Running Head: SUCCESS FOR ALL

The Long-Term Effects and Cost-Effectiveness of Success for All

Geoffrey D. Borman

University of Wisconsin-Madison

Gina M. Hewes

Johns Hopkins University

Abstract

Several renowned early interventions have compelling evidence of enduring achievement effects for at-risk children: Perry Preschool; the Abecedarian Project; and the Tennessee class-size experiment. The costs and potential for national dissemination of such model programs, though, represent key practical concerns. This study examines the long-term outcomes and costs of another popular early intervention: Success for All. Relative to controls, Success for All students completed eighth grade at a younger age, with better achievement outcomes, fewer special education placements, fewer retentions, and at the same educational expense. Further cost-effectiveness comparisons to the three prominent interventions suggest that Success for All is deserving of similar recognition as a sound educational investment that provides strong and lasting educational benefits. None of these exemplary programs, though, can be expected to be the "great equalizer."

Few widespread American policies for improving educational equality and productivity have convincing evidence of long-lasting benefits to students and society. This has been a common criticism of the nation's large-scale compensatory education initiatives, Head Start (Lee, Brooks-Gunn, & Schnur, 1988; McKey et al., 1985) and Title I (Borman & D'Agostino, 2001; Carter, 1984), and of many other early childhood interventions (Barnett, 1992; White, 1985). In some cases, the programs have lacked the intensity and quality to make a difference in the long-term. In many other cases, though, poor research designs and the difficult proposition of tracking samples of children across 10 to 15 years, or longer, has stood in the way of establishing lasting benefits. Barnett (1995) noted two model projects that have provided particularly compelling data showing long-term positive effects on outcomes, such as achievement, grade retentions, special education placements, high school graduation, and socialization: the Abecedarian Project and Perry Preschool studies. More recently, a longitudinal followup to another well-designed study, the Tennessee Student/Teacher Achievement Ratio (STAR) experiment, provided support for the long-term efficacy of class-size reductions in the early elementary school years (Finn, Gerber, Achilles, & Boyd-Zaharias, 2001; Krueger & Whitmore, 2001; Nye, Hedges, & Konstantopolous, 2000). Studies of these three model interventions provide important evidence of the sustained effects that early educational policies and programs can have on students' academic outcomes through middle school and, in some cases, into adulthood. Moreover, they inspire strong hope for the American ideal for education, as expressed by Horace Mann, to be the "great equalizer," or "balance wheel of the social machinery."

From a more mundane perspective, policymakers and practitioners also are interested in the feasibility and costs associated with replication of these programs and their outcomes. The level of funding and quality of services provided by the large-scale public preschool programs

generally do not match those found for the carefully designed, model programs that yielded the long-term effects. Therefore, a frequently voiced concern about Head Start and other state-funded programs that have been inspired by the Perry Preschool and Abecedarian efforts is whether they can produce in a routine manner the same positive outcomes (Gomby, Larner, Stevenson, Lewit, & Behrman, 1995). Similarly, the implementations and outcomes of class-size reduction initiatives that have emerged in response to the STAR findings, including the prominent $1.5 billion per year California effort, have not necessarily matched the results of the smaller and highly controlled Tennessee experiment (Bohrnstedt & Stecher, 1999). These sobering outcomes raise questions regarding both financing and replicating these exemplary programs.

In this article, we present the results from the first analysis of the long-term benefits and costs of another important and widely replicated early intervention: Success for All. Currently implemented in approximately 2,000 schools serving over 1 million children throughout the United States, Success for All is a comprehensive school reform program that focuses on promoting early school success among educationally at-risk students. As a comprehensive school reform model, the program also includes components designed to promote stronger links between the school and the home and to help address social, behavioral, and health issues that may be interfering with students' general academic performance. Success for All was developed by Robert Slavin, Nancy Madden, and colleagues at the request of the Baltimore City Public School System, and was piloted in one Baltimore elementary school during the 1987-1988 school year. Four additional Baltimore schools implemented Success for All during the subsequent year, 1988-1989. Our analyses track the educational outcomes through the eighth grade for the original Success for All students and for a quasi-experimental untreated control

group composed of students from matched comparison schools. In addition, we examine the differential costs associated with Success for All and control students' schooling through eighth grade. We conclude by placing the cost-effectiveness results from this study in the context of similar findings from the Perry Preschool, the Abecedarian Project, and the Tennessee class-size reduction initiative.

A Conceptual Framework for Understanding the Enduring Effects of Early Interventions

What features of early interventions are related to long-lasting cognitive benefits and why would we expect to find enduring effects from them? A review by Ramey and Ramey (1998) of the major findings of rigorous studies provides consistent answers, which help explain the enduring effects of the Perry Preschool program, Abecedarian Project, and Tennessee class-size reduction effort. The conceptual framework, *biosocial developmental contexualism*, derived from this review predicts that fragmented, weak efforts in early intervention are not likely to succeed, whereas intensive, high-quality, ecologically pervasive interventions can and do. This framework highlights six principles: developmental timing; program intensity; direct provision of learning experiences; program breadth and flexibility; individual differences in program benefits; and environmental maintenance of development. These six principles, along with select references to how the three model early interventions serve as exemplars of these principles, are discussed below.

The work of Bloom (1964), among others, stresses the great malleability of human development in the early years of life. The principle of developmental timing supports the notion that interventions that begin earlier are especially advantageous in that they may capitalize on this malleability and help alter at-risk children's long-range developmental trajectories before they have had the chance to diverge substantially from the trajectories of more advantaged

children. The Abecedarian Project and other early education programs showing some of the largest effects enrolled children during infancy and continued for several years. The Perry Preschool and Tennessee STAR interventions also started early, serving children at the ages of 3 to 4 and 5, respectively. Although the optimal timing for early intervention is open to some debate, the sustained effects on reading achievement through mid-adolescence for the earlier interventions, Abecedarian (Ramey et al., 2000) and Perry Preschool (Scheinhart, Barnes, & Weikart, 1993), were larger than the effects for STAR's reductions in kindergarten through third grade class sizes (Krueger & Whitmore, 2001).[1]

Ramey and Ramey (1998) also point out that early interventions that provide highly intensive services (indexed by variables including the number of hours per day, days per week, and weeks per year that the intervention is offered) produce greater sustained effects than do less intensive programs. Perry Preschool and Abecedarian provided a range of intensive services for both children and parents. Similarly, those children from the Tennessee STAR study with more years of exposure to reduced class sizes showed stronger sustained effects than students with fewer years in a small class (Finn et al., 2001; Nye et al., 2000). All three interventions also serve as prime examples of Ramey and Ramey's (1998) third principle, in that they provided direct educational experiences to participating students rather than relying exclusively on intermediary routes, such as parent training, to change children's competencies.

The fourth principle suggests that those interventions that offer a comprehensive approach, including a strong educational program, social services, family support, and individualized assistance, tend to produce larger effects than do interventions with a narrower focus. The Perry Preschool and Abecedarian programs provide strong prototypes of this principle as both offered educational programs tailored to meet the individual needs of children along with

parent outreach components and other efforts to help meet children's social, psychological, and health needs. Although no data was collected in Tennessee to document how reduced class sizes affected teachers' interactions with children, there is other evidence suggesting that smaller classes encourage more individualized attention to students' academic and personal problems (Stasz & Stecher, 2000). Furthermore, findings from STAR revealed that children who attended small classes achieved improved social and behavioral outcomes in comparison to their counterparts in regular-sized classes (Finn, Fulton, Zaharias, & Nye, 1989).

The benefits accruing to students from these interventions tend to be driven by the principle of individual differences in program benefits. Some children show greater benefits from participation than others and these differences appear to be related to aspects of children's initial risk conditions. An important example of this principle comes from the Tennessee STAR study, which documented that minority students typically benefited more from reductions in class size than non-minority children (Finn & Achilles, 1999; Krueger & Whitmore, 2001).

A final caveat affecting long-term outcomes is suggested by Ramey and Ramey's (1998) sixth principle, which suggests that the initial positive effects of early intervention will diminish over time to the extent that there are not adequate follow-ups to maintain children's social, behavioral, and academic gains. The only intervention that assigned some students to receive follow-up services, the Abecedarian Project, showed that children receiving a longer duration of services, during both preschool and elementary school, achieved better long-term reading outcomes than students with shorter durations of participation (Campbell & Ramey, 1994; 1995). Over time, though, most analyses of these three programs showed slighter effects after the interventions had ceased.

Success for All

Success for All is, arguably, the most widely implemented, widely researched, and widely critiqued educational program in the United States. Today, schools may purchase the program from the not-for-profit Success for All Foundation as a comprehensive package, which includes materials, training, ongoing professional development, and a highly specified "blueprint" for implementing and sustaining the model. Schools that elect to adopt Success for All implement a schoolwide program for students in grades pre-K to five that organizes resources to attempt to ensure that every child will reach the third grade on time with adequate basic skills and will continue to build on those skills throughout the later elementary grades. Rather than special education, retentions in grade, and other forms of remediation, the program emphasizes prevention and early, intensive intervention designed to detect and resolve learning problems as early as possible, before they become serious.

Students in Success for All schools spend most of their day in traditional, age-grouped classes, but are regrouped across grades for reading lessons targeted to specific performance levels. Teachers assess each student's reading performance at eight-week intervals and make regrouping changes based on the results. Instead of being placed in special classes or retained in grade, students who need additional help receive one-on-one tutoring to get them back on track.

The use of cooperative learning methods also helps children develop academic skills and encourages them to engage in teambuilding activities and other tasks that deal explicitly with the development of interpersonal and social skills. In addition, a Success for All school establishes a Family Support Team, serving to increase parents' participation in school generally, to mobilize integrated services to help Success for All families and children, and to identify and address particular problems such as irregular attendance, vision correction, or problems at home. Finally, each Success for All school designates a full-time Program Facilitator who oversees the daily

operation of the program, provides assistance where needed, and coordinates the various components. These are the main features of Success for All, both as originally conceived and as currently disseminated.[2]

Similar to the Abecedarian Project, Perry Preschool, and Tennessee class-size reduction initiatives, Success for All is relatively costly. A recent review of 26 school reform models identified Success for All as one of the most expensive whole-school reforms, with estimates of first-year personnel, materials, and training costs of between $70,000 and $270,000 for a typical school (Herman et al., 1999). In estimating the costs of implementing three reforms, Success for All, the Accelerated Schools model, and the School Development Program, King (1994) also concluded that the per-school cost estimates for Success of All, which ranged from $261,060 to $646,500 per year, were the highest. However, the Success for All developers argue that schools with high enrollments of poor children generally have sufficient supplemental federal and state Title I funding and personnel to implement a credible form of the model, and these resources often are further augmented by reallocated funds and personnel from special education, desegregation settlements, and other sources (Slavin et al., 1994).

Previous studies of the program's outcomes have focused primarily on its immediate achievement effects. The prototypical Success for All evaluation has used a quasi-experimental, untreated control group design employing matched student samples from two similar schools. Although there are exceptions (cf Jones, Gottfredson, & Gottfredson, 1997), the majority of these studies has documented positive achievement outcomes and reduced retention and special education placement rates for Success for All students (Slavin & Madden, 2001). Previous analyses of the short-term effectiveness of the program in the 5 Baltimore elementary schools revealed effect sizes above 0.50 on individually administered reading tests in grades 1 through 3

(Madden, Slavin, Karweit, Dolan, & Wasik, 1993). The most educationally at-risk students, who were identified as the lowest-scoring 25% on the pretest, showed the strongest effects from the intervention, with effect sizes at or close to 1.00 (Madden et al., 1993). These outcomes from the pilot Success for All schools, and their successful replication in numerous other schools, are impressive. Nevertheless, the long-term outcomes of the intervention remain untested and the overall cost-effectiveness of the program is open to debate.

Hypotheses of the Current Study

We hypothesized that, similar to the studies of the Perry Preschool program, the Abecedarian Project, and the Tennessee class-size reduction effort, we would find sustained effects on important academic outcomes for students who had participated in the Success for All program. Further, although Success for All is a somewhat costly intervention, we expected that the model's focus on prevention would not be substantially more expensive than the traditional emphasis on remediation, primarily in the form of retention and special education, to which control students would, most likely, be exposed.

Why would we expect Success for All to promote long-term positive effects for students? The robust short-term effects for the program present a compelling empirical case to support this expectation, but the application of Ramey and Ramey's (1998) six-principle framework provides an added theoretical basis. First, Success for All stresses early intervention and prevention, with services beginning in pre-kindergarten or kindergarten. Second, the program stresses intensity, from its daily 90-minute reading periods to its multi-year approach. Third, learning experiences are delivered to students directly, efficiently, and effectively in classrooms that are regrouped according to students' current achievement levels. Fourth, program breadth and flexibility are exemplified by the range of individualized services offered to parents and to students by the

Family Support Team, classroom teachers, and one-on-one tutors. With respect to the fifth principle, individual differences in program benefits, Success for All appears to have the most pronounced impacts on the students with the greatest academic needs. The current study helps us examine Ramey and Ramey's (1998) sixth principle: environmental maintenance of development. That is, after discontinuation of the Success for All intervention during the elementary grades, do children maintain their academic gains through middle school into mid-adolescence and, if so, at what cost?

Method

Data

All data, from the 1986-1987 through 1998-1999 school years, were abstracted from computerized files provided by the Baltimore City Public School System (BCPSS). All student background, transcript, and achievement data were collected by the BCPSS as part of its annual district-wide testing and data collection programs. We utilized two yearly files, the Pupil Information File (PIF) which contains basic data such as race/ethnicity, gender, grade level, school, and special education participation information, and a data file containing student-level standardized test results. We used data from the district-administered California Achievement Test (CAT) and Comprehensive Test of Basic Skills, Fourth Edition (CTBS/4). The CAT provided the pretest reading data used in our analyses and the CTBS/4 provided our eighth-grade reading and math achievement outcomes.

Sample

The sample included students from the original Success for All elementary schools from Baltimore, Maryland, Abbottston, City Springs, Dr. Bernard Harris, Harriet Tubman, and Dallas F. Nicholas, Sr., and their five matched control schools, respectively Tench Tilghman,

Collington Square, Patapsco, Harlem Park, and Charles Carroll of Carrollton. These schools were matched on the demographic characteristics of their students and received similar base levels of funding and supplemental federal Title I funds. Therefore, the clearest difference between the Success for All and matched control schools was the added resources associated with the Success for All implementation.[3]

One other more subtle difference between the two groups of schools was that the Success for All schools were among the lowest performing schools in Baltimore and were singled out by the district as in need of improvement. A possible element of self-selection may be involved in that the staff from these low-performing schools expressed interest in the program and voted to adopt it. The control schools, on the other hand, were not offered the opportunity to adopt the model because they were not among the schools that were identified as most in need of improvement. Thus, while there may be some degree of school-level self-selection into the Success for All group, it is difficult to determine the exact degree of the difference in motivation for reform between the Success for All and control school staff because the control schools never had the opportunity to decline or adopt the program. It is even possible, though perhaps not likely, that control schools had more motivation for reform. The procedures followed in the original Success for All schools are the same as those that are currently used. That is, school staff interested in adopting the model first express an interest in it, then they attend a presentation describing the program, and finally they vote to decide whether or not they will adopt it. Therefore, in terms of motivation for reform, these first 5 schools are reasonably representative of current schools implementing Success for All.

Beginning with the baseline year of Success for All implementation, which was 1987-1988 for Abbottston and 1988-1989 for the other 4 Success for All schools, we identified four

independent cohorts of first-grade students from the 1987-1988, 1988-1989, 1989-1990, and 1990-1991 academic years. Combining these four cohorts of students yielded a total sample of 1,388 Success for All and 1,848 control students. These quasi-experimental Success for All and control samples were formed in a manner akin to an "intention-to-treat" (ITT) sample from a randomized experiment. That is, we analyzed the effects of initial first-grade enrollment in a Success for All or control school regardless of students' subsequent enrollment in the Success for All or control school. This ITT definition avoided the potential biases that would result by systematically excluding from our analyses those students who transferred out of their original schools. Using this definition, though, students did not necessarily receive the full Success for All intervention.[4] In these respects, our estimates of program effects are somewhat conservative. The ITT estimates are also realistic, in that they provide policymakers an indication of the program effect that can be expected in typical urban settings in which students often do change schools and, therefore, would not necessarily experience the full Success for All program.

The ITT samples were reduced in our analyses due to listwise deletion of cases with missing data on the background variables and outcomes. For analysis of eighth-grade test score outcomes, the final analytical sample sizes were 581 and 729 for Success for All and control students, respectively. The 581 Success for All students in this sample were enrolled in the 5 Success for All schools from one to six years and averaged 3.84 years of participation. Similarly, the 729 control students were enrolled in the 5 control schools from one to eight years and averaged 3.76 years of attendance. All other analyses were based on transcript data from the PIF and included 735 Success for All and 995 control students.[5] These students were enrolled in Success for All and control schools over the same range of years as noted above, with Success for All students attending their baseline schools for 3.92 years, on average, and control students

attending their baseline schools for an average of 3.74 years. Listwise deletion of cases with missing values did not cause differential attrition rates by program condition, leaving 42% of the baseline sample of 1,388 Success for All students and 39% of the 1,848 baseline controls for our analyses of eighth-grade test score outcomes χ^2 (1, N = 1310) = 1.91 and 53% and 54% of the respective baseline Success for All and control samples for our analyses of transcript outcomes χ^2 (1, N = 1730) = 0.14.

The data attrition from our study occurred for several reasons. Over 50% of the Success for All and control students who were dropped from our analyses remained in the Baltimore school system through eighth grade but had missing data on one or more measures. About one-quarter of both Success for All and control students who were dropped from our analyses had left the Baltimore City schools before eighth grade and, thus, had no outcome data in the city's database. These students entered private schools within the city, transferred to a public or private school outside of the city, state, or country, or otherwise left the city's public school system. Due to the more frequent retentions for control students, a somewhat higher percentage of controls (16% from the transcript analysis and 12% from the test score analysis) than Success for All students (11% from the transcript analysis and 9% from the test score analysis) had not yet completed the eighth grade by the spring of 1999. These students, of course, had no eighth grade outcomes to analyze as of yet. Depending on the analytical sample, from 4% to 8% of the Success for All and control students' whereabouts were unknown. Finally, 1% or fewer of Success for All or control students were deceased, institutionalized, or incapacitated.

Although rates of Success for All and control student attrition were statistically equivalent and the reasons for their attrition were similar, there is a possibility that the background characteristics of the Success for All students who dropped out of our analyses were

systematically different from the backgrounds of the control students who dropped from our analyses. Such differences could bias our estimates of the program effects and compromise the internal validity of the study. In addition, it is possible that those students who dropped from our analyses, including both Success for All and controls, were systematically different from those who remained in the analytical samples. Differences of this sort could limit to whom we might generalize our results and, thus, might compromise the study's external validity.

We assessed the potential for both types of biases. First, to address the issue of internal validity, we contrasted the background characteristics of those Success for All students who dropped out of our analyses to the backgrounds of the control students who dropped from our analyses. The Success for All students and control students who were dropped from the analyses were statistically equivalent on all background characteristics, with the exception of the statistically significant Success for All-control difference on the reading pretest score for those who dropped out of the transcript analysis sample, $t(868) = -5.97$, $p < .001$ (two-tailed), $d = -0.40$, and the achievement test analysis sample, $t(1291) = -6.94$, $p < .001$ (two-tailed), $d = -0.38$. The magnitudes of these differences, though, were essentially the same as the magnitudes of the Success for All-control pretest differences for the samples that we retained for our transcript analyses ($d = -0.35$) and achievement analyses ($d = -0.35$). Therefore, the backgrounds of the Success for All students who dropped out of our samples were essentially the same as the backgrounds of the control students who were dropped.

Second, with respect to the study's external validity, we contrasted the background variables of all students who dropped out of our study to the backgrounds of the students who we were able to retain in our analyses. The students retained in our transcript analysis samples and the students who dropped from our transcript analysis were statistically equivalent on all

background variables. However, for the achievement analyses, those who were retained in the analysis had higher pretest scores than those who were not retained for both the Success for All sample, $t(1136) = -4.68$, $p < .001$ (two-tailed), and for the control sample, $t(1463) = -3.70$, $p < .001$ (two-tailed). These results suggest that our analyses of the transcript outcomes have greater external validity than our analyses of the achievement outcomes. However, given the consistency of the Success for All outcomes across the transcript and achievement analyses, it does not appear that this discrepancy had serious consequences for the general direction and magnitudes of the effect estimates. In fact, because the Success for All program tends to have stronger effects on the lowest achievers, our analyses of the achievement outcomes may underestimate the program effect.

We also identified a subsample of the lowest achieving students for analysis. We defined the low-achieving subsample as those students whose pretest scores, from the year prior to the cohort year, fell within the bottom quartile of their cohort's distribution. Because the 1990-1991 cohort had higher pretest scores than the other cohorts, we identified the bottom 25% within each cohort separately, but regardless of Success for All or control status. From the sample of students with valid background data and eighth-grade test score results, we identified 172 Success for All and 106 control low achievers for the achievement analyses. The final low-achieving subsamples of students for the transcript analyses were 257 and 187 for the Success for All and control conditions, respectively.[6]

Baseline data for the analytical samples are displayed in Table 1. Sample sizes, means, and standard deviations are provided for the full sample and low-achieving sample and for the samples used in the achievement and transcript analyses. The samples were comprised of students who were overwhelmingly poor, as indicated by free or reduced price lunch eligibility,

and African American. Students' baseline ages, recorded during September of first grade, were relatively typical, as were the gender splits.

We found differences between the Success for All and control group students on only one key background characteristic: reading pretest. In the overall analytical sample, the difference favored control students by one quarter of one standard deviation. Attempts to match Success for All and control students one on one failed, as statistically significant pretest differences remained in the optimally matched samples.[7] Therefore, attempts to analyze the one-on-one matched samples were not pursued and, instead, all analyses included the full control and Success for All samples while statistically controlling for pretest differences. In addition, in the case of one analytical sample, the low-achieving sample used for the PIF transcript analyses, we found one relatively small but statistically significant difference between Success for All and control group students on baseline age. Analyses of the transcript outcomes for low achievers, thus, also included baseline age as a covariate.

Insert Table 1 about here

Measures

Student Outcomes. Success for All is designed specifically to affect three outcomes: achievement, grade-level progression (or retention), and special education placements. Our analyses included measures of all three outcomes through mid-adolescence. We used Total Reading and Total Math scale scores from the district-wide administration of the CTBS/4 during the fall of eighth grade. Because this is the only time at which all Baltimore City students are

tested after elementary school, we took a student's CTBS/4 score from when he or she first started eighth grade, regardless of when or at what age a student reached eighth grade.

Our analyses of special education placements and retentions in grade included four variables: the number of years a student was placed in special education during elementary school; the number of years a student was placed in special education during middle school; whether or not a student was retained during elementary school; and whether or not a student was retained during middle school. The special education variables were continuous measures and the retention variables were dichotomous outcomes. Retentions and special education placements during elementary school were assumed to be affected by students' improved academic outcomes, and corresponding decreased needs for these interventions, and by educators' reliance on the Success for All interventions (e.g., one-on-one tutoring and services offered through the Family Support Team) instead of the traditional options of retention and special education referral.[8] In contrast, middle school students' retentions and special education placements occurred after leaving the Success for All elementary school program and, thus, are not directly affected by the availability of alternative Success for All treatment options.

Pretest Covariate. The CAT pretest was administered during the spring prior to each student's cohort year. For the 1988, 1989, and 1990 cohorts, pretest was defined as the student's CAT Total Reading score. Because a Total Reading score was not available for most students from the 1991 cohort, we derived a similar score by taking the mean of three CAT reading subscales, Phonics Analysis, Structural Analysis, and Vocabulary ($\alpha = .88$). However, this resulted in a mean pretest scale score for the 1991 cohort that was considerably higher than the mean pretest scores for the other three cohorts. Therefore, we standardized the pretest scores

within each cohort by converting them to z-scores. It is the z-score of the CAT pretest that is used in all analyses.[9]

Cost Estimates. Our second research question concerned the relative costs of schooling for Success for All and control students. Several cost estimates were required for this comparison. First, we established the costs of Success for All using the ingredients method (Levin & McEwan, 2001). Summaries of the costs derived from this analysis are provided by school and by year in Table 2. Detailed descriptions of the program ingredients were provided by Slavin, Madden, Karweit, Dolan, and Wasik (1992). The authors reported the numbers of tutors, whether Program Facilitators were full-time or half-time employees, and additional support staff employed at each school. We used these data to determine the number of personnel required for the Success for All program at each school.

We derived information regarding the current market values for these personnel salaries from the U.S. Department of Labor, Bureau of Labor Statistics' (1999) 1998 Occupational Employment Statistics (OES) national survey. Costs for Success for All tutors and Program Facilitators were estimated using the OES reported median annual salary for teachers ($36,110), the costs for the lead Family Support Team member were based on the OES median salary for counselors ($38,650), and costs for additional Family Support Staff were based on the OES median salary for teacher aides ($16,280). To express all salaries in current dollar amounts, all of these 1998 salaries were then adjusted to constant 2000 dollars using gross domestic product implicit price deflators. Finally, we calculated the additional costs of fringe benefits as 26.35% of instructional staff salary and 27.99% of support staff salary for each of these two categories of personnel (U.S. Bureau of the Census, 1997).

In addition to personnel costs, implementation of the Success for All program at the Baltimore schools involved expenses for training, materials, and ongoing professional development. Current market value estimates of the costs of these goods and services were derived from pricing information available on the Success for All internet site, www.successforall.net: $70,000 to $85,000 for the first year of implementation; $26,000 to $30,000 for the second year; and $23,000 to $25,000 for the third and later years. The midpoint of each of these ranges was used as the estimate for the Baltimore schools.[10]

The Success for All programs operating at the five Baltimore schools focused on grades 1 through 5. Consequently, the school-by-school and year-by-year marginal costs listed in Table 2 were divided by each school's yearly student enrollment in grades 1 through 5 to obtain the yearly per-pupil costs of Success for All used in our analysis. We assigned these school-specific and year-specific per-pupil costs to all students enrolled in grades 1 through 5 in a Success for All school. These marginal cost estimates served as the core of the cost analyses of Success for All and also provide valuable information regarding the current expenses associated with replicating the Success for All interventions as originally configured in the five Baltimore schools.[11]

As Levin (2002) has noted, there are great challenges in determining the costs of comprehensive school reform models and other interventions that cannot be readily identified as distinct interventions or as "add-ons" to the regular school program. The comprehensive models may include both reallocations of existing resources and added resources that are pertinent to whole-school reform. Many of the comprehensive school reform developers, including the Success for All developers (Slavin et al., 1994), argue that all or most of these costs can be covered by reallocations of existing funds, personnel, and other resources that already exist

within schools. As Odden & Archibald (2000) argued, this method of "resource reallocation" can make implementations of programs like Success for All essentially "costless."

Our approach treated all ingredients that are demanded by Success for All as having a cost. Some of these costs were funded by the Baltimore schools through reallocation of resources. For instance, three of the Success for All schools, Dallas Nicholas, Bernard Harris, and Harriet Tubman, reassigned their Title I reading instructors as Success for All reading tutors (Slavin et al., 1992). The matched control schools also had Title I reading teachers, but in the control schools, the Title I instructors continued teaching traditional reading pullout classes rather than offering tutoring help, as specified by the Success for All model. One could consider these instances as costless redeployments, in that there are no additional marginal costs for tutors at these Success for All schools. We do not have the evidence available to know whether the reassignments of these staff may have sacrificed some other beneficial use of their time and services, though. For instance, these sacrifices may have included the academic benefits to a larger group of Title I students than served through the one-on-one Success for All tutoring component or some other measured or unmeasured benefits we have not considered. In the absence of such information, we have included the full cost of all personnel demanded by Success for All in our estimates.

In terms of reallocations, one must also consider the potential costs associated with Success for All that may be related to its emphasis on literacy instruction to the potential detriment of student outcomes in other subject areas. Levin (2002) suggested that if one is to assume that these potential reallocations are "costless," one should measure changes in outcomes that would be most likely to be affected by the reallocation along with the outcomes for those to which the resources have been addressed. Therefore, in addition to reading achievement, we also

consider achievement outcomes for the other major elementary school subject area, mathematics.[12] This evidence concerning the potential long-term effects for math, and for the other general academic outcomes that we assess, including retentions and special education placements, helps us determine whether the focus on reading skills is detrimental to other academic outcomes. Based on the evidence and assertions of other scholars of early literacy learning, we suspect that Success for All's strong literacy component will not have negative consequences for other subject areas and may, instead, actually improve outcomes in other subject areas. This belief is supported by the work of Whitehurst & Lonigan (2001), who noted that literacy skills provide a critical part of the foundation for children's overall academic success. Children who read well read more and, as a result, acquire more knowledge in various academic domains (Cunningham & Stanovich, 1998).

Beyond the resources related to Success for All, three additional estimates were needed to calculate the cost of each student's education. First, as a measure of the overall market value of each child's basic education program, we used the annual average per-pupil current expenditure for the United States, which was $6,911 for the 1999-2000 school year (U.S. Department of Education, 2002).[13] Second, as an estimate of the overall market value of special education services, we utilized the annual average additional per-pupil expense of special education services for students with non-severe handicapping conditions (i.e., specific leaning disabilities, other health impairments, emotional disturbance, and speech and language impairments). This estimate of $5,312, after adjusting to 2000 dollars, was derived from 1985-1986 data collected by Moore, Strang, Schwartz, & Braddock (1988) and reported by (Parrish, 2001). This is the most recent available expenditure information for non-severe special education expenditures and

it is likely that it underestimates the current cost of these services as it is well-documented that special education expenditures have increased dramatically since that time (Parrish, 2001).[14]

Finally, we established the yearly cost of retaining a child in grade. For Success for All students, this cost was simply the annual average per-pupil current expenditure of $6,911. For control students, though, we applied a discount rate of 5% to this expenditure to obtain a present value of $6,565.45. Rather than spending money on Success for All to help at-risk students in the present, retention delays the intervention—repeating a grade—for a year. Because control schools did not spend present resources on Success for All, but instead utilized future funds to accommodate retained students, they incurred the discounted cost. Though there is widespread agreement that costs occurring in the future are less of a burden than costs occurring in the present (Levin & McEwan, 2001), there is less agreement on the specific discount rate that should be applied. Our choice of 5% is informed by Levin and McEwan (2001) and Barnett (1996), who, respectively, suggested discount rates of 3% to 5% and 3% to 7%.

Based on each student's yearly data, from grades 1 through 8, regarding special education assignments, school assignments, and retentions we estimated the costs associated with each year of his or her educational program. A combination of three possible values was involved in each student's yearly educational cost estimate: (a) the current base per-pupil expense of $6,911 or a present value of $6,565.45 for years in which control students were retained ; (b) the additional non-severe special education per-pupil expenditure of $5,312 and; (c) the school-specific and year-specific per-pupil cost of Success for All.[15] The sum of each student's yearly cost of schooling through the end of middle school, that is, through the successful completion of eighth grade, was the final figure that we analyzed as an outcome variable.

Insert Table 2 about here

Analytical Approach

The two primary inferential methods that we used were standard analysis of covariance (ANCOVA) for continuous outcome variables, and logistic regression for dichotomous outcome variables. In each analysis we controlled for pretest score. In one case, for our analysis of low-achieving students' transcript outcomes, we also controlled for a statistically significant Success for All-control difference for baseline age. For the continuous outcomes, each adjusted mean difference between Success for All and control students that we obtained from the ANCOVAs was divided, or standardized, by the pooled posttest standard deviation for the outcome. The resulting standardized differences, or effect sizes, provide summaries of the magnitude of each effect and are interpretable as the number of standard deviation units separating Success for All students from control students on the covariate-adjusted outcomes.

Adjusted means, effects sizes, and odds ratios for the dichotomous variables were produced from logistic regression output. The adjusted means were calculated by saving the residuals from the logistic regression analysis and then adding the overall mean to each residual to create a fitted average value for each case. The group means on these fitted average values then became the adjusted group mean on the dichotomous variable. Using the adjusted means in the arcsine transformation, suggested by Lipsey and Wilson (2001), produced an effect size estimate that is algebraically equivalent to the standard means-based effect size.

An additional summary measure of effect for the dichotomous variables is the odds ratios derived from the logistic regression analyses. Taking the antilog of a coefficient from logistic regression yields an odds ratio, which can be interpreted as the multiplicative change in the odds

for a one-unit increase in the predictor. Since the treatment indicator in our analyses is dichotomous and coded 0 or 1, where 1 = Success for All and 0 = control, a one-unit change simply represents the multiplicative difference in odds between the Suceess for All and control groups. An odds ratio has a lower bound of 0 and no upper bound. A value of one represents parity between the groups; while an odds ratio less than one indicates that Success for All students are less likely than control students to have the attribute in question, and an odds ratio greater than one indicates that SFA students are more likely than control students to have that attribute.

Results

Table 3 displays the results for the analyses of achievement outcomes and transcript outcomes for both the total sample and the low-achieving sample. All statistically significant Success for All-control differences revealed by the ANCOVAs and logistic regression analyses are indicated in the next-to-last column displaying the pretest-adjusted effect size values. After controlling for the kindergarten pretest differences, Success for All students had higher eighth grade CTBS/4 reading and math scale scores, than did control students, Fs (1, 1307) = 29.84 and 4.40, ps < .001 and .05, respectively. Expressing the pretest-adjusted eighth grade test scores in grade equivalents, Success for All students held a six-month advantage over control students in reading, 5.7 versus 5.1, and a three-month advantage in math, 6.2 versus 5.9. The magnitude of the Success for All reading effect size exceeded one quarter of one standard deviation. These results suggest that the achievement effects of the program, especially in the subject area of reading, attained both statistical and practical significance.

The long-term special education outcomes in Table 3 also show statistically significant differences between Success for All students and control students, with Success for All students

spending fewer years than control students enrolled in special education during the elementary school grades and middle school grades, *F*s (1, 1730) = 15.43 and 16.59, respectively, *p*s < .001. After controlling for the pretest, Success for All students, on average, spent about half of 1 academic year (0.55) during elementary school in special education, compared to over three quarters of a year (0.82) for control students. This result is not surprising, in that the Success for All model improves the academic skills of students and, thus, prevents the need for special education assignments. At the same time, though, it provides viable alternatives for educating students who continue to have special needs. Therefore, the special education differences between Success for All and control students in elementary school may be interpreted as a result of the improved academic performance of Success for All students, the reform initiative itself, or, most likely, some combination of both.

Insert Table 3 about here

Special education placements after the elementary grades, in middle school, are not open to this combination of interpretations. In fact, if it were the case that the Success for All model of offering individualized help and other services in lieu of special education referrals were ill-advised, we might expect former Success for All students to have higher rates of special education placement after they left the program, on the theory that some of the students that were not referred to special education under Success for All actually were in need of referral. On the contrary, we see that Success for All students continued to be less likely than control students to participate in special education in the middle school grades. After controlling for the pretest, the average Success for All student spent approximately half of 1 school year (0.49) during middle

school in special education versus over two thirds of a school year (0.70) for the typical control student.

The findings for retentions in elementary and middle school are consistent with the special education findings, and bear the same importance for judging the effectiveness of Success for All. During the elementary grades, nearly all Success for All students avoided retention (0.91) and about three quarters (0.77) of control students avoided being retained. The odds ratio of 3.41 for this outcome indicates that Success for All students were 3.41 times more likely than controls to have avoided retention during elementary school. In middle school, though, Success for All and control students were successfully promoted at a statistically equivalent rate, 0.88 for both groups, with an odds ratio of 1.06. Like the outcomes for special education, Success for All students' lower rate of being retained during the elementary school years most likely reflects a combination of both improved student outcomes and the schools' use of the alternatives Success for All offers for helping children who have fallen behind.

Regarding the middle school outcomes, again, we might expect to see a higher frequency of retentions for Success for All students than for control students if the program simply served to promote students who should have been retained. Instead, the results in Table 3 show that there is no difference in retention frequency for Success for All and control students during the middle school grades. Associated with the less frequent retentions of Success for All students, we also found that they successfully completed eighth grade at a younger age than control students, $F\,(1, 1730) = 20.94$, $p < .001$. The difference between the pretest-adjusted means indicated that Success for All students, on average, finished eighth grade 1.5 months ahead of their control group peers.

Low Achievers

We repeated the above analyses on the subsample of initially low-achieving students scoring in the bottom 25% of their respective cohort on the pretest.[16] The results are shown at the bottom of Table 3. Consistent with previous analyses of the short-term outcomes for the Baltimore Success for All students (Madden et al., 1993; Slavin et al., 1990), and consistent with Ramey and Ramey's (1998) principle of individual differences in program benefits, the effect size magnitudes for all outcomes for the low-achieving sample exceeded those found for the full sample. Despite the smaller sample sizes and corresponding loss of statistical power, five of the eight outcomes achieved conventional levels of statistical significance.

For CTBS/4 reading, initially low-achieving Success for All students had a higher adjusted mean scale score than did initially low-achieving control students, $F\,(1, 275) = 7.83$, $p < .01$. Expressing the covariate-adjusted outcomes in the grade equivalent metric, Success for All students were 7 months ahead of controls, with group means of 4.8 and 4.1, respectively. Although the Success for All students' eighth grade math scores were nearly one quarter of one standard deviation higher than control students' outcomes, the difference did not attain statistical significance, $F\,(1, 275) = 3.57$.

After adjusting for the pretest and baseline age for our analyses of the low-achievers' transcript outcomes, we found that the Success for All students spent fewer elementary school years than control students in special education, $F\,(1, 440) = 6.77$, $p < .05$, and Success for All students were 3.44 times more likely than controls to have avoided retention during elementary school. The special education and retention outcomes during the middle school years also favored Success for All students. The difference for special education placements was statistically significant, $F\,(1, 440) = 8.77$, $p < .01$, but the difference for the retention outcome,

which resulted in an odds ratio of 1.12 that favored Success for All students, was not. The fewer overall retentions among Success for All students helped contribute to a statistically significant age advantage at the completion of eighth grade, F (1, 440) = 14.95, $p < .001$. That is, after adjusting for the pretest and age at baseline**,** Success for All students completed the eighth grade 2.6 months earlier than did control students. These outcomes have all of the same implications discussed previously for the full sample.

The Cost of Success for All

Our final transcript-based outcome is the estimated per-pupil costs of schooling through the end of middle school, or through the successful completion of eighth grade. The adjusted average costs from the ANCOVA for both the total sample and low-achieving sample are displayed in Table 3.[17] These estimates include base per-pupil expenditures, special education expenditures, discounted (for control students) and non-discounted (for Success for All students) costs of retention, and school- and year-specific Success for All costs for each year of the students' schooling. After adjusting for the pretest, the costs associated with Success for All students' schooling through eighth grade were statistically equivalent to the costs of control students' schooling, F (1, 1730) = 0.00.

The costs of schooling for low-achieving students were considerably higher. Controlling for the pretest and baseline age for our analyses of the low-achievers, the adjusted cost for Success for All students was lower than the cost for control students, but this difference was not statistically significant, F (1, 440) = 2.64. The magnitude of the difference was equivalent to an effect size of –0.15, or $2,682.18, which suggests that the preventative approach of Success for All was somewhat less costly than the remedial approach of more frequent special education placements and retentions that was characteristic of the low-achieving control group.

Discussion

The findings from this study have important implications for educational policy, theory, and practice. From a policy perspective, our results indicate that a nationally disseminated elementary school program may deliver enduring educational benefits to the students it serves at no additional cost. Specifically, Success for All students complete eighth grade at a younger age, with better achievement outcomes, fewer special education placements, and less frequent retentions in grade at a cost that is essentially the same as that allocated to educating their control-group counterparts. The results also provide another clear, supporting example of the theoretical framework advanced by Ramey and Ramey (1998), who argue that high quality, intensive, ecologically pervasive approaches, like the comprehensive Success for All model, tend to promote meaningful sustained effects on students' academic outcomes. More generally, this study suggests that the replicable educational practices of prevention and early intervention, as modeled by Success for All, are more educationally effective, and equally expensive, relative to the traditional remedial educational practices of retention and special education.

These findings also may be appreciated by providing a context for understanding their importance relative to the outcomes for three other valuable model early interventions, the Perry Preschool Program, Abecedarian Project, and the Tennessee STAR study's class-size reductions to 15 students. In Table 4, we summarize the per-pupil costs and reading and math achievement outcomes from the present study and the costs and reading and math outcomes of the three other interventions. We also present the ratio of achievement effects to cost per $1000 in the last two sets of columns in Table 4. The first two columns present reading and math effects per $1000 that are based on the full costs for each of the four interventions. The last two columns are based on the full costs for Success for All, but reduced costs for the other three interventions.

These two reading and math cost-effectiveness ratios for the four interventions provide a sensitivity analysis that corresponds to two distinct assumptions regarding how we may conceptualize the programs' relative effects. As indicated by our theoretical framework, one perspective is that reductions in class size, Success for All, the Perry Preschool, and the Abecedarian Project all represent *comprehensive* educational programs. Accordingly, we have presented a range of educational outcomes for Success for All and have compared the relative effects of the four programs on the basis of two key educational outcomes: reading and math achievement. Although the four programs are likely to affect outcomes on a range of academic and non-academic outcomes, for comparison purposes we focus on these two key academic outcomes.

The tabulated results in the first of the final two sets of columns, thus, provide similar data related to the overall current marginal per-pupil costs of replicating each of the four model interventions along with the long-term reading and math achievement outcomes attained by the intention-to-treat samples from the original demonstration projects at a similar point during mid-adolescence. We have selected these two outcomes, reading and math achievement, for two main reasons. First, from a practical standpoint, these are the primary outcomes that are consistently documented across the studies of all four interventions. Second, and more importantly, we consider reading and math achievement outcomes as two of the most important outcomes at eighth grade. We also believe that these outcomes are strong predictors of other future educational and adult outcomes. Indeed, as Jencks and Phillips (1998) argued, reducing the test score gap is probably both necessary and sufficient for substantially reducing inequality in educational attainment and earnings.

We also concede, though, that the preschool programs and the Tennessee STAR class-size reduction effort, to a lesser extent, have generated evidence of sustained effects on a broader range of academic and other outcomes that are not reflected in Table 4, which focuses on reading and math achievement. Our underlying theory of Success for All as a comprehensive model suggests that it should have some of these same sustained benefits, but we do not have sufficient empirical evidence to support this hypothesis at this time. Success for All may or may not have the sustained social and economic benefits demonstrated by the preschool programs and may or may not have meaningful sustained effects on other academic outcomes. These are questions that can only be answered by further research.

Therefore, the second assumption depicted in the final set of columns in Table 4 is that statistically significant improvements in math and reading are the only likely goals and outcomes of Success for All, but, in contrast, represent only a fraction of the objectives and benefits of the other interventions. To provide a reasonable comparison under this assumption, we assume that 100% of the cost of Success for All is allocated to improving these outcomes, but less than 100% of the cost of the other interventions is directed toward math and reading. We contacted the developers of both preschool programs to obtain an estimate of the percentage of program time and resources that were devoted to improving math and reading achievement, as opposed to improving children's other academic or non-academic outcomes. Larry Schweinhart (personal communication, November 13, 2002) estimated 50% of the Perry resources were devoted to math and literacy outcomes, and Joseph Sparling and Craig Ramey (personal communication, November 21, 2002) estimated that 40% of the Abecedarian resources were allocated to improving math and literacy outcomes.

For the class size-reduction estimate, we based our estimate on reasoning presented by Levin, Glass, & Meister (1987), who argued that "reduction in class size is an overall educational intervention that should affect all of the educational activities during the school day" (p. 64). As a result, the authors suggested that only a portion of the marginal cost of class size reduction should be viewed as an educational intervention to improve outcomes in a single subject area, such as reading or math. Levin et al. stated that about one third of the school day is devoted to a single subject area, reading or math, and the other two thirds of the day is spent on other activities and objectives. Here, since we are looking at effects for both math and reading, we used two-thirds of the full cost for the reduced-cost comparisons.

The average annual per-pupil marginal cost for Success for All across all years and schools ($795), is multiplied by the average years of participation in the program by our intention-to-treat sample, 3.84 years, to arrive at a total per-pupil cost, $3,054.66. The sustained reading and math effect sizes are taken directly from our results previously shown in Table 3. The reading and math effect sizes through age 14 of the Perry Preschool program are taken from Schweinhart, Barnes, & Weikart (1993). These are the intention-to-treat effects, or the effects of initial assignment to the two-year program. On average, these students from poor and minority backgrounds participated in Perry Preschool for 1.74 years (L. Schweinhart, personal communication, May 29, 2002). Similar effects through age 15 are summarized for the Abecedarian Project based on data reported by Ramey et al. (2000). Most of the disadvantaged children who took part in the Abecedarian study began attending the program within a few months after birth and continued through the beginning of kindergarten. On average, children from the sample employed in the Ramey et al. (2000) analyses participated in the Abecedarian program for 5 years (P. Burchinal, personal communication, June 3, 2002). The annual per-pupil

expenditures associated with replicating the Abecedarian and Perry Preschool efforts rely on previously reported marginal cost estimates provided by the Developmental Center for Handicapped Persons (1987) and Barnett (1992), respectively, and are converted to constant 2000 dollars.

For the Tennessee STAR study, we show reading and math achievement effects that are based on the work of Krueger and Whitmore (2001) and further information provided by one of the authors (D.M. Whitmore, personal communication, June 6, 2002). Similar to the Success for All sample and the samples involved in the two preschool studies, these data show the sustained effects through eighth grade for African American students initially assigned to a small class. On average, over the four years from kindergarten through third grade, the students from this intention-to-treat sample spent 2.10 years in a small class. The current cost of replicating the STAR study's class size reductions was drawn from Levin et al.(1987), and includes the costs of physical space, furnishings, energy needs, insurance, maintenance, and the salary of a teacher. Converting the Levin et al. estimate of $28,138 in 1980 dollars to constant 2000 dollars using gross domestic product implicit price deflators, the cost is $54,172. If we assume an average class-size reduction in the Tennessee STAR study of 40%, from 25 to 15 students, we obtain a per-pupil expenditure estimate of $1,444.59 for a small class of 15 students (0.40 * $54,172 / 15 = $1,444.59).[18]

Although all but one of the reading and math effect estimates are relatively larger than the respective long-term Success for All effect sizes, the full costs of the other interventions also tend to be more expensive. Considering the ratios of effects to cost per $1,000 per-pupil expenditure, Success for All produced a reading effect size of 0.09, the Tennessee STAR class size reductions produced an effect of 0.07, Perry Preschool yielded an effect of 0.03, and

Abecedarian produced a reading effect size of 0.01. With respect to math achievement, the Tennessee STAR study produced the highest cost-effectiveness ratio, 0.07, Success for All showed the second strongest result, 0.04, followed by the two preschool interventions.

These cost-effectiveness ratios, though, do not take into consideration that the investments in the preschool interventions and the Tennessee STAR reductions in class size have focused on other areas of children's development that are not necessarily measured by reading and math achievement alone. The last set of columns in Table 4 take this into account by comparing the appropriate reduced costs of the three models to the full cost of Success for All. When comparing the full Success for All cost to the reduced reading and math costs for the three other models, the ratios of effects per $1000 show a similar pattern of outcomes for reading, with Success for All having a slight edge over the other three interventions. The picture is different for math achievement, however, as Success for All has the second lowest cost-effectiveness estimate. These results provide a helpful context for understanding the significance of Success for All in the context of findings from other highly regarded educational interventions.

Insert Table 4 about here

Another key policy issue to consider is the large-scale replicability of the programs and their effects. Success for All and Perry Preschool are the two interventions of the four that are available as nationally disseminated models. Studies from diverse localities suggest that the educational effects of the original Success for All pilot programs tend to be replicated with a good deal of consistency, but that these effects depend on the quality of the implementation (Slavin & Madden, 2001). Implementation is not a trivial matter, as Success for All requires

educators throughout a school to rethink and actively change many of their practices. After all, it is a whole-school reform model. If teachers do not accept the changes that the model suggests, it is not likely to succeed in improving practices and is not likely to affect student outcomes. Before adopting Success for All, the developer requires that 80% of the faculty agree, by secret ballot, to follow through with the implementation. If this support wanes, or if systemic support through the district or state tails off, the reform is likely to fail. This has typically been the case in circumstances in which Success for All has failed, including the Memphis, Tennessee school district, which recently dropped Success for All from more than 40 of its schools, and the Miami-Dade County school district, which dropped the program from all but seven of the 45 schools that once ran it. Large-scale replication, though, clearly is helped by the Success for All Foundation's growing national infrastructure for supporting increasing numbers of schools that wish to adopt the model and by recent federal policy changes, which have made more resources available to finance the continued scale-up of comprehensive school reform programs like Success for All.

Similarly, the educational approach used in the Perry Preschool classrooms and home visits is widely implemented today, primarily through the use of federal Head Start funds, as the High/Scope Curriculum (Epstein, 1993). Unfortunately, though, the significant resources necessary to replicate the Perry Preschool program, as it was originally designed in Ypsilanti, typically have not been available through public programs (Kagan, 1991; Barnett, 1995). There are other recent examples of resource-intensive, high-impact preschool programs, including the Chicago Child Parent Centers (Reynolds, Temple, Robertson, Mann, 2001), that have shown enduring effects on achievement and other important student outcomes. Examples such as this are significant in showing that the general concept of the intensive and relatively costly Perry

Preschool model can be successfully funded and replicated. More public commitment, through programs such as Head Start and Title I, or private support, through community organizations and foundations, is needed to establish the large-scale national replication of the pilot program's effects.

Widespread efforts to deliver the Abecedarian model of highly intensive educational childcare to children beginning shortly after birth have not been fully realized either. The Abecedarian project did inspire the U.S. Congress, in its reauthorization of the Head Start Act in 1994, to develop the Early Head Start program, which covers the first three years of life. Since its inception, Early Head Start has grown to a nationwide effort of 635 community-based programs serving 45,000 children. Similar to the comparison between Head Start and Perry Preschool, though, the Early Head Start program has not provided the same high-intensity services that the Abecedarian children received. Again, though the research evidence from the Abecedarian project clearly demonstrates that highly intensive early intervention can make a profound and enduring difference for the children who participate, the considerable monetary investments and capacity-building efforts to establish a similarly intensive national network of programs have not been undertaken.

On the surface, the reductions in class size modeled by the Tennessee STAR study would seem to be the most easily replicated intervention of the four. In recent years, the federal government has made available billions of dollars to reduce class sizes in the early grades. State-led efforts, such as California's massive initiative, also have begun recently. At least two noteworthy differences, though, set apart the Tennessee STAR model from these national and state-level initiatives. First, the Tennessee STAR class-size reductions occurred in only those schools that had the facilities to accommodate the new classrooms needed to reduce class sizes.

Second, the experiment operated in a relatively smaller number of schools and, therefore, did not create tremendous demands for new teachers. As suggested by California's recent statewide initiative, scaling up class-size reductions to larger numbers of schools has resulted in higher than anticipated costs, shortages of classroom space and qualified teachers, and smaller than anticipated achievement effects (Bohrnstedt & Stecher, 1999). In addition, rather than improving equality of opportunity, Bohrnstedt and Stecher report that the California effort has exacerbated disparities between districts serving many minority and poor students and districts serving few minority and poor students. Therefore, in areas that require considerable capital improvements to make available the additional classroom space needed to reduce class sizes and where there are potential shortages of qualified teachers, class-size reduction policies may not enjoy the level of success experienced in Tennessee.

The final point to consider regarding these four interventions focuses on the research designs of the supporting studies. Because this study, unlike the other three studies of Abecedarian, Perry Preschool, and Tennessee STAR, used a quasi-experimental rather than an experimental design, causal inferences about Success for All are somewhat more tentative. The findings from this quasi-experimental study, though, are unusually robust, in that the pretest showed a statistically significant difference favoring control students but the posttest showed a statistically significant difference favoring Success for All students. This posttest difference in reading achievement was statistically significant in the ANCOVA analyses and remained statistically significant even when analyzed as a simple, unadjusted mean posttest difference for both the total sample, $t(1308) = 2.74$, $p < .01$ (two-tailed), and low-achieving subsample, $t(276) = 2.89$, $p < .01$ (two-tailed). Bracht and Glass (1968) noted the desirability of basing causal inferences on interaction patterns like this, which result in a switching of mean differences. Cook

and Campbell (1979) also argued that these results are generally more conclusive than any other outcome from a no-treatment control group design because of the implausibility of alternative explanations due to scaling differences, "ceiling" effects, regression to the mean, or selection-maturation problems. Therefore, the fact that this study used a quasi-experimental design while the others employed true experimental designs may be of less significance than one might typically expect.

Implications and Limitations of the Analyses

It is the unfortunate reality that limited funds force policymakers and practitioners to choose between educational interventions. It is tempting, therefore, to compare these model programs on the basis of their cost-effectiveness, replicability, and general strength of their supporting research and draw a summative conclusion regarding the efficacy of one over the others. This is not the primary implication of our analyses and cost-effectiveness comparison. Instead, these findings suggest that, relative to the three other model early interventions, Success for All is deserving of similar recognition as a sound educational investment that provides strong and lasting educational benefits, at least through eighth grade.

We suggest that decisions among the interventions be driven by more complex decision-making processes and by careful analysis of the local context in which the program is to be implemented. For instance, we have outlined some contextual factors that may hinder the replication of each of the model programs. These factors should be considered by local policymakers when choosing among alternative approaches to improving the education of educationally at-risk children. Local funding shortfalls would prevent faithful replication of the two preschool programs. Teacher shortages and a lack of additional classroom space might

complicate class size reductions. Finally, a lack of commitment among teachers and principals to alter their practices and reform their schools may derail attempts to implement Success for All.

Beyond the various assumptions of our cost-effectiveness analyses, there are potential caveats or limitations to consider. These tabulated comparisons also do not go beyond eighth grade and, therefore, exclude some important information, most notably from the Perry Preschool intervention, that shows lasting effects into adulthood on outcomes including social adjustment and economic success (Barnett, 1992; Schweinhart et al., 1993). Further, there is convincing evidence from cost-benefit analyses, provided by Barnett (1985) and Schweinhart and Weikart (1986) that suggests that these long-term benefits have considerably outweighed the substantial costs associated with the original preschool program. The profound merit of these outcomes should not be overlooked. Success for All and the Tennessee STAR class-size reductions have yet to show these important sustained economic benefits for students and society.

In addition, though we have standardized the cost-effectiveness results for the four model interventions as an effect per $1000 investment in each program, these cost-effectiveness ratios do not necessarily suggest that spending on one of the less expensive approaches could be increased to produce the larger effects of the more expensive models. In other words, it is an open question whether increased spending on class-size reductions, and class size reductions below 15 students, or increasing spending on Success for All, and even more intensive program services, could achieve the same effects of the preschool interventions. Likewise, it is not clear from these results that less intensive and less costly versions of the preschool programs could achieve the same sustained effects of Success for All and class-size reductions.

Conclusion

Ramey and Ramey's (1998) principle of environmental maintenance of development suggests that a choice of one program over another or a reliance on, for instance, only preschool intervention without elementary school and later school-based programs is misguided. Indeed, as Ramey and Ramey point out, no developmental theory is based on the assumption that positive early learning experiences are alone sufficient to ensure that children perform well throughout their lives. Our results and the results from the other model interventions consistently support these ideas.

At mid-adolescence, all students who participated in the four model interventions enjoyed outcomes that were superior to those of controls but, relative to the initial effects of the programs, these advantages had tended to wane over time, and the participants had not generally attained normative academic outcomes. For instance, for the mid-adolescent achievement outcomes, Campbell & Ramey (1995) reported that Perry Preschool attendees scored between the 15th and 17th percentiles and control students scored below the 10th percentile. Abecedarian Project children scored between the 38th and 41st percentiles and controls scored between the 28th and 30th percentiles. The Success for All students from our analysis had eighth-grade reading and math percentile scores of 20 and 17, respectively, and controls scored at the 14th and 15th percentiles, respectively.

It is not likely that any one of these interventions could serve as the "great equalizer," or as the educational equivalent to the polio vaccine, which provides a child with protection for a lifetime all in one early dose. To compensate for poor schools, suboptimal health care, economic hardship, and other contextual conditions known to have adverse effects on at-risk students' development, educational interventions must be more akin to flu shots, which are administered throughout one's life as new risks arise within the environment.[19] Rather than choosing one

intervention over the others, the best policy may be to expand high-quality preschool programs, to continue scaling up elementary-school-based class-size reduction initiatives and Success for All programs, and to implement additional middle-school and high school interventions to sustain the effects of the earlier programs.

References

Barnett, W. S. (1985). Benefit-cost analysis of the Perry Preschool Program and its policy implications. *Educational Evaluation and Policy Analysis, 7,* 333-342.

Barnett, W.S. (1992). Benefits of compensatory preschool education. *Journal of Human Resources, 27,* 279-312.

Barnett, W.S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children, 5*(3), 25-50.

Barnett, W.S. (1996). *Lives in the balance: Age-27 benefit-cost analysis of the High/Scope Perry Preschool Program.* Ypsilanti, MI: High/Scope Press.

Bergstralh, E.J., & Kosanke, J.L. (1995). *Computerized matching of cases to controls* (Technical Rep Ser No. 56). Rochester, MN: Section of Biostatistics, Mayo Clinic.

Bloom, B.S. (1964). *Stability and change in human characteristics.* New York: JohnWiley & Sons.

Bohrnstedt, G.W., & Stecher, B.M. (Eds.) (1999). *Class-size reduction in California: Early evaluation findings, 1996-1998.* (CSR Research Consortium, Year 1 Evaluation Report). Palo Alto, CA: American Institutes for Research.

Borman, G.D., & D'Agostino, J.V. (2001). Title I and student achievement: A quantitative synthesis. In G.D. Borman, S.C. Stringfield, & R.E. Slavin (Eds.), *Title I: Compensatory education at the crossroads* (25-57). Mahwah, NJ: Erlbaum.

Bracht, G.H., & Glass, G.V. (1968). The external validity of experiments. *American Educational Research Journal, 5,* 437-474.

Brewer, D. J., Krop, C., Gill, B.P., & Reichardt, R. (1999). Estimating the cost of national class size reductions under different policy alternatives. *Educational Evaluation and Policy Analysis, 21,* 179-192.

Campbell, F.A. & Ramey, C.T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development, 65,* 684-698.

Campbell, F.A. & Ramey, C.T. (1995). Cognitive and school outcomes for high-risk African-American students at middle adolescence: Positive effects of early intervention. *American Educational Research Journal, 32,* 743-772.

Carter, L.F. (1984). The Sustaining Effects Study of compensatory and elementary education. *Educational Researcher, 13* (7), 4-13.

Cook, T.D., & Campbell, D.T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston: Houghton-Mifflin.

Cunningham, A.E., & Stanovich, K.E. (1998). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33,* 934-945.

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

Epstein, A.S. (1993). *Training for quality: Improving early childhood programs through systematic inservice training.* Ypsilanti, MI: High/Scope Press.

Finn, J.D. & Achilles, C.M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21,* 97-109.

Finn, J.D., Fulton, B.D., Zaharias, J., & Nye, B.A. (1989). Carry-over effects of small classes. *Peabody Journal of Education, 67,* 75-84.

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

Gomby, D.S., Larner, M.B., Stevenson, C.S., Lewit, E.M., & Behrman, R.E. (1995). Long-term outcomes of early childhood programs: Analysis and recommendations. *Future of Children, 5*(3), 6-24.

Herman, R., Aladjem, D., McMahon, P., Masem, E., Mulligan, I., O'Malley, A.S., Quinones, S., Reeve, A., & Woodruff, D. (1999). *An educators' guide to schoolwide reform.* Arlington, VA: Educational Research Service.

Jencks, C., & Phillips, M. (1998). *The black-white test score gap*. Washington, DC: Brookings.

Jones, E.M., Gottfredson, G.D., & Gottfredson, D.C. (1997). Success for some: An evaluation of a Success for All program. *Evaluation Review, 21,* 643-670.

Kagan, S.L. (1991). Excellence in early childhood education: Defining characteristics and next-decade strategies. In *The care and education of America's young children: Obstacles and opportunities. Ninetieth yearbook of the National Society for the Study of Education*, S.L. Kagan (Ed.), pp. 237-258. Chicago, IL: National Society for the Study of Education.

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

Krueger, A.B., & Whitmore, D.M. (2001). *Would smaller classes help close the Black-White achievement gap?* Working paper #451. Princeton, NJ: Princeton University. [On-line]. Available: http://www.irs.princeton.edu/pubs/pdfs/451.pdf.

Lee, V.E., Brooks-Gunn, J., & Schnur, E. (1988). Are Head Start effects sustained? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Child Development, 61,* 495-507.

Levin, H.M. (2002). *The cost effectiveness of whole school reforms. ERIC Clearinghouse on Urban Education, Urban Diversity Series 114.* New York: Teachers College, Columbia University.

Levin, H.M., Glass, G.V., & Meister, G.R. (1987). Cost-effectiveness of computer-assisted instruction. *Evaluation Review, 11*, 50-72.

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

Lipsey, M.W., & Wilson, D.B. (2001). *Practical Meta-Analysis.* Thousand Oaks, CA: Sage.

McKey, R.H., Condelli, L., Ganson, H., Barret, B.J., McConkey, C., & Plantz, M.C. (1985). *The impact of Head Start on children, families, and communities.* Washington, DC: U.S. Government Printing Office.

Madden, N.A., Slavin, R.E., Karweit, N.L., Dolan, L.J., & Wasik, B.A. (1993). Success for All: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal, 30,* 123-148.

Moore, M.T., Strang, E.W., Schwartz, M., & Braddock, M. (1988). *Patterns in special education services delivery and cost.* Washington, DC: Decision Resources Corporation.

Nye, B., Hedges, L.V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis, 21,* 127-143.

Odden, A., & Archibald, S. (2000). *Reallocating resources: How to boost student achievement without asking for more.* Thousand Oaks, CA: Corwin.

Parrish, T.B. (2001).Who's paying the rising cost of special education? *Journal of Special Education Leadership, 14,* 4-12.

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.

Ramey, C.T., & Ramey, S.L. (1998). Early intervention and early experience. *American Psychologist, 53,* 109-120.

Reynolds, A.J., Temple, J.A., Robertson, D.L., Mann, E.A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest—A 15-year follow-up of low-income children in public schools. *Journal of American Medical Association, 285*, 2339-2346.

Rosenbaum, P.R. (1989). Optimal matching for observational studies. *Journal of the American Statistical Association, 84,* 1024-1032.

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.

Schweinhart, L.J., & Weikart, D.P. (1986). The effects of the Perry Preschool Program on youths through age 15 - A Summary. In Consortium for Longitudinal Studies (Ed.), *As the twig is bent…lasting effects of preschool programs* (pp. 71-101). Hillsdale, NJ: Erlbaum.

Slavin, R.E. & Madden, N.A. (with Chambers, B., Daniels, C., French, M., Haxby, B., Simons, K., & Wasik, B.) (2001). *One million children: Success for All.* Thousand Oaks, CA: Corwin.

Slavin, R.E., Madden, N.A., Dolan, L.J., Wasik, B.A., Ross, S.M., & Smith, L.M. (1994). 'Whenever and wherever we choose' The replication of 'Success for All.' *Phi Delta Kappan, 75,* 639-647.

Slavin, R.E., Madden, N.A., Karweit, N.L., Dolan, L.J., & Wasik, B.A. (1990). *Success for All: effects of variations in duration and resources of a schoolwide elementary restructuring program.* Center for Research on Effective Schooling for Disadvantaged Students, report no. 2. Baltimore, MD: Johns Hopkins University.

Slavin, R.E., Madden, N.A., Karweit, N.L., Dolan, L.J., & Wasik, B.A. (1992). *Success for All: A relentless approach to prevention and early intervention in elementary schools.* Arlington, VA: Educational Research Service.

Slavin, R.E., Madden, N.A., Karweit, N.L., Livermon, B.J., & Dolan, L.J. (1990). Success for All: First-year outcomes of a comprehensive plan for reforming urban education. *American Educational Research Journal, 27,* 255-278.

Slavin, R.E., Madden, N.A., Dolan, L.J., & Wasik, B.A. (1993). *Success for al in the Baltimore city public schools—Year 6 report.* Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Stasz, C., & Stecher, B.M. (2000). Teaching mathematics and language arts in reduced size and non-reduced size classrooms. *Educational Evaluation and Policy Analysis, 22,* 313-329.

U.S. Bureau of the Census (1997). *Public education finances: 1996-1997.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Education, National Center for Education Statistics (2002). *Statistics in brief: Revenues and expenditures for public elementary and secondary education: School year 1999-2000.* [On-line]. Available: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2002367.

U.S. Department of Labor, Bureau of Labor Statistics (1999). *Occupational employment and wage estimates, 1998.* [On-line]. Available: http://stats.bls.gov/oeshome.htm.

White, K.R. (1985). Efficacy of early intervention. *Journal of Special Education, 19,* 401-416.

Whitehurst, G.J., & Lonigan, C.J. (2001). Emergent literacy: Development from prereaders to readers. In S.B. Neuman & D.K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 11-29). New York: The Guilford Press.

Footnotes

[1] Several researchers have analyzed the data regarding lasting benefits of the Tennessee STAR intervention, including Finn et al. (2001), Nye et al. (1999), and Krueger & Whitmore (2001). Only Krueger and Whitmore, though, provided separate intention-to-treat (ITT) effect estimates through eighth grade for minority students. The analyses of the Abecedarian and Perry Preschool interventions, and the Success for All program reported here, are based on similar student samples from minority backgrounds and are based on the same intention-to-treat design.

An ITT analysis specifies how to handle individuals who switch from the treatment condition to the control condition, or vice-versa, over the course of a randomized experiment. ITT requires that individuals be analyzed in the groups they were randomized into, regardless of whether they complied with the treatment they were given. This type of analysis tends to generate a more conservative, and generally more realistic, estimate of the treatment effect than an analysis of only those individuals who complied with their original treatment assignment. If one were to analyze only compliant individuals, the control and experimental groups defined by compliance would no longer be formed on the basis of randomization and are, thus, subject to biases. Also, the groups defined by compliance may not accurately represent the practical impact of the treatment, as similar noncompliance is likely to occur in implementations of the program beyond the experiment.

[2] Compare, for example, the original Success for All pilot program as described by Slavin et al., (1990) and a more recent description by Slavin and Madden (2001) of the current model that is available.

[3] The added, or marginal, resources for implementing Success for All came from various sources. The funding for the Abbottston Success for All program was provided by Chapter 2 funds, City

Springs paid for its implementation through a grant from a private foundation, and the remaining Success for All schools used funds from a U.S. Department of Education dropout prevention grant (Slavin, Madden, Karweit, Dolan, & Wasik, 1990).

[4] Though Success for All students did not receive the full program, the average student did receive 75% of the five-year intervention. As a comparison, the Abecedarian students received 100% of that five-year preschool program, the average Perry Preschool student received 87% of that two-year program, and Tennessee STAR minority students received, on average, 53% of the four-year class-size reduction initiative.

[5] It was more likely for students to have missing achievement data than missing transcript information and this resulted in greater data attrition from the achievement analyses. This difference was primarily due to the fact that students had to be in attendance at school during the district testing in order to have complete achievement data, but no such discrete condition was required to obtain transcript information, such as whether or not a student was retained or placed in special education.

[6] We performed analyses similar to those for the full analysis samples that investigated potential threats to internal and external validity due to differential attrition. With respect to internal validity, the low-achieving Success for All students and low-achieving control students who were dropped from the analyses were statistically similar on all background factors, with one exception: low-achieving control students who dropped from the analyses were less likely to be male than low-achieving Success for All students who dropped from the analyses ($X^2 = 6.38$, $p < .05$). In terms of external validity, we found no differences between low-achieving students who dropped from the transcript analyses and those who were included in the transcript analyses. We did, however, find some differences between low-achieving students retained and those dropped

for the achievement analyses. Those who were included in the achievement analysis had higher pretest scores than those who were not included, for both the Success for All sample, $t(389) = -2.99$, $p < .01$ (two-tailed), and for the control sample, $t(283) = -2.59$, $p < .05$ (two-tailed). Also, for the Success for All sample only, low-achieving students retained for the achievement analyses were less likely to be male than were low-achieving students dropped from the analyses ($X^2 = 16.81$, $p < .001$).

[7] One-on-one computerized matching of Success for All and control students was attempted using an algorithm developed by Bergstralh and Kosanke (1995), which is based on the optimal matching procedures described by Rosenbaum (1989). Statistically significant Success for All-control pretest differences remained even after the optimal one-to-one Success for All-control pretest matches were identified by the computerized matching procedure.

[8] The avoidance of special education placements and retentions has always been a Success for All *suggestion*, but of course the Success for All developers do not run the schools and cannot establish this as an explicit policy. Clearly, the Success for All focus is on early identification, tutoring, targeted family support services, grouping by achievement level, and so on, both to prevent children from needing to be retained or assigned to special education and to give schools a broader range of options should a child be performing at a very low level.

[9] We conducted a sensitivity analysis, which is available upon request from the authors, to assess whether this within-cohort transformation of the pretest scores to z-scores may have influenced our results. After removing the non-conforming 1991 cohort from the sample, we reestimated all analyses using (1) the original CAT scale scores and (2) the within-cohort standardized scores. No differences were found between the outcomes generated under the two conditions. Our use of the standardized scores and the use of different pretest and posttest instruments, though, prevents

us from characterizing or interpreting the achievement outcomes in terms of absolute gains or changes in achievement on a common scale. Instead, the outcomes indicate differences in eighth grade achievement levels after prior levels of relative achievement have been taken into account. [10] These amounts represent the actual costs that schools incur by contracting with the Success for All Foundation to facilitate the implementation of the program. Of course, the amounts charged by the developer cover not only the actual costs of the services provided to the school, but also offset Success for All overhead costs, including expenses related to facilities and administration. [11] There were considerable differences between the cost estimates we derived based on the ingredients method and the original Success for All costs estimated by Slavin et al. (1992). The primary difference between our estimates and the estimates of Slavin et al. are due to the way in which the salaries and benefits of Success for All tutors at Dallas Nicholas, Dr. Bernard Harris, and Harriett Tubman were treated. As Slavin and his colleagues reported, these schools did not have to hire new personnel to fill the tutoring positions. Instead, existing staff members from the schools were reassigned to these positions. Therefore, on the face of it, there were no marginal costs for tutors at these schools. As explained in our Method section, though, our analysis treats the tutors' salaries and benefits as real costs to the schools rather than "costless" reassignments.

To compare our cost estimates to the original 1988 dollar amounts reported by Slavin et al. (1992) (i.e., $400,000 for Abbottston and City Springs and $40,000 for the remaining three Success for All schools), consider our school-by-school Year 1 cost estimates converted to 1988 dollars: Abbottston, $348,701.56; City Springs $461,931.32; Dallas Nicholas $145,374.53; Dr. Bernard Harris, $180,448.18; Harriett Tubman, $180,448.18.

[12] With regard to other subject areas, previous research on the schools involved in our analyses showed statistically significant short-term advantages for Success for All students relative to controls in both science and social studies (Slavin, Madden, Wasik, & Dolan, 1993).

[13] There are several options when considering which per-pupil expenditure figure to use. The first decision regards current expenditures versus total expenditures. The former includes all day-to-day operational expenses and the latter includes these plus expenditures associated with capital outlays, debt repayment, and other programs besides those relating to preschool through twelfth grade. The current expenditure is the more appropriate choice for our purposes because this figure is more reliable than total expenditures, which are subject to variations in accounting decisions for capital expenses and debt repayment. Also, total expenditures include costs for programs not associated with primary and secondary education, such as adult education programs or community services (U.S. Department of Education, 2002), which were not appropriate to include in our estimates of educational costs. A second factor to consider in estimating per-pupil expenditure is whether to divide current expenditures for the year by fall enrollment or by average number of students in daily attendance. We ran the analyses using both calculations, and the results were essentially the same in both cases. In the analyses we report here, we present the results using the larger divisor (fall enrollment), which produces a lower dollar estimate of schooling costs for control students, who spent more years in grades one through eight. Therefore, this decision favors control students.

[14] The choice of the non-severe special education expenditure over other options, including the simple average or median special education expenditure, is based on the assumption that Success for All is most likely to either: (a) improve the academic performance and behavior of students who are at risk for special education assignment to the extent that they no longer qualify for such

services; or (b) present viable alternative interventions, including one-on-one tutoring and services offered through the Family Support Team, that educators in Success for All schools may employ in lieu of special education assignments for those students who do qualify these services. In both cases, students' placements in the most costly special education services are not likely to be affected by Success for All. Therefore, by using the average expenditure figure for special education instead of the figure for non-severe expenditures, we may overstate the costs saved by reducing special education placements through Success for All.

In analyses not shown here, we did assess empirically whether the Success for All versus control differences in special education assignments were limited to the least severe categories. The Baltimore City school system's data set defines seven levels of special education participation: (1) indirect or consultative services; (2) up to 5 hours per week; (3) up to 15 hours per week; (4) more than 15 hours per week; (5) enrollment in a special school or center; (6) residential placement in a facility; (7) residential placement in a hospital. For each student who received one or more years of special education services, we calculated an average special education level experienced by the child while in a Success for All or control school. Although control students experienced more frequent placements, the special education students in Success for All and control schools had statistically equivalent special education levels and similar distributions across categories. These data regarding the intensity of special education participation do not precisely distinguish between less and more severe handicapping conditions and less and more costly services. The analysis does, though, provide some evidence to suggest that our reliance on the non-severe expenditure information rather than the overall average expenditure may underestimate the special education cost savings of Success for All.

[15]There were 47 control students who attended a Success for All school during one or more school years after the initial cohort year. These students were assigned the regular cost of schooling and special education costs if necessary, but not the additional cost for Success for All. This method, of course, produced a lower cost for control students than if we had included the additional cost of Success for All.

[16]Using the entire analytical sample, we also specified ANCOVA models with a Success for All by pretest interaction effect. These interaction effects, though, were not statistically significant.

[17]The mean unadjusted cost of Success for All students' schooling was $66,227.87 (*SD* = 13691.09) and the mean unadjusted cost for control students was $64,220.21 (*SD* = 14336.31). For the low-achieving sample, the unadjusted mean was $72,888.66 (*SD* = 16938.77) for Success for All students and $75,257.77 (*SD* = 17807.50) for control students.

[18] Brewer, Krop, Gill, & Reichardt (1999) provided a more recent 1998-1999 per-pupil estimate of the average costs of reducing class sizes to15 students. As these authors concede, though, this estimate of $981, or $998 in constant 2000 dollars, does not take into account the additional expenses associated with providing new classroom facilities to make the class-size reductions possible. Though somewhat dated, the Levin et al. (1987) estimate provides a more realistic estimate of the full costs schools are likely to face when reducing class sizes. Neither of the estimates, though, takes into account the potential costs associated with higher salaries due to greater demand for teachers.

[19] This point assumes that there are cumulative effects from successive interventions. Relatively little good evidence regarding cumulative exists, though. Perhaps the best evidence of cumulative effects of interventions comes from research on the Abecedarian preschool program. As part of the study, children were randomized into four intervention conditions: (a) preschool

only (infancy to age 5); (b) elementary intervention only (age 5-8); (c) preschool and elementary interventions (infancy to age 8); and (d) a no-treatment control group. Results from this study generally supported an intensity hypothesis, in that scores on academic achievement and cognitive measures increased as duration of treatment increased (Campbell & Ramey, 1994; 1995).

Author Note

This article was written under funding from the Office of Educational Research and Improvement, U.S. Department of Education (Grant No. OERI–R-117-D40005). However, any opinions expressed do not necessarily represent positions or policies of OERI. The authors thank Spyros Konstantopoulos, Hank Levin, Nancy Madden, Stephen Meyer, Allan Odden, Bob Slavin, and four anonymous reviewers for their helpful feedback on previous drafts of this article.

Table 1

Baseline Data for Analytical Samples.

	Success for All			Control		
	N	*M*	*SD*	*N*	*M*	*SD*
Full Sample						
Achievement Analysis	581			729		
CAT Reading Pretest		-0.04***	0.99		0.29	0.84
Female		0.56	0.50		0.56	0.50
African American		0.99	0.07		0.99	0.10
Free/Reduced Lunch		0.91	0.29		0.90	0.29
Age		6.27	0.39		6.28	0.36
Transcript Analysis	738			995		
CAT Reading Pretest		-0.17***	1.04		0.17	0.91
Female		0.50	0.50		0.53	0.50
African American		0.99	0.12		0.99	0.08
Free/Reduced Lunch		0.92	0.27		0.91	0.29
Age		6.30	0.42		6.32	0.40
Low-Achieving Sample						
Achievement Analysis	172			106		
CAT Reading Pretest		-1.28	0.42		-1.18	0.50
Female		0.53	0.50		0.46	0.50
African American		0.98	0.13		1.00	
Free/Reduced Lunch		0.92	0.27		0.93	0.25
Age		6.23	0.39		6.29	0.47
Transcript Analysis	257			187		
CAT Reading Pretest		-1.34	0.53		-1.26	0.48
Female		0.44	0.50		0.45	0.50
African American		0.98	0.14		0.99	0.07
Free/Reduced Lunch		0.93	0.26		0.91	0.28
Age		6.27*	0.42		6.37	0.49

Note: * $p < .05$; *** $p < .001$.

Table 2

Program Ingredients and Costs of Success for All by School and by Year.

	Abbottston	City Springs	Dallas Nicholas	Dr. Bernard Harris	Harriet Tubman
Certified Reading Tutors FTEs	6	9	2	3	3
Salary per Certified Reading Tutor	$37,291.00	$37,291.00	$37,291.00	$37,291.00	$37,291.00
Benefits per Certified Reading Tutor	9,826.18	9,826.18	9,826.18	9,826.18	9,826.18
Tutors Sub-Total:	$282,703.08	$424,054.62	$94,234.36	$141,351.54	$141,351.54
Family Support Staff FTEs	2	2.5	0	0	0
Salary per Social Worker	$30,940.00	$30,940.00			
Benefits per Social Worker	8,660.11	8,660.11			
Salary per Family Support Staff	16,812.00	16,812.00			
Benefits per Family Support Staff	4,705.68	4,705.68			
Family Support Staff Sub-Total:	$61,117.79	$71,876.63			
Program Facilitator FTEs	1	1	0.5	0.5	0.5
Salary per Facilitator	$37,291.00	$37,291.00	$18,645.50	$18,645.50	$18,645.50
Benefits per Facilitator	9,826.18	9,826.18	4,913.09	4,913.09	4,913.09
Facilitator Sub-Total:	$47,117.18	$47,117.18	$23,558.59	$23,558.59	$23,558.59
Training, Implementation, Materials, Year 1	$77,500.00	$77,500.00	$77,500.00	$77,500.00	$77,500.00
Training, Implementation, Materials, Year 2	$28,000.00	$28,000.00	$28,000.00	$28,000.00	$28,000.00
Training, Implementation, Materials, Year 3	$24,000.00	$24,000.00	$24,000.00	$24,000.00	$24,000.00
Total Cost, Year 1:	$468,438.05	$620,548.42	$195,292.95	$242,410.12	$242,410.12
Total Cost, Year 2:	$418,938.05	$571,048.42	$145,792.95	$192,910.12	$192,910.12
Total Cost, Years 3+:	$414,938.05	$567,048.42	$141,792.95	$188,910.12	$188,910.12

Table 3

Achievement and Transcript Outcomes through Eighth Grade.

	Success for All *M* [a]	Control *M* [a]	Pooled *SD*	*d*
Full Sample				
Achievement Outcomes				
CTBS/4 Total Reading scale score	716.97	703.71	46.42	0.29***
CTBS/4 Total Math scale score	723.29	718.08	47.13	0.11*
Transcript Outcomes				
Years of special education in elem. school	0.55	0.82	1.48	-0.18***
Years of special education in midl. school	0.49	0.70	1.16	-0.18***
Never retained in elem. school	0.91	0.77	0.38	0.27***
Never retained in midl. school	0.88	0.88	0.33	0.01
Age at 8th grade	14.20	14.33	0.58	-0.22***
Educational expenditures	65054.59	65090.44	14096.17	-0.00
Low-Achieving Sample				
Achievement Outcomes				
CTBS/4 Total Reading scale score	692.45	676.49	46.48	0.34**
CTBS/4 Total Math scale score	700.26	689.73	45.04	0.23
Transcript Outcomes				
Years of special education in elem. school	1.30	1.78	1.94	-0.25*
Years of special education in mid. school	1.04	1.44	1.41	-0.28**
Never retained in elem. school	0.78	0.54	0.46	0.33***
Never retained in mid. school	0.84	0.82	0.37	0.03
Age at 8th grade	14.33	14.55	0.67	-0.33***
Educational expenditures	72756.80	75438.98	17329.72	-0.15

Note * $p < .05$; ** $p < .01$; *** $p < .001$.

[a]Success for All and control mean columns display covariate-adjusted means.

Table 4

Per-Pupil Expenditures and Sustained Effects for Four Educational Interventions.

	Annual PPE[a]	Years of Intervention	Total PPE[a]	Sustained Effect Size		Effect per $1000[a] (Full Cost of SFA)			
						Full Cost Others		Reduced Cost Others	
				Reading	Math	Reading	Math	Reading	Math
Success for All	$795	3.84	$3,054.66	0.29	0.11	0.09	0.04	0.09	0.04
Class Size Reduction to 15	$1445	2.10	$3,034.50	0.14	0.19	0.05	0.06	0.07	0.09
Perry Preschool	$8929	1.74	$15,536.46	0.51	0.49	0.03	0.03	0.07	0.06
Abecedarian Project Preschool	$10,496	5.00	$52,480.00	0.53	0.65	0.01	0.01	0.03	0.03

Note: PPE = per-pupil expenditure.

[a]All expenditure data were converted to constant 2000 dollars using gross domestic product implicit price deflators.