

Narrative Versus Meta-Analytic Reviews: A Rejoinder to Graham's Comment

Harris Cooper and Nancy Dorr
University of Missouri–Columbia

We examine Graham's (1995) concerns about meta-analysis regarding (a) the use of poor-quality studies and (b) an overemphasis on quantitative comparisons of substantively disparate literatures. First, many meta-analysts eschew making questionable global judgments of quality so as to exclude studies on an a priori basis. Instead, they demonstrate their concern for research quality by including methods variables in a search for influences on study outcomes. Further, our meta-analysis (Cooper & Dorr, 1995) demonstrated the independence of decisions about (a) what studies to include in a review and (b) whether to use quantitative synthesis techniques by using the same evidential base Graham used for her narrative review. Second, we agree with Graham that substantively disparate literatures ought not be compared. However, we argue that literatures that might be defined as disparate for one purpose could be comparable for another. Regardless, her concern is irrelevant to our comparison of the two reviewing methods.

Graham's (1995) response to our comment provides some important addenda to both our meta-analysis (Cooper & Dorr, 1995) and her original box score review of research on race differences in motivation (Graham, 1994). With most of her points, as with her initial review, we have no contention. However, we think a few corrections should be made to her new exposition, both in regard to our work and to meta-analysis in general.

The Issue of Quality Control

Graham states that meta-analysts possess a "reluctance to distinguish between low- and high-quality studies" (p. 509) and "tend to downplay the question of poor-quality studies as a barrier to a good review" (p. 511). We disagree. Rather, we would suggest that many meta-analysts (but not all and not in every circumstance; see, for example, Ottenbacher & Cooper, 1983) eschew the process of discarding studies from consideration based on unitary, dichotomous judgments of quality. Underpinning this decision is a healthy skepticism about the reliability and objectivity of global judgments of research quality. Instead, these meta-analysts assess empirically the impact of multiple dimensions of methodology on the outcomes of studies, and thereby display a great sensitivity to issues of research quality.

Regardless, the issue of quality control is moot in the present instance. In our meta-analysis, the evidential base was identical to that constructed by Graham herself. Her quality criteria were our quality criteria. Further, we examined the

same methodological influences on study outcomes. Thus, Graham cannot claim that our meta-analysis downplayed the issue of research quality without simultaneously drawing into question her own decision criteria.

Graham (1995, p. 511) asserts that she opted for a narrative review rather than meta-analysis in part because she did not want to include studies with poor methods. Here, she has confounded inclusion criteria with the use of quantitative procedures to combine the results of studies. Our meta-analysis demonstrated that the studies Graham reviewed narratively were, in fact, amenable to meta-analysis. Thus, the larger point is that decisions about (a) what studies to include in a synthesis (based on methodological or any other criteria) and (b) whether to use meta-analytic procedures are largely independent.

Finally, we disagree with Graham over whether or not unpublished studies have a place in research synthesis. She states, "I chose to exclude dissertations because *in most cases* I judged the quality of the research design and/or conceptualization to be unacceptably low" (p. 512, italics added). Our concern is for those cases in which high-quality dissertations might have been excluded solely because of publication status. Inclusion of these studies might have bolstered or refuted Graham's published conclusions. Research can go unpublished for reasons other than quality. Publication status alone is insufficient justification for excluding a study.

What Becomes Figure and What Remains Ground

Graham acknowledges that we "documented that race differences in need for achievement (favoring Whites over Blacks) were just as reliable as those in self-concept of ability (favoring Blacks over Whites)" (p. 513). In contrast, her narrative review found "no consistent differences between Blacks and Whites in the strength of the achievement motive, whereas African Americans were equal to or higher than their White counterparts in self-concept of ability" (pp. 512–513). Graham expresses concern that our comparison across content areas might focus *substantive* attention on the relative magnitudes of these effects and thereby obscure their different theoretical literatures, subject populations, and measurement instruments. We share Graham's concern about mixing apples and oranges. However, we withhold any categorical judgment about whether or not such comparisons might prove enlightening. In some contexts and for appropriate purposes, comparing apples and oranges makes good sense. Further, we doubt whether ill-conceived comparisons are strictly the province of meta-analysis; they most certainly occur in narrative reviews as well.

Lastly, we hasten to point out that our meta-analysis was concerned with a comparison of methods, not substance. Graham herself equated the topics of need for achievement and self-concept when she established the same box score criteria for a significant finding in both areas. We accepted her equivalence and then demonstrated that the box score method and/or problems in its application led to an erroneous conclusion. That was our point of comparison, no more or less.

Conclusion

We would like to conclude by stressing a major point of agreement between Graham and ourselves. She states that "a delicate balance must be achieved between the reporting of statistics (e.g., comparing average effect sizes) and

substantive interpretation” (p. 512). We concur. And further, the relative newness of quantitative methods in research synthesis has led in some instances to their overemphasis and to their confusion with other issues of method and substance. However, meta-analytic procedures will eventually become much less intimidating and mystical. Then, the research community will be no more likely to accept the conclusions of a review of systematic empirical replications without meta-analysis than it is now likely to accept the conclusions of an empirical, two-group comparison without a *t*-test or analysis of variance.

References

- Cooper, H., & Dorr, N. (1995). Race comparisons on need for achievement: A meta-analytic alternative to Graham's narrative review. *Review of Educational Research*, 65, 483–508.
- Graham, S. (1994). Motivation in African-Americans. *Review of Educational Research*, 64, 55–117.
- Graham, S. (1995). Narrative versus meta-analytic reviews of race differences in motivation: A comment on Cooper and Dorr. *Review of Educational Research*, 65, 509–514.
- Ottensbacher, K., & Cooper, H. (1983). Drug treatments of hyperactivity. *Developmental Medicine and Child Neurology*, 25, 353–357.