Meta-analysis and Gifted Education

William Asher

This article discusses meta-analysis in the context of research in gifted education. It provides the rationale for encouraging meta-analytic reviews, rather than narrative reviews, to synthesize the research in a given area. The article summarizes the results of reported meta-analyses retrieved from electronic databases and provides guidelines for conducting a meta-analysis and suggested resources for additional information.

Introduction

Educators are professionals. Their practices are based on conceptual, theoretical knowledge. A problem is that our knowledge in the behavioral sciences and education is, for the most part, somewhat ambiguous. We rely on sampling theory to describe populations from samples and also have sampling error. Our measurement systems can never be completely accurate because our test and observation reliabilities cannot be perfect. (The Spearman-Brown formula requires an infinite set of scores or observations for the resulting reliability to be 1.00.) Thus, we will always have standard errors of measurement. In our experimental research, we rely on statistical analyses to accept or reject the null hypothesis of no effect of our treatments. In education, particularly in gifted and talented education, our sample sizes tend to be small. Also, our experimental curricular reform treatments, while advantageous, are not often spectacularly better than our current practices, which have been developed over the years by thoughtful practitioners. We have known for some time (Hedges & Olkin, 1980) that, with total samples of 70 or less and experimental group effects higher than standard educational treatments of less than three tenths of a standard deviation (about 12% of the experimental sample greater than the mean of the control group sample), the use of standard statistical analyses will likely lead the researcher to the wrong inference of the

William Asher is Professor of Education at Purdue University, West Lafayette, IN.

Journal for the Education of the Gifted. Vol. 27, No. 1, 2003, pp. 7–19.

actual evidence. Further, the more research and statistical analyses done under these circumstances, the more likely it is that the inaccurate inference will be drawn!

Given these problems with how our theories in gifted and talented education are developed, how do we as classroom teachers, curriculum developers, college faculty, and researchers make the best determination as to what is the best theory in our field? The best and most accurate (but not perfect) method is meta-analysis, which quantitatively combines the results of the reported research. Gene Glass (1976), a professor of education, developed meta-analysis. It was further developed by another professor of education, Larry Hedges (Hedges & Olkin, 1985). Meta-analysis combines the results of all the available experiments in an unbiased manner and weights them by sample size to arrive at the best estimate of the effectiveness of an educational procedure. By combining the samples from all the studies, typically the degrees of freedom for the statistical tests are so large that the problem of being unable to distinguish true differences among the experimental and standard treatments (a statistical Type II error) is negligible. Type II statistical errors are the problem that has plagued gifted and talented education research (and most all educational, psychological, and medical research) since the inception of the application of statistical analyses to these fields.

In reviewing the qualitative literature, we are able to identify likely variables that influence our selection procedures and instructional methods. In the quantitative literature, we often counted the number of studies with statistically significant results and the number with nonsignificant results and drew a conclusion. We now know that this "box-score" method of summarizing the results in the literature is flawed for two reasons. First, given the sample sizes and the typical strength of the treatments and the abundance of Type II errors, nonsignificant results are generated. Second, it has become apparent in recent years (via meta-analyses) that the editors and reviewers for journals have tended to accept for publication in our literature those results that were statistically significant, thus biasing our knowledge of the real world of education and psychology. (Meta-analysts make major efforts to obtain all the published literature, dissertation research, ERIC documents reporting research, and unpublished reports from various sources. Then, they code the studies for published versus unpublished and often find a distinct difference in the results.)

Thus, we can conclude that the published literature in the field of gifted and talented education is inherently flawed. Our prior methods of attempting to review it are also flawed. The best method of summarizing the literature is via meta-analysis, and meta-analysis methods are starting to appear in introductory research methods texts (Vockell & Asher, 1995), which should be consulted for recommendations about how to do one well.

Asher (1986) asserted that meta-analysis allowed researchers and educators in the area of gifted to overcome interpretation obstacles of imprecise measurement and small samples. He reviewed the technique and responded to critiques of the approach.

There is one other point to be made about the more traditional methods of summarizing the literature. Traditional statistical analysis methods tend to simply give "yes" or "no" results. Attempts to infer the strength of the results also tend to be flawed. Statistical significance at the *p* equals .001 and .01 levels, rather than the .05 levels, are often more the result of the size of the sample, rather than the size of the effect.

Meta-analysis yields *effect sizes* between types of treatment. Effect sizes are simply our well-known standard scores that have ranges from a mean of zero to about plus three as an upper limit and minus three as a lower limit. Typically, standard scores in our usual measurement systems are labeled *z* scores. In meta-analyses, effect sizes are labeled *d* scores or, after unbiasing, *g* scores. (Some analysts prefer transforming everything into correlations and analyzing these.) They are the standardized-difference score between the two means, the experimental-group mean minus the comparison-group mean. This result is standardized by dividing the difference between the two means by a pooled standard deviation of the two or more groups.

It is worthwhile for practitioners to examine the effect size strength, rather than just the statistical significance, of an experimental effect to determine its usefulness in an educational practice. Further, the experimental effect size and cost can be compared to that of the typical curriculum to yield a cost effectiveness comparison that is beneficial in all fields of endeavor, as well as education.

Meta-analyses are useful to researchers in another sense, as well. Meta-analyses, with their extensive, usually computerized searches of the literature, show gaps in the literature where further research needs to be done.

Meta-analysis is an exemplification of the highest standards of science. Meta-analyses are entirely replicable. The literature databases (ERIC, Psychological Abstracts, Dissertation Abstracts), recent journals hand-searched, journal articles and book reference listings searched, and personal inquiries made can be repeated and sometimes extended with more experimental studies. Typically, codings are made of the experiments' conditions and aspects of the curricula, school type, ages or grade levels of the students, grouping methods, teacher's attributes and special skills, research methods, decade when the study was done, publication type, and many aspects of the research that relate to particular theoretical interests and questions. These codings are based on described operational definitions and can be replicated independently, changed, or added to as desired.

The meta-analyses methods are fairly standard, although they have evolved over time as the methods become more sophisticated (Cooper & Hedges, 1994; Hedges & Olkin, 1985). Thus, replication and the ability to test alternative hypotheses via coding of the experimental conditions are implicit in meta-analysis. Generalizations and, thus, broadened theory are quite possible and are further enhanced because questions about possible differences among the research studies can be developed and tested explicitly. If it can be shown that studies done in urban, suburban, and rural settings do not differ in their effect sizes, the results are more generalizable. The same may be true across several grade levels or variations in the curricular treatments. Explicit statistical tests of hypotheses across the research studies are a major advantage of meta-analyses and add markedly to the ability to generalize results and theory. Further, these statistical tests are not as flawed by Type II statistical errors of inference as in the single study results because of the large degrees of freedom typically available.

The set of study results can also be analyzed for outliers beyond the overall confidence limits established over all the studies. Then these high- and low-result studies can be examined qualitatively for possible reasons their results differ so much from those that are reasonably homogeneous. It should be noted that this set of homogeneous results are really the *best evidence* available about the effectiveness of an instructional method or condition. The literature review method, which purports to be best evidence, is really a qualitative evaluation of the quality of the experimental intervention, the criteria used to assess it, and the research and analysis methods. Typically, only one or a few experiments pass the best-evidence analyst's criteria; and, therefore, all the problems of instability and variability of sampling and measurement can distort the best-evidence results.

Thus, I commend to your attention the results of a number of meta-analyses done in the area of gifted and talented education. Searching the ERIC database and using the search term "meta-analysis," combined with "gifted" or "talented," developed these. The results were 8 documents using "talented" and 25 documents using "gifted." From the PSYCHINFO database, again using "talented" with "meta-analysis" as search terms, 4 additional citations were located. However, not all were meta-analyses with statistical tests as generally defined by Glass (1976) and Cooper and Hedges (1994). The value associated with the modern technique of meta-analysis has become so prestigious that some writers have used the term *meta-analysis* to associate their work with this method. As I have indicated, since the results of these meta-analyses constitute some of the best bases for theory and practice in the field of gifted and talented, a brief summary and citations will be given for the major meta-analysis work in grouping and acceleration, pull-out programs, and self-concept.

Meta-analysis in Gifted Education

Husband and wife team James and Chen-Lin Kulik have together and independently done meta-analyses in a number of educational and psychological areas and have made major contributions to the gifted and talented literature with a series of meta-analyses on the topic of grouping and acceleration. James Kulik (1993) summarized the results of two sets of meta-analyses of five kinds of ability grouping. His work was supported by the National Research Center on the Gifted and Talented at the University of Connecticut. He analyzed (a) high-, middle-, and low-class groupings; (b) cross-grade groupings; (c) within-class groupings; (d) accelerated classes; and (e) enriched classes. He concluded that, in general, the groupings with the strongest benefits were those with a great deal of adjustment of the curriculum for highly talented learners. As a rule, no grouping method in and of itself provided strong benefits as such, but high-aptitude students usually benefited academically. Kulik (1992) essentially reported the same analysis as above, except in more detail with respect to the grade equivalent for the various types of groups for the many levels of ability.

Using 31 studies, C.-L. Kulik and J. Kulik (1984) did another meta-analysis on ability grouping with elementary school students. The meta-analytic results indicated that ability grouping does have a statistically significant, positive effect on academic performance, though it is small. However, when high-ability students were put in special classes with enriched instruction, the effect sizes were moderate for achievement. The effect on self-concepts was trivial.

C.-L. Kulik (1985) reported the effects of interclass ability grouping on achievement and self-esteem. She analyzed 85 studies from both elementary and secondary schools. For general achievement, the effect size of .15 is not large, but, as the result of her ability to code types of features across studies, she could determine that honors programs designed for talented students had clear, positive effects. Self-esteem results were near zero. These results clearly show the major advantages of a quantitative meta-analysis. They give the strength of the results and not merely the statistical significance, or lack thereof. Further, various conditions across studies can be systematically evaluated, as well.

J. Kulik and C.- L. Kulik (1992) reported on meta-analytic findings on gifted and talented ability-grouping programs: multilevel classes, cross-grade grouping, within-class grouping, enriched classes, and accelerated classes. Multilevel classes have had little or no effect on achievement. Enrichment and acceleration with the largest degree of curricular adjustment have had the largest effect on student learning. Further, they could determine that students of lower aptitude are not harmed academically or emotionally. They also reported that gifted students gained little from programs with minimal instructional modification (multilevel classes), more from cross-grade and within-class modifications, and the most from enrichment and acceleration. Again, the advantage of the meta-analysis was in presenting the strength of the effects, not merely their statistical significance. (As noted, even these statistical results are flawed as the result of small sample sizes typical in gifted and talented educational research.)

J. Kulik and C.-L. Kulik (1984a, 1984b) conducted a meta-analysis of 26 studies that investigated the effects of accelerated instruction on public school students. As part of the meta-analysis, the Kuliks coded eight variables that described the features of each study. Accelerants outperformed nonaccelerants of equivalent age and intelligence by nearly one grade level academically. Their nonintellectual outcomes were not consistent among the few analyses conducted. Their meta-analysis of educational acceleration was also summarized by Higham (1984) in an article in the first issue of *Academic Talent*. The first volume of this newsletter published by The Johns Hopkins University Center for the Advancement of Academically Talented Youth (1985) was also published as an ERIC document.

Vaughn, Feldhusen, and Asher (1991) reported a meta-analysis reviewing nine studies of pull-out programs for gifted students in grades one through nine. They concluded that there were significant, positive effects on achievement, critical thinking, and creativity, but not self-concept.

Goldring (1990) reported a meta-analysis of the effects on achievement of gifted students in special homogeneous classes. The gifted students in the special classes achieved more.

Hoge and Renzulli (1993) explored via a meta-analysis the differences in the self-concepts of gifted and nongifted children. They investigated the effect of labeling children as gifted and then placing them in special programs with respect to their self-concept. They noted that gifted students generally have higher academic self-concepts

Rogers (1991a), also supported by the National Research Center on the Gifted and Talented, summarized grouping practices reported in 13 research studies on the effects of ability grouping; mixed-ability, cooperative grouping; and grouping for acceleration. She reported her summarization methods as best-evidence and meta-analysis. However, the meta-analysis method was clearly not a meta-analysis in the usual sense with statistical tests, and concerns about best evidence have been expressed above. She reported that the research is scant and weak for socialization and psychological adjustment, which did show where there are gaps in the literature. Rogers (1991b) essentially reported the same work, but added a series of guidelines for practices. Rogers (1988), again not really using meta-analysis, but a broad-ranging survey of research publications, examined the types of research used to study cognitive processing development in the gifted. In the 522 research publications she reviewed, the most frequent designs were correlational and case studies.

Lynch (1986) reported the results of the U.S. Department of Education's Joint Dissemination Review Panel (JDRP), which reported 165 effect sizes (out of 232 submissions) for various educational programs and interventions, content areas, classroom settings, and types of students. Large effect sizes were reported for gifted participants. However, they cautioned that effect-size data should not be interpreted simplistically. Several tables were provided.

Gagné (1989) analyzed 13 validation studies of the use of peer nomination forms to identify gifted students. He suggested that there were methodological weaknesses in current procedures and proposed a four-step plan of research to improve them. Begin and Gagné (1994) reviewed 50 variables in the literature that predicted people's attitudes toward gifted education. No single variable was a major explanatory factor. Again, this work was not a meta-analysis as such, but was, instead, a narrative review.

The Shaughnessy (1984) report also was not a meta-analysis, but a narrative literature review of motivation for the gifted and talented using the theories of Ellis, Maslow, Skinner, Piaget, Bloom, Rogers, May, and Frankl. Shaughnessy (1985, 1990) again narratively reviewed the literature of cognitive structures of the gifted and made an attempt at a meta-analysis of Sternberg's triarchic model and Gardner's six types of intelligence.

Johnson, Johnson, and Smith (1997) reported about enriching college instruction through intellectual conflict. This research summary had little to do with meta-analysis. Menchaca and Ruiz-Escalante (1995) suggested instructional strategies for migrant children. Again, any meta-analyses played a minor role in these suggestions. Haury (1993) was concerned with teaching science via inquiry. The report had little to do with meta-analysis results as such. Ellis (1984) summarized research on class size. He cited the meta-analysis done by Glass, McGaw, and Smith (1981), which had little to do with the gifted. Friedman (1995) conducted a meta-analysis of gender differences and correlations on the space variables in mathematics. With gifted youth, the correlations were higher with females than with males.

Discussion

So, in these studies, it is reasonable to say that the meta-analysts reported the best evidence available from the extensive literature in gifted and talented education to form bases for a theory of gifted and talented instruction. Narrative reviews of a selected few studies in a subarea are bound to have larger variances because of the smaller samples, and they are also more prone to the subjective selection biases of the reviewer. These meta-analysts performed powerful statistical tests across the set of studies using coding of the variables to test alternative hypotheses about ability to generalize across variations in instruction, curricula, age-grade groups, and other variables.

These meta-analysts also met two of the major concerns of good science: replication and alternative hypotheses tested. The studies involved in a meta-analysis can be retrieved, reanalyzed, new codes of across-studies added and tested, and new studies added—all with powerful statistical analyses. The methods are relatively straightforward, though not without complexities. For the calculations of effect sizes from obscurely reported statistics in a publication, see Glass et al. (1981, Chapter 5). See Hedges and Olkin (1985) for modern methods of statistical analyses. The use of computer programs, such as

DSTAT (Johnson, 1993), for pooling standard deviations and so forth is useful. *Comprehensive* meta-analysis by Borenstein and Rothstein (1999) facilitates the unbiasing of raw statistics from the publications, the weighting of the data, and the now-recommended random model for statistical tests and confidence limits. (These unbiasing methods, weighting of the various original sample sizes, and random model statistical analyses were often not available in the earlier meta-analyses.) An excellent summary of the many advantages, concerns, and issues in the use of meta-analysis is found in the *Handbook of Research Synthesis* (Cooper & Hedges, 1994).

The results of meta-analyses are valuable for those who are making decisions about gifted and talented programs in the schools, for those who teach in higher education who wish to present the best available theories of methods in gifted and talented education, and for researchers who are interested in gaps in the literature and new areas where further research and meta-analyses can be done. These researchers can also add new variable codings and studies to prior meta-analyses to further strengthen them and to test alternative hypotheses in the interest of advancing the theory of gifted and talented education. Graduate students can do meta-analyses with a class in the method and help from faculty experienced in the nuances of the methods and those knowledgeable in the field of gifted and talented. (One major advantage of this work is that no human subjects committee approval is required, just many long hours in the library.)

Major advances in the theory of instruction of the gifted and talented have been made via the use of meta-analyses, as indicated by the extensive literature presented here. This is an ongoing effort in which many in the field should be involved.

Asher (1991) gave educational practitioners a practical guide to read, interpret, and conduct meta-analysis research critically in gifted education. This chapter also gives calculation, coding, and analysis examples. In general, define your problem and search all of the literature to find all of the experiments related to the problem. Retrieve them. Extract the effect sizes. Code the qualitative aspects of the studies to allow analyses of possible differences among the studies. Have a second person also code the studies and check the replicability of your codings. If there are a few or none, then the overall results of the meta-analysis are more generalizable. Note that the degrees of freedom of the statistical tests across all the studies will be quite large. Thus, the power of the tests to reject the null hypotheses is quite good. There should be little acceptance of "no differences" by default. Typically, there will be several operational definitions of concepts and variables and data collection methods across the studies. Therefore, the generalizability of the results of the meta-analysis will be greater than any single study.

As a word of caution, note that a meta-analysis can be poorly done; therefore, the methodology should be checked. Further, the results of the process of meta-analysis has become so prestigious and fashionable that studies labeled as such often are not meta-analyses.

References

Asher, J. W. (1986). Conducting research with meta-analysis: A new direction for gifted education. *Gifted Child Quarterly, 30*, 7–10.

Asher, J. W. (1991). Meta-analysis. In N. K. Buchanan & J. F. Feldhusen (Eds.), *Conducting research and evaluation in gifted education: A handbook of methods and applications* (pp. 220–241). New York: Teachers College Press.

Begin, J., & Gagné, F. (1994). Predictors of attitudes toward gifted education: A review of the literature and a blueprint for future research. *Journal for the Education of the Gifted, 17*, 161–179.

Borenstein, M., & Rothstein, H. (1999). Comprehensive meta-analysis: A computer program for research synthesis (Version 1.0.16) [Computer Software]. Englewood, NJ: Biostat.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

Ellis, T. I. (1984). *Class size* (ERIC Digest, Number 11). Eugene, OR: ERIC Clearinghouse on Educational Management, University of Oregon, Center for Advanced Technology. (ERIC Document Reproduction Service No. ED259454)

Friedman, L. (1995). The space factor in mathematics: Gender differences. *Review of Educational Research, 65*, 22–50.

Gagné, F. (1989). Peer nominations as a psychometric instrument: Many questions asked but few answered. *Gifted Child Quarterly, 33*, 53–58.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3–8.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Goldring, E. B. (1990). Assessing the status of information on classroom organizational frameworks for gifted students. *Journal of Educational Research, 83*, 313–326.

Haury, D. L. (1993). *Teaching science through inquiry* (ERIC/CSMEE Digest). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education. (ERIC Document Reproduction Service No. ED359048)

Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis. *Psychological Bulletin, 88*, 359–364.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Higham, S. (1984). A summary of a meta-analysis of educational acceleration. *Academic Talent, 1*(1), 5–6.

Hoge, R. D., & Renzulli, J. S. (1993). Exploring the link between giftedness and self-concept. *Review of Educational Research, 63*, 449–465.

Johns Hopkins University, Center for the Advancement of Academically Talented Youth. (1985). *Academic Talent, 1(1–4), Winter 1984–Fall 1985*. (ERIC Document Reproduction Service No. ED264690)

Johnson, B. T. (1993). Software for meta-analytic review of research literatures (Version 1.11) [Computer software]. Hillsdale, NJ: Erlbaum.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1997). *Academic controversy: Enriching college instruction through intellectual conflict* (ERIC Digest). Washington, DC: ERIC Clearinghouse on Higher Education, George Washington University. (ERIC Document Reproduction Service No. ED409828)

Kulik, C.-L. C. (1985, August). *Effects of interclass ability grouping on achievement and self-esteem*. Paper presented at the meeting of the American Psychological Association, Los Angeles. (ERIC Document Reproduction Service No. ED263492)

Kulik, C.-L. C., & Kulik, J. A. (1984, August). *Effects of ability grouping on elementary school pupils: A meta-analysis*. Paper presented at the meeting of the American Psychological Association, Toronto. (ERIC Document Reproduction Service No. ED255329)

Kulik, J. A. (1992). *An analysis of the research on ability grouping: Historical and contemporary perspectives* (Research-based decision making series). Storrs: University of Connecticut, National Research Center on the Gifted and Talented. (ERIC Document Reproduction Service No. ED350777)

Kulik, J. A. (1993, Spring). An analysis of the research on ability grouping. *National Research Center on the Gifted and Talented Newsletter, 3*, 8–9. (ERIC Document Reproduction Service No. ED367095)

Kulik, J. A., & Kulik, C.-L. C. (1984a). Effects of accelerated instruction on students. *Review of Educational Research, 54*, 409–425.
Kulik, J. A., & Kulik, C.-L. C. (1984b). Synthesis of research on effects of accelerated instruction. *Educational Leadership, 42*, 84–89.
Kulik, J. A., & Kulik, C.-L. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36*, 73–77.
Lynch, K. B. (1986, April). *Effect sizes of programs applying to the joint dissemination review panel*. Paper presented at the meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED273656)
Menchaca, V. D., & Ruiz-Escalante, J. A. (1995). *Instructional strategies for migrant students* (ERIC Digest). Charleston, WV: ERIC Clearinghouse on Rural and Small Schools. (ERIC Document Reproduction Service No. ED388491)
Rogers, K. B. (1988, April). *Research methodology in the cognitive development of gifted learners: A preliminary analysis*. Paper presented at the meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED303938)
Rogers, K. B. (1991a). *The relationship of grouping practices to the education of the gifted and talented learner: Executive summary*. Storrs: University of Connecticut, National Research Center on the Gifted and Talented. (ERIC Document Reproduction Service No. ED343329)
Rogers, K. B. (1991b). *The relationship of grouping practices to the education of the gifted and talented learner*. Research-based decision making series. Storrs: University of Connecticut, National Research Center on the Gifted and Talented. (ERIC Document Reproduction Service No. ED343330)
Shaughnessy, M. F. (1984). *Motivating the gifted, talented, creative. A meta-analysis*. (ERIC Document Reproduction Service No. ED288313)
Shaughnessy, M. F. (1985, August). *Cognitive structures of the gifted: Theoretical perspectives, factor analysis, triarchic theories of intelligence, and insight issues*. Paper presented at the Sixth World Conference on Gifted and Talented Children, Hamburg, West Germany. (ERIC Document Reproduction Service No. ED268746)
Shaughnessy, M. F. (1990). Cognitive structures of the gifted: Theoretical perspectives, factor analysis, triarchic theories of intelligence, and insight issues. *Gifted Education International, 6*, 149–151.

Vaughn, V. L., Feldhusen, J. F., & Asher, J. W. (1991). Meta-analyses and review of research on pull-out programs in gifted education. *Gifted Child Quarterly, 35*, 92–98.

Vockell, E. L., & Asher, J. W. (1995). *Educational research* (2nd ed.). Englewood Cliffs, NJ: Macmillan.