

July 2007

MDRC Working Papers on Research Methodology

Empirical Benchmarks for Interpreting Effects Sizes in Research

Carolyn J. Hill, Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey

In this article we argue that there is no universal guideline or rule of thumb for judging the practical importance or substantive significance of a standardized effect size estimate for an intervention. Instead one must develop empirical benchmarks of comparison that reflect the nature of the intervention being evaluated, its target population, and the outcome measure or measures being used. We apply this approach to the assessment of effect size measures for educational interventions designed to improve student academic achievement. Three types of empirical benchmarks are presented: (1) normative expectations for growth over time in student achievement; (2) policy-relevant gaps in student achievement, by demographic group or school performance; and (3) effect size results from past research for similar interventions and target populations. Our analysis draws from a larger ongoing research project that is examining the calculation, interpretation, and uses of effect size measures in education research. The more general message — that effect sizes should be interpreted using relevant empirical benchmarks — is applicable to any policy or program area, however.