

INTERPRETING THE EFFECTS OF TITLE I SUPPLEMENTAL EDUCATIONAL SERVICES

Georges Vernez and Ron Zimmer, RAND
October 2007

An ongoing challenge in the evaluation of the effects of programs to improve student achievement is how to describe the estimated effect size in terms accessible to non-technical readers. Many education journals encourage researchers to report their study results in standardized effect sizes¹ to allow comparisons across interventions and help facilitate summaries of findings for a specific intervention. Standardized units also facilitate comparisons across districts or states that use different tests.

In an evaluation of Title I school choice and supplemental educational services conducted as part of the National Longitudinal Study of No Child Left Behind, Zimmer et al. found positive effects averaging 0.08 of a standard deviation unit in both reading and math for students participating in supplemental services for the first time.² The effect was cumulative when students participated in supplemental services multiple times (in two or more years) with effect sizes growing to 0.15 and 0.17 of a standard deviation for reading and math, respectively. How significant these effect sizes are educationally is discussed below.

Interpreting effect size estimates for education programs

Researchers often have used a guideline provided by Jacob Cohen (1988) as benchmark for interpreting the magnitudes of effects in the social sciences. His guideline suggests that effect sizes of 0.20, 0.50, and 0.80 of a standard deviation unit should be considered small, medium, and large, respectively. However, in education policy interventions, most effect sizes that have been measured do not exceed 0.25, and it is rare for an education intervention to have an effect size as large as 0.80.

To better interpret the magnitude of the gain in student achievement of an education intervention, three alternative potential benchmarks may be referred to: the gain in student achievement of an additional year of learning, the size of the achievement gap between minority and white students, or the gains obtained by other education interventions (Hill et al., 2007).

Effect size for one additional year of learning

Hill et al. (2007) examined annual achievement test score gains from one grade to the next for students in grades K-12 from a national norming sample of seven major standardized tests in reading and six in math. They computed the achievement that students gained over each year, which includes the learning gained in school and learning and maturation gained outside of school, minus any learning lost during school vacations. A great deal of variation was observed across grades:

- The average annual gains for students between 4th and 5th grade were 0.40 and 0.56 of a standard deviation in reading and math, respectively.
- The average annual gains for students between 7th and 8th grade were 0.26 and 0.32 of a standard deviation in reading and math, respectively.

¹ Effect size is calculated as the difference between outcome means between treatment and control groups, divided by the within group standard deviation of one of the groups, which is assumed to be common across the two groups. It should be noted that educational studies often use either achievement “levels” or annual achievement “gains” as outcome measures and the effect sizes across these two measures may not be the same. Therefore, comparison of effect sizes between studies using achievement “levels” and “gains” may be imperfect.

² The study used longitudinal student-level data for 2001-02 to 2004-05 from nine large urban districts.

- The average annual gains for students between 10th and 11th grade were 0.19 and 0.14 of a standard deviation in reading and math, respectively.

Therefore, whether a certain effect size is large or small may depend upon the grade range the intervention is targeted. An effect size of 0.10 would be comparable to one-fifth to one-quarter of an additional year of learning for a 4th grader, but up to one-half to two-thirds of an additional year of learning for a 10th grader.

Size of the achievement gap between minority and white students

Because one of the more pressing educational policy challenges is the achievement gap between minority and white students, this gap is a relevant benchmark for judging the size of effects of educational interventions. Hill et al. (2007) examined the achievement gap between black and white students as well as between Hispanic and white students using the National Assessment of Educational Progress (NAEP).

- For black students, the NAEP achievement gap ranged from 0.67 to 0.83 of a standard deviation unit in reading and 0.94 and 1.04 of a standard deviation unit in math, depending upon the grades examined.
- For Hispanic students, the achievement gap ranged between 0.53 and 0.77 of a standard deviation unit in reading and 0.68 and 0.85 of a standard deviation in math.

In this context, an effect size in the range of 0.10 would represent between one-tenth to one-fifth of the achievement gap, depending upon the subject and the grade.

Effect sizes for selected education interventions

Hill et al. (2007) reviewed dozens of meta-analyses of studies that used random and quasi-experimental designs to evaluate a wide range of educational interventions and programs. Based on this review, they estimated overall average effect sizes of 0.23, 0.27, and 0.24 of a standard deviation for elementary, middle, and high school students.

Similarly, the often-cited randomized class reduction experiment in Tennessee measured an effect size of 0.15 to 0.25 of a standard deviation (Krueger, 1999). Somewhat lower effect sizes have been measured in meta-analyses of comprehensive school reform models (0.09 to 0.15) (Borman et al., 2002) and out-of-school programs (0.06 to 0.13 for reading and 0.09 to 0.17 for math) (Lauer et al., 2004).

Relative to the experience gained so far with education interventions designed to increase student achievement, the interpretation of their effect sizes should be interpreted differently than suggested by Cohen (1988) for the social sciences more generally: 0.25 of a standard deviation might be considered a large effect, 0.15 a medium effect, and 0.05 to 0.10 a small effect.³

Interpreting effect sizes for Title I supplemental educational services

As previously mentioned, the evaluation of Title I school choice and supplemental educational services conducted by Zimmer et al. (2007) found positive effects averaging 0.08 of a standard deviation unit in both reading and math for students that participated in supplemental services during one school year and effect sizes of 0.15 to 0.17 for students that received supplemental services during two or more years. About 90 percent of the student participants included in the achievement gains analysis were elementary and middle school students.

³ This interpretation is supported by Lipsey and Wilson's (1993) review of meta-analyses across psychological, educational, and behavioral outcomes, which concluded that effect sizes of 0.10 to 0.20 should not be seen as trivial.

Compared to the benchmarks discussed above, the reading and math gains from participating in Title I supplemental educational services during one school year can be interpreted as approximately equivalent to:

- 1.7 to 2.4 months of additional classroom instruction for a 4th grade student and 2.5 to 2.7 months of additional classroom instruction for a 7th grade student.⁴
- a one-tenth to one-seventh decrease in the gap between minority and white students,⁵ or
- a small effect relative to class size reduction but equivalent to the lower-bound of effects of comprehensive school reform or of out-of-school programs.

The equivalent effects estimated above are nearly twice as large for students that received supplemental educational services during two or more years.

It should be noted that the Zimmer et al. study did not control for the amount of supplemental educational services that each student received (because most of the sample districts did not provide this information); students were considered participants if they attended at least one supplemental services session during the year. Consequently, the effect sizes measured may underestimate the effects of supplemental services for students who completed all supplemental services sessions during the year.

References

- Borman, G.D., G.M. Hewes, L.T. Overman, and S. Brown (2002). *Comprehensive School Reform and Student Achievement: A Meta-Analysis*. Report # 59, Center for Research on the Education of Students Placed At Risk. Available at: www.csos.jhu.edu/CRESPAR/TechReports/Report59.pdf.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed., Hillsdale, N.J.: Lawrence Erlbaum.
- Hill, C.J, H.S. Bloom, A. R. Black, and M.W. Lipsey (2007). *Empirical Benchmarks for Interpreting Effect Sizes in Research*, MDRC Working Papers on Research Methodology, New York, N.Y.: MDRC. Available at: www.mdrc.org/publications/459/full.pdf.
- Krueger, A.B. (1999). "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics*, vol. 114, no. 2.
- Lauer, P.A., M. Akiba, S.B. Wilkerson, H.S. Apthorp, D. Snow, and M. Martin-Glenn (2004). *The Effectiveness of Out-of-School-Time Strategies in Assisting Low-Achieving Students in Reading and Mathematics: A Research Synthesis*, Washington D.C.: U.S. Department of Education, Institute of Education Sciences. Available at: www.mcrel.org/topics/products/151/.
- Lipsey, M.W., and D.B. Wilson (1993). "The Efficacy of Psychological, Educational, and Behavioral Treatment: Confirmation from Meta-Analysis." *American Psychologist*, 48, 1181-1209.
- Zimmer, R., B. Gill, P. Razquin, K. Booker, and J.R. Lockwood III (2007). *State and Local Implementation of the No Child Left Behind Act: Volume I—Title I School Choice, Supplemental Educational Services, and Student Achievement*, Washington D.C.: U.S. Department of Education, Policy and Program Studies Service. Available at: www.ed.gov/about/offices/list/oepd/ppss/reports.html#title.

⁴ 0.08 effect size measured by the study divided by the 0.40 one-year learning gain in reading and the 0.56 one-year learning gain in math for a 4th grade student multiplied by 12 months.

⁵ 0.08 effect size measured by the study divided by the lower-bound 0.53 Hispanic gap for reading and the upper-bound 1.04 black gap for math, respectively.