Advances in Developing Human Resources

http://adh.sagepub.com/

Meta-Analysis Research and Theory Building

Baiyin Yang

Advances in Developing Human Resources 2002 4: 296
DOI: 10.1177/1523422302043005

Chapter 4

Meta-Analysis Research and Theory Building

Baiyin Yang

The problem and the solution. Although meta-analysis has been widely recognized as a powerful empirical research method, it has not been well acknowledged that it can be a valuable tool for theory-building purposes. This chapter describes common meta-analytic approaches and outlines a process of building theory in applied disciplines through this approach. Contributions of meta-analysis to theory building are discussed in terms of its unique roles in the relationship between theoretical constructs and empirical evidences. Both merits and disadvantages of the meta-analytic approach in theory-building research are discussed.

Niemi (1986) defined meta-analysis as "the application of statistical procedures to collections of empirical findings from individual studies for the purpose of integrating, synthesizing, and making sense of them" (p. 5). Clearly, meta-analysis is a special approach to reviewing the research literature on a topic; it reviews and synthesizes empirical studies in the literature. Merriam and Simpson (2000) maintained that literature review is a crucial step in the research process and its purpose is to summarize and integrate previous work and, thus, to offer suggestions for future studies. Whereas most literature reviews tend to be descriptive and narrative, a carefully designed meta-analysis should be inferential and conclusive. It goes beyond the conventional literature review with the aid of sophisticated statistical methods. Consequently, meta-analysis is more than a narrative review of the literature. For example, hundreds of studies have examined the factors that influence the transfer of learning (Baldwin & Ford, 1988; Ford & Weissbein, 1997; Holton, Baldwin, & Naquin, 2000). These studies have used not only diverse theoretical definitions, procedures, research methods, and samples but also identified different predictive variables affecting the learning transfer from various domains of study such as training design, individual differences, and organizational environment. Consequently, it is not uncommon that some of the research findings on learning transfer are at odds with each other, and researchers tend to have conflicting interpretations and conclusions. A meta-analysis would be desirable to integrate

Advances in Developing Human Resources Vol. 4, No. 3 August 2002 296-316

results from the existing studies to reveal patterns of causal relationships between learning transfer and its influential determinants.

Meta-analysis method uses formal statistical techniques to sum up a body of separate but similar empirical studies. The purpose of meta-analysis is to synthesize and organize the existing empirical findings into a coherent pattern. Glass, McGaw, and Smith (1981) distinguished among the primary analysis, secondary analysis, and meta-analysis of research:

> Primary analysis is the original analysis of data in a research study. . . . Secondary analysis is the reanalysis of data for the purpose of answering the original research question with better statistical techniques, or answering new questions with old data. . . . Meta-analysis of research invites one who would integrate numerous and diverse findings to apply the full power of statistical methods to the task. . . . It is the statistical analysis of the summary findings of many empirical studies. (p. 21)

In other words, meta-analysis is "analysis of analysis" (Glass, 1976, p. 3). Glass et al. further identified three characteristics of meta-analysis. First, meta-analysis is quantitative and uses numbers and statistical techniques for organizing and extracting valuable information that is nearly incomprehensible by other methods. Second, meta-analysis does not tend to evaluate the quality of existing studies. However, meta-analysis attempts to record various aspects of research methodologies for the existing studies to identify their relationship to study findings. Third, meta-analysis aims to compare existing studies and to seek general conclusions across studies.

Meta-Analysis as a Research Method

There are a number of books and methodological articles on meta-analysis, and most of them offer similar steps to conducting meta-analysis (Durlak & Lipsey, 1991; Glass et al., 1981; Rosethal & DiMatteo, 2001; Wolf, 1986). Though there is no single correct way to conduct a meta-analysis, there are certain procedures essential to meta-analytic research. Rosethal and DiMatteo (2001) suggested three basic principles that should guide meta-analysis: accuracy, simplicity, and clarity. A typical meta-analysis has the following steps.

Defining variables of interest and formulating the research question(s). Suppose a researcher is interested in the relationships between training design, learning style, and their impacts on learning transfer; the researcher might want to conduct a meta-analysis to examine the impacts of some influential variables on transfer of learning. Consequently, learning transfer can be the dependent or response variable, whereas those influential variables of interest will be treated as independent variables or predictors. A meaningful research then can be formulated such as the following one: Do trainees' learning styles moderate the impact of training design on learning transfer?

Searching literature and identifying adequate empirical studies in a systematic way. The next task of the meta-analysis is to search related literature and to identify all the published (and often the unpublished) empirical studies related to variables of interest. Using the previously mentioned hypothetical study as an example, the researcher needs to identify all of the available empirical studies in the literature that have studied the impacts of training design and learning style on learning transfer. It is necessary to read each of the studies and associated research methods and thus to assess how variables of interest were operationalized and measured. For example, the concepts of training design, learning style, and transfer of learning are theoretical constructs that have been frequently used by human resource development (HRD) professionals to represent certain observable organizational behaviors. HRD scholars and practitioners as well are often interested in developing and verifying theoretical models that depict their relationships to guide and inform HRD practice. The researcher of such a study needs to thoroughly understand theoretical meanings of these constructs and operational definitions and measurement in different empirical studies.

One challenge faced by the researcher is the multifaceted nature of these constructs. Consequently, different studies might have attached diverse interpretations to the same construct and thus operationalized differently. Another challenge often comes from the fact that existing studies might have used different measurements for the same construct. This might be charged by some criticisms of meta-analysis for combining and comparing apples with oranges (Rosethal & DiMatteo, 2001). A meta-analysis researcher thus needs to be fully aware of the differences among all included empirical studies. These differences include but are not necessarily limited to the following: different types of sample (age, gender, ethnicity, etc.), treatment situations, instruments with different psychometric properties, and the study and/or publication time. In fact, a well-done meta-analysis should take these differences into account by treating them as possible moderator variables. Suppose there are 50 studies in the literature that have investigated the impacts of training design and learning style on learning transfer, and 20 of these studies used male subjects, 10 of them included female subjects, and another 20 had participants with all genders; let us further assume that the researcher has a hunch that gender has an indirect impact on learning transfer with the interaction with learning style. Then, the variable of gender should be included in the meta-analysis even though it was not considered in the previous studies.

Coding previous studies and selecting appropriate index of effect size. Based on the research question(s) and appropriate conceptualization, the researcher needs to code variables of interest into the meta-analysis. Durlak and Lipsey (1991) noticed that "it is impossible to specify all the variables that should be coded in any meta-analysis" (p. 303). However, suggestion has been made to code those substantive and methodological characteristics that might influence study findings. Also, existing theories should play an important role in the selection of coding variables and the determination of coding method.

Durlak and Lipsey (1991) contended that meta-analyses "have varied from coding just a few variables to coding over a hundred variables per study" (p. 303) and suggested using research questions as a guide for variable selection and coding. One of the common key informational items that should be recorded for each empirical study is the effect size or correlation between variables of interest. Sample size of each study is another commonly recorded variable.

Analyzing the data collected from previous empirical studies. There are three major approaches to analyzing data in a meta-analysis. The first is known as the vote-counting method, where researchers sort the results of each existing study into one of three categories: positive significant, nonsignificant, and negative significant. This is a descriptive approach as the conclusions are drawn based on the resulting tallies. Wolf (1986) concluded that "the vote-counting approach is no longer recommended because of the poor statistical properties associated with it" (p. 13).

The second approach is called combined test, where researchers analyze the results of the same research hypothesis from different primary studies to conduct a summary overall test of the hypothesis. Suppose there are dozens of empirical studies that have examined the impact of learning style on transfer of learning; some of them have demonstrated the significant relationship, and others have failed to do so. A meta-analysis can be conducted to test the statistical significance of the combined results across these primary studies. There are a number of statistical tests available for conducting a combined test in meta-analysis; their results tend to be consistent with each other (Wolf, 1986).

Closely related to the combined test is a method of estimating the magnitude of the effect size across existing studies. Suppose there are 20 studies that have examined the impacts of collaborative learning as a training method on transfer of learning in similar organizational settings but have revealed different effect sizes. A meta-analysis is needed to calculate a grand effect size to achieve a conclusion about the extent to which transfer of learning is accounted for by collaborative learning method. There are two main families of effect size: the *d* family for group differences and the *r* family for correlational relationships. Although these two types of effect sizes are based on different research designs and thus need to be interpreted accordingly, there are formulas developed for the purpose of converting various summary statistics into commonly used metrics such as in the form of the Pearson product–moment correlation (Cohen, 1988; McGaw & Glass, 1980). If possible, appropriate corrective adjustments to the effect sizes should be applied due to unreliable, invalid, and limited range of measures and relative small sample sizes (Durlak & Lipsey, 1991).

The third approach to meta-analysis is to explore and examine possible interaction and/or mediator effects. This approach starts with examining the variability among the effect sizes of the existing studies. It is possible that the variability of effect sizes is attributed to sample characteristics (e.g., gender and race) or other influences such as geographical location and the time the research study was conducted. In the previously mentioned fictional example, the researcher might want to examine if trainees' gender, race, and other variables have mediated the relation between training design and transfer of learning. Statistical tests such as the chi-square test can be used to test homogeneity of effect sizes across different types of studies. More sophisticated tests such as the homogeneity test (generally called Q statistic) should be used (Hedges & Olkin, 1985). This approach allows researchers to examine the viability of any conceptual grouping of the existing studies. The variability among effect sizes points to the possibility of an existing mediator variable that might explain the variability in the effect sizes. If studies with samples of female trainees yielded significantly higher effect sizes on average than those with male samples, then it can be inferred that gender mediates the impact of training design on transfer of learning. A regression method can also be used to test if the impact of a mediator variable is statistically significant or not, where the effect size is used as the response variable and the mediator as one of the predictors. The regression method is particularly useful when some simple grouping variables are found to insufficiently explain the heterogeneous nature of effect sizes between many empirical studies. In this approach, variables coded from various characteristics of previous studies are used to identify predictor variables to explain effect size.

The regression method can also be used to explore and examine the interaction effect of interested variables on the variability of effect size. Suppose some studies on learning transfer have been conducted for training with the collaborative learning method, and others have used the conventional training method. Researchers may suspect that there is an interaction effect of gender and training design on the transfer of learning. In this fictional example, the sample characteristics (i.e., gender) and treatment (i.e., training design) should be coded for each of the studies in the literature, and these two variables and their interaction term will be treated as predictors with the effect size as a response variable in a multiple regression analysis.

Interpreting the results and drawing appropriate research conclusions. Durlak and Lipsey (1991) suggested three cautions in interpreting meta-analysis results and drawing conclusions. First of all, nonsignificance should be interpreted adequately. They warned that "null results might accurately reflect the true state of affairs, but they can also be artifactual" (p. 323). Meta-analysis researchers should be aware of confounding factors suppressing the real impacts on substantive variables. Also, researchers need to be aware of the limited statistical power for small sample size of the analysis (i.e., total and valid numbers of studies included in meta-analysis). Secondly, meta-analysis researchers should restrict their generalizations to the literature reviewed. Lastly, researchers should recognize limitations in the database and thus interpret the results in relation with available studies.

Relationship Between Empirical Study and Theory

As a step toward understanding the relationship between empirical study and theory and outlining the role of meta-analysis in theory building, I have modified Benson and Hagtvet's (1996) conceptual framework of measurement and placed a series of empirical studies in relation with theoretical and empirical domains (see Figure 1). This framework captures the relations between theory, measurement, and data analysis. Three related research domains—theoretical, empirical, and measurement—can be identified at either the theoretical or empirical levels. Theoretical and empirical levels correspond to two major components of the theory-building process theoretical and research parts (Dubin, 1978).

Theory and Theoretical Constructs

At the theorizing level, theoretical constructs are the main focus. According to Bacharach (1989), a theory represents a "system of constructs and variables in which the constructs are related to each other by propositions and the variables are related to each other by hypothesis—the whole system is bounded by the theorist's assumptions" (p. 496). Many social and behavioral theories, if not all, consist of several theoretical constructs as their major components and explicitly state the relationships between and boundaries of these constructs in the format of hypotheses or propositions.

For example, an existing simple transfer theory may posit that training design has both positive direct impacts on the outcome (e.g., transfer of learning) and indirect impacts via the learners' preferred learning style. In other words, this hypothetical theory assumes that learning style mediates impacts of training design on transfer of learning. Three major elements of this theory, training design, learning style, and transfer of learning, are constructs, and the interrelationships between these constructs have been postulated prior to the application of meta-analysis. The top of Figure 1 depicts such relationships, and these constructs are represented by Cs (C_1 for the transfer of learning, C_2 for the learning style, and C_3 for the training design).

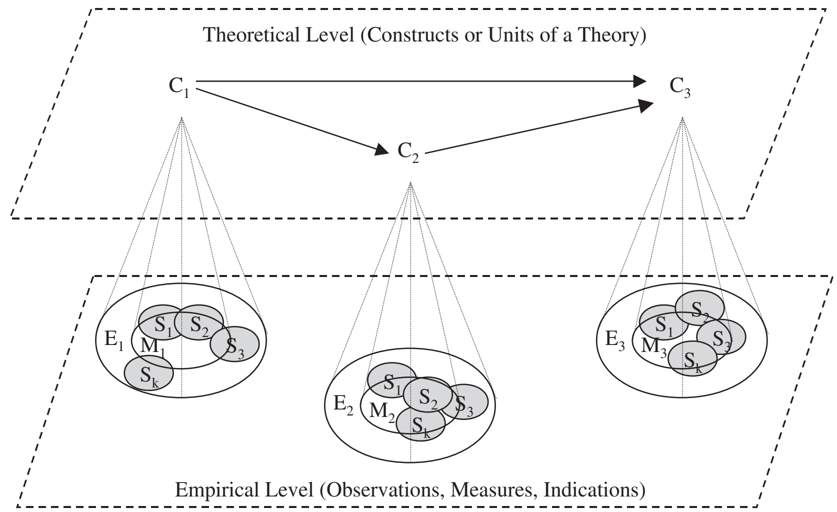

FIGURE 1: Empirical Basis of Theory Building

Note: C = theoretical constructs (or units); E = empirical domains for the constructs; M = measurements of empirical domains that operationally bond the constructs; S = different empirical studies of practice.

They are constructs because they represent some general abstract ideas not yet directly touched or measured. However, they can be inferred from certain commonalities among observable phenomena.

Most theories have abstract constructs to illustrate certain social and organizational phenomena, and they are postulated to explain, to predict, and to control the empirical world. Scholars of all kinds are working back and forth between theoretical and empirical levels to develop, operationalize, test, and apply their favored theories. Functionally, this means that meta-analysis can contribute to and potentially fulfill the requirements of four of the five phases of the general method of theory building in applied disciplines (Lynham, 2002). As also depicted in Figure 3, they are

- conceptual development
- operationalization
- confirmation or disconfirmation, and
- continuous refinement and development.

Empirical and Measurement Domains

At the empirical level, all observed evidences directly related to a construct compose an empirical domain for the construct (i.e., each of the Es in

Figure 1). "The empirical domain comprises all the possible ways to measure the construct as suggested by the definition of the theoretical domain" (Benson & Hagtvet, 1996, pp. 85-86). Each theoretical construct has its own empirical domain that contains all the potential observations that are assumed to represent the construct. However, not all empirical evidences can be easily accessed and measured, and not all observations are equally reliable and valid in inferring the implied underlying constructs. Consequently, researchers have to develop and validate specific sets of measurement items as appropriate indicators for the theoretical constructs. A representative sample of observations from the empirical domain constitutes the measurement domain for a construct (i.e., each of the Ms in Figure 1). Benson and Hagtvet (1996) pointed out that "as a specific instrument, M represents an operational definition of the theoretical domain by specifying exactly which observables comprise the construct" (p. 86). In other words, theoretical constructs are approximated in the empirical world (Bacharach, 1989). For example, learning style as a construct has to be operationalized and measured in a format of the assessment instrument (e.g., self-report or peer rating). Other theoretical constructs or concepts such as learning transfer and intelligence have to be approximated by reliable and valid instruments to confirm related theories with adequate empirical evidences.

The relation between measurement and theoretical domains illustrated in Figure 1 can also illustrate the concepts of reliability and validity. Reliability refers to the consistency of a set of measurement items and is normally estimated either by the stability of the measures over a time period for the same sample in different occasions (i.e., test-retest reliability) or by the intercorrelations among the set of measures (i.e., internal consistency such as Cronbach's alpha). Therefore, reliability of a set of measures for a theoretical construct is reflected by the consistency of these measures within the measurement domain. Validity indicates the accuracy of a set of measures for the theoretical construct being assessed. A set of measurement items is deemed to be valid for a construct if these items are proved to be adequate representatives (at the empirical level) for the underlying construct (at the theoretical level). Demonstrated in Figure 1, validity reflects the degree to which the measurement items are close to the central point projected by the underlying construct from the theoretical level.

Even for those commonly used theoretical constructs such as intelligence and learning style that seem to be simple at the first glance, not all researchers have agreed on a single measurement tool or method. For example, how to assess human intelligence remains controversial after more than a century's continuous efforts. There are numerous assessment instruments developed for the measurement of learning style, and yet none of them tends to be highly satisfactory (James & Blank, 1993). There tend to be many theoretical conceptualizations and theories developed to explain a social and organizational phenomenon (e.g., motivation); often, there are diverse approaches to operationalizing and measuring one theoretical construct. Actually, it is common and beneficial to the scholarly field as well to have multiple approaches to operationalizing and measuring one theoretical construct. Scholars studying the same theoretical construct with different research approaches can add to the knowledge base from diverse perspectives because a single construct may imply several meanings and can thus be perceived as different observable behaviors at the empirical level, and they can also advance the theoretical development by adding different pieces of a puzzle together. Consequently, multiple empirical studies are needed to verify a theory and related theoretical constructs (Ss in Figure 1). S_1, S_2, . . . , and S_k represent a total number of k empirical studies around a set of related theoretical constructs.

Data Analysis Strategy

Empirical studies have at least two distinctive and yet related tasks. Accordingly, these tasks also require appropriate data analysis strategies and statistical techniques. Though these tasks are sometimes indistinguishable and not explicitly stated in some studies, it is important to recognize their unique functions in theory building. The first task is to develop and validate a set of measures (i.e., measurement items) from the empirical domain to represent underlying theoretical constructs. Researchers normally first identify a set of observations for the phenomena being investigated as an initial item pool (i.e., items, indicators, or manifest variables) and then examine the intercorrelations among these items to identify and select those representative items (both reliable and valid) for the underlying theoretical constructs (i.e., latent variables). It is common that some initially identified items are not fine representatives for the empirical domain and thus cannot be selected as adequate measures for the underlying constructs. A validation study involves such effort to identify adequate measurement items and to discard less adequate items based on the examination of the relations between observable indicators. Several statistical techniques are available for the validation study, such as correlational design, multitrait-multimethod design, exploratory and confirmatory factor analyses, and multifaceted measurement designs (Benson & Hagtvet, 1996). The essence of the validation study is to evaluate the adequacy of a set of measurement items in inferring underlying theoretical constructs and thereby to enhance the construct validity of the theoretical framework of a theory.

The second task of empirical studies is to verify the proposed relationships between theoretical constructs of interest with the patterns of empirical evidences from adequate observable behaviors assessed on a set of reliable and valid measurement items. This involves the process of confirming or disconfirming a proposed theory with empirical evidence. Some studies overlook the first task and immediately jump to the second task in testing a proposed theory. Such studies tend to offer limited and weak empirical support for theory building. The heart of testing a theory with adequate empirical evidence is to examine if a theoretical construct of interest such as transfer of learning fits into a network of expected relationships with other constructs (e.g., training design and learning style). To illustrate the role of empirical studies for theory verification, Figure 1 can be used to represent a hypothetical theory that assumes learning style (i.e., C_2 in Figure 1) is a mediator for the impacts of training design (C_1) on transfer of learning (C_3). This theory will not be adequately tested unless the following four hypotheses are satisfactorily accepted: (a) There is a valid measure for training design, Instrument A; (b) there is a valid measure for learning style, Instrument B; (c) there is a valid measure for transfer of learning, Instrument C; and (d) scores on Instruments A, B, and C are correlated in a way as illustrated in Figure 1. Hypotheses 1 through 3 should be supported by validation studies, and Hypothesis 4 is specially used to verify the theory being tested. Nevertheless, all of these hypotheses should be tested and supported to verify the theory because the evidence for Hypothesis 4 cannot be the proof for Hypotheses 1 through 3. Scores on Instruments A, B, and C could be correlated, but their correlation could be due to the constructs other than those proposed by the theory. Consequently, the validity of an instrument used to assess theoretical constructs, as illustrated by the correspondences between theoretical and measurement domains in Figure 1, should never be overlooked for testing theory in any empirical study.

Another frequently overlooked area in testing theory with empirical evidence comes from the data analysis strategy. Although most conventional data analysis techniques such as analysis of variance and multiple regression analysis have to use an instrument's raw scores to represent the underlying construct, the methodological problems associated with this approach should be recognized. Statistically significant test-based raw scores can be influenced by measurement error and by other factors (Bollen, 1989; Crocker & Algina, 1986). Although we can calculate disattenuating partial correlations with reliability estimates, this procedure can only be used to examine a simple bivariate relationship and is not applicable to examine multivariate relationships, which tends to be a more desirable analysis technique for complex social and organizational phenomena. Fortunately, a multivariate statistical technique called structural equation modeling (SEM) has been developed to overcome methodological problems existing in most conventional data analysis techniques. According to Benson and Hagtvet (1996),

> SEM is a general data analytic technique that subsumes many statistical and psychometric procedures (e.g., analysis of variance and covariance, correlation, regression, factor analysis, and reliability estimation) and has been applied in many disciplines in the behavioral and medical sciences. (p. 100)

SEM is a superb statistical tool for testing theory because it allows researchers to examine the relationships between a set of different theoretical constructs while taking into account associated measurement errors. For instance, a regression analysis or path analysis can be used to test the previously mentioned fictional theory illustrated in Figure 1, but these techniques can only reveal the correlation coefficients for the raw scores instead of true scores. Consequently, these techniques can partially examine the relationships between the constructs of interest because the impacts of measurement errors are not known. On the other hand, the SEM technique will provide strong empirical evidence if it is employed because it reveals correlation coefficients for the true scores of theoretical constructs being investigated while partialling out measurement errors.

Empirical Study and Theory Building

Figure 1 not only describes the relations between theory, research design, and data analysis but also captures the essence of the applied theory-building process from the empirical research perspective. Lynham (2002) identified five major phases for the process of applied theory-building research. The first phase is conceptual development, where initial theoretical ideas and conceptual frameworks are formulated. The outcome of a theory-building effort in this phase is a preliminary and tentative theoretical framework outlined at the top level of Figure 1. The second phase of a theory-building effort is operationalization, where theoretical ideas are linked to empirical and operational indicators. Therefore, the linkages between theoretical and empirical levels in Figure 1 reflect the operationalization phase. A theory-to-research strategy of theory building tends to start at the theoretical level by developing an explicit theory and then to verify it at the empirical level. On the other hand, a research-to-theory strategy of theory building is likely to start at the empirical level by discovering possible systematic patterns from empirical evidences and then to formulate these patterns in theoretical statement. The third phase of theory-building research suggested by Lynham is called confirmation or disconfirmation. Researchers engaged in this phase need to collect empirical data with appropriate design and analytical tools to confirm or reject an initially formulated and operationalized theoretical framework. Some of the empirical domain and a considerable portion of the measurement domain illustrated in Figure 1 are involved in this phase. The fourth theory-building research phase is application, where a confirmed theory is put into practice (Lynham, 2002). This phase can be viewed as an extension of verified measurement domains of related theoretical constructs at the empirical level. Finally, the fifth phase in the theory-building research process is ongoing refinement and development. This component immerses into both the theoretical and research parts of applied theory building and can be identified either at the theoretical level or empirical level presented in Figure 1.

Figure 1 also demonstrates the role of meta-analysis in theory building. Suppose there are a number of existing empirical studies (i.e., $S_1, S_2, \ldots, S_k$) for a set of identifiable theoretical constructs. These studies could have been guided by the same or similar theoretical frameworks, or they could have used the same or different instruments to measure the theoretically equivalent constructs. Then, a meta-analysis will offer unique contributions to the knowledge base by testing and building the theory directly related to these empirical studies. It can accomplish several vital tasks that cannot be done by other research and theory-building methods. First of all, a combined test in meta-analysis offers an accurate estimate for the correlations of interested constructs and thus provides a grand test for the existing theory. As it has been illustrated in Figure 1, different studies may have included different empirical evidences and thus cannot offer a complete picture for the relations between theoretical constructs of interest. A meta-analysis can be used to synthesize and integrate the existing empirical findings, and the combined test provides more reliable and valid results than any of a single study. Second, meta-analysis can be used as an evaluative tool to assess existing theories and thus promotes the advancement of theory building. There may be several competing theories for the explanation of a same-response variable, and they may include diverse explanatory variables; meta-analysis can be used to evaluate relative contributions of different variables and thus test the applicability and boundary of existing theories. Such results can be used either for evaluation of existing competing theories or for the theory refinements and modifications. Third, meta-analysis offers a unique technique in identifying moderator variables and interaction effects. Theory building is an ongoing process, and different theories might be applicable to different situations. As a technique of analysis of analyses, a sound meta-analysis not only reveals the situations where certain theories are applicable through the review of existing empirical studies but also identifies underlying reasons for such variability by searching for significant moderator variables and possible interaction effects that might have been ignored by the existing theories. The above tasks can be accomplished by carefully conducted analyses specially dedicated to the purpose of theory building. The next section describes the process of meta-analysis as a theory-building method.

Using Meta-Analysis for Theory Building

Although meta-analysis has been widely recognized as a powerful empirical research method, it has not been well acknowledged as a valuable tool for theory-building research. Figure 2 illustrates a new model of the utility of meta-analysis for theory building in applied disciplines. The process of theory building that uses the meta-analytic technique consists of five steps: (a) Review existing theory and identify variables of interest, (b) search existing empirical studies and code variables of interest, (c) examine the variability of effect sizes for the variables of interest, (d) conduct appropriate statistical test(s) to adequately explain the variability of effect sizes, and (e) confirm and disconfirm current theory and/or search for alternative theory. The five layers of the decision-making process presented in Figure 2 correspond to these five steps. The correspondence between the five steps of theory building from meta-analysis and the phases of the general method of theory-building research in applied disciplines (Lynham, 2002) is later shown in Figure 3.

The first step is to review existing theory (or theories) on an interested topic and to identify variables of interest. An initial research question might be formed in this step to serve as a guide for theory refining and building effort. As indicated in Figure 3, this step corresponds to the conceptual development phase of the general method of theory-building research in applied disciplines (Lynham, 2002). For example, Eagly, Makhijani, and Klonsky (1992) conducted a meta-analysis on the evaluation of male and female leaders. This meta-analysis stemmed from a gender role theory and built on a prior meta-analysis of sex difference in leadership style. Gender role theory suggests that people expect their own and others' behaviors to be appropriate to the relative gender roles. In an organizational setting, people's expectations about the behavior that is appropriate for leaders and managers match their expectations about males more closely than their expectations about females. Consequently, a research question was formulated to examine if female leaders were perceived somewhat less favorably than equivalent male counterparts. This research question helped the researchers in identifying appropriate predictive variables for the leadership evaluation in this meta-analysis. Because gender was one of the main variables in the research question, several study attributes such as gender of leaders in the existing studies, gender of subordinates, and the gender distribution in leadership role were all included and subject to examination in the study.

The second step of theory building from meta-analysis is to search existing empirical studies in the literature and to code these variables of interest. As per Figure 3, this step matches the operationalization phase in Lynham's (2002) general method of theory-building research in applied disciplines. The main purpose of this theory-building phase is to link some abstract theoretical ideas to observable indicators at the empirical level. For example, leadership style is an abstract theoretical construct that tends to associate with leadership evaluation. In the meta-analysis testing the relation between gender and leadership evaluation, Eagly et al. (1992) recoded this construct

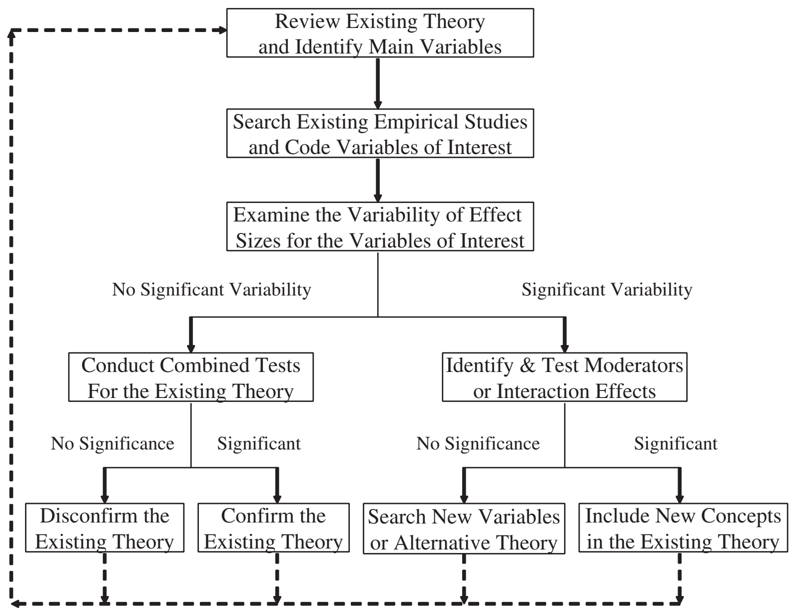

FIGURE 2: The Process of Meta-Analysis as Theory-Building Research

into two categories, masculine and feminine leadership styles, based on the available measurement. Although the masculine or feminine leadership styles could not be discerned in some of the existing studies, the majority of leader portrayal was captured and classified as either masculinity or femininity. Another closely related variable, leadership style portrayed, was deemed to be one of the explanatory variables and coded in one of five categories (interpersonally orientated, task oriented, autocratic, democratic, and other or mixed). This variable was subsequently dummy coded (autocratic style or other styles) and entered into a regression analysis to predict the effect size of leadership evaluation between male and female leaders. In sum, selecting and coding meaningful explanatory variables should be guided by substantive theoretical ideas and appropriate analysis techniques.

The next three steps of theory building from meta-analysis tend to belong to the phase of confirmation or disconfirmation in Lynham's (2002) general method of theory-building research in applied disciplines. These steps are distinguishable in meta-analysis because they represent a series of statistical testing and decision making with regard to the theory being examined. These tests are directly related to the researchers' decision of accepting or rejecting the theory being examined.

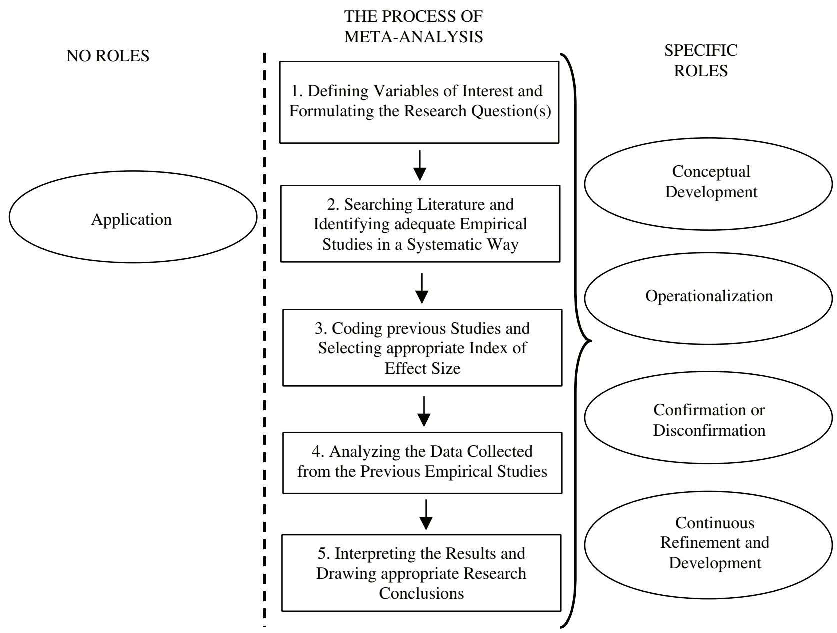

FIGURE 3: The Role of Meta-Analysis in the Context of the General Method of Theory-Building Research in Applied Disciplines

The third step of theory building from meta-analysis is to examine the variability of effect sizes based on conceptualized characteristics (normally between meaningful groups) of existing empirical studies. A significant result implies the impact of some moderator variables and thus calls for the refinement of the existing theory by including these variables, if further statistical tests demonstrate that they are significant. On the other hand, a nonsignificant result of the homogeneity test rejects the possibility of possible moderator variables being examined. Nevertheless, further analyses such as combined tests should be followed to confirm or disconfirm the current theory guiding the existing empirical studies. For example, to examine the multiple impact of leadership style portrayed (autocratic vs. other styles) on leadership evaluation in the previously mentioned meta-analysis, Eagly et al. (1992) conducted a homogeneity (Q statistic) test to confirm if the effects of gender on the leadership evaluation were significantly different across existing empirical studies. A statistically significant difference was revealed and thus indicated the rejection of the hypothesis of homogeneous effect sizes. The researchers further formulated several categorical models to examine the homogeneity of effect sizes for these meaningful study attributes such as gender distribution in leadership role, gender of subordinates, and type of organizational context. Leadership style portrayed was one of the attributes that were found to be statistically significant.

The fourth step of theory building from meta-analysis is to conduct more appropriate statistical analysis based on the result of the homogeneity test. Suppose the homogeneity test is not statistically significant; then, the result implies that the study attributes fail to demonstrate significant impacts on the effect size being investigated. In other words, the result tends to reject the hypothesis that study attribute variables contribute to the effect sizes of the current explanatory variable(s) on the response variable. Such results rule out the possibility of including additional explanatory variables from the study attributes in the existing theory. Nevertheless, a combined test is desirable to examine the average effects of present explanatory variables. This combined test is valuable in assessing the overall efficacy of the existing theory.

Suppose the homogeneity test has demonstrated statistical significance; then, the result shows that study attributes are likely to have significant impact on the effect size being investigated. Such results indicate that study attribute variables may contribute to the effect sizes of the current explanatory variable(s) on the response variable of interest. Therefore, meta-analysis researchers should further examine possible effects of moderating variables or interaction effects. In the example discussed in the previous paragraphs, Eagly et al. (1992) examined the moderating impacts of several study attributes on the relative evaluation of female and male leaders with multiple regression analysis. The regression equation included most of the theory-relevant predictors and resulted in a multiple R of .59.

The fifth step of theory building from meta-analysis is to draw theoretical implications based on the statistical analyses conducted in the previous step. Possible consequences associated with the two types of statistical tests conducted in the previous step are a result of such statistical analyses. When the combined test is used to examine the overall impact of several explanatory variables, the overall test can be either nonsignificant or significant. The nonsignificant consequence suggests that the variables included in the current theory fail to adequately explain the variability of the response or dependent variable. Therefore, the theory has been disconfirmed, and continuous theoretical development and refinement should take place. On the other hand, the significant consequence indicates the confirmation of the existing theory. Nevertheless, researchers should not solely rely on the significance test in judging the adequacy of the theory being tested (Cohen, 1994). Both statistical indicators (e.g., effect size, power, and the total variance explained by the predictors) and substantive theoretical ideas should be used to guide the search for additional explanatory variables and refinement of the existing theory.

Similarly, there are two possible consequences associated with the test of moderators or interaction effects—statistically nonsignificant or significant results. The purpose of such tests is to identify additional significant predictors to be included in the existing theory. The nonsignificant result suggests that the moderators or interaction terms being examined fail to contribute to the variability of effect sizes of the interested dependent variable across existing empirical studies. This result thus disconfirms the hypothesis of including additional explanatory variables in the existing theoretical framework. Consequently, researchers may need to verify the result and continuously identify other explanatory variables with theoretical meanings. This will lead to the next phase of theory building—continuous refinement and development.

The result of statistical significance provides positive evidence to confirm the theoretical ideas being tested. In the case of testing moderators in meta-analysis, the significant result implies that the initial theoretical ideas are confirmed, and these significant variables should be included in the theory. In the meta-analysis study of leadership evaluation for male and female leaders, Eagly et al. (1992) built a multiple regression model to predict the variability of gender effect on leadership evaluation. The model was found to be statistically significant and accounted for nearly 35% of the variations of gender effect on leadership evaluation (multiple $R = .59$). The significant predictors in this model included leadership style portrayed, gender distribution in role, gender of subordinates, and type of organizational context. Overall, the results supported the researchers' initial hypothesis that female leaders received slightly more negative evaluations than did equivalent male counterparts. The significant moderator variables such as leadership style and organizational context add new perspectives to the gender role theory. Specifically, results from a homogeneity test and multiple regression analysis revealed that female leaders were devalued relative to male leaders when leadership was carried out in stereotypically masculine style, particularly when the leadership style was autocratic or directive. The devaluation of the female leaders was severe when they occupied male-dominated roles.

Although statistically significant results such as the one in the above example allow researchers to reject the null hypothesis and thus to accept the tentative research hypothesis, the confirmed theory is never complete and needs further refinement and development. One statistical indicator that signals the need for finding additional explanatory variables is the magnitude of the relationship between the dependent variable and a set of predictors. Such statistics as effect size, bivariate and multiple correlation coefficients, and coefficient of the determination can provide useful information to assess the extent to which a confirmed theory explains the social and organizational phenomena. Take the above mentioned meta-analysis as an example; the obtained R^2 for the multiple regression model was slightly smaller than .35, and this means that nearly 35% of the variability of effect size of the leadership evaluation was accounted for by study attribute variables related to gender. Put in another way, more than 65% of the variance of the gender effect on leadership evaluation was unexplained and due to factors other than those that had been included in the regression equation. The gender factors significantly contributed to the female leaders' devaluation, thereby confirming the initial theoretical ideas. But it would be incomplete to conclude that female leaders' devaluation relative to their male counterparts' was largely due to the gender factors and that the confirmed theory offers us a clear picture about the phenomenon of leadership evaluation between male and female leaders. There might be other important variables that have contributed to the gender effect on the female leaders' devaluation such as leader's capability, tenure and leadership experiences, validity of the evaluation instrument, and organizational politics. Therefore, it is the challenge for researchers to continuously search and verify other influential variables that might have explained the more than 65% of the variance of the gender effect of leadership evaluation. This outcome enables contribution to the fifth phase of continuous refinement and development in the general method model (Lynham, 2002).

As indicated in the above discussion, meta-analysis can be used as a theory-building research method. As shown in Figure 3, meta-analysis research can be used to contribute to four of the five phases of the general method of theory-building research in applied disciplines, namely, conceptual development, operationalization, confirmation or disconfirmation, and continuous refinement and development.

Strengths and Weaknesses of Theory Building From Meta-Analysis Method

Although meta-analysis has its unique features as a valuable research technique, the process of theory building from meta-analysis tends to be identifiable with other approaches. There are both advantages and disadvantages associated with a meta-analytic approach to theory building.

One advantage of a meta-analytic approach to theory building is its capacity to integrate and synthesize current empirical studies on a particular topic. As is illustrated in Figure 1, there may be a series of empirical studies for a set of theoretical constructs. Meta-analysis allows researchers to integrate the existing empirical findings with some sophisticated tools such as combined tests. Because different existing studies may come from various empirical areas, a combined test tends to cumulate the existing findings in a scientific way (weighted or unweighted cases) and thus offers results with more generalizability.

A second advantage of meta-analysis for theory building comes from its nature of analysis of analysis. Meta-analysis not only cumulates results from individual studies but also can be used to test complex theories involving many variables. Because social and organizational phenomena tend to be complex, different theories from various domains have been put forward to explain such phenomena. There might be several competing theories or theoretical frameworks within one research domain. For example, researchers can identify different theories and associated variables from training design (i.e., education), learners' individual characteristics (i.e., psychology), and organizational environment (i.e., sociology and organization theory) to explain the variability of transfer of training. Meta-analysis offers a useful method to evaluate the relative impacts of existing predictors on the dependent variable and thus provides aggregated empirical results for reviewing and judging existing theories or conceptual models.

A third advantage of using meta-analysis is its tendency to offer guidelines for variable selection and research design in future research. Meta-analysis reviews the selected literature with empirical evidences and thus provides a broad and updated outlook about the relations between theoretical ideas and empirical studies. The utility of such broad and updated outlooks is multiple. Researchers can use such information to reflect on the existing design and find some promising variables for future studies. They can also use such information to develop new conceptual and theoretical ideas based on empirical evidence revealed in meta-analysis such as moderators and interaction effects. In sum, meta-analysis allows researchers to develop and verify new theoretical ideas based on possible attributes and characteristics of all possible existing studies. That is to say, meta-analysis can follow a "research-then-theory-strategy" of theory building (Reynolds, 1971). Comparing other approaches with the "research-then-theory-strategy" in theory building, the main advantage of meta-analysis is that it is based on a number of proved empirical studies (i.e., published or other ways of being thus judged) instead of on single research.

A fourth advantage of using meta-analysis as a theory-building method comes from its role in continuous refinement and development of the existing theory. By identifying and testing those influential moderators and possible interaction effects, meta-analysis offers concrete conclusions about including newly proved variables or discarding old, less influential variables in the existing theories and conceptual models.

Although there are several advantages to using meta-analysis as a theory-building tool, some disadvantages should be acknowledged. The first disadvantage of meta-analysis is that the fundamental parameters of the theory used to explain social or organizational phenomena have been set up by the existing studies. The meta-analysis researcher cannot include and test variables that have not been examined in the existing studies. The research thus cannot confirm or disconfirm a distinctive theory that is beyond the thinking of the existing theoretical frameworks and empirical studies.

The second disadvantage of meta-analysis lies in its theory-building strategy and is due to the nature of the analysis of the existing analyses. The meta-analytic researcher cannot operationalize new theoretical ideas beyond the variables and study attributes that have been included in the existing studies. Consequently, a meta-analytic approach to theory building tends to be more applicable to a research-then-theory than a theory-then-research strategy of theory building. Meta-analysis, therefore, cannot be used to develop and validate a groundbreaking theory.

Conclusions

This chapter first describes major elements of meta-analysis as a research method and then identifies its unique role under the interrelationships between theory, research, and data analysis. Meta-analysis situates at the empirical level and offers distinctive contributions to the knowledge base by integrating and synthesizing existing empirical studies.

The chapter presents a process for using meta-analysis to inform theory-building research and to guide researchers in building explicit theory based on existing theory. This process, outlined in Figure 2, highlights clear steps in conducting meta-analysis for the purpose of informing theory building. It is evident that this process parallels with the general method of theory-building research in applied disciplines (Lynham, 2002) (see Figure 3). Several advantages of the meta-analytic approach to theory building were identified. The main strength of meta-analysis is in its powerful capability to confirm and refine a theory with all available empirical findings. However, its contribution to theory building tends to be constrained by the predetermined parameters in the existing theory and empirical studies. Meta-analysis is limited in developing and testing a revolutionarily new theory.

References

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review*, *14*(4), 496-515.

Baldwin, T. T., & Ford, K. J. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, *41*, 63-105.

Benson, J., & Hagtvet, K. (1996). The interplay among design, data analysis, and theory in the measurement of coping. In M. Zeidner & N. S. Endler (Eds.), *Handbook of coping: Theory, research, applications* (pp. 83-106). New York: John Wiley.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.

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

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997-1003.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: CBS College Publishing.

Dubin, R. (1978). *Theory building* (Rev. ed.). New York: Free Press.

Durlak, J. A., & Lipsey, M. W. (1991). A practitioner's guide to meta-analysis. *American Journal of Community Psychology*, *19*, 291-332.

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, *111*, 3-22.

Ford, K. J., & Weissbein, D. A. (1997). Transfer of training: An update review and analysis. *Performance Improvement Quarterly*, *10*, 22-41.

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, CA: Sage.

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

Holton, E. F., III, Baldwin, T. T., & Naquin, S. S. (Eds.). (2000). Managing and changing learning transfer systems. *Advances in Developing Human Resources*, *8*.

James, W. B., & Blank, W. E. (1993). Review and critique of available learning style instruments for adults. In D. D. Flannery (Ed.), *Applying cognitive learning theory to adult learning* (New Directions for Adult and Continuing Education, No. 59, pp. 47-57). San Francisco: Jossey-Bass.

Lynham, S. A. (2002). The general method of theory-building research in applied disciplines. *Advances in Developing Human Resources*, *4*(3), 221-241.

McGaw, B., & Glass, G. (1980). Choice of the metric for effect size in meta-analysis. *American Educational Research Journal*, *17*, 325-337.

Merriam, S. B., & Simpson, E. L. (2000). *A guide to research for educators and trainers of adults* (2nd ed.). Malabar, FL: Krieger.

Niemi, R. G. (1986). Series editor's introduction. In F. M. Wolf (Ed.), *Meta-analysis: Quantitative methods for research synthesis* (Sage University Papers: Quantitative Applications in the Social Sciences, 59, Series No. 07-059). Beverly Hills, CA: Sage.

Reynolds, P. D. (1971). *A primer in theory construction*. New York: Macmillan.

Rosethal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature review. *Annual Review of Psychology*, *52*, 59-82.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis* (Sage University Papers: Quantitative Applications in the Social Sciences, 59, Series No. 07-059). Beverly Hills, CA: Sage.