Citation:

Czienskowski, U. (2003). Meta-analysis – not just research synthesis. In R. Schulze, H. Holling & D. Böhning (Eds.), *Meta-analysis: New developments and applications in medical and social sciences* (pp. 141-152). Hogrefe & Huber.

9

Meta-Analysis – Not Just Research Synthesis!

Uwe Czienskowski

Max Planck Institute for Human Development
Berlin

Summary

Meta-analysis represents an advanced methodological approach to the (quantitative) synthesis of different studies within a research field. However, meta-analytical integration is mostly not pursued further after several moderators have been identified that are responsible for much of the heterogeneity of results across primary research. In this chapter, the necessity of completing a meta-analytical integration of previous research by independently conducting primary research is stressed. It is shown that this approach to meta-analysis allows one to distinguish between merely potential moderators and real ones. This approach particularly considers meta-analysis a tool for the generation of new hypotheses as well as for the design of precise and sensitive decision studies. As an example, research on the self-reference memory effect is presented to demonstrate how to use meta-analysis not only to integrate a research field, but also to identify theoretical and empirical shortcomings within primary research. Discussing several possible objections against meta-analysis, it is concluded that meta-analysis, if adequately conducted and interpreted, is not only a tool for research integration, but may also be used in a theoretically fruitful way.

9.1 INTRODUCTION

During the last decades, meta-analysis has frequently been proven to be a collection of useful statistical techniques for the quantitative integration of results from different fields (e.g., Cooper & Hedges, 1994; Hedges, 1992). In comparison with other approaches to assessing the state of research in a specific area (e.g., narrative review, simple vote counting; see Bushman, 1994) meta-analysis relies upon statistical indices that represent the magnitude of an empirical effect, investigated by means of, for example, experimental or correlational techniques on a common scale of measurement that is independent of a study's sample size. Using the notion of this so-called *effect size*(ES; e.g., Richardson, 1996; Rosenthal, 1994; Tatsuoka, 1993), different meta-analytical approaches have been developed, depending not only on the type of ES but also on the kind of investigation predominantly used within a research field. For example, collections of studies using Cohen's *d* (standardized difference of means; see Cohen, 1988) as ES for experiments or quasi-experiments may be integrated using procedures described by Hedges and Olkin (1985), whereas research described best by variance-compound-directed ESs, for example, estimated ω^2 (Hays, 1994) or the correlation coefficient *r*, might be integrated by applying a "psychometric" meta-analysis as described by Hunter and Schmidt (1990; see also Johnson, Mullen, & Salas, 1995; Schmidt & Hunter, 1999). Furthermore, different procedures for the combination of ESs for categorical data, for example, *rate ratios* or *odds ratios* (see Fleiss, 1994), are widely used in medicine and epidemiology (e.g., Petitti, 1994). In general, meta-analytical integration is directed to present an average ES for a group of studies investigating the same empirical effect. In the simplest case, the mean ES for *i* studies to be integrated can be computed as a sum of *i* weighted ESs, divided by the sum of *i* weights (Shadish & Haddock, 1994). It is generally assumed that there are only two possible sources of variation of ESs: Variation can occur by chance if all studies share a common population ES, and additional systematic variation between studies can arise if they do not. In the latter case, categorical variables (moderators) are investigated to determine if they are responsible for this systematic variation. This is known as the moderator analysis approach. Another strategy to cope with unexpected systematic variation is known as the random effects model. This statistical model does not assume that each study effect estimates the same population effect, but rather that each single effect represents a random variable with its own distribution (e.g., Raudenbush, 1994; Shadish & Haddock, 1994).

In this chapter, the moderator analysis approach will be discussed in more detail. It will be shown that an independent investigation of moderators is necessary to cope with uncertainty of the state of potential moderators, especially if data from experimental studies have been integrated meta-analytically. The usefulness of directly manipulating moderating variables in subsequent experimentation has already been demonstrated, for example, in a study conducted by Bornstein, Kale, and Cornell (1990; see also Eagly & Wood, 1994). However, although this study has been inspired by a previous meta-analysis (Bornstein,

1989), it is not directed to an experimental evaluation of moderators that have been drawn from meta-analytical integration. Furthermore, the use of research syntheses for theoretical progress has been discussed extensively by Cook et al. (1992) as well as Miller and Pollock (1994).

In this chapter, it will be demonstrated how information on a tentatively hypothesized moderating variable can be used for a direct evaluation of its actual meaning. In addition, it will be shown that by tying meta-analysis to primary experimental research, more general problems of meta-analytic approaches can be solved in a simple way. Most important, it will be demonstrated how meta-analysis as well as primary experimental research inherit the specific advantages of each other by this link, and how this link may lead to theoretical progress that cannot be obtained without the interplay of meta-analytical integration and experimental validation.

9.2 MODERATORS IN RESEARCH INTEGRATION: AN EXAMPLE

Suppose we conduct a meta-analysis on a specific memory effect that has been investigated in, say, 72 different experiments. Most of these experiments support the idea that presenting an orientation task like "Does the following word describe you?" leads to better recall for subsequently presented words than the orientation task "Does the following word describe Bill Clinton?". This recall difference is known as the so-called "self-reference effect" in memory (SRE). Suppose further that our meta-analysis supports the conclusion that the first condition (self-reference) actually does result in better memory performance than the second (other-reference). The average ES for this comparison is about $r = .25$ (we conducted a meta-analysis following the approach of Hunter & Schmidt, 1990). Our analysis reveals noteworthy heterogeneity, so that a moderator analysis seems to be indispensable. Fortunately, we are able to identify two variables, hardly compared within single studies but quite often between studies: intimacy with the person referred to in the other-reference condition (*high* vs. *low*) and type of material presented (*adjectives* vs. *nouns*). Further analysis has revealed that both variables seem to moderate our previously noted results. The magnitude of recall enhancement under a self-referential instruction is only marginal when compared with a high-intimacy target person in the other-reference condition, but substantial if a low-intimacy person is referred to. However, this effect is only observable for adjectives; it disappears when nouns are to be recalled. To sum up, on a meta-analytical level we observed a pattern indicating an interaction between intimacy and word type (see Figure 9.1).

9.3 WHAT IS THE REAL MEANING OF A MODERATOR?

The question remains, however, whether a difference between two or more groups of studies that has been identified by means of a moderator analysis

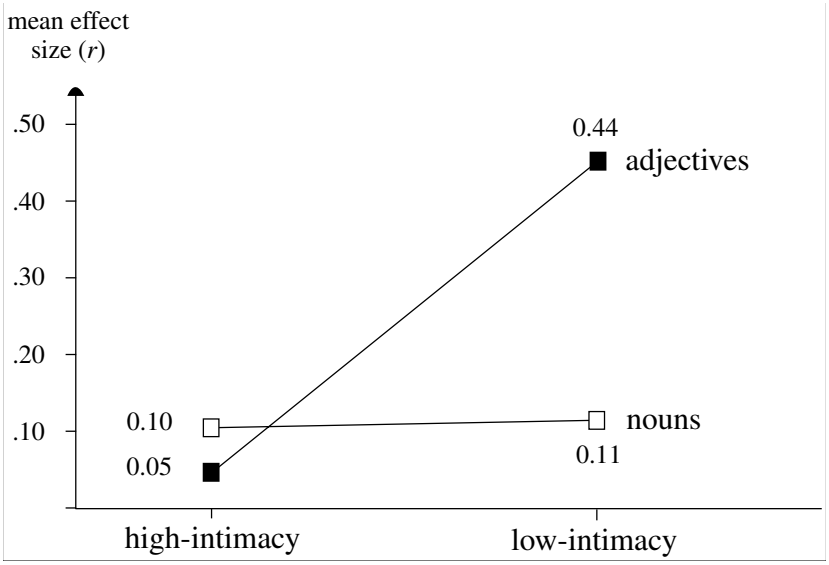

Figure 9.1 Interaction of SRE for intimacy and word type.

actually does represent a valid difference, or whether it merely represents an explanation by chance. This uncertainty can be reduced by selecting a more appropriate statistical model; that is, to interpret an effect size as a random effect allows a higher degree of generalizability than to consider it as fixed. Still, the basic problem remains unsolved even if we regard an effect size as random: A successful categorization of a collection of effect sizes by a specific variable that differentiates *between* different studies does only allow a post-hoc explanation of some of the variation between effect sizes that differs from random error. But in this case a meta-analytical approach is basically correlational (see Hall & Rosenthal, 1991); that is, no causal relationship can be established with this procedure. Even if a moderating variable can differentiate sufficiently between subgroups of effect sizes, the conclusion cannot be drawn that this variable has actually caused these differences. Since causation can generally be inferred only if based on experimental manipulations with results supporting this relation, a strong requirement can be formulated concerning the state of a moderator: A moderating effect of a variable that has been identified post-hoc based on meta-analytical results should be treated as a tentatively accepted potential moderator. Its state as a real moderator has to be evaluated by means of independent follow-up experiments. If this validation procedure is omitted, a scientific explanation of the differences between primary studies by means of the supposed moderator under study is not justified, even if a statistical explanation of the observed heterogeneity has been obtained by meta-analysis.

9.4 TESTING MODERATOR HYPOTHESES EMPIRICALLY

At this point we can pick up the thread again and discuss in more detail how to proceed with the result pattern presented in Figure 9.1. It has been noted that the meta-analysis on the difference in memory performance between recalling words under a self-referent task and an other-referent task seems to be moderated by at least two moderating variables, intimacy and word type. Taking both as potential moderators as discussed above, we may now use the meta-analytical results directly. The average ES for a group of integrated studies actually represents the most exact estimate of a population effect available, since it covers more data than a single study could provide. Furthermore, we are not required to conduct an exploratory study to estimate an expected effect. Relying on the given ES estimate, we are now able to design a decision study that enables us to decide with maximum precision whether our potential moderator does actually have explanatory power for the observed heterogeneity of ESs or whether the meta-analytical results should be taken merely as chance hits. Let us consider the result that the memory advantage of a self-referential orientation task compared with low-intimacy other-referential encoding is strong for adjectives ($r = .44$), but only marginal for nouns ($r = .11$). For the moment we will put aside a discussion of this puzzling result and the interesting question of whether the contrast between adjectives and nouns is meaningful at all. We will return to these topics below.

If we are interested in testing for the above-mentioned result pattern, we only have to specify the smallest difference to be detected between two conditions that should be shown as different and to specify appropriate levels for the first-order and second-order errors (e.g., .05 and .20). Then the required sample size for an adequately designed experiment (e.g., Keppel, 1991) can be computed or taken by power tables as published by Cohen (1988). In the case at issue, a group size of about 22 persons is required to test for the difference of $r = .39$ between two SREs (low- and high-intimacy others) for adjectives both exact and sufficiently sensitive ($\alpha = .05, \beta = .20$). If only the SRE with low-intimacy other persons ($r = .44$) is to be tested, $N = 18$ are required, provided that the references will be compared within subjects (as is mostly done in primary research).

As the reader may already have guessed, our examples are not fictitious but are results drawn from a meta-analysis on the self-reference effect that has actually been published (Czienskowski, 1997). In comparison with the meta-analysis on the SRE by Symons and Johnson (1997), the main aim of the integration by Czienskowski was to identify subordinate moderating patterns, using a hierarchical breakdown strategy (Hunter & Schmidt, 1990). Furthermore, it was attempted to test some predictions generated by this meta-analysis by means of further experiments. For example, Czienskowski (1997, 1998) reports experiments showing that intimacy with a person referred to seems actually to be a central factor that determines the magnitude of the investigated advantage of self-referential encoding. If an extremely low-intimacy other-reference condition is used, Czienskowski (1998, Exp. 2) reports a much stronger SRE

($r = .61$). Further experiments indicate that for average low-intimacy other-referents the SRE is approximately the same as predicted from meta-analysis ($r = .43$ resp. $r = .41$), but no remarkable difference can be obtained for the contrast between self-reference and high-intimacy other-reference (Czienskowski, 1997, Exp. 2; 1998, Exp. 1). Moreover, other results seem to confirm the prediction that only the use of adjectives, but not nouns, can produce the difference reported above. Problems related to the investigation of an SRE using nouns will be discussed below.

To sum up, it seems very promising to take an apparent moderator from meta-analysis as a merely potential moderator, which has to be tested independently. Since an empirical test of a prediction generated by moderator findings may also fail, this requirement is not at all trivial but indispensable to protect a meta-analysis against the obvious problem of "capitalizing on chance".

9.5 IS META-ANALYSIS USEFUL FOR THEORY DEVELOPMENT?

Although this leading question is answered affirmatively by most researchers who rely on meta-analytical techniques (e.g., Hall, Rosenthal, Tickle-Degnen, & Mosteller, 1994; Cook et al., 1992), it is often taken for granted that meta-analysis represents a powerful statistical toolbox that can be used to integrate different studies but that has no influence on the development of scientific theories. Actually, an ignorant use of meta-analytic tools may result in incorrect conclusions and remarkable confusion. But this is not only true for meta-analysis but for all advanced statistical technologies. To show how an adequate meta-analytical approach may in fact be used in a theoretically fruitful way, I will now focus on a special problem of the meta-analytical results referred to above.

As previously noted, it seems puzzling that a strong SRE is obtained for a low-intimacy other-referent condition if adjectives are used, but not if nouns are used. A closer look at the meta-analytical database suggests that these results actually may be an artifact. The database of studies included in the meta-analysis does not contain any study that compares adjectives with abstract nouns, but only with rather concrete nouns. On the other hand, one study that investigates a somewhat different kind of SRE (i.e., the recall difference between a self-reference and a merely semantically directed orientation task) using at least partly abstract nouns (Bock, 1986), indicated a very strong SRE. Since especially the distinction between concreteness and abstractness represents a central dimension for the explanation of memory performance (e.g., Gee, Nelson, & Krawczyk, 1999; Holcomb, Kounios, Anderson, & West, 1999; Marschark & Surian, 1992; Paivio, Walsh, & Bons, 1994), it can reasonably be assumed that the meta-analytical results could be confounded with a further but uncontrolled factor called "concreteness of word type". To examine this assumption further, Czienskowski (1997) reports an experiment that compares self-reference and low-intimacy other-reference using adjectives and matched

abstract nouns. For both materials a strong SRE has been found, which is completely incompatible with results yielded by the meta-analysis. In a second experiment, the assumption was tested that the SRE is obtained if abstract nouns but not concrete ones are used, and moreover, that this effect holds only for low-intimacy and not for high-intimacy other-referents. Planned simple effects analyses and simple comparisons are reported that support the expected result pattern. Simple comparisons between low-intimacy other-reference on the one hand and high-intimacy others and self on the other hand indicate strong memory differences (about $r = .41$) only for abstract nouns, whereas concrete nouns do not produce any detectable difference. More important might be that this effect is only due to a reduced recall performance for low-intimacy other-referents under the abstract noun condition. When exclusively explaining the SRE by referring to special features of self-referent encoding processes (e.g., as Rogers, 1981, does), this result still remains puzzling, because then it cannot be explained why an SRE should only occur when using abstract nouns. Moreover, the results reported seem to indicate that the SRE, at least for the comparison of self and others, is not an effect of enhanced self-referent encoding but of reduced recall for abstract material if low-intimacy others are referred to.

Focussing now on the fact that the critical difference causing the result pattern just discussed seems to be the concreteness or abstractness of the stimulus word presented, a reasonable assumption might be that the so-called SRE is actually a subordinate effect occurring only if other conditions are absent that could support encoding processes. More precisely, in the above case a concreteness effect as described, for example, by the dual-coding theory (DCT; Paivio, 1971, 1991) seems to be superior, whereas a self-reference effect (which rather seems to be an effect of intimacy or familiarity) takes place only if the concreteness of a stimulus is too low for promoting memory encoding. Czienskowski and Giljohann (2002) report two experiments indicating that a recall advantage of self-reference and high-intimacy other-reference can only be detected when abstract nouns are presented. With concrete nouns, the recall under self-reference is substantially lower (about $r = .30$) than for both other-reference conditions. The results confirm the expectation that only the absence of a possibility to encode information pictorially may result in a strong and unequivocal self-reference effect and support the view that the self-reference effect is not a general memory effect, but only a subordinate one. It may buffer the decrease of memory performance if pictorial coding is not possible, but it is not able to compensate this.

9.6 META-ANALYSIS AS A TOOL: IDENTIFYING THEORETICAL DEFICIENCIES AND NEW HYPOTHESES

Collecting the evidence from different sources, we are now able to conclude that an adequate application of meta-analysis in a rather more developed field of empirical research does not necessarily represent mere integration but can

also be used for the theoretical refinement or even reformulation of existing and frequently tested theoretical assumptions. In the case discussed above, the quantitative integration of studies investigating a rather prominent effect of cognitive social psychology has revealed both theoretical problems as well as new empirical hypotheses that might not have been detected without the application of meta-analytical methods. Admittedly, we can imagine a situation in which these problems and new hypotheses might have been developed by an attentive researcher interested in the SRE and aware of the DCT without making use of meta-analysis. However, in this case the predictions and the tests conducted subsequently would be much more imprecise than the expectations generated by research integration because only rough predictions of expected effects are possible. Thus, the state of the obtained results would remain rather unclear. Moreover, the evidence for a primary result is inferior compared to the evidence given by a decision study based on meta-analytical predictions drawn from a rather large set of primary studies.

It is quite obvious that this approach is not only applicable to the quantitative integration of research, but also, for example, if more evidence is required for an appropriate use of a therapy, for the development of educational strategies, or for a decision between two competing theories of memory. On the contrary, different fields of primary research could profit from this approach. In general, the determination of an adequate sample size for an experiment requires fixing an ES as precisely as possible if it is to be tested adequately by an experiment (i.e., both controlling the risk of rejecting both types of hypotheses, i.e., H_0 and H_1, falsely). Since theories in behavioral sciences are mostly not able to specify the exact magnitude of an effect to be tested, the integration of meta-analytical procedures into the process of designing primary studies should be seen as a opportunity to avoid wasting effort on conducting uneconomical (i.e., the sample size is too large for an effect to be tested) or meaningless (i.e., the sample size is too small for sensitively detecting an effect) experiments. The integration of meta-analytical procedures seems to be particularly favorable, too, just because several studies (e.g., Cohen, 1962; Sedlmeier & Gigerenzer, 1989) have reported that the average power of experiments published in certain journals only amounts to about .50 or even less. Thus, the use of meta-analytical tools can been seen as an indispensable supplement to the use of other design tools as, for example, power analysis, at least in a field of research that is rather developed.

9.7 CONCLUSION

In the previous parts of this chapter, a perspective on meta-analysis has been developed that is motivated mainly by requirements predominantly stated within the realm of primary experimental research. Hence, the orientation is directed to fields that can be investigated experimentally, at least in principle. With this restriction in mind, I can now discuss some conclusions that could bring about a new evaluation of several objections directed against meta-

analytical approaches. I will concentrate on some topics that are generally referred to as main problems of meta-analysis (e.g., Glass et al., 1981; Beelmann & Bliesener, 1994).

Let us first investigate the so-called apples-and-oranges problem. It is stated that, because the main feature of a study is its theoretical background, different operationalizations indicate different concepts so that different studies cannot be compared. But this conjecture is not convincing, since it implies the impossibility of modifying theories with the help of statistical meta-analysis. However, theories are quite often affected or even falsified by data. Since an empirical hypothesis cannot be rejected by any a priori argument, no a priori evaluation of the empirical relevance of any potential moderator hypothesis is possible. If a moderator variable can be identified that is able to explain the difference between studies not merely statistically, but also within an independent study, a theory not predicting this effect must be characterized as deficient (for sure, many possible moderators are theoretically irrelevant or even trivial, but since a meta-analyst is expected to be an expert in the field to be integrated, these irrelevant "moderators" will probably be sorted out early). Thus, the conjecture that meta-analysis is not able to provide more theoretical information than primary research can be refuted. Moreover, from a primary research point of view it could actually be advantageous to analyze some theoretical relations across different studies before more primary research is conducted. A meta-analytic integration of research and its use for the design of new studies does allow a goal-directed and precise search for theoretical relations to be identified empirically. By comparison, without any information from research synthesis, tests of these effects would be imprecise at best, but mostly these effects would not be identifiable at all.

A more general problem of meta-analysis could be its epistemological assumption of the possibility of accumulating scientific knowledge. From the author's point of view, the method of meta-analysis is indifferent to the problem if accumulation within the progress of science is really possible. Actually, it does not seem very useful to integrate empirical results of studies from different research fields, even if the results seem to be similar or at least comparable to each other. However, within a field in which a specific research question is investigated, a meta-analytical integration across different studies may be used to acquire higher precision of empirical statements. If meta-analysis is thus simply taken as a statistical tool, basically neither better nor worse than other statistical tools widely used (or misused) in behavioral sciences, there is no need to emphasize some of its problems more than, say, the problem of applying analysis of variance on ordinal dependent variables. Meta-analysis can be misused as much as other statistical procedures can, but there are many research problems that can be solved by this approach, therefore rejecting the importance of meta-analytical research integration would be throwing the baby out with the bathwater.

Some further problems of meta-analytic approaches, for example the "garbage in – garbage out problem" (i.e., a meta-analysis that integrates across poorly designed studies will probably be biased), can be resolved in a straight-

forward manner by explicitly testing potential moderators. The crucial question is whether the relevance of a possible moderator can be confirmed empirically. An empirical test will yield significant differences between different levels of the moderator if the effect size estimation from meta-analysis does actually represent a population effect, but it will fail if no population difference exists. This means that if moderators are indicated on the basis of biased samples or simply by chance, these potential moderators will be rejected if their test fails, otherwise the effect seems to be real, even if indicated by poorly designed studies.

Finally, to sum up the arguments given in this chapter, we can state that meta-analytical methods are powerful techniques and should be seriously considered if their integration into the methodological toolbox, especially the toolbox of primary research, could have advantages. On the other hand, research synthesis could benefit from this link, too, because some severe shortcomings that could affect the meaning of a research synthesis can be avoided if the presented techniques are applied. This approach seems especially useful if the real meaning of moderating variables is to be understood.

REFERENCES

Beelmann, A., & Bliesener, T. (1994). Aktuelle Probleme und Strategien der Meta-Analyse [Recent problems and strategies of meta-analysis]. *Psychologische Rundschau*, *45*, 211–233.

Bock, M. (1986). The influence of emotional meaning on the recall of words processed for form or self-reference. *Psychological Research*, *48*, 107–112.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, *106*, 265–289.

Bornstein, R. F., Kale, A. R., & Cornell, K. R. (1990). Boredom as a limiting condition on the mere exposure effect. *Journal of Personality and Social Psychology*, *58*, 791–800.

Bushman, B. J. (1994). Vote-counting procedures in meta-analysis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 193–213). New York: Russell Sage Foundation.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, *65*, 145–153.

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

Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., Light, R. J., Louis, T. A., & Mosteller, F. (1992). *Meta-analysis for explanation: A casebook.* New York: Russel Sage Foundation.

Cooper, H., & Hedges, L. V. (1994). Research synthesis as a scientific enterprise. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 3–14). New York: Russell Sage Foundation.

Czienskowski, U. (1997). Selbstbezug: Eine besonders wirksame Enkodierungsstrategie? Meta-Analyse und experimentelle Moderatorprüfung [Self-referencing: A

very effective encoding strategy? Meta-analysis and experimental validation of moderators]. *Zeitschrift für Experimentelle Psychologie*, *44*, 361–393.

Czienskowski, U. (1998). Kann experimentelle psychologische Forschung von Meta-Analysen profitieren? [Can experimental psychological research benefit from meta-analyses?]. In K. C. Klauer & H. Westmeyer (Eds.), *Psychologische Methoden und Soziale Prozesse* (pp. 210–229). Lengerich: Pabst Science Publishers.

Czienskowski, U., & Giljohann, S. (2002). Intimacy, concreteness, and the 'self-reference effect'. *Experimental Psychology*, *49*, 73–79.

Eagly, A. H., & Wood, W. (1994). Using research syntheses to plan future research. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 485–500). New York: Russell Sage Foundation.

Fleiss, J. L. (1994). Measures of effect size for categorial data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245–260). New York: Russell Sage Foundation.

Gee, N. R., Nelson, D. L., & Krawczyk, D. (1999). Is the concreteness effect a result of underlying network interconnectivity? *Journal of Memory and Language*, *40*, 479–497.

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

Hall, J. A., & Rosenthal, R. (1991). Testing for moderator variables in meta-analysis. Issues and methods. *Communication Monographs*, *58*, 437–448.

Hall, J. A., Rosenthal, R., Tickle-Degnen, L., & Mosteller, F. (1994). Hypotheses and problems in research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 17–28). New York: Russell Sage Foundation.

Hays, W. L. (1994). *Statistics for psychologists.* New York: Holt, Rinehart & Winston.

Hedges, L. V. (1992). Meta-analysis. *Journal of Educational Statistics*, *17*, 279–194.

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

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual coding, context availability and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 721–742.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

Johnson, B. T., Mullen, B., & Salas, E. (1995). Comparison of three meta-analytic approaches. *Journal of Applied Psychology*, *80*, 94–106.

Keppel, G. (1991). *Design and analysis: A researcher's handbook.* Englewood Cliffs, NJ: Prentice-Hall.

Marschark, M., & Surian, L. (1992). Concreteness effects in free recall: The roles of imaginal and relational processing. *Memory and Cognition*, *20*, 612–620.

Miller, N., & Pollock, V. E. (1994). Meta-analytic synthesis for theory development. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 457–483). New York: Russell Sage Foundation.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart & Winston.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*, 255–287.

Paivio, A., Walsh, M., & Bons, T. (1994). Concreteness effects on memory: When and why? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 1196–1204.

Petitti, D. (1994). *Meta-analysis, decision-analysis, and cost-effectiveness analysis.* Oxford: Oxford University Press.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–321). New York: Russell Sage Foundation.

Richardson, J. T. E. (1996). Measures of effect size. *Behavior Research Methods, Instruments, & Computers*, *28*, 12–22.

Rogers, T. B. (1981). A model of the self as an aspect of the human information processing system. In N. Cantor & J. F. Kihlstrom (Eds.), *Personality, cognition and social interaction* (pp. 193–214). Hillsdale, NJ: Lawrence Erlbaum.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

Schmidt, F. L., & Hunter, J. E. (1999). Comparison of three meta-analysis methods revisited: An analysis of Johnson, Mullen, & Salas (1995). *Journal of Applied Psychology*, *84*, 144–148.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, *105*, 309–316.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–281). New York: Russel Sage Foundation.

Symons, S. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*, 371–294.

Tatsuoka, M. (1993). Effect size. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences* (pp. 461–479). Hillsdale, NJ: Lawrence Erlbaum.

Acknowledgments

I would like to thank Hubert Feger, Bärbel Knäuper, Barbara Krahé, and Mansur Lalljee for helpful comments concerning this chapter.