

Comments on

Measuring Effect Sizes: The Effect of Measurement Error

Brian A. Jacob

University of Michigan & NBER

April 23, 2008

Overview

- Very nice paper
 - Careful technical work, previously ignored topic, valuable contribution
- Few qualms with estimation of ME variance
- Interpretation/implications is trickier
- Likely not the biggest concern with VA

Some Technical Points

- Main Findings: 17% of variance due to ME => inflate effect sizes by 4x
- Some concerns, with a caveat that none of this will likely matter much
- Several assumptions are likely suspect
 - True scores and ME uncorrelated
 - Variance of true score constant across students
 - No persistence in ME across grades
- Only explore one source of heterogeneity in ME
- Alternative approaches to getting estimate of ME variance: what about other test-retest studies?

Thoughts on Interpretation

- ME corrections will not affect significance of estimates
 - MA degree, teacher “ability” (still some conflicting evidence)
- Mistakes in administrative data files are another potentially important source of ME (e.g., Springer & Ballou)
- Correct effect size depends to some extent on the question one asks

The Use of Effect Sizes?

- Effect size is largely a *relative* concept
- One could inflate estimates for most other educational interventions in a similar manner
 - TN STAR: 0.2 => 0.8
 - Chicago summer school: 0.1 => 0.4
 - Hill et al. (2007): 0.3 => 1.2
- ME corrections will not influence assessment of the *relative* cost effectiveness of alternative interventions
- Might policymakers care about the cost-benefit in an absolute sense?
 - How to benchmark the “value” of an educational improvement? Returns to achievement?

Other concerns (research opportunities) involving VA

- Consistency
- Reliability (precision)
- Stability
- Variation in effectiveness across context (very difficult to identify)
- Middle and high schools
- Certain teacher characteristics (experience, teacher ability, subject matter knowledge)
- Relationship between “effort” and VA (common support)

Fadeout

- Explored a bit, but much more work needed
 - VA fadeout relative to other interventions
- Jacob, Lefgren and Sims (2008)
 - Use IV approach to characterize fadeout
- $Y(i,t) = a + bY(i,t-1) + XB + e$
 - B-OLS $\sim .66$
 - B-IV using $Y(i,t-2) \sim .97$
 - B-IV using $VA(i,t-1) \sim .2$
- B-IV-VA possibly biased upward
- Reading effects are slightly more persistent than math
- 2 yr persistence is roughly .12