Teachers College Record
The Voice of Scholarship in Education

Early Childhood Education: Comment on Camilli and Colleagues' Meta-Analysis and a Preliminary Synthesis of Recent Randomized Controlled Trials

by Kevin M. Gorey — December 21, 2009

This commentary critically appraises Camilli, Vargas, Ryan and Barnett's (2010) meta-analysis of the cognitive effects of early education interventions. It also presents a related synthesis of recent randomized controlled trials.

Updating my own meta-analysis of cognitive and associated benefits of early education interventions (Gorey, 2001), I found Camilli, Vargas, Ryan and Barnett's (2010) meta-analysis among 21 relevant research reviews that were published or appeared in grey literatures in the interim. Their synthesis of 123 studies was arguably the field's broadest; mine of 35 studies, the methodologically narrowest. Yet our central findings were near exact systematic replicates. Camilli and his colleagues found that the substantial immediate cognitive benefits of better resourced preschools (Gorey, 2001; Nelson, Westhues, & MacLeod, 2003) assessed with higher quality research designs were still quite large five years later and seemed to remain practically significant even into high school (Barnett, 2008; Gorey, 2001; Nelson et al. 2003). Converting Camilli and colleagues' effect size (ES) estimates to Cohen's $\underline{U}_3$ statistics (1988), it can be inferred that approximately three-quarters of the children taught in smaller preschool classes scored better on intelligence and academic achievement measures in 5th grade than did the typical child taught in larger classes. And the small group advantage was still probably experienced by two-thirds of them in high school (my secondary synthesis of Camilli et al. 2010, Table 7). Such a long-term effect might be interpreted as modest (66% of the intervention group did better than the median control group child), but it actually seems a huge effect in terms of the entire population standing to benefit from well-resourced preschool programs. Imagine the aggregate societal benefit of significantly increasing the academic success and consequent life chances of one to two of every ten of the millions of children who enter school in America each year. Such is the practical policy implication of Camilli and colleagues' research synthesis.

Unweighted Effects and Composite Measurement of Research Quality

In addition to its exhaustive search of published and unpublished research literatures that seemed not only to assure its external validity, but also the robustness of its findings to publication bias, Camilli and colleagues' meta-analysis was sound in another important way. Rather than focusing on mere main effects, the Achilles' heel of many such analyses (Greenland, 1987), it explored the potential moderation of intervention effects by key characteristics of the child participants and their families, the preschool programs and study research designs as exhaustively as its primary study reports allowed. However, I did wonder about threats to validity that may have intruded because of two of its methodological features: unweighted effects that did not account for large sample-size discrepancies and composite measurement of study design quality that confounded objective information with subjective judgements.

Camilli and colleagues seemed concerned themselves with reporting "less optimal" unweighted effects. Given the magnitude of this field's investments as well as the great potential policy significance of sound research syntheses, the "safest" analytic plan would probably also be the most statistically valid one. Camilli and colleagues seemed to assume that intervention ESs were significantly associated with the size of study samples. Why not empirically test the sample size-ES relationship and report its statistical and practical significance? If significant, why not either include the sample as a moderator variable (e.g., small versus large samples based on defensible criteria) or report unweighted and weighted ESs together at least for exemplary outcomes? This analyst is simply looking for increased empirical comfort that small studies more prone to sampling error and related confounding did not overwhelm the potentially more valid findings of larger studies (Cooper, Patall, & Lindsay, 2009; Konstantopoulos, 2008). A replication of unadjusted ESs with adjusted ESs for key outcomes would go a long way toward providing such comfort and greatly increase confidence in Camilli and colleagues' findings.

One essential service of any practically useful meta-analysis is its interpretation of research findings in light of the research methods that produced them. Such was at the heart of Camilli and colleagues' meta-analytic plan. They coded the most critical methodological characteristics of their sample of primary studies: randomized, matched or empirically assessed pre-test equivalence of study groups, provided evidence of program fidelity, assessed sample attrition and corrected for any observed attrition bias. Additionally, coders assessed whether they thought that such information was adequately provided in each primary study report. Camilli and colleagues' finding that the intervention effects observed by "high quality" studies were significantly larger than those observed by lower quality studies was encouraging. However, our ability to understand the practical meaning of this apparent intervention effect moderation by research quality is limited in two ways, both related to the use of a composite measure of study design quality (Greenland, 1987; 1998). First, because Camilli and colleagues' multi-item quality measure was ultimately discretely defined as either high quality or not, useful information about the moderating effects of specific design features (e.g., fidelity of intervention and control conditions separate from differential attrition of study groups) was probably lost. Second, the composite measure confounded objective information with subjective judgements. Exploring the potential moderating influence of each research design feature separately would seem a more informative analytic strategy. Addendum analyses that address these issues would probably engender more confidence in Camilli and colleagues' meta-analytic findings and may provide even clearer future programmatic and research direction.

Need for High Fidelity and Powerful Randomized Controlled Trials

Though we analyzed distinctly different sampling frames in different ways, Camilli and colleagues (2010) and I (Gorey, 2001) identified the same need for new randomized *controlled* trials (RCT). This field's apparent lack of scientific advancement is indeed regrettable. When I was engaged in my synthesis a decade ago it was clear that there had not yet been any systematic investment in such methods. Some small randomized trials had been accomplished. But for the most part these were single site experiments accomplished by very interested investigators. And they tended to be so small and uncontrolled that they could not produce any more confident knowledge than that produced by previous quasi-experimental and observational research. What is the field's status today? Camilli and colleagues put it quite accurately when they suggested that "few, if any" high quality RCTs have yet been accomplished. More details are needed though in providing direction to a research agenda that will probably require substantial diverse investments from federal and state governments, arms-length funders such as the National Science Foundation (Vinovskis, 2001), as well as private foundations.

Attempting to update and systematically replicate this field's earlier meta-analyses of predominantly observational research with a purposively restricted sampling frame, I searched for large RCTs (minimum analytic samples of 3,300 [Fleiss, Levin, & Paik, 2003]) powerful enough to confidently detect modest but practically significant effects, and allow for stratification or adjustment across at least two hypothetically important characteristics (e.g., socioeconomic status and an indicator of program endowment) that had been externally peer-reviewed. No relevant studies were identified. Eliminating the sample size and peer-review criteria, two randomized studies were identified. They were the multi-site Head Start Impact Study ([HSIS] U.S. Department of Health and Human Services, 2005) and a local test of otherwise well-endowed full-day versus half-day pre-kindergarten programs (Robin, Frede, & Barnett, 2006). Their cognitive effects are summarized in Table 1.

Table 1: Cognitive Effects of Head Start and State-Funded Pre-Kindergarten Programs: Summary of Recent Randomized Controlled Trials

Summary Statistics	Head Start-1-Year Follow-Up[a]	
Analytic sample	3,900	
Cohen's $\underline{d}$-index	0.23	
Cohen's $\underline{U}_3$ (%)	59.1	
	District Pre-Kindergarten-2-Year Follow-Up[b]	
	Half-Day	Full-Day
Analytic samples	217	77
Cohen's $\underline{d}$-index	0.65	1.15
Cohen's $\underline{U}_3$ (%)	74.2	87.5

click to enlarge

[a] Sample-weighted average of the two cognitive measures of both the 3 and 4-year old groups (U.S. Department of Health and Human Services, 2005).

[b] Averages of four cognitive measures (Robin, Frede, & Barnett, 2006).

The display of analogous ESs (Cohen's $\underline{d}$-indexes and $\underline{U}_3$s) seems a close replication of the short-term meta-analytic findings of Camilli and colleagues (2010) and Gorey (2001), ranging from modest, but arguably practically significant Head Start impacts to large cognitive effects of the best resourced preschool programs, in this instance a full-day, extended year program. It seems though that the limitations of these randomized trials may severely limit their "controlled" ability to answer this field's important questions. In addition to substantial treatment diffusion and consequent lack of program infidelity (Barnett, 2008; Ludwig & Phillips, 2008), the HSIS's attrition has been selective: respectively, Head Start (12%) versus control group (23%); c^2 (1, N = 4,667) = 98.02, p < .001. Such a methodological effect (d = 0.28, Cooper, 2009) is actually larger than the tested intervention effect of 0.23 so its biasing influence seems probable. And the provocative study of full/half-day preschool seemed underpowered. It was too small to ensure the pre-test equivalence of its study groups or to powerfully test (a < 0.05, power > 0.80; Fleiss et al, 2003) its seemingly large effects. It essentially screened an important hypothesis.

Conclusion

Camilli and colleagues called it right. There is a need for *new* randomized controlled trials in this field, to put the short to long-term main effects of early education interventions as well as hypothetical program endowment moderations (e.g., class size, program intensity and duration, and teacher education) to confident tests. Moreover, we should not wait for the longer-term findings of existing trials, even for the HSISs. It is hard to imagine that these ongoing, necessarily statistically adjusted correlation analyses will be able to transcend the methodological problems. Notwithstanding the sometimes legitimate arguments, philosophical and practical, against RCTs (Cook, 2002; Schutz, 2008), it seems that a small series of well planned RCTs will be essential to either confidently affirm or refute this field's central notions that have been astutely observed by early educators and researchers over the past nearly 50 years.

References

Barnett, W. S. (2008). *Preschool education and its lasting effects: Research and policy implications.* Boulder, CO and Tempe, AZ: Education and Public Interest Center & Education Policy Research Unit.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112.*

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educational Evaluation and Policy Analysis, 24,* 175-199.

Cooper, H. M. (2009). *Research synthesis and meta-analysis: A guide for literature reviews* (4th ed.). Thousand Oaks, CA: Sage.

Cooper, H. M., Patall, E. A., & Lindsay, J. J. (2009). Research synthesis and meta-analysis. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 344-370). Thousand Oaks, CA: Sage.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). Hoboken, NJ: Wiley.

Gorey, K. M. (2001). Early childhood education: A meta-analytic affirmation of the short- and long-term benefits of educational opportunity. *School Psychology Quarterly, 16,* 9-30.

Greenland, S. (1998). Meta-analysis. In K. J. Rothman & S. Greenland (Eds.), *Modern Epidemiology* (2nd ed., pp. 643-673). Philadelphia, PA: Lippincott-Raven.

Greenland, S. (1987). Quantitative methods in the review of epidemiologic literature. *Epidemiologic Reviews, 9,* 1-30.

Konstantopoulos, S. (2008). An introduction to meta-analysis. In J. Osborne (Ed.), *Best practices in quantitative methods* (pp. 177-194). Thousand Oaks, CA: Sage.

Ludwig, J., & Phillips, D. A. (2008). Long-term effects of Head Start on low-income children. *Annals of the New York Academy of Science, 1136,* 257-268.

Nelson, G., Westhues, A., & MacLeod, J. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention and Treatment, 6,* 31.

Robin, K. B., Frede, E. C., & Barnett, W. S. (2006). *Is more better? The effects of full-day vs. half-day preschool on early school achievement.* New Brunswick, NJ: National Institute for Early Education Research.

Schutz, S. (2008). Chasing randomized trials—The research to nowhere. Comment on Camilli et al. *Teachers College Record.*

U.S. Department of Health and Human Services, Administration for Children and Families (2005). *Head Start impact study: First year findings.* Washington, DC.

Vinovskis, M. (2001). *Revitalizing federal education research and development: Improving the R & D centers, regional educational laboratories, and the "new" OERI.* Ann Arbor, MI: University of Michigan Press.

Cite This Article as: *Teachers College Record*, Date Published: December 21, 2009
http://www.tcrecord.org ID Number: 15879, Date Accessed: 3/18/2010 3:31:48 AM