

# The Quantitative Integration of Research: An Introduction to Meta-Analysis

Michael Lynn, Cornell University  
Brian Mullen, Syracuse University

*Meta-analysis refers to a set of procedures for statistically summarizing, integrating and comparing the results of previous research. This article elucidates the place of meta-analysis in hospitality research. First, a brief overview of meta-analytic procedures is provided. Second, the superior precision, rigor and accuracy of meta-analytic over narrative reviews is illustrated by selecting a recent narrative review in the Hospitality Research Journal, conducting a meta-analysis of the same studies and comparing the conclusions of the meta-analysis with those of the narrative review. Third, the authors illustrate the use of meta-analysis to test new, theoretical hypotheses by describing a recent meta-analysis of potential interest to the hospitality industry. Lastly, the article discusses several issues concerning the application of meta-analysis in hospitality and other research.*

Scientists and policy makers alike often attempt to use the results of previous research to determine what we already know, to plan investment of resources in subsequent research endeavors and to formulate plans of action to address critical problems. For a long time, the means by which the results of previous research were most often integrated was what we call a narrative review. In these types of reviews, the reviewer reads and thinks about a collection of relevant studies and then writes some narrative account of whether the hypothesis under consideration seems to be supported by the evidence. However, recent years have seen a burgeoning interest in what is called meta-analysis. In meta-analytic reviews, the results of previous research are statistically integrated and compared. Meta-analysis is not a single statistical procedure which distills a domain of research into one simple answer. Rather, meta-analysis embodies a general conceptual approach to the problems of summarizing, integrating and comparing the results of previous research and a constellation of different statistical techniques, developed and suited for these purposes.

The purpose of this paper is to elucidate the place of meta-analysis in hospitality research. To begin, an overview of meta-analysis will be presented. Next, a brief example of the difference between a narrative review and a meta-analytic integration will be examined. Then, an example of the use of meta-analytic techniques to test theoretical issues of interest to the hospitality industry will be considered.

Lastly, several issues concerning the application of meta-analysis in hospitality and other research will be discussed.

### **Meta-Analysis: An Overview**

In coining the term meta-analysis, Gene Glass (1976) identified three levels of analysis-i.e., primary analysis, secondary analysis and meta-analysis. Primary analysis refers to the original statistical analysis of raw data collected by the original researcher. Secondary analysis refers to the analysis of data by someone other than the original researcher, with theoretical goals and/or analytic techniques that may differ from those of the original researcher. Meta-analysis refers to the statistical integration of the results of several independent studies. Meta-analysts treat the results of independent studies, rather than the responses of individual subjects, as their units of analysis.

Something that novices at meta-analysis often fail to recognize is that the fundamental issues, the basic approach and many of the techniques of meta-analytic statistics have been around for a very long time. The early work of Fisher (1932), Pearson (1933), Thorndike (1933), Snedecor (1946) and Stouffer (1949) represent seminal efforts at meta-analytic statistical integration of the results of independent studies. Rosenthal's (1961, 1963) summary of research on experimenter expectancy effects is one of the earliest and most comprehensive efforts to integrate the results of many separate studies in a given domain of research. However, it was not until Glass (1976) labeled this perspective as "meta-analysis" that this approach received the popularity and the currency that it enjoys today.

Procedurally, there are several distinct steps in the development of a responsible and informative meta-analytic integration (for more detailed presentation, see Mullen, 1989; Mullen and Rosenthal, 1985; Rosenthal, 1991). First, the hypothesis test to be examined must be carefully and precisely defined. The specific operationalizations of the independent variables and the dependent variables must be clearly articulated. At this stage, the "size of the net" is determined: General and inclusive definitions of the independent and dependent variables can render a meta-analytic database of several hundred studies (e.g., Smith and Glass, 1977), whereas specific and restrictive definitions of the independent and dependent variables can render a meta-analytic database of five or six studies (e.g., Driskell and Mullen, 1990; Mullen, Mullen and Carey, 1993). Clearly, a larger meta-analytic database would seem to warrant more confidence than a smaller meta-analytic database. However, it should be emphasized that any gain in statistical power from a larger meta-analytic database is easily offset by a lack of theoretical focus.

After the well-defined hypothesis test has been identified, the relevant studies must be located and retrieved. Relevant studies may be located in published academic journals, scholarly textbooks, unpublished papers presented at conferences, unpublished theses and dissertations and published and unpublished technical reports. Several distinct strategies are used to locate relevant studies. The ancestry approach uses the bibliographies and reference sections of relevant studies which have already been retrieved to locate earlier relevant studies. The descendancy approach uses indexing sources (such as the Social Science Citation Index) to locate subsequent relevant studies which have cited earlier relevant studies. Abstracting services (such as Psychological Abstracts and Educational Resources Information Center) allow the user to identify studies associated with keywords and phrases. All three approaches can be conducted via on-line computer databases. The "invisible college" approach refers to the informal network of scientists working on a given problem. Letters, phone calls and conversations with researchers most active in a particular research domain can sometimes uncover new, unpublished studies at various stages of being "in the works."

Once the relevant studies have been retrieved, the appropriate tests of the well-defined hypothesis must be derived from the study report. Sometimes this is perfectly straightforward. However, very often researchers will report imprecise or imperfect tests of the relevant hypothesis. For example, sometimes researchers will identify the difference between two means as being simply "significantly different," without reporting a statistical test of the difference. Similarly, researchers will sometimes report an imprecise F-test based on more than one degree of freedom in the numerator, which actually tests the ill-defined hypothesis of whether there are any differences of any kind among three or more conditions. In these instances of improperly reported hypothesis tests, the precise test of the well-defined hypothesis must be reconstructed. Sometimes, this can be accomplished by working backwards from reported means and standard deviations, from reported means and Mean Square errors and/or from related-but-different F-tests. Other times, the only way to derive the precise test of the well-defined hypothesis is by contacting the original researchers and requesting the requisite statistical information.

After well-defined hypothesis tests have been located or reconstructed, they must be placed on common metrics. One study may report a t-test, a second study may report a chi-square, a third study may report a correlation coefficient and so on. These different statistics are on different metrics. That is, one cannot add the t-test from the first study, the chi-square from the second study, the correlation coefficient from the third study and divide by three to obtain the average statistical result. Each primary level statistical test of the hypothesis must be transformed to more standard common metrics in order

for any integration of results to begin. The two common metrics for statistical results are significance levels (Z and one-tailed p) and effect size (Z Fisher and r). These two study outcomes address two related but different questions about the results of a particular hypothesis test. Significance levels indicate the likelihood that these results are due to chance and effect sizes indicate the magnitude or strength of the effect of one variable on the other. Both are informative, in different ways, regarding different facets of the evidence regarding tests of a hypothesis.

Once placed on these common metrics, the results of separate hypothesis tests can be combined, compared and examined for the fit of predictive models. Meta-analytic combinations of significance levels and effect sizes provide a gauge of the overall combined probability and strength of the effects. Meta-analytic diffuse comparisons provide a gauge of the extent to which study outcomes are heterogeneous or inconsistent. Meta-analytic focused comparisons provide a gauge of the extent to which these effects increase or decrease as a function of some theoretically relevant or practically important predictors. There are three distinct statistical approaches to this phase of meta-analysis-i.e., the Hedges and Olkin (1985) techniques, the Rosenthal (1991) techniques and the Hunter and Schmidt (1990) techniques. The differences between these techniques are complex and describing them is beyond the scope of this paper. However, interested readers can find an excellent conceptual and empirical comparison of these different techniques in a recent article by Johnson, Mullen and Salas (1995). These authors found that the Hedges and Olkin techniques and the Rosenthal techniques produced results that were similar to one another and that were consistent with statistical conventions like the Law of Large Numbers. In contrast, the results from the Hunter and Schmidt techniques differed from those of the other meta-analytic techniques and did not conform with statistical convention. Thus, we recommend that hospitality researchers employ either the Hedges and Olkin (1985) or the Rosenthal (1991) techniques. In the sections that follow, we use the Rosenthal techniques to illustrate the unique strengths and advantages of meta-analytic reviews.

### **A Narrative Review versus a Meta-Analytic Integration**

Meta-analysis permits scholars to draw more precise, more rigorous and, in many cases, more accurate conclusions from research literatures than can be drawn from more traditional narrative reviews. To highlight these advantages and strengths of meta-analytic reviews, we next compare a meta-analysis with a narrative review of the same research studies. A recent article in the Hospitality Research Journal by Lynn and Graves (1996) contained a brief narrative review of research on the

relationship between tip size and service evaluations that is typical of this type of review. The authors of this article wrote:

Existing research does not permit any firm conclusions about tipping's relationship to service evaluations. Lynn and Grassman (1990) did find a positive relationship between restaurant customers' evaluations of the service encounter and their self-reported tip amounts. However, their study involved a small sample ( $n=103$ ) from only one restaurant, which raises questions about the generalizability of their results. Furthermore, numerous other studies have failed to find a significant relationship between tipping and service evaluations (Bodvarsson and Gibson, 1994; Crusco and Wetzel, 1984; Lynn, 1988; Lynn and Latane, 1984; May, 1978), which raises a question about the replicability of Lynn and Grassman's results. Clearly, there is a need for additional assessments of tipping's relationship to service evaluations. (Lynn and Graves, 1996, p.3).

In order to provide a concrete comparison between narrative and meta-analytic reviews, we conducted a meta-analysis of the same tipping studies included in Lynn and Grave's (1996) narrative review. Table 2 presents the relevant statistical results reported in each of the reviewed articles/theses as well as the effect sizes and z-scores associated with those statistical results. The effect sizes and z-scores in the table were calculated from the statistics reported in the original papers. They were then meta-analytically combined and compared as described below. [The formulas used in this meta-analysis are presented in the appendix, but readers should consult Mullen (1989) and Rosenthal (1991) for more detailed explanations of these formulas and their uses].

The statistical combination of effect sizes from these studies produced an  $r$  of .13. This means that, on average, service evaluations accounted for about 2 percent of the variation in tip sizes in the studies reviewed, with larger tips accompanying more favorable evaluations.

The statistical combination of significance levels (Z-scores) from these studies produced a combined  $Z$  of 3.71, which was highly significant,  $p < .0001$ . This finding means that a set of effects like those in the reviewed studies are unlikely to be due to chance alone. [Note that the significance of the combined effects cannot be attributed solely to Lynn and Grassman (1990), who found the largest and most significant effect in the review, because the combined  $Z$  remained significant when their effect was omitted,  $Z=2.38$ ,  $p < .009$ .]

**Table 2**  
**Studies Examining the Relationship Between**  
**Tip Size and Service Evaluations**

Study	Statistic [Direction]	N	one-tailed p	r	z
Lynn & Grassman (1990)	F(1, 96)=15.32 [+]	102	.0001	.37	3.76
Bodvarsson & Gibson (1994)	t(691)=1.01 [+]	700	.16	.04	1.00
Crusco & Wetzel (1984)	partial r(108)=.13 [+]	114	.09	.13	1.35
Lynn (1988)	r(205)=.09 [+]	207	.10	.09	1.29
Lynn & Latane (1984)	r(167)=.12 [+]	169	.06	.12	1.55
May (1978)	r(182)=.01 [+]	184	.45	.01	.13

A diffuse comparison of effect sizes indicated that they were significantly heterogeneous,  $X^2(5)=12.15, p<.04$ . This means that the differences between the studies' effect sizes are unlikely to be due to chance alone. A disjoint cluster analysis of the effect sizes indicated that Lynn and Grassman's (1990) effect size was significantly different at the .05 level from all the others, which were not significantly different from one another. Thus, something unique about Lynn and Grassman's study appears to moderate the size of the relationship between tip size and service evaluations. Unfortunately, it is not clear from these studies what that moderator might be.

Comparing the conclusions of this meta-analysis with those of Lynn and Grave's (1996) narrative review of the same literature is illuminating. Noting that only one of six studies had found a significant relationship between tip size and service evaluations, Lynn and Graves (1996) claimed that the existing research permitted no firm conclusions about this relationship. They suggested that the inconsistent findings raised serious questions about the relationship's reliability and generalizability that only additional data collection could answer. This is an extremely reasonable reading of the literature, but it is not entirely accurate. When meta-analyzed, the existing data indicated that the relationship between tip size and service evaluations was positive and statistically significant. In addition, the meta-analysis indicated that the tipping service relationship was moderated by some unknown factor associated with Lynn and Grassman's study, but that the reliability of the combined effect was not dependent on their study.

Of course, the limitations of Lynn and Grave's (1996) narrative review are not unique to this example. Narrative reviews often lead researchers to draw inaccurate conclusions from research literatures, because they focus the reviewer's attention on the significance level of each separate study's effect. Significance levels are certainly important, but they are only part of the story (see Mullen,

1989; Rosenthal, 1991). Moreover, examining significance tests on a study by study basis fails to capitalize on the statistical power that can be achieved by statistically combining the significance levels across studies. In contrast to narrative reviews, meta-analysis focuses the reviewer's attention on effect sizes as well as significance levels and it allows the reviewer to rigorously combine and compare these statistics across studies.

### **An Example of the Use of Meta-Analysis to Test New Hypotheses**

The previous example illustrated the precision, rigor and accuracy with which meta-analysis allows reviewers to summarize a research literature. However, meta-analysis can be used for more than just this. By coding differences in study characteristics and relating them to the studies' effect sizes, meta-analysts can also test theoretical hypotheses that were not tested in the separate studies being reviewed. In order to illustrate this aspect of meta-analysis, we next present a recent meta-analysis of group cohesiveness effects on group performance. Given the hospitality industry's emerging emphasis on human resource programs (e.g., Berger, Fulford and Krazmien, 1993; Bitner et al., 1990; Kinicki et al., 1992; Mahesh, 1988), this review is of potential interest to the industry for its content as well as for its illustration of meta-analytic techniques.

Festinger (1950) described group cohesiveness as "the resultant forces which are acting on the members to stay in a group," and most subsequent research on group cohesiveness has tended to accept this description. The expectation that a cohesive group will exhibit successful task performance is dramatically confirmed in Hollywood's depiction of our favorite groups from TV and film: the crew of the starship Enterprise, the personnel of *Baywatch*, the staff of *Fantasy Island* and even the crew of the *Love Boat* typify the paradigmatically cohesive and successful small group.

While the association between cohesiveness and performance seems reasonable and is consistent with anecdotal evidence drawn from history and the popular media, this seemingly straightforward phenomenon has generated a considerable amount of theoretical controversy. There have been several narrative reviews of research on the group cohesiveness-performance effect (e.g., Hogg, 1992, Chapter 8; Lott and Lott, 1965; Mudrack, 1989; Steiner, 1972; Summers, Coffelt and Horton, 1988; Tziner, 1982) that have rendered various and conflicting conclusions. For example, Steiner (1972, p. 33) confidently asserted that: "These findings do not support the view that group productivity and cohesiveness tend to be positively related." Alternatively, Summers et al. (1988, p. 631) asserted with equal confidence that: "In general, cohesion promotes productivity." In light of this contradiction and

inconclusiveness, it seems particularly important to determine in precise terms the significance and the magnitude of the relation (or the lack thereof) between cohesiveness and performance.

Another central concern in research on the cohesiveness-performance effect is the relative contributions of specific components of cohesiveness. In Festinger's (1950; Festinger, Schachter and Back, 1950) seminal formulation, cohesiveness was posited to result from interpersonal attraction, liking for or commitment to the group task and group prestige or pride. However, no direct comparisons of the effects of these three classic components of cohesiveness have been made at the level of individual studies. Many researchers seem to assume that the more a group's cohesiveness incorporates any of these three components, the stronger the effect of cohesiveness on performance. For example, Schachter (1951, p. 192) wrote that "whether cohesiveness is based on friendship, the valence of the activity mediated by the group, or group prestige, the consequences of increasing group cohesiveness are identical". Despite researchers' assumptions to the contrary, it is possible that not all three components of group cohesiveness will be equally important in determining increased performance. This possibility is important because the relative contributions of the three components may provide insight into the mechanisms by which cohesiveness impacts performance. In addition, the relative contributions of the three components of cohesiveness may highlight optimally effective interventions for enhancing productivity.

Thus, two questions loomed large in the research literature on the cohesiveness-performance effect: (1) Does the weight of available evidence support the existence of an effect of group cohesiveness on performance? and (2) To what extent do the three distinct components of cohesiveness contribute to the cohesiveness-performance effect? In an effort to examine these questions about the cohesiveness-performance effect, Mullen and Copper (1994) conducted a meta-analytic integration of tests of the cohesiveness-performance hypothesis. A total of 49 papers yielded 66 separate tests of the cohesiveness-performance effect, representing the responses of 8,702 subjects.

Regarding the first question, the combined results of these 66 tests of the cohesiveness-performance effect revealed a significant,  $Z = 8.492$ ,  $p < .00001$ , small,  $Z_{Fisher} = 0.254$ ,  $r = .248$  effect. Of the 66 hypothesis tests, 61 (or 92%) reported a positive direction of effect. A rather substantial fail-safe number of  $N(fs(p=.05)) = 3,766.5$  indicated that over 3,700 additional studies averaging no cohesiveness-performance effect would be needed before these results could be ascribed to sampling error. Thus, there seems to be substantial and rather unequivocal support for the general relation between cohesiveness and performance. In application to hospitality research, these results provide compelling justification for the consideration of the cohesiveness of hotel staff groups and service personnel teams.

Regarding the second question, three predictor variables were developed to reflect the extent to which a given hypothesis test's operationalization of cohesiveness involved each of the three components of cohesiveness. Then, the magnitude of the cohesiveness-performance effect was compared across different levels of each of these predictor variables. Neither interpersonal attraction,  $r = -.003$ ,  $Z = 0.046$ ,  $p > .48$ , nor group pride,  $r = -.139$ ,  $Z = 2.069$ ,  $p < .02$ , were associated with larger cohesiveness-performance effects (indeed, the magnitude of the cohesiveness-performance effect was actually smaller when a study's operationalization of cohesiveness involved more group pride). However, commitment to task was a significant positive predictor of the cohesiveness-performance effect,  $r = .284$ ,  $Z = 5.413$ ,  $p < .00001$ . Thus, commitment to task emerges as the critical component of cohesiveness in the cohesiveness-performance effect. In application to hospitality research, these results provide compelling justification for an emphasis on commitment to task, rather than interpersonal attraction or group pride, in considerations of the cohesiveness of hotel staff groups and service personnel teams.

The major point of presenting this meta-analysis was to illustrate the theory testing and theory construction uses of meta-analysis (see Mullen, Salas and Miller, 1991). Too many scholars continue to caricature meta-analysis as being simply and only a technique for obtaining average effect sizes; at this point it should be clear that meta-analysis is much more than that. The meta-analytic integration considered above allowed the research literature on the group cohesiveness-group performance effect to move forward in considering fundamental issues regarding mechanism and theoretical explanation. Previous researchers had typically treated cohesiveness as a unidimensional construct (based upon the assumption established early on that the three components of cohesiveness were equipotent). However, Mullen and Copper's (1994) integration documented that it was commitment to task and not interpersonal attraction or group pride, that seems to be the prime mover in the cohesiveness-performance effect.

### **Questions about the Application of Meta-Analysis**

The preceding sections of this paper provided a brief overview of meta-analysis and illustrated the unique strengths and advantages of this methodological tool. In the paragraphs that follow, several issues that commonly arise when researchers conduct meta-analyses are considered. Specifically, questions are answered about when reviewers should meta-analyze a research literature, how broadly meta-analysts should "cast their nets," what types of study characteristics meta-analysts should code as potential predictors of effect size and what information meta-analysts should include in their written reports.

### *When Is Meta-Analysis Appropriate?*

Meta-analysis is appropriate any time researchers want to summarize or integrate a body of research. The prototypic meta-analysis integrates the results of 50 or more studies and is reported in a stand-alone article (e.g., Mullen and Copper, 1994), but meta-analyses need not match this prototype. To begin with, meta-analyses need not involve large numbers of studies to be useful. In fact, informative meta-analyses have been conducted on as few as 5 or 6 study results (e.g., Driskell and Mullen, 1990; Mullen, Mullen and Carey, 1993). What is important in meta-analysis is not the size, but the comprehensiveness, of the sample of studies reviewed. Consider the meta-analysis of the tipping-service relationship reported above. Despite being based on only six studies, that analysis would be very informative if those six studies represented all of the published and non-published research on this topic. However, since those studies do not represent all of the unpublished tests of the tipping-service relationship, the meta-analytic results reported may be biased-the meta-analysis reported was for purposes of illustration only.

In addition, meta-analyses need not be limited to stand-alone reports. Meta-analyses can be fruitfully used in the introduction and/or discussion sections of primary research reports. In fact, it is not uncommon to see meta-analyses in the discussion sections of primary research reports, where they are used to demonstrate that the differing results of multiple reported studies produce a significant effect when combined.

### *How Broadly Should Meta-Analysts "Cast Their Nets"?*

Decisions about what hypothesis tests and literatures to include in a meta-analysis ultimately depend on the analyst's goals for the review. However, some general considerations may prove helpful in making these decisions. Casting a wide net by defining hypothesis tests in abstract terms that encompass a variety of operationalizations and by including studies from a variety of contexts and disciplines brings advantages in terms of large meta-analytic databases, multiple operationalism and enhanced generalizability. However, it also 'raises questions about the comparability of the included effects. All too often, an overly inclusive approach has contributed to critics' caricature of meta-analysis as being a thoughtless adding of "apples and oranges." Of course, more specific and focused approaches avoid the "apples and oranges" problem, but bring the liabilities of smaller meta-analytic databases, mono-method bias and reduced generalizability. The proper balance with respect to this tradeoff will vary across meta-analyses, but experience suggests that it is generally better to direct one's meta-analytic efforts to the summary and integration of tests of hypotheses that are operationally equivalent.

At the very least, researchers deciding to cast a broader net should be prepared to separately examine and compare those groups of effects based on different operationalizations.

#### *What Study Characteristics Should be Used as Predictors?*

The specific study characteristics that meta-analysts should code as potential predictors of the studies' effect sizes will vary with the literature being meta-analyzed. However, there are several classes of study characteristics that meta-analysts should systematically consider as sources of potentially interesting predictor variables (Stock, 1994). Among these classes of study characteristics that meta-analysts should consider are report, context, subject and methodological characteristics. Report characteristics include author, year and publication status of a report. These characteristics should be coded when the reviewer wants to assess potential investigator biases, history effects and publication biases respectively. Context characteristics include physical, geographic, cultural and other aspects of the setting in which a study was conducted. Coding these characteristics allows a reviewer to assess the generalizability of an effect across the coded contexts. Subject characteristics include demographic and psychographic traits of the subjects. These characteristics should be coded when the reviewer wants to assess the generalizability of an effect across different subject populations. Finally, methodological characteristics include such details as research design, sampling procedure, subject response mode and measurement instrument. Coding these characteristics allows the reviewer to assess potential methodological biases in the literature. Meta-analysts should look to each of these types of study characteristics when seeking interesting potential predictors of effect size.

#### *What Information Should Be Included in Reports of Meta-Analyses?*

Reports of meta-analyses should generally employ the structure used in primary research articles-i.e., they should include introduction, method, results and discussion sections (Halvorsen, 1994). The introduction should describe the question/issue that the meta-analysis addresses and the importance of that question/issue. The method should describe the decisions and procedures involved in: (1) searching the literature for relevant studies, (2) determining the uncovered studies' eligibilities for inclusion in the meta-analysis, (3) extracting hypothesis tests from the included studies, (4) calculating effect sizes and significance levels, and (5) coding study characteristics. The results should include descriptions of the primary studies being meta-analyzed (usually presented in tables) as well as the outcomes of the meta-analytic combinations, diffuse comparisons and focused comparisons. In addition, the results section often contains graphic displays of the data such as funnel plots and scatter plots.

Finally, the discussion should summarize the major findings of the meta-analysis, discuss their implications and point out directions for future research.

### **Conclusion**

Meta-analysis refers to a set of procedures for statistically summarizing, integrating and comparing the results of previous studies. Meta-analytic techniques have been around for many years, but to date, only two meta-analyses have been published in hospitality journals (i.e., Crouch, 1994, 1995). Thus, meta-analysis represents a new and underutilized tool in hospitality research. In this article, the authors provided a brief overview of meta-analysis, illustrated its unique strengths and discussed several issues concerning the application of meta-analysis. It is hoped that this material will encourage hospitality researchers to seek more information about meta-analysis and will prompt them to use meta-analytic procedures when reviewing research literatures.

### **References**

- Berger, F., Fulford, M.D. & Krazmien, M. (1993). Human resources management in the 21st Century: Predicting partnerships for profit. *Hospitality Research Journal: The Futures Issue*, 17, 87-102.
- Bitner, M.J., Booms, B.H., & Tetreault, M.S. (1990). The service encounter: Diagnosing favorable and unfavorable incidents. *Journal of Marketing*, 54, 71-84.
- Bodvarsson, O .B. & Gibson, W .A. (1994). Gratuities and customer appraisal of service: Evidence from Minnesota restaurants. *Journal of Socio-Economics*, 23, 287-302.
- Crouch, G.I. (1994). Demand elasticities for short-haul versus long-haul tourism. *Journal of Travel Research*, 33, 2-7.
- Crouch, G.I. (1995). A meta-analysis of tourism demand. *Annals of Tourism Research*, 22, 103-118.
- Crusco, A.H. & Wetzel, C.G. (1984). The Midas touch: The effects of interpersonal touch on restaurant tipping. *Personality and Social Psychology Bulletin*, 10, 512- 517.
- Driskell, J.E. & Mullen, B. (1990). Status, expectations and behavior: A meta-analytic review and test of theory. *Personality and Social Psychology Bulletin*, 16, 541 -553.
- Eagly, A.H., & Cali, L.L. (1981). Sex of researchers and sex-typed communications as determinants of sex-differences in influenceability: A meta-analysis of social influence studies. *Psychological Bulletin*, 90, 1-20.
- Festinger, L. (1950). Informal social communication. *Psychological Review*, 57, 271-282

- Festinger, L., Schacter, S. & Back, K. (1950). *Social pressures in informal groups: A study of human factors in housing*. Stanford, CA: Stanford University Press.
- Fisher, R.A. (1932). *Statistical methods for research workers* (4th ed.). London: Oliver and Boyd.
- Glass, G.V. (1976). Primary, secondary and meta-analysis of research. *Educational Research*, 5, 3-8.
- Halvorsen, K.T. (1994). The reporting format. In H. Cooper & L. Hedges (Eds), *The handbook of research synthesis*, New York: Russell Sage Foundation
- Hedges, L.V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hogg, M.A. (1992). *The social psychology of group cohesiveness: From attraction to social identity*. New York: New York University Press.
- 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 major meta-analytic approaches. *Journal of Applied Psychology*, 80, 94-106.
- Kinicki, A.J., Carson, K.P., & Bohlander, G.W. (1992): Relationship between an organization's actual human resource efforts and employee attitudes. *Group and Organization Management*, 17, 135-152.
- Lott, A.J. & Lott, B.E. (1965). Group cohesiveness as interpersonal attraction: A review of relationships with antecedent and consequent variables. *Psychological Bulletin*, 64, 259-309.
- Lynn, M. (1988). The effects of alcohol consumption on restaurant tipping. *Personality and Social Psychology Bulletin*, 14, 87-91.
- Lynn, M. & Grassman, A. (1990). Restaurant tipping: An examination of three "rational explanations". *Journal of Economic Psychology*, 11, 169-181.
- Lynn, M. & Graves, J. (1996). Tipping: An incentive/reward for service? *Hospitality Research Journal*, 20, 1-14.
- Lynn, M. & Latane, B. (1984). The psychology of restaurant tipping. *Journal of Applied Social Psychology*, 14, 551-563.
- Maccoby, E., & Jacklin, C. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
- Mahesh, V.S. (1988). Effective human resources management: Key to excellence in service organizations. *Vikalpa*, 13, 9-15.
- Marks, G., & Miller, N. (1987). Ten years of research on the false consensus effect: An empirical and theoretical review. *Psychological Bulletin*, 102, 72-90.

- May, J.M. (1978). Tip or treat: A study of factors affecting tipping behavior. *Unpublished masters thesis*, Loyola University of Chicago.
- Mudrack, P.E. (1989). Group cohesiveness and productivity: A closer look. *Human Relations*, 9, 771-785.
- Mullen, B. (1989). *Advanced BASE Meta-analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mullen, B., & Copper, C.M. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin*, 115, 210-227.
- Mullen, B., & Hu, L. (1988). Social projection as a function of cognitive mechanisms: Two meta-analytic integrations. *British Journal of Social Psychology*, 27, 333-356.
- Mullen, B., & Rosenthal, R. (1985). *BASE Meta-analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mullen, B., Salas, E., & Miller, N. (1991). Using meta-analysis to test theoretical hypotheses in social psychology. *Personality and Social Psychology Bulletin*, 17, 258-264.
- Mullen, M., Mullen, B. & Carey, M. (1993). The association between beta-agonist use and death from asthma: A meta-analytic integration of case control studies. *Journal of the American Medical Association*, 270, 1842-1845.
- Pearson, K. (1933). On a method of determining whether a sample size  $n$  supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika*, 25, 379-410.
- Rosenthal, M. (1961). On the social psychology of the psychological experiment: With particular reference to experimenter bias. In H. W. Riecken (Chair), On the social psychology of the psychological experiment. Symposium conducted at the meeting of the *American Psychological Association*, New York.
- Rosenthal, R. (1963). On the social psychology of the psychological experiment: The experimenter's hypothesis as unintended determinant of experimental results. *American Psychologist*, 57, 268-283.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.
- Schachter, S. (1951). Deviation, rejection and communication. *Journal of Abnormal and Social Psychology*, 46, 190-207.
- Schmidt, F.L. (1992). What do data really mean? Research findings, meta-analysis and cumulative knowledge in psychology. *American Psychologist*, 47, 1173-1181.
- Smith, M.L. & Glass, G. (1977). Meta-analysis of psychotherapy outcomes. *American Psychologist*, 32, 752-760.
- Snedecor, G.W. (1946). *Statistical methods* (4th ed.). Ames, IA Iowa State College Press.

- Steiner, I. (1972). Group cohesiveness, productivity and strength of formal leadership. *Journal of Social Psychology*, 87, 99-105.
- Stock, W.A. (1994). Systematic coding for research synthesis. In H. Cooper & L. Hedges (Eds), *The handbook of research synthesis*, New York: Russell Sage Foundation.
- Stouffer, S. A. (1949). *The American soldier. (Vol. I: Adjustment during army life)*, Princeton: Princeton University Press.
- Summers, I., Coffelt, T. & Horton, R.E. (1988). Work-group cohesion. *Psychological Reports*, 63, 627-636.
- Thorndike, R.L. (1933). The effect of the interval between test and retest on the constancy of the IQ. *Journal of Educational Psychology*, 24, 543-549.
- Tziner, A. (1982). Differential effects of group cohesiveness types: A clarifying overview. *Social Behavior and Personality*, 70, 227-239.

## Appendix

Formulas used in the Rosenthal (1991) approach to meta-analysis. Readers should consult Mullen (1989) or Rosenthal (1991) for fuller explanations of these formulas and their uses. A user-friendly software program, called "Advanced BASIC Meta-Analysis," that simplifies the use of these formulas is available from Lawrence Erlbaum and Associates (Mullen, 1989).

### *Formulas Used in Calculating Z-Scores*

- for  $t(df)$ :  $Z = (df(\log(1 + (t^2/df))))^{1/2}(1 - (1/(2df)))^{1/2}$
- for  $F(1,df)$ :  $Z = (df(\log(1 + (F/df))))^{1/2}(1 - (1/(2df)))^{1/2}$
- for  $\chi^2(1)$ :  $z = (x^2)^{1/2}$
- for  $r$ :  $t = (r(N-2)^{1/2}) / (1-r^2)^{1/2}$  then use formula for  $t$  above

### *Formulas Used in Calculating Effects Sizes*

- from  $t(df)$ :  $r = [t^2/(t^2 + df)]^{1/2}$
- from  $F(1,df)$ :  $r = [F/(F + df)]^{1/2}$
- from  $\chi^2(1)$ :  $r = (\chi^2/N)^{1/2}$
- from  $Z$ :  $r = (Z^2/N)^{1/2}$

### *Formula for Transforming r to Z FISHER*

$$Z \text{ FISHER} = .5 \{ \log[(1+r)/(1-r)] \}$$

*Formula for Transforming Z FISHER to r*

$$r = (e^{2Z_{FISHER}} - 1) / (e^{2Z_{FISHER}} + 1)$$

*Formulas for Combining Significance Levels*

$$Z = (\sum w_j Z_j) / (\sum w_j^2)^{1/2}$$

where:  $w_j$  = weight assigned to the results of hypothesis test j

$Z_j$  = Z associated with significance level of hypothesis test j .

*Formula for Combining Effects of Sizes*

$$Z_{FISHER} = (\sum w_j Z_{FISHER}) / \sum w_j$$

where:  $w_j$  = weight assigned to the results of hypothesis test j

$Z_{FISHER}$  = Z FISHER associated with effect size for hypothesis test j

*Formula for Fail-Safe N*

$$NFS (Z_{FISHER_e}) = (k(\bar{Z}_{FISHER} - Z_{FISHER_e})) / (Z_{FISHER_e})$$

where: k = number of hypothesis tests

$\bar{Z}_{FISHER}$  = mean Z FISHER for k hypothesis tests

$Z_{FISHER_e}$  = Z FISHER associated with specified criterion p – value

*Formula for Diffuse Comparison of Significance Levels*

$$\chi^2(k-1) = \sum (Z_j - \bar{Z})^2$$

where:  $Z_j$  = Z associated with significance level of hypothesis test j

$\bar{Z}$  = mean Z

k = number of hypothesis tests

*Formula for Diffuse Comparison of Effects Sizes*

$$\chi^2(k-1) = \sum (N_j - 3)(Z_{FISHER_j} - \bar{Z}_{FISHER})^2$$

where:  $N_j$  = N associated with hypothesis test j

$Z_{FISHER_j}$  = Z FISHER associated with hypothesis test j

k = number of hypothesis tests

$\bar{Z}_{FISHER}$  = mean z FISHER

### *Formula for Focused Comparison of Significance Levels*

$$Z = (\sum \lambda_j Z_j) / ((\sum \lambda_j^2)^{1/2})$$

where:  $\lambda_j$  = contrast weight assigned to results of hypothesis test j

$z_j$  = Z associated with significance level of hypothesis test j

### *Formula for Focused Comparison of Effect Sizes*

$$Z = (\sum \lambda_j Z_{FISHER_j}) / ((\sum (\lambda_j^2) / (N_j - 3)))^{1/2}$$

where:  $\lambda_j$  = contrast weight assigned to results of hypothesis test j

$Z_{FISHER_j}$  = FISHER'S Z associated with effect size of hypothesis test j

$N_j$  = sample size associated with hypothesis test j

## **Glossary of Meta-Analytic Terms**

**Abstracting Services:** A strategy for locating studies in which collections of abstracts are searched using key words or phrases.

**Ancestry Approach:** A strategy for locating earlier studies from the bibliography and references sections of already-located studies.

**Browsing:** To read at random; a locating strategy that obtains studies from peripheral sources by chance exposure.

**Combined Effect Sizes:** The mean effect size for the k-Included hypothesis tests.

**Combined Significance Levels:** The probability that the significance levels of the k-included hypothesis tests might have been obtained if the null hypothesis were true.

**Descendency Approach:** A strategy for locating studies that obtains subsequent studies that have cited earlier, already located studies.

**Diffuse Comparison of Effect Sizes:** A technique that indicates the extent to which there is a significant amount of heterogeneity in effect sizes among the k-included hypothesis tests.

**Diffuse Comparison of Significance Levels:** A technique that indicates the extent to which there is a significant amount of heterogeneity in significance levels among the k-included hypothesis tests.

**Direct Codings:** Quantitative predictors that are readily available from the written report of the study.

**Direction of Effect:** The directional pattern of the relationship between two variables (i.e., whether Y increases or decreases as a function of X). Direction of effect is often coded as being in the expected direction (+) or in the unexpected direction (-).

**Disjoint Cluster Analysis:** A technique that identifies non-overlapping clusters of hypothesis tests, based on the effect size and sample size associated with each hypothesis test.

**Effect Size:** The strength or magnitude of an effect.

**Fail-safe Number:** An estimate of the number of unretrieved studies averaging null results, or no effect, that would be needed to bring the combined results of a body of research to some specified minimum level.

**File Drawer Problem:** The possibility that unknown, unpublished studies might exist, whose results fail to support the pattern established by published findings.

**Focused Comparison of Effect Sizes:** A technique indicating the extent to which the effect sizes of the  $k$ -included hypothesis tests are significantly predictable in some systematic, specifiable manner.

**Focused Comparison of Significance Levels:** A technique indicating the extent to which the significance levels of the  $k$ -included hypothesis tests are significantly predictable in some systematic, specifiable manner.

**Funnel Plot:** A graphic display with sample size on the vertical axis and effect size on the horizontal axis. This plot should take the shape of a full, inverted funnel if there is no publication bias in operation in the research domain.

**Hypothesis Test:** A primary-level statistical test of some well defined, directional statement of the effect of one variable on another.

**Invisible College:** The informal network of scientists working on the same problem: a strategy for locating studies that obtains studies from individual researchers who are active in the field.

**Judges' Ratings:** The average of numbers assigned to each hypothesis test's procedure by judges, representing the extent to which a predictor would have been involved or engaged in each hypothesis test.

**Meta-Analysis:** The statistical integration of the results of independent studies.

**On-Line Computer Searches:** A set of strategies for locating studies that uses computerized literature databases to implement ancestry approach, descendancy approach, and abstracting services.

**$p$  (Probability Value):** A number, ranging from 0.0 to 1.0, which represents how improbable a statistic would be if the hypothesis being tested was true.

**Predictor:** Some attribute of the procedure, measurements, manipulations, situation, or context of hypothesis tests that that may account for variability in results of the hypothesis tests.

**Primary Analysis:** The original statistical analysis of data by the researcher who collected the data.

**$r$  (product-moment correlation coefficient):** A number representing effect size that ranges from -1.00 (perfect inverse relation) to +1 .00 (perfect direct relation).

**Scatterplot:** A graphic display of a correlation. Meta-analytic scanner plots usually present effect size on the vertical axis and a predictor on the horizontal axis.

**Secondary Analysis:** The analysis of data by someone other than the researcher who collected the data, for purposes or with analytic strategies other than those of the original researcher.

**Significance Level:** The degree of improbability which is deemed necessary to cast sufficient doubt upon the truth of a hypothesis to warrant its rejection.

**$Z$  (Standard Normal Deviate):** A standardized metric that has a mean of zero and a variance of one.

**$Z$ Fisher (Fisher's  $Z$  transformation of  $r$ ):** A standardized metric representing effect size that makes the sampling distribution of  $r$  approximately Gaussian (or normal).