

Differing Program Environments and Limited Replications Constrain Generalization:
The Case of Published Evaluations of School Finance Policies

Danielle V. Handel and Eric A. Hanushek
Stanford University

Well-identified evaluation studies promise unbiased estimates of programmatic impacts, but unbiasedness does not imply precision or generalizability. To deal with this, the obvious solution is obtaining multiple estimates of the same parameter, but this is often difficult to do because many evaluations are very expensive and difficult to produce. Further, there are few incentives for researchers or journals to produce such replications. It may be possible to combine estimates across somewhat different policy environments in order to obtain better impact parameter estimates, but doing this in general requires an understanding of the environmental factors that are held constant in the impact analysis.

This paper uses recent research into educational resource-achievement evaluations to provide a “laboratory” for considering how to generalize from the results of studies that are developed from different policy initiatives. This analysis is particularly important because the results of research in this area often quickly and directly enter into policy discussions about school finance policy. School finance policy is generally determined by individual state legislatures, but the courts have also been very influential. In both places, research on resources and educational outcomes has been cited frequently.

The most basic school finance question that continues to be discussed is whether just adding money – i.e., changing the budget constraint for schools – leads to better student outcomes. Even asking such a question is strange, because the simplest microeconomic theory would dismiss it out of hand. The underlying issues in the school finance discussions, however, are not standard textbook problems of adjusting a budget constraint. School budgets are produced with a range of institutional and regulatory constraints: the uses that any additional money can be put to are restricted by specific spending requirements, by state and federal laws, by local teacher contracts, and by a myriad of limitations that might make best use of any funds in the local school district impossible. Thus, this fundamental consideration of the relationship of funding and achievement becomes an empirical question, one which has recently again become a heavily researched topic.

A large literature developed after the Coleman Report (Coleman et al. (1966)) to address the question of how different school inputs including total resources affected student outcomes. This early production function study suggested that school resources had little to do with student performance, and it led to a wide range of studies that delved into the determinants of student outcomes. Because decision making on school funding – both in state legislatures and in the courts – naturally related to consideration of the role of funding in ensuring quality schooling, this literature had direct policy linkages from the outset even though it was not motivated directly by overall issues of school finance. But, the pace of this line of research into education production functions slowed noticeably by the 1980s and 1990s as new insights waned.

The early research was, however, replete with low quality studies even for the time. Thus, many of the results were surrounded by uncertainty. The prior production function evidence included a number of studies that would not meet current quality standards for empirical analysis.

The more recent path of empirical research toward more convincing identification of program impacts revived research into the relevant line of school finance studies. The search for more credible empirical evidence on the impacts of various school and other inputs has led to a new literature. The recent evidence provides a new look at the longstanding issues of resources and outcomes while paying much greater attention to the identification of key policy parameters.

Interestingly, the overall empirical results of the recent analyses tend to mirror those from the older production function work. When the recent estimates of the key impact parameters are standardized, the wide variation in estimated effects becomes very apparent. While the specific focus of the two lines of research has been somewhat different, both lines of research point to large heterogeneity in the impact of resources on outcomes. The individual impact studies provide a range of specific impact parameters. Once these are put on a common scale, the heterogeneity of results underscores a necessity of focusing on how resources are used.

The variation in findings across studies also raises questions about what generalizations are appropriate from either line of research. Both the historic and more contemporary studies include a meaningful proportion of estimated resource parameters that are not statistically significant. But the much of the variation in estimated parameters arises from differences in the underlying true parameters and not due to noise. There is currently little explanation about the mechanisms underlying these widely different point estimates for the impact of resources, making it difficult to generalize from available research.

This article will investigate the range and generalizability of estimates of the impact of resources on achievement. It will categorize results by the institutional background behind the observed resource changes.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, Robert L. York (1966). *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office.