DOCUMENT RESUME

ED 251 722 CG 017 850

AUTHOR Schmidt, Frank L.
TITLE Meta-Analysis: Implications for Cumulative Knowledge in the Behavioral and Social Sciences.
PUB DATE 25 Aug 84
NOTE 14p.; An invited address at the Annual Convention of the American Psychological Association (92nd, Toronto, Ontario, Canada, August 24-28, 1984).
PUB TYPE Viewpoints (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Meta Analysis; *Psychology; Research Methodology; Research Problems; *Social Science Research

ABSTRACT

The most important problem in psychology and the social sciences today is the failure to produce cumulative knowledge. This situation has led many to conclude that cumulative knowledge and general principles and theories may be impossible to establish in psychology, a conclusion suggesting that psychology can never be a science--only at best a technology producing answers limited in their applicability to specific situations. This notion has contributed to reductions in funding for research. Psychological researchers have relied on a general two-step procedure to produce cumulative knowledge. First, individual scientists conduct numerous studies; second these studies are integrated subjectively and non-quantitatively, published, and reviewed. However, the information processing task in reviewing a body of studies so that general principles can be drawn from them may be too complex for the unaided human mind. The most important recent development, therefore, is the development of quantitative methods which lift the information processing burden from the reviewer by quantitatively integrating findings across studies, while simultaneously correcting for the effects of statistical and measurement artifacts which distort study findings. Meta-analysis, the term for this process, corrects for the distorting effects of artifacts such as sampling error, measurement unreliability, and range restrictions. The response to these new methods by researchers has been gratifying, as evidenced by meta-analytic studies in journals, books and convention presentations. (JAC)

Meta-Analysis: Implications for Cumulative Knowledge in the Behavioral and Social Sciences*

Frank L. Schmidt
U.S. Office of Personnel Management
and
George Washington University

The most important problem in psychology and the social sciences today is the failure to produce cumulative knowledge. In one research area after another, psychologists have conducted numerous studies, and in almost every case it has been found that different studies give different results. Some studies report significant findings, some do not. Some studies support the hypothesis, some do not. This situation has led many to conclude that cumulative knoweldge and general principles and theories may be impossible to establish in psychology and the social sciences.

If this conclusion is true, then psychology can never be a science--only at best a technology producing answers limited in their applicability to specific situations. It follows then that psychological research can never provide solutions to broad social problems business, education, and society--as had been expected by research funding organizations such as foundations and government agencies. The resulting frustration is expressed in a speech given in 1970 by Walter Mondale:

> What I have not learned is what we should do about these problems. I had hoped to find research to support or to conclusively oppose my belief that quality integrated education is the most promising approach. But I have found very little conclusive evidence. For

---

*Invited Division 5 address, 92nd Annual APA Convention, Toronto Canada, Aug. 25, 1984.

> every study, statistical or theoretical, that contains a proposed solution or recommendation, there is always another, equally well documented, challenging the assumptions or conclusions of the first. No one seems to agree with anyone else's approach. But more distressing I must confess, I stand with my colleagues confused and often disheartened.

Our continuing inability to produce cumulative knowledge and general principles has, I believe, led to widespread disillusionment with psychological research and consequent reductions in funding for our research by governmental and other bodies.

Psychological researchers have relied on a general two-step procedure to produce cumulative knowledge. First, individual scientists conduct numerous studies on a given hypothesis--the hypothesis that perceptual speed predicts success in clerical work, for example, or the hypothesis that reducing class size increases pupil learning. Second, this body of studies is integrated subjectively and non-quantitatively in a narrative review which is published in a review journal such as Psychological Bulletin. This paradigm for the production of knowledge has failed. Almost invariably, the reviewer reports that there are inconsistencies, conflicts, and contradictions between findings of different studies. The reviewer then concludes that further research is needed and that conclusions about general principles are not yet possible.

Some have hypothesized that the conflicting findings mean that the phenomena we study are much more complex than we originally thought, and that the complexity of findings is due to the presence of many unsuspected interactions (i.e., moderator variables). But there is another, more

parsimonious possibility: the possibility that the information processing task imposed in reviewing a body of studies is too complex for the unaided human mind. The reviewer must simultaneously execute two very difficult tasks. He or she must subjectively integrate findings across numerous studies--sometimes hundreds of studies. At the same time, he or she must keep in mind and allow for the effects of statistical and methodological artifacts which distort study findings--for example, sampling error, measurement unreliability, and restriction in range. These tasks are beyond the unaided information processing capacities of any human being.

Failure to produce cumulative knowledge is the most important problem today in psychology and the social sciences. The most important recent development is therefore the development of quantitative methods which solve this problem. These methods lift the information processing burden from the reviewer by quantitatively integrating findings across studies while simultaneously correcting for the effects of statistical and measurement artifacts which distort study findings. As Glass (1976) has pointed out, these techniques represent the belated application of the same statistical methods that we routinely use to analyze data within studies to the problem of integrating findings across studies. Application of these methods can clarify confusing research literatures and allow the establishment of general principles.

Gene Glass at the University of Colorado was the first to systematically address this problem. He also originated the term "meta-analysis". He and his associates advanced a meta-analytic method composed of the following steps:

1. All effect sizes are expressed in SD units (or alternatively, in correlation form).

2. The mean effect size across studies is computed. This represents the expected magnitude of a treatment condition or the expected size of a correlation. For example, Smith and Glass (1977) found psychotherapy has an average effect size of .68 standard deviation units of the control group. White (1976) found that the mean correlation between SES and academic achievement is .25.
3. Properties or characteristics on which studies differ are coded and then correlated with effect sizes in an effort to find the causes of differences between studies in reported effect sizes. Studies may differ on age or sex of subjects, methodology used, and other variables. Although numerous variables are typically coded, the general finding has been that few are correlated with study outcomes. Because the sample size for this type of analysis is the number of studies--not the number of people--there are often severe problems of capitalization on chance and low statistical power. For example, there may be 70 studies and 50 study characteristics.

Glass and his associates have applied his methods of meta-analysis to a varitey of heretofore confused research literatures. In almost every case, the research literature has been clarified and general principles have been established. As one example, Glass and Smith (1979) have applied meta-analysis to the vast, conflicting, and heretofore uninterpretable literature on the effects of class size on pupil achievement. Based on 725 studies, their results revealed a very definite monotonic relation between class size and achievement, with the achievement difference ranging up to .90 SD units for the smallest (N = 1) vs. the largest (N = 40) classes. Further, the effect sizes were larger for the better controlled studies.

This is the kind of generalization that is needed as a foundation for both theory development and social policy decision-making.

Concurrently with Glass' work and independently of it, John Hunter and I and our associates developed our own meta-analysis procedures. Our procedures extend meta-analysis by providing methods of correcting for the distorting effects of artifacts such as sampling error, measurement unreliability, and range restriction while integrating findings across studies (Schmidt and Hunter, 1977; Schmidt, Hunter, Pearlman, and Shane, 1979; Pearlman et al., 1980). Methodological contributions in this area were also made by Callender and Osburn (1980) and Raju and Burke (1983). This procedure was originally developed to integrate employment test validities across studies, but has since been generalized for application to all research areas (Hunter, 1979; Hunter, Schmidt, and Jackson, 1980). Steps in this procedure are as follows:

1. Effect sizes are expressed as correlations or d-values and the average effect size is computed across studies. This mean effect size is then corrected for the attenuating effects of instrument unreliability and range restriction. This is a step not included in Glassian meta-analysis.
2. One then determines whether the variance in effect sizes across studies is due solely to statistical and measurement artifacts. This step is also not included in Glassian meta-analysis. If one can reject the hypotheses that the observed variance of effect sizes is greater than the variance expected from artifacts, one concludes that the mean corrected effect size estimates the true effect size, and a general principle has been established. The

mean corrected effect size then incorporates and summarizes the results of all previous studies.

3. If one cannot reject the hypotheses that the variance of effect sizes is greater than the expected from artifacts, one then determines whether any of the study characteristics are correlated with effect size. This step we borrowed from Glass and his associates (while recognizing and warning against the severe problems of capitalization on change and low statistical power).
4. If the remaining variance is still too large to be accounted for by artifacts, it is adjusted for the effects of these artifacts, and this adjusted variance is used to set confidence or credibility intervals around the mean effect size. Again, this is a step not included in Glassian meta-analysis.

To date, this procedure has been applied to over 500 research literatures in employment selection, each one representing a predictor-job performance combination. These predictors have included nontest procedures such as evaluations of education and experience and interviews, as well as ability and aptitude tests. In many cases, artifacts accounted for all variance across studies; the average amount of variance accounted for by artifacts has been approximately 80%. As an example, consider the relation between quantitative ability and overall job performance in clerical jobs (Pearlman et al., 1980). This substudy was based on 453 correlations computed on a total of 39,584 people. 77% of the variance in observed validities was traceable to artifacts, leaving a neglible variance of .019. The mean effect size was .47. Thus integration of this massive amount of data leads to the general and generalizable principle that the correlation between quantitative ability

and clerical performance is .47, with very little if any true variation around this value. Like our other findings, this finding shows the old belief that validities are situationally specific to be false. Cumulative, generalizable knowledge is possible.

Current Status of Meta-Analysis

The response to these new methods by researchers has been very gratifying. Here is some evidence of this:

1. Meta-analytic studies are appearing more and more frequently in psychological and educational journals--and not just in review journals like Psychological Bulletin. They are also appearing in the primary research journals--even though meta-analysis are reviews. Meta-analysis appears to have increased the status of reviews.
2. Hardly a week goes by that I do not receive several meta-analysis studies from journal editors to review. I'm sure this is true of other reviewers also.
3. There are now four books on meta-analysis:
   A. Glass, McGaw, and Smith (1981)
   B. Hunter, Schmidt, and Jackson (1982)
   C. Cooper (1984)
   D. Rosenthal (1984)
5. A chapter on meta-analysis has appeared in the Annual Review of Psychology (Green & Hall, 1984).
6. A count revealed that there are at least 26 meta-analysis presentations on the program at this convention (not counting this one). This count includes only those with meta-analysis in the title and is therefore almost certainly on the low side.

8. Our meta-analysis procedures have now been applied to numerous topics outside the area of validity generalization in employment selection. Some examples include:
   1. Correlates of Role Conflict and role ambiguity (Fisher, Gettelson, 1984, and Jackson & Schuler, in press).
   2. Effects of realistic job previews (Premack and Wanous, 1984; Cascio and McEvoy, 1984).
   3. Evaluation of Fielder's theory of leadership (Peters, et al., 1984).
   4. Accuracy of self-ratings of ability and skill (Mabe & West, 1982).
   5. Relation of LSAT scores to performance in law schools (Linn & Dunbar, 1981).
   6. Relation of job satisfaction to absenteeism (Terborg, et al., 1982).
   7. Ability of financial analysts to predict stock growth (Coggin & Hunter, 1983).
   8. Premorbid functioning and recidivism in Schizophrenia (Stoffelmeyr, Dillavou, & Hunter, 1983).
9. Examples of Application of Glassian meta-analysis methods include:
   1. Effectiveness of computer assisted instructions (Kubik, Cohen, & Ebeling, 1979).
   2. Effects of social facilitation on task performance (Bond & Titus, 1983).
   3. Sex differences in influenceability (Eagly & Carli, 1981).
   4. Effects of psychotherapy (Smith & Glass, 1977; Shapiro & Shapiro, 1982).

5. Sex bias in counseling and psychotherapy (Smith, 1980).
6. Effects of drug therapy on psychological disorders (Miller, 1977).
7. Sex differences in ability to use nonverbal cues (Hall, 1978).
8. Class size and student achievement (Glass & Smith, 1979).
9. Effects on cooperation and competition on group achievement (Johnson, et al., 1981).

In a talk on the need for meta-analysis that I presented four years ago at the 1980 APA convention, I made the following statement:

"At one time in history of psychology and the social sciences, the pressing need was for more empirical studies examining the problem in question. But now large numbers of research studies have accumulated on many research questions. The need today is increasingly becoming not additional empirical data but some means of making sense of the vast amounts of data that have accumulated. Unless we can do this, there is little hope of producing the cumulative generalizable knowledge essential for theory development and for the solution of social problems. Quantitative techniques like those described here provide a solution to this problem. Their application and exploitation is therefore the most pressing research need of the 1980's."

Based on developments since then, I believe we are on the way toward meeting that need.

References

Bond, C. F., & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. Psychological Bulletin, 94, 265-292.

Callender, J. C., & Osburn, H. G. (1980). Development and test of a new model for validity generalization. Journal of Applied Psychology, 65, 543-558.

Cascio, W. F., & McEvoy, G. M. Extension of utility analysis to turnover reduction strategies. Presented in the Symposium "Overcoming the Futilities of Utility Applications. (Steven Wroten, Chair) American Psychological Association Convention, Toronto, Canada, August 1984.

Coggin, T. D., & Hunter, J. E. (1983). Problems in measuring the quality of investment information: The perils of the information coefficient. Financial Analysts Journal, May/June, 3-10.

Cooper, H. (1984). The integrative research review: A social science approach. Beverly Hills: Sage.

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

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-Analysis in Social Research. Beverly Hills, Calif.: Sage Publications.

Green, B. F., & Hall, J. A. (1984). Quantitative methods for literature reviews. Annual Review of Psychology, 35, 37-53.

Glass, G. V., & Smith, M. L. Meta-analysis of research on the relationship of class-size and achievement. Evaluation and Policy Analysis, 1979, 1, 2-15.

Hall, J. A. Gender effects in decoding nonverbal cues. Psychological Bulletin, 1978, 85, 845-857.

Hunter, J. E., & Schmidt, F. L. (1981). Cumulating findings across studies: Correction for sampling error, a proposed moratorium on the significance test, and critique of current multivariate reporting pratices. Invited address, American Psychological Association Convention, 1981.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). Meta-Analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage Publications.

Eagly, A. H., Carli, L. L. 1981. Sex of researchers and sex-typical communications as determinants of sex differences in influenceability: a meta-analysis of social influence studies. Psycholgical Bulletin, 90: 1-20.

Fisher, C. D., & Gitelson, R. (1983). A meta-analysis of the correlates of role conflict and ambiguity. Journal of Applied Psychology, 68, 320-333.

Jackson, S. E., & Schuler, R. S. (in press). A meta-analysis and conceptual critique of research on role abmiguity and role conflict in work settings. Organizational Behavior and Human Performance.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., Skow, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: a meta-analysis. Psychological Bulletin. 87, 47-62.

Kulik, J. A., Cohen, P. A., & Ebeling, B. J. Effectiveness of programmed instruction in higher education: A meta-analysis of findings. Educational Evaluation and Policy Analysis, in press.

Linn, R. L., Harnisch, D., & Dunbar, S. B. (1981). Validity generalization and situational specificity: An analysis of the prediction of first year grades in law school. Applied Psychological Measurement, 5, 281-289.

Mabe, P. A., III, & West, S. G. Validity of self evaluations of ability: a review and meta-analysis. Journal of Applied Psychology, 1982, 67, 280-296.

Miller, T. L. The effects of drug therapy on psychological disorders. Ph.D. Dissertation, University of Colorado, 1977.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict training success and job proficiency in clerical occupations. Journal of Applied Psychology, 65, 373-406.

Peters, L. H., Harthe, D., & Pohlman, J. (1984). Fieldler's contigency theory of leadership: an application of the meta-analysis procedures of Schmidt and Hunter. Paper under review for publication.

Raju, N. S., & Burke, M. J. (1983). Two new procedures for studying validity generalization. Journal of Applied Psychology, 68, 382-395.

Rosenthal, R. (1984). Meta-analysis procedures for social research. Beverly Hills, CA: Sage.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 62, 529-540.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Shane, G. S. (1979). Further tests of the Schmidt-Hunter Bayesian validity generalization procedure. Personnel Psychology, 32, 257-281.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences and validity of aptitude tests in selection: A red herring. Journal of Applied Psychology, 66, 166-185.

Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old Theories and New Research Findings. American Psychologist, 36, 1128-1137.

Shapiro, D. A., Shapiro, D. (1982). Meta-analysis of comparative therapy outcomes studies: a replication and refinement. Psychological Bulletin. 92, 581-604.

Smith, M. L. 1980. Sex bias in counseling and psychotherapy. Psychological Bulletin. 87:392-407.

Smith, M. L., Glass, G. V., & Miller, T. I. 1980. The Benefits of Psychotherapy. Baltimore: Johns Hopkins Univ. Press.

Smith, M. L., & Glass, G. V. Meta-analyses of psychotherapy outcome studies. American Psychologist, 1977, 32, 752-760.

Stoffelmayr, B. E., Dillavou, D., & Hunter, J. E. (1983). Premorbid functioning and outcome in schizophreniz: A cumulative analysis. Journal of Consulting and Clinical Psychology, 51, 338-352.

Terborg, J. R., & Lee, T. W. Extension of the Schmidt-Hunter validity generalization procedure to the prediction of absenteeism behavior from knowledge of job satisfaction and organizational commitment. Journal of Applied Psychology, 1982, 67, 280-296.

Walberg, H. J., & Haertel, E. H. (Guest Editors). Research integration: The state of the art. Special Issue of Evaluation in Education, 1980, 4, 1-42.

Wanous, J. P., & Premack, S. (1984). Meta-analysis of realistic job preview experiments. Presented as part of the Symposium, "Use of meta-analysis in industrial/organizational psychology." Hannah R. Hirsh, Chair. The 92nd Annual APA Convention, Toronto, Canada, August 27.

White, K. R. 1982. The relation between socioeconomic status and academic achievement. Psychological Bulletin. 91, 61-81.