Issue Paper

Potential Impacts of Chalkboard Project's Early Age Initiatives on Student Achievement

Introduction

RESEARCH ON EDUCATIONAL "BEST PRACTICES" increasingly suggests that primary-grade initiatives can significantly improve overall academic performance and reduce the subsequent need for more expensive, and less effective, interventions. The most widely studied early age interventions can provide sustained achievement gains through eighth grade, and possibly beyond. Evaluations of specific programs using randomized controlled trials, the "gold standard" for evaluation research, provide the foundation from which to predict the future benefits of the Chalkboard Project's Early Age Initiatives (EAI).

The limited number of rigorous studies and variety of unique programs make precise quantitative predictions difficult. However, the EAI proposal maximizes the probability of success by combining elements from the most successful and most studied primary-grade interventions. If implemented effectively, the anticipated achievement gains could significantly increase the proportion of students that meet or exceed important benchmarks and improve Oregon's overall educational performance relative to other states. Although less definitive, some research has linked variations on particular Chalkboard initiatives to additional benefits such as fewer children held back in school, improved high school attendance, and increases in Scholastic Aptitude Test (SAT) scores and in the proportion of students taking the SAT.

The EAI include both a significant class size reduction for kindergarten and first grade and a significantly expanded tutoring program for low-achieving students. The U.S. Department of Education cites both types of intervention as among those with the strongest empirical support for significant educational benefits. By investing in proven reforms, Oregon maximizes the probability of experiencing the greatest possible return. Quantifying the total impact of combining multiple interventions is difficult. Nonetheless, positive interactions between successful initiatives could result in benefits beyond those described for each individual intervention below.

● ECONorthwest's analysis indicates that fully implementing the EAI class size reductions could increase Oregon's average National Assessment of Educational Progress (NAEP) scores by up to 4 points in math and reading, boosting Oregon's rank among states into the top third for both fourth and eighth grade and into the top ten for eighth grade math. For fourth grade math and reading, and eighth grade math, Oregon's scores could meet or surpass those of Washington, a state that consistently outperforms Oregon on the NAEP.

● Fully implementing the EAI early reading tutors proposal could reduce the proportion of students failing to meet state reading benchmarks by up to 10% by eighth grade, an increase in the pass rate of nearly four percentage points.

Class Size Reductions

At present, the average kindergarten and first grade (K-1) class size is about 20 in Oregon, with about 25 percent of K-1 students in classes larger than 24 students and only about 8 percent in classes of 15 or fewer students. The EAI would reduce Oregon's K-1 class sizes to a maximum of 15 students, similar to the reduction implemented in the first and most widely evaluated class size reduction experiment in the nation, the Tennessee Student/Teacher Achievement Ratio (STAR) study. Evaluations of the STAR experiment suggest that reducing class sizes to an average of 15 students in grades K-1 can significantly improve student achievement through at least eighth grade.

Encouraging results from the STAR study led California to reduce kindergarten through third grade class sizes to 20 or fewer students. Despite rapid implementation, achievement scores increased marginally if at all and measures of teacher quality actually declined. Although the outcome in California cautions against excessive optimism, two features of the EAI proposal increase the likelihood that Oregon students would receive significant benefits. First, California reduced class sizes to 20 students or smaller, despite evidence that reductions resulting in class sizes larger than 15 are unlikely to have a significant impact. Second, the Chalkboard Project proposes to phase-in the reductions over almost a decade, allowing the supply of qualified teachers to grow with demand.

Although numerous other states have implemented ambitious class size reduction initiatives, the STAR experiment remains the most studied. Successful implementation of a STAR-like reduction in Oregon could significantly improve the ranking of Oregon's students relative to those from other states on achievement tests such as the NAEP and increase the proportion of students meeting and exceeding state and national benchmarks.

If the reductions generate the maximum effect consistent with evaluations of the STAR study, Oregon's average NAEP scores could rise by up to four points. Fourth grade scores could rise from below average into the top quarter of states in math and to above average in reading. Improvement in average eighth grade scores could rise from the top third in math into the top ten and from below average into the top third in reading.[1] Implementation details could significantly affect actual outcomes, however, particularly if class size reductions are one element of a package of reforms.

While the NAEP analysis gives an indication about how class size reductions could improve Oregon's standing relative to other states, NAEP data is highly aggregated and hence, this analysis does not allow for intrastate variation in the impact of reducing class sizes. More detailed test score data from Oregon's Norm Referenced Intelligence Test (NRIT) assessments does allow for a more refined analysis. Applying the same analysis to district-level NRIT scores as was applied to NAEP scores suggests that reducing class sizes could increase average third grade math and reading NRIT scores by between 0.6 and 1.4 points, and increase eighth grade scores by between 0.5 and 1.5 points.

A one point increase may not represent an educationally significant improvement for a single student, but a small increase in average scores could indicate much larger increases for a subset of students. Indeed, evidence from the STAR experiment suggests that low-performing and at-risk students benefit more from smaller class sizes than their higher-achieving peers. Under a plausible set of assumptions, the proposed class size reductions could increase the proportion of eighth grade students at or above the state standards for math and reading by up to 4 percentage points,

---

[1] In 2005, the average fourth grade score was 238 in math (ranked 28th) and 217 in reading (ranked 34th). The average eighth grade score was 282 in math (ranked 16th) and 263 in reading (ranked 27th).

representing approximately 12 percent of the students who would otherwise fail to meet the performance standard. Class size reductions could reduce the number of third grade students failing to meet standard by up to 17 percent, although the proportion meeting the standard would increase by only about 2 percentage points because the overall passing rate is higher in lower grades.

Early Reading Tutors

With fewer students competing for the teacher's attention, class size reductions benefit every affected student. Regardless, some students will inevitably fall behind, and achievement gaps in early grades generally widen over time. The tutoring component of the EAI addresses these gaps by providing intensive one-to-one reading tutoring for low-achieving students.

Schools can select from a wide array of established tutoring approaches. However, the variety of methods and the scarcity of large-scale evaluations make it difficult to estimate the impact of any specific tutoring program. Nonetheless, analyses of the widely adopted Success for All curriculum (one of the most studied early reading programs) imply that providing an intensive regime of early age interventions, including tutoring, for at-risk students can increase academic achievement well beyond elementary school. Unfortunately, analyses of tutoring as an independent intervention typically rely on small samples and idiosyncratic methods, leading to conflicting or ambiguous results.

The type of training tutors receive is one important dimension along which programs differ. For example, the Success for All program relies on certified teachers to provide individual reading instruction, while programs such as Start Making A Reader Today (SMART) rely exclusively on volunteer tutors.

An evaluation of the SMART program found that SMART participants demonstrate significant improvements in reading ability after receiving one or two 30-minute tutoring sessions per week over two school years. On the other hand, evaluations of other tutoring programs are less conclusive, suggesting that certified reading tutors and a comprehensive approach to reading instruction are important to success. These findings contrast with the relatively flexible SMART approach involving community volunteers and minimal tutor training, characteristics seen as beneficial by SMART proponents.

Despite the conflicting evidence, both types of program show some promise for improving students' reading skills. The EAI take a mixed approach to tutoring, splitting a significant increase in paid FTE equally between SMART coordinators and certified teachers.

Results from studies of the Success for All programs suggest that the tutoring component of the EAI would significantly improve low-achieving students' performance on the NAEP and Oregon assessment tests, assuming that the volunteer tutors are at least as effective as highly trained tutors. The initiatives could reduce the proportion of students failing to meet the state reading standard by up to 10 percent by eighth grade, an increase in the passage rate of nearly four percentage points. This analysis assumes that the EAI would provide tutors for about 25,500 of the lowest performing readers in kindergarten through third grade. SMART tutors already provide tutoring to about 10,500 students per year. This analysis assumes that the early reading tutors proposed in the EAI would reach an additional 25,500 students of the approximately 40,000 in the bottom quartile of reading achievement.[2]

---

[2] This analysis assumes that the 375 paid, certified tutors will tutor 18 students per day for a total of 6,750 students per year and that each of the 375 SMART coordinators will recruit 40 SMART tutors, each of whom tutors 1.25 students on average, for a total of 18,750 students. The assumptions about SMART tutors are based on current SMART program data. SMART currently serves about 10,500 students

Appendix
Class Size Reduction Analysis

STATE AND DISTRICT OF COLUMBIA average scores from the 2004-05 NAEP determine state rankings for this study. Krueger's 1999 analysis of data from the STAR experiment is the primary source used to predict the impact of the proposed class size reductions. Krueger's study evaluates class size effects in kindergarten through third grade. Other analyses of the STAR data suggest that small classes can improve student achievement through 8th grade and beyond.

The STAR experiment reduced class size from an average of about 23 students to an average of about 15 students. This reduction is comparable to that proposed for Oregon, and evaluations of the STAR experiment provide the best foundation for an analysis of the EAI. This paper predicts the independent impact of fully implementing the proposed reductions, holding all other determinants of educational achievement constant at their 2004-05 levels. This implicitly assumes that Oregon test scores would increase consistent with the aggregate "STAR effects" found by Krueger and others, despite the fact that Oregon's students and STAR participants likely differ in important determinants of educational achievement (*e.g.*, race and socioeconomic status).

Krueger presents numerous estimates for the impact of the STAR class size reduction. Three of these were selected as the basis for high, medium, and low impact scenarios. In each case, the "STAR effect" is measured as an increase in average scores as a fraction of the standard deviation of the pooled (at the classroom level) scores from STAR students in both small and regular classes. Average NAEP scores in Oregon are adjusted by a similar fraction of the standard deviation of NAEP scores in Oregon.[3] No attempt is made to account for sampling variability in either Krueger's estimates or the NAEP scores.

For each scenario, the relevant effect is discounted to reflect the fact that a significant fraction of kindergarten and first grade (K-1) classes in Oregon have fewer than 15 students. The average K-1 class size was 20.03 students during the 2004-05 school year. About 16% were smaller than 16 students and about 30% had more than 23 students. It is doubtful whether students would receive significant advantages from a class size reduction from, for example, 18 to 15 students.

Existing research provides little guidance about the benefits of reductions of a specific size. If benefits accrue only with reductions of seven or more students, the "STAR effect" should be diminished by about 60% when applied to Oregon. On the other hand, assuming that one seventh of the benefits accrue for each one-student reduction in size, the discount would be approximately 30%. The effects described below are discounted by 45% when the rankings are recalculated, halfway between these extremes.

Low-impact scenario

Krueger's Table IX, column (1) implies students with one year in a small class scored, on average, 4.08 percentile points higher than students with no small class experience. The estimate is an average over the four grades studied (kindergarten through third grade) and does not account for student or class characteristics such as the fraction of classmates receiving a

---

[3] Note that the statewide NAEP standard deviations are conceptually different from those used to calculate the "STAR effect" from the STAR data.

free lunch. This is also one of the smallest estimates presented by Krueger. Oregon's fourth grade ranking is recalculated by increasing the average score by 0.10 standard deviations, the equivalent of 4.00 percentile points for normally distributed test scores. Finn et al, find that in terms of standard deviations, STAR participants from small classes exhibited decreasing relative achievement gains by eighth grade. Hence, Oregon's eighth grade scores were increased by 0.08 standard deviations, consistent with one of the lower estimates in Finn *et al*.[4]

Medium-impact scenario

Krueger's Table IX, column (1) implies an increment in average scores of 1.19 percentile points for each additional year in small classes. Hence, this scenario applies two years of "STAR effects" or an increase in average scores of 0.13 standard deviations, equivalent to 5.00 percentile points, to Oregon's average scores for both grades.

High-impact scenario

STAR participants appeared to show the greatest achievement gains after first grade, the last grade affected by the class size reduction proposed for Oregon. Krueger's Table V, column (1) estimates the impact of small classes for each study grade separately without controlling for student characteristics and is one of the largest effects found in the studies reviewed for this analysis. The high impact scenario should be seen considered an optimistic upper bound rather than a likely outcome. Krueger's estimate implies that two years in small classes would produce average scores that are 8.57 percentile points higher than they would be otherwise. Under this scenario, Oregon's scores increase by 0.21 standard deviations, equivalent to 8.50 percentile points.

Early Reading Tutors Analysis

The proposed tutoring program would target low achieving students. Hence, this section of the analysis focuses on the lower tail of the distribution of students as measured by Oregon's NRIT reading test scores. The "tutoring effects" used in the scenarios are drawn from the Borman and Hewes 2002 evaluation of the Success for All program. Of particular relevance, Borman and Hewes perform separate analyses for students scoring in the bottom quartile of all students. The projected rankings described in this analysis are calculated using the proportion of students at or above the NAEP basic achievement level (hereafter "basic"). The method used to adjust Oregon's scores is similar to the method used in the class size reduction analysis. Essentially all of the caveats mentioned above apply to this section as well.

Tutoring is assumed to increase the score for every student scoring below the Oregon reading benchmark by an amount proportional to the relevant "Success for All effect" and to the standard deviation of NRIT reading scores in that student's district. The aggregate effect of the increases is then discounted because the Chalkboard proposal is not sufficient to serve the entire bottom quartile of readers, and existing SMART tutoring programs already serve about 7% of K-3 students in Oregon. In particular, this analysis assumes that the 375 paid, certified tutors will tutor 18 students per day for a total of 6,750 students per year and that each of the 375 SMART coordinators will recruit 40 SMART tutors, each of whom tutors 1.25 students on average, for a total of 18,750 students. The assumptions about SMART tutors are based on current SMART program data.

---

[4] The estimate comes from an analysis of math scores. Students showed greater relative gains in reading than in math. In general, Krueger and Finn et al present similar estimates for the "STAR effect" where their estimates are comparable.

In aggregate, these assumptions imply that the percent of students meeting the state reading benchmark will increase in proportion to the district-wide standard deviation of reading scores, adjusted for the assumed proportion of students the proposed tutors could serve.

Low-impact scenario

This scenario assumes that the "tutoring effect" is equal to half of the difference between math and reading improvement that Borman and Hewes attribute to Success for All. Estimates from Table 3 in Borman and Hewes imply a total difference of 0.11 standard deviations, resulting in a "tutoring effect" of 0.055 for this scenario.

Medium-impact scenario

This scenario assumes that the "tutoring effect" is equal to the full 0.11 standard deviation difference described above.

High-impact scenario

This scenario assumes that the "tutoring effect" is equal to one half of the entire estimated impact of Success for All on standardized reading scores found in Borman and Hewes. The increase of 0.17 standard deviations is more appropriately considered an optimistic upper bound for the impact of a package of educational reforms. Existing evidence suggests that improvements of this magnitude are not likely to be generated through tutoring alone.

References

Baker, Scott, Russell Gersten, and Thomas Keating, "When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training," *Reading Research Quarterly*, Oct.-Dec. 2000, vol. 35, no. 4, pp. 494-519.

Borman, Geoffrey D. and Gina M. Hewes, "The Long-Term Effects and Cost-Effectiveness of Success for All," *Education and Policy Analysis*, Winter 2002, vol. 24, no. 4, pp. 243-266.

Bohrnstedt, G.W., and B.M. Stecher, eds., *What we have learned about class size reduction in California*, 2002, California Department of Education.

Finn, Jeremy D., Susan B. Gerber, Charles M. Achilles, and Jayne Boyd-Zaharias, "The enduring effects of small classes," *Teachers College Record*, April 2001, vol. 103, no. 2, pp. 145-183.

Krueger, Alan B., "Experimental estimates of education production functions," in *The Economics of Schooling and School Quality Volume II: Efficiency, Competition and Policy*, Eric A. Hanushek, ed., Edward Elgar Publishing, Ltd., Northampton, MA., 2003, pp. 40-75.

Krueger, Alan B. and Diane M. Whitemore, "The effect of attending a small class in the early grades on college-test taking and middle school test results: evidence from project star," *The Economic Journal*, Jan. 2001, vol. 111, pp. 1-28.

Mishel, Lawrence, and Richard Rothstein, eds., *The Class Size Debate*, Economic Policy Institute, Washington, DC, 2002.

Slavin, Robert E. and Barbara A. Wasik, "Preventing early reading failure with one-to-one tutoring: A review of five programs," *Reading Research Quarterly*, Apr.-June 1993, vol. 28, no. 2, pp. 179-200.

Torgerson, Carole J., Sarah E. King, and Amanda J. Sowden, "Do Volunteers in Schools Help Children Learn to Read? A Systematic Review of Randomised Controlled Trials," *Educational Studies*, vol. 28, no. 4, 2002, pp. 433-444.

Data Sources

National Assessment of Educational Progress, National Center for Education Statistics, United States Department of Education.

Oregon Norm-Referenced Intelligence Test scores and other data, Oregon Department of Education.