editoriales: Rafael O. Serrano (ASU) & Lucia Terra (UBC)

Hugo Aboites

UAM-Xochimilco, México

Claudio Almonacid Avila

UMCE, Chile

Alejandra Birgin

FLACSO-UBA, Argentina

Mariano Fernández Enguita

Universidad de Salamanca. España

Roberto Leher

UFRJ, Brasil

Pia Lindquist Wong

CSUS, USA

Alma Maldonado

University of Arizona, USA

Imanol Ordorika

IIE-UNAM, México

Miguel A. Pereyra

Universidad de Granada, España

Romualdo Portella de Oliveira

Universidade de São Paulo, Brasil

José Ignacio Rivas Flores

Universidad de Málaga, España

José Gimeno Sacristán

Universidad de Valencia, España

Susan Street

CIESAS Occidente, México

Daniel Suárez

LPP-UBA, Argentina

Jurjo Torres Santomé

Universidad de la Coruña, España

Armando Alcántara Santuario

CESU, México

Dalila Andrade de Oliveira

UFMG, Brasil

Sigfredo Chiroque

IPP, Perú

Gaudêncio Frigotto

UERJ, Brasil

Nilma Lino Gomes

UFMG, Brasil

María Loreto Egaña

PIIE, Chile

José Felipe Martínez Fernández

UCLA, USA

Vanilda Paiva

UERJ, Brasil

Mónica Pini

UNSAM, Argentina

Paula Razquin

UNESCO, Francia

Diana Rhoten

SSRC, USA

Daniel Schugurensky

UT-OISE Canadá

Nelly P. Stromquist

USC, USA

Antonio Teodoro

Universidade Lusófona, Lisboa

Lílian do Valle

UERJ, Brasil