Chapter II

Literature Review

2.1. Introduction

This chapter will review concept of meta-analysis, the steps for doing meta-analysis, and strengths and weaknesses of meta-analysis. It also introduces the approaches toward meta-analysis and review recent researches on comparison of some of these approaches with different criteria.

2.2. What is Meta-Analysis?

Research use three basic types of analysis; primary analysis are used when the investigator collects data and analyses the results; Secondary analysis use data from a previous study to answer new questions, and finally, Meta-analysis combines the results of multiple primary studies to address research questions. Meta analysis is systematic reviews of the Literature (Conn, 2004). As defined, literature review is a broad term that includes two major types of reviews, non-systematic review and systematic review (see Figure 1).

Figure (1). Literature Reviews
As can be seen in figure 1, non-systematic reviews provide background information for research articles. This type of review does not minimize bias, thus risks offering misleading conclusions and misdirecting practice. In contrast, systematic reviews involve an attempt to discover inconsistencies and account for variability in studies that appear similar by applying the rules of the scientific process to extract and pool findings (Bowman, 2007).

Systematic reviews are a form of research. The population is all of the literature, the sample is the portion of the literature that is relevant to the review, and the data are the findings from each study (Cooper, 1998). A systematic review is a scientific tool that can be used to appraise, summarize, and communicate the results and implications of otherwise unmanageable quantities of research. Systematic reviews are of particular value in bringing together a number of separately conducted studies, sometimes with conflicting findings, and synthesizing their results. Hence, they are often called overviews.

Systematic reviews may be quantitative or qualitative and the stages of each type of review are identical (Bowman, 2007). Indeed, systematic reviews may or may not include a statistical synthesis. When a review strives to comprehensively identify and track down all the literature on a given topic (also called “systematic literature review”) systematic review is qualitative and called overview. But when a specific statistical strategy is taking for assembling the results of several studies into a single estimate, systematic review would be quantitative. As Rosenthal et. al., (2006) defined meta-analysis is a set of procedures for conducting quantitative
reviews of an existing research literature. Meta-analysis is a set of statistical methods for combining quantitative results from multiple studies to produce an overall summary of empirical knowledge on a given topic. It is used to analyze central trends and variations in results across studies, and to correct for error and bias in a body of research. Results of the original studies usually are converted to one or more common metrics, called effect sizes, which are then combined across studies. This allows us to synthesize results from studies that use different measures of the same construct or report results in different ways (Littell et al., 2008).

2.3. Steps in Meta-analysis

The steps in a typical meta-analysis include data collection, evaluation, analysis and interpretation, significance testing, and drawing conclusions, just as in other research types, see Table (1). Hence, meta-analysis can be called a primary research investigation in itself with unique innovative characteristics in relation to research design (Cooper, 1998). Perhaps the most obvious difference between meta-analysis and other primary research techniques is that studies (rather than people) are treated as the unit of analysis.
### Figure (2). Steps in the Systematic Review, Meta-Analysis and Survey Research

<table>
<thead>
<tr>
<th>Step</th>
<th>Systematic review &amp; Meta-analysis</th>
<th>Survey research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topic information</strong></td>
<td>Central question, hypotheses, objectives</td>
<td>Central questions, hypotheses, objectives</td>
</tr>
</tbody>
</table>
| **Overall study design** | Protocol development  
Specify problems/conditions, populations, settings, interventions, and outcomes of interest  
Specify study inclusion and exclusion criteria | Protocol development  
Specify key constructs, information needs  
Specify sample characteristics |
| **Sampling**          | Develop a sampling plan  
Sampling unit is study  
Consider universe of all potentially relevant studies  
Obtain studies | Develop a sampling plan  
Sampling unit may be the individual, household, or group  
Identify sampling frame  
Sample units |
| **Data collection**   | Data are derived (extracted) from studies onto standardized forms | Data are collected from individuals via self-administrated surveys or interviews |
| **Data analysis**     | Descriptive data(examine study qualities, samples, and intervention characteristics: compute effect sizes)  
Pool effect sizes and assess heterogeneity (meta-analysis) | Descriptive data(examine qualitative and categorical data, frequencies and distributions on continuous variables) Measures of central tendency and variability |
| **Reporting**         | Description of results in narrative, tables, and graphs  
Interpretation and discussion  
Implication for policy, practice and further research | Description of results in narrative, tables, and graphs  
Interpretation and discussion  
Implications for policy, practice, and further research |
Although, the meta-analysis includes several phases that are parallel to those of primary researches, however, specifically, meta-analysis involves the following steps based on Streiner (2003):

*Step 1: Defining the question*

It may seem that defining the question to be addressed by the meta-analysis is a simple and straightforward task. If it looks easy and problem-free, there are major problems ahead, and this is no exception. Questions such as: Does psychotherapy work? Does computer assisted instruction lead to more learning than traditional instruction? Is mastery learning better than traditional learning? are too broad to yield meaningful conclusions. Does "psychotherapy," for example, mean reading a self-help book, or spending a few sessions with a college counselor, or completing a multi-year psychoanalysis? It is quite possible that particular psychotherapy approach that works for some people may not be effective for others. If the results of different studies are combined, misleading results may be drawn. To avoid vague generalities researcher must make the focus of his meta-analysis much more explicit by establishing criteria for including or excluding studies. However, the criteria cannot be too restricting, or you might not find a sufficient number of studies; and even if researcher does find enough studies, he might not find anything interesting or illuminating. The more focused the question, the more useful and more accurate the results of the meta-analysis.
Step 2: Detailing the selection criteria

Once the question has been defined, meta-analyst should prepare a checklist so that the criteria can be applied in a uniform fashion. The list need not be long, but should include all of the reasons for accepting or rejecting studies. For example, if the question reads: Is spending a few sessions with a college counselor effective for adults with minor problem, and researcher wants to look only at this type of psychotherapy, during selecting the studies if a study used various form of psychotherapy but the results were reported separately for each form, then this study should be included in meta-analysis. But if the results were combined the study would be rejected.

Step 3: Doing the search

The next step is to actually find the studies. What is intended is to locate all the studies done that contain data relevant for the goal of meta-analysis, whenever they are enough homogeneous with the other as to be integrated. There are many sources where the several procedures for the search are described (Botella and Gambara, 2002; Cooper and Hedges, 1994). However, the tendency is to rely more and more on the electronic bases, as they are increasingly wide in the amount and the variety of their input sources. Computerized databases, such as PsychINFO, ERIC and the like have made searching easier in this regard. If we add the increasing ease of access and the more and more sophistication of the criterion for searching, the result is that today the tendency is to put the main accent on that source. Within the past few years, other resources have been developed that are
extremely useful. One of these resources is the Cochrane Databases of Systematic Review, which is an on-line set of meta-analyses. There are a number of advantages to this database. The main advantage is that someone has already done the work for meta-analyst. Second strict methodological criteria have been used in the selection of the primary articles, so researcher can be sure that all of the studies in the review have met exacting standards.

Finally, an excellent source is unpublished studies (e.g., doctoral dissertations, congress abstracts, etc.) in order to minimize the publication bias (Cooper, 2010; Lipsey and Wilson, 2001).

Step 4: Selecting the studies

This step consists of applying the selection criteria advised in Step 2 to the studies found in Step 3. The most important point of this step is to avoid any suspicion that studies were rejected because they failed to show what the reviewer wanted, rather than not meeting the criteria. The best way to ensure this is to have two or more independent reviewer evaluate each study.

Step 5: Appraising the studies

Step 4 addressed the minimal criteria for an article to be included in the meta-analysis. However, all studies are not created equal. Cook and Campbell’s (1979) framework can be used for differentiating between the internal and external validity of a study. Internal validity refers to how well the study itself was conducted, and the degree to which we can believe the findings; external validity relates to the ability to generalize the results from the study sample to the
population at large. Numbers of checklists have been developed over the years that allow researchers to evaluate the design and execution of a study. The researcher has to decide at what point based on the evaluating checklist violation of internal and external validity jeopardize the study.

*Step 6: Abstracting the results*

Key elements of each study now have to be abstracted from the articles and entered into a spreadsheet, or a program specifically designed to do meta-analyses. What should be abstracted? Once collected the studies that are going to be part of the meta-analysis and once adopted an effect size index, a detailed exploration of the studies is in order. The objectives are to obtain from each study the information needed to calculate the effect size and code it according to a series of categories previously defined. The main question is, of course, how many and which must be those characteristics. It must include all the characteristics suspicious of having something to do with what researcher is studying. But if he only study those characteristics it is possible to let out the variable with the largest covariation with the effect size. The characteristics quantified and coded are usually classified in three groups: substantive, methodological and extrinsic (Lipsey, 1994).

*Step 7: Calculating effect sizes*

One major problem in combining various studies is that they often use different outcome measures. The issue is to find a common yardstick, so that the results are all reported using the same metric. The most commonly used measure
is the effect size (ES). Effect size which is mean of control group minus mean of treatment group divided by standard deviation, expresses the results in standard deviation units allows reviewers to use the table of the normal curve to figure out what proportion of people in the treatment group did better than the average person in the control group.

**Step 8: Combining the studies**

Once the ES has been derived for each study, we have to summarize (or “pool”) them in some way to get an estimate of the mean; that is, an overall estimate of the effectiveness or ineffectiveness of the intervention. The simplest way is to add them up and divide by the number of ESs; after all, that is what we mean by the “mean.”

**Step 9: Selecting the type of analysis**

There are two general approaches to analyzing the results of meta-analyses: a fixed-effects model and a random-effects model. A fixed-effects model assumes that there is a “true” effect size that underlies all of the studies, and that they differ among each other only because of sampling error. A random-effects model makes the assumption that there is a population of effect sizes, from which the studies in the meta-analysis are a random sample (Hedges & Vevea, 1998). The reason that this distinction is important is that, in many situations, the two types of analyses yield different results. A fixed-effects model is less conservative and may give statistically significant results in some situations when a random-effects model will not. So, which model is it appropriate to use and when? Fixed-effects model is
appropriate if we want to draw conclusions about the particular set of articles in
the meta-analysis. That is, it does not allow us to say anything about studies that
may have been missed or those that will be done in the future. On the other hand, a
random-effects model is perhaps more realistic in two regards. First, by saying that
there is a population of effect sizes, the model acknowledges the fact that studies
differ with respect to the sample, the procedures used and other aspects of the
design, all of which may result in different findings. Second, it allows generalizing
from this particular set of articles to studies of this phenomenon in general; studies
that did not include and studies yet to be done. In most situations, and especially if
the test of homogeneity is significant, it would be wise to go with a random-effects
model (Streiner, 2003).

2.4. Putative advantages or strengths of Meta-analysis

Why should one consider using meta-analysis to summarize and analyze a
body of research studies rather than conventional research reviewing techniques?
If science view as the accumulation and refinement of information and knowledge
(Hunter et al., 1982; Pillemer and Light, 1980), it then becomes critical to
establish guidelines for reliable and valid reviews, integrations, and syntheses of
studies examining similar research questions (Cooper, 1982; Jackson, 1980).
Procedures employed in meta-analysis permit quantitative reviews and syntheses
of the research literature that address these issues.

Potential problems with traditional literature reviews that are addressed in
meta-analysis include (1) selective inclusion of studies, often based on the
reviewer's own impressionistic view of the quality of the study, (2) differential subjective weighting of studies in the interpretation of a set of findings, (3) misleading interpretations of study findings, (4) failure to examine characteristics of the studies as potential explanations for disparate or consistent results across studies, and (5) failure to examine moderating variables in the relationship under examination.

What is needed are approaches that will integrate results from existing studies to reveal patterns of relatively invariant underlying relations and causalities, the establishment of which will constitute general principles and cumulative knowledge (Hunter et al., 1982). The "fundamental problem," as Glass et al. (1981) refer to it, is the inability of the human mind to address this task reliably and validly given the enormous amount of data that must be gathered, processed, assimilated, and synthesized in many disciplines. It is ironic that the traditional review of scientific data has typically been done in an unscientific, impressionistic fashion. There are basically four reasons that constitute the primary advantages of meta-analysis.

First, meta-analysis procedures impose a useful discipline on the process of summarizing research findings. Good meta-analysis is conducted as a structured research technique in its own right and hence requires that each step be documented and opens to scrutiny. It involves specification of the criteria that define the population of study findings at issue, organized search strategies to identify and retrieve eligible studies, formal coding of study characteristics and
findings, and data analysis to support the conclusions that are drawn. By making
the research summarizing process explicit and systematic, the consumer can assess
the author's assumptions, procedures, evidence, and conclusions rather than take
on faith that the conclusions are valid.

Second, meta-analysis represents key study findings in a manner that is more
differentiated and sophisticated than conventional review procedures that rely on
qualitative summaries or "vote-counting" on statistical significance. By encoding
the magnitude and direction of each relevant statistical relationship in a collection
of studies, meta-analysis effect sizes constitute a variable sensitive to findings of
different strength across studies. By contrast, using statistical significance to
differentiate studies that find effects from those that do not is potentially quite
misleading. Statistical significance reflects both the magnitude of the estimated
effect and the sampling error around that estimate, the latter almost entirely a
function of sample size. Thus, studies with small samples may find effects or
relationships of meaningful magnitude that are not statistically significant because

Third, meta-analysis is capable of finding effects or relationships that are
obscured in other approaches to summarizing research. Qualitative, narrative
summaries of findings, while informative, do not lend themselves to detailed
scrutiny of the differences between studies and associated differences in their
findings. The systematic coding of study characteristics typical in meta-analysis,
on the other hand, permits an analytically precise examination of the relationships
between study findings and such study features as respondent characteristics, nature of treatment, research design, and measurement procedures. Furthermore, by estimating the size of the effect in each study and pooling those estimates across studies (giving greater weight to larger studies), meta-analysis produces synthesized effect estimates with considerably more statistical power than individual studies. Thus, meaningful effects and relationships upon which studies agree, and differential effects related to study differences, are both more likely to be discovered by meta-analysis than by less systematic and analytic approaches.

Fourth, meta-analysis provides an organized way of handling information from a large number of study findings under review. When the number of studies or amount of information extracted from each study passes a fairly low threshold, note-taking or coding on index cards cannot effectively keep track of all the details. The systematic coding procedures of meta-analysis and the construction of a computerized database to record the resulting information, by contrast, have almost unlimited capability for detailing information from each study and covering large numbers of studies. It hasten to add, however, that meta-analysis does not require large numbers of studies and, in some circumstances, can be usefully applied to as few as two or three study findings. The claimed strengths of meta-analysis are also presented in Figure (2).
**Figure (3). Claimed strengths of Meta-analysis**

<table>
<thead>
<tr>
<th>Strength</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can reveal research designs as moderators of study results</td>
<td>Cooper</td>
</tr>
<tr>
<td>Can determine if the effect of the intervention is sufficiently large in</td>
<td>Lipsey and Wilson</td>
</tr>
<tr>
<td>practical as well as statistical terms</td>
<td></td>
</tr>
<tr>
<td>Can allow more objective assessment of evidence and thereby</td>
<td>Egger and Smith</td>
</tr>
<tr>
<td>reduce disagreement.</td>
<td></td>
</tr>
<tr>
<td>Can clarify heterogeneity between study results</td>
<td>Egger and Smith</td>
</tr>
<tr>
<td>Can suggest promising research questions for future study</td>
<td>Egger and Smith</td>
</tr>
<tr>
<td>Can assist accurate calculation of sample size needed in future</td>
<td>Egger and Smith</td>
</tr>
<tr>
<td>studies.</td>
<td></td>
</tr>
<tr>
<td>Can increase precision of literature reviews</td>
<td>Cooper and Hedge</td>
</tr>
<tr>
<td>Can, through consistent coding of primary study characteristics and</td>
<td>Cooper and Hedge</td>
</tr>
<tr>
<td>use of multiple judges, reduce bias in judgments about the “quality” of</td>
<td></td>
</tr>
<tr>
<td>individual studies.</td>
<td></td>
</tr>
<tr>
<td>Can, through various statistical formulae, provide confidence</td>
<td>Cooper and Hedge</td>
</tr>
<tr>
<td>interval calculation of effect size estimates.</td>
<td></td>
</tr>
<tr>
<td>Can, using different assumptions or alternative statistical models,</td>
<td>Cooper and Hedge</td>
</tr>
<tr>
<td>clarify and interpret the range of possible conclusions about the</td>
<td></td>
</tr>
<tr>
<td>“quality” of segments of the literature review.</td>
<td></td>
</tr>
<tr>
<td>Can overcome problems of traditional literature reviews involving</td>
<td>Wolf</td>
</tr>
<tr>
<td>(a) selective inclusion of studies, (b) subjective weighting of studies</td>
<td></td>
</tr>
<tr>
<td>and their interpretation, (c) failure to examine study characteristics as</td>
<td></td>
</tr>
<tr>
<td>source of disparate or consistent results across studies and (d)</td>
<td></td>
</tr>
<tr>
<td>failure to address influence of moderating variables in the</td>
<td></td>
</tr>
<tr>
<td>relationship being examined.</td>
<td></td>
</tr>
<tr>
<td>Can, through systematic use of threats-to-inference framework,</td>
<td>Campbell-Tanley &amp; Noble</td>
</tr>
<tr>
<td>reveal structural flaws and sources of bias in research procedures.</td>
<td></td>
</tr>
</tbody>
</table>

*Source: Noble, (2006).*
2.5. The weaknesses of Meta-analysis

Meta-analysis is not without disadvantages and it is the subject of harsh criticism from some quarters (Sharpe, 1997). One disadvantage of meta-analysis is simply the amount of effort and expertise it takes. Properly done, a meta-analysis with more than a handful of study findings is labor intensive and takes considerably more time than a conventional qualitative research review. Additionally, many aspects of meta-analysis require specialized knowledge, especially the selection and computation of appropriate effect sizes and the application of statistical analysis to them.

Another concern about meta-analysis relates to its structured and somewhat mechanical procedures, which, in other regard, can be viewed as strengths. For some applications (and some critics say for all applications), the relatively objective coding of data elements and effect sizes from research studies, and the type of analysis to which such data lend themselves, may not be sensitive to important issues, e.g., the social context of the study, theoretical influences and implications, methodological quality, more subtle or complex aspects of design, procedure, or results, and the like. To draw on the survey research analogy used earlier, meta-analysis is a structured, closed-ended questionnaire approach to summarizing research findings. Some survey applications require a more open-ended approach, e.g., unstructured interviews or focus groups, to deal with the complexity or subtlety of certain topics. It may well be that some research issues also require a more qualitative assessment and summary than meta-analysis can
provide. Of course, there is no reason in principle why both meta-analytic and qualitative reviews cannot be done on the same body of research findings, with overall conclusions drawn from both. One approach to meta-analysis, what Slavin (1986, 1995) calls *best evidence synthesis*, attempts to do just that, combining qualitative and quantitative reviewing techniques in the same research review. Perhaps the most persistent criticism of meta-analysis has to do with the mix of studies included. Critics argue, with some justification, that mean effect sizes and other such summary statistics produced by meta-analysis are not meaningful if they are aggregated over incommensurable study findings. There would be little sense, for instance, in constructing the distribution of effect sizes for a mix of study findings on methadone maintenance for drug abusers, gender differences in social skills, and effects of unionization on employee morale. On the other hand, few would object to a meta-analysis of findings from virtual replications of the same study. Most of the criticism on this point, however, has been well short of such obvious extremes. The gray area in between becomes controversial when a meta-analyst includes study findings that are clearly not replications, but are claimed to relate to a broader theme. As noted earlier, Smith and Glass (1977) included a wide range of studies of psychotherapy in their pioneering meta-analysis on the grounds that the issue of interest to them was the overall effectiveness of psychotherapy. However, they were stridently criticized by researchers who saw vast differences between the different therapies and different outcome variables and felt it was misleading to report average effectiveness over
such distinct approaches as behavioral therapy, psychodynamic therapy, and gestalt therapy and such diverse outcomes as fear and anxiety, self-esteem, global adjustment, emotional-somatic problems, and work and school functioning.

The problem comes in, of course, when the different types of study findings are averaged together in a grand mean effect size. Meta-analysts who wish to deal with broad topics are increasingly approaching their task as one of comparison rather than aggregation. Where distinctly different subcategories of study findings are represented in a meta-analysis, they can be broken out separately and the distribution of effect sizes and related statistics can be reported for each, permitting comparison among them. Additionally, technical advances in meta-analysis have made it possible to statistically test for homogeneity to determine if a grouping of effect sizes from different studies shows more variation than would be expected from sampling error alone. This provides an empirical test for whether studies show such disparate results that it may not be plausible to presume that they are comparable. Put differently, contemporary meta-analysis is increasingly attending to findings, rather than aggregating results together into a grand average. This emphasis provides a more careful handling of distinctly different subgroups of study findings and runs less risk of vexing critics who are concerned about such differences.

A related and more troublesome issue is the mixing of study findings of different methodological quality in the same meta-analysis. Some critics argue that a research synthesis should be based only on findings from the highest quality
studies and should not be degraded by inclusion of those from methodologically flawed studies. Indeed, some approaches to meta-analysis set very strict methodological criteria for inclusion, e.g., the best evidence synthesis approach mentioned earlier (Slavin, 1986). What makes this point controversial is that there are problematic trades-offs and judgment calls whichever way the meta-analyst goes. One difficulty is that, aside from a few simple canons, there is relatively little agreement among researchers on what constitutes methodological quality. Moreover, few research areas provide studies that all reviewers would agree are methodologically impeccable in sufficient numbers to make a meaningful meta-analysis. Many areas of research, especially those that deal with applied topics, provide virtually no perfect studies and the ones closest to textbook standards may be conducted in circumstances that are unrepresentative of those in which the meta-analyst is most interested. For instance, methodologically rigorous studies of psychotherapy are more likely in demonstration projects and university clinics than in routine mental health practice. Thus, much of the knowledge we have on some issues resides in studies that are methodologically imperfect and potentially misleading. The meta-analyst must decide how far to go with inclusion of findings from studies that are judged interpretable but flawed, knowing that relaxed methodological standards may result in a derisive reproach of "garbage in, garbage out," while stringent standards are likely to exclude much, or most, of the available evidence on a topic.
Two approaches have emerged on this issue. One is to keep the methodological criteria strict and accept the consequences in regard to the limitations thus imposed on the proportion of available and relevant study findings that may be included. In this instance, the meta-analyst has assurance that the synthesis is based on only the "best" evidence but its results may summarize only a narrow research domain and have little generality. The other approach is to treat methodological variation among studies as an empirical matter to be investigated as part of the meta-analysis (Greenland, 1994). In this case, less stringent methodological criteria are imposed, but the meta-analyst carefully codes methodological characteristics that may the variance of effect size distributions rather than the means of those distributions. That is, the primary question of interest often has to do with identifying the sources of differences in study influence the study findings. One phase of statistical analysis then investigates the extent to which various methodological features are related to study findings (e.g., random vs. nonrandom assignment in treatment studies). If a questionable methodological practice has no demonstrable relationship to study findings, the corresponding findings are included in the final analysis, which thus gains the benefit of the evidence they contain. If, however, studies with the questionable practice show results significantly different from those without that problem, they can then be excluded from the final results or used only with statistical adjustments to correct for their bias.
Glass et al. (1981) have also grouped criticisms on meta-analysis into four categories:

1. Logical conclusions cannot be drawn by comparing and aggregating studies that include different measuring techniques, definitions of variables (e.g., treatments, outcomes), and subjects because they are too dissimilar.

2. Results of meta-analyses are non-interpretable because results from "poorly" designed studies are included along with results from "good" studies.

3. Published research is biased in favor of significant findings because non-significant findings are rarely published; this in turn leads to biased meta-analysis results.

4. Multiple results from the same study are often used which may bias or invalidate the meta-analysis and make the results appear more reliable than they really are, because these results are not independent.

The first criticism has been referred to as the "apples and oranges problem," in that it is argued that diversity makes comparisons inappropriate. For example, in one of the first meta-analyses, Smith and Glass (1977) synthesized the results of approximately 400 evaluations of the efficacy of psychotherapy and found (1) that the average therapy client is better off than 75 percent of untreated individuals and (2) virtually no difference between behavioral and non-behavioral therapies. Presby (1978) criticized Smith and Glass for ignoring "important differences among the non-behavioral therapies, for example, the superior effects of rational-
emotive therapy (RET) as compared to the others in that class. These differences are cancelled in the use of very broad categories, i.e., mixing 'apples and oranges,' which leads to the erroneous conclusion that research results indicate negligible differences among outcomes of different therapies." Similarly, Slavin (1983) took exception with the definitions of cooperation, competition, and achievement used in a meta-analysis conducted by Johnson et al. (1981). These definitions and criteria clearly affect the type of study to be included in the research synthesis, which may affect the results that follow. This issue may be dealt with empirically by coding the characteristics for each study and statistically testing whether these differences are related to the meta-analytic results. Even the most prevalent problem of differing operational definitions and measurement procedures for dependent variables may be examined empirically. The second criticism can also be handled empirically within meta-analyses by coding the quality of the design employed in each study and examining whether the results differ for poorly and well designed studies. A review of meta-analyses that have been done thus far suggests that the magnitude of the effect is unrelated to the worthiness of the design in some research domains but not in others. Even though no significant differences in effect size between poorly and well designed studies may be found in a meta-analysis, there may be considerably more effect size variation among "poorly" designed than well-designed studies (i.e., a significant difference in variance).
The third criticism, relating to the non-typicality and bias in favor of significant results in published research studies, can be addressed in several ways. One approach is to review results in books, dissertations, unpublished papers presented at professional meetings, and the like and compare them to the results for published articles. Another approach is to estimate the number of additional studies with non-significant results that would be necessary to reverse a conclusion drawn from the meta-analysis, thus providing some estimate of the robustness and validity of the findings.

The fourth criticism concerns the number of results from the same experimental study that should be used. Some meta-analysts (e.g., Kulik, 1983) choose to perform separate analyses for each different outcome (criterion or dependent variable), while others, including Glass, choose to lump them into the same analysis. Alternatively, some reviewers choose to limit themselves to a fixed number of results, perhaps two, from each study while others take the average of all results from the same study. Again, this is an empirically answerable issue that may influence the obtained results.

Other criticisms of meta-analysis include the assertion that interaction effects are ignored at the expense of main effects (Cook and Leviton, 1980; Slavin, 1983). Again, this can be addressed by examining the potential mediating effects of substantive and methodological characteristics of studies. Cooper and Arkin (1981) suggest the possibility of focusing future meta-analyses on particular effects for clearly articulated interaction hypotheses.
What may be the most important caveat is that "meta-analysis can have mischievous consequences because of its apparent 'objectivity,' 'precision,' and 'scientism.' To naive readers these lend social credibility that may be built on procedural invalidity" (Cook and Leviton, 1980). However, it has been pointed out (Cooper and Arkin, 1981) that this statement is true for any innovative methodology and that this problem resides within the particular use and user rather than in the approach per se. It should be noted that there are great differences in the quality of meta-analyses regarding issues of validity and reliability. Slavin (1983) maintains: "What traditional reviews usually do that meta-analyses do not is discuss the studies being reviewed, looking for patterns and inconsistencies, and placing more weight on studies that use strong designs than on numbers of studies falling on one or the other side of an issue." He advocates the use of meta-analysis to "enhance rather than replace an intelligent discussion of the critical issues." It is essential that sufficient information on the coding procedures for studies in a meta-analysis be presented so that the methodological rigor of the particular application can be determined. Excellent detailed summaries of how to approach the location, retrieval, and coding of studies is presented in Glass et al. (1981) and Hunter et al. (1982) and is not included here. It has also been suggested that nonparametric rather than parametric statistics are more appropriate in the quantitative analyses of results from independent studies because distributions of effect sizes are often highly skewed (Slavin, 1983; Kraemer and Andrews, 1982).
After a brief summary of the advantages and disadvantages of meta-analysis, common approaches of conducting a meta-analysis, and the issues and considerations inherent in each approach will explain.

2.6. Meta-analysis approaches

According to Hunter Schmidt (2004), approaches of meta-analysis are classified in three groups: purely descriptive approaches that their purpose is simply to summarize and descriptive the studies in a research literature. Second group are those usually address just sampling error, and the last are psychometric approaches that their purpose is to estimate as accurately as possible the construct level relationships in the population.

2.6.1. First category (Descriptive meta-analysis approach (Glassian and related approaches))

Purely descriptive approaches (Glass approaches and study effects meta-analysis approaches) paint a descriptive picture of what is in the research literature but do not attempt to analyze, correct for, or otherwise address any of the artifacts that distort study findings.

For Glass, the purpose of meta-analysis is descriptive; the goal is to paint a very general, broad and inclusive picture of a particular research literature (Glass, 1977). The questions to be answered are very general; Glassian approaches often combine studies with somewhat different independent variables and different dependent variables. As a result, some have criticized these approaches as
combing appeals and oranges. Glassian meta-analysis has three primary properties:

i. A strong emphasis on effect size rather than significance levels.

ii. Acceptance of the variance of effect sizes at face value.

iii. A strongly empirical approach to determining which aspects of studies should be coded and tested for possible association with study outcomes.

The major criticisms of Glassian approaches are the following:

i. In Glassian approaches include all studies in the meta-analysis regardless of methodological quality (Bangert & Downs, 1986, Salvin, 1986) is a serious problem.

ii. Glass approaches mix very different independent variables in the meta-analysis, thereby masking important differences in the mean outcomes for different independent variables.

iii. The lost major criticism is that Glassian approaches mix measures of very different dependent variables. For example variables like attitudes, beliefs, disciplinary behavior and academic achievement may all be included in the same meta-analysis.

Study effect meta-analysis approaches as a response to criticisms

One variation on Glass’s approaches has been labeled study effects meta-analysis by Bangert-Drowns (1986). These approaches differ from Glass’s procedures in several ways. First only one effect size from each study is included
in the meta-analysis. Second, this procedure calls for the meta-analysis to make at least some judgments about study methodological quality and to exclude studies with deficiencies judged serious enough to distort study outcomes. This procedure seeks to determine the effect of a particular treatment on a particular outcome rather than to paint a broad Glassian picture of research area (Hunter & Schmidt, 2004).

2.6.2. Second category

Approaches that address any the artifact of sampling error. These include the homogeneity test based approaches of Hedges and Olkin (1985), Rosenthal and Rubin (1982a, 1982b), and also a “bare-bones” meta-analysis approach. The most important characteristic of these approaches is that they don’t address the effects of artifacts other than sampling error. In particular they do not address measurement error. Meta-Analysis approaches focusing only on sampling error are as follows:

1. Homogeneity Test-based approaches

2. Bare-Bones approaches

2.6.2.1. Homogeneity test based approaches

Homogeneity test based approaches have been advocated independently by Hedges (1982 Hedges & Olkin 1985) and by Rosenthal and Rubin (1982a, 1982b). These approaches proposed that chi-square statistical tests be used to decide whether study outcomes are more variable than would be expected from sampling error alone. If these chi-square tests of homogeneity are not statistically
significant, then the population correlation or effect size is accepted as constant across studies and there is no search for moderators. Although Hedges (1982b) and Hedges and Olkin (1985) extended the concept of homogeneity tests to develop a more general procedure for moderator analysis based on significance testing.

Criticisms of these approaches

First of all chi-square test of homogeneity typically has low power to detect variation beyond sampling error (Hedges & Pigott, 2001). Hence, the meta-analyst will often conclude that the studies are homogeneous when they are not.

Another problem is that the chi-square test has a type I bias (Schmidt & Hunter, 2003). Under the null hypotheses, the chi square test assumes that all between study variance in study outcomes is sampling error variance; but there are other purely artificial sources of variance between studies in effect sizes.

The other problem is that homogeneity test-based meta-analysis represents a return to the practice that originally led to the great difficulties in making sense out of research literatures reliance on statistical significance tests (Hunter & Schmidt, 2004).

2.6.2.2. Bare-Bones Meta-Analysis Approaches

The other approach to meta-analysis in this group that attempt to control for the artifact of sampling error is what we referred to earlier as bare bones meta-analysis. This approach can be applied to correlations, d values, or any other effect size statistic for which the standard error is known. Because there are always other
artifacts (such as measurement error) that should be corrected for, we have consistently stated in our writing that the bore-bones approaches are incomplete and unsatisfactory (Hunter & Schmidt 2004).

2.6.3. Third category (Psychometric Meta-analysis Correction for multiple artifacts)

Psychometric approach refers to the approaches that address and correct for the effects of not only sampling error but also a variety of other artifacts that distort study results. These approaches estimate the results in a methodological unflawed manner. These approaches called psychometric meta-analysis approaches. They follow the purpose that Rubin (1990) stated meta-analysis approaches should serve. These approaches correct not only for sampling error (an unsystematic artifact) but for other, systematic artifacts, such as measurement error, range restriction or enhancements, dichotomization of measures, and so forth. These other artifacts are said to be systematic because, in addition to creating art factual variation across studies, they also create systematic downward biases in the results of all studies.

Psychometric meta-analysis corrects not only for the art factual variation across studies, but also for the downward biases. Psychometric meta-analysis is the only approach that takes into account both statistical and measurement artifacts (Hunter, Schmidt 2004).

As we know sampling error and measurement error have a unique status among the statistical and measurement artifacts with which meta-analysis must
deal. They are always present in all real data. Other artifacts, such as range restriction, artificial dichotomization of continuous variables, or data transcription errors, may be absent in a particular set of studies being subjected to meta-analysis.

Meta-analysis has clearly shown that no single primary study can ever resolve an issue or answer a question. Research findings are inherently probabilistic (Kulik, 1984), and therefore the results of any integration of findings across studies can control. Sampling error and other artifacts provide a foundation for conclusions. And yet meta-analysis is not possible unless the needed primary studies are conducted. In fact meta-analysis represents an improvement for the primary researcher in one respect—all available relevant studies are included in a meta-analysis and hence every study has an effect (Hunter & Schmidt, 2004).

According to Mullen, B., Jhonson, T.B., and Salas, E. (1995) there are three major meta-analytic approaches which are different from the statistical point of view. They are as follows:

1. **Hedges and Olkin Techniques:**

   The techniques in this approach to meta-analysis have been developed since the early 1980 and thus represent a relatively new entry into the meta-analytic arena. In this approach, study outcomes usually are converted into standard deviation units, or $g$ values, which are then corrected for bias (overestimate of the population effect size, which occur especially for small study samples). Finally, these transformed values, which are termed $d$ values, are combined, their
homogeneity is examined and their variability is explained by using models with continuous or categorical moderators.

2. Rosenthal and Rubin Techniques

This meta-analytic approach is the oldest set of techniques in the meta-analytic arena. The basic logic of this approach is to convert study outcomes to standard normal metrics, which are $Z$s associated with one-tailed probabilities for significance levels and Fisher’s $r$ to $z$ transformation for effect sizes. These indexes are then combined to produce weighted means and are examined in diffuse and focused comparisons.

3. Hunter et al. Techniques

Stemming from the validity generalization tradition within industrial–organizational psychology, this meta-analytic approach differs from the Hedges and Olkin and the Rosenthal and Rubin approaches. It does not attempt to correct the biases in the effect size indexes of $r$ or $g$ before deriving mean effect size or before applying moderators to these indexes. However, this approach is more sophisticated than other frameworks in its efforts to correct effect size indexes for potential sources of error (e.g., sampling error and measurement error) before meta-analytically integrating the effect sizes across studies. This unique focus of the Hunter and Schmidt approach is seldom fully used, however, because in most literatures, very few studies report sources of error as a matter of course. Yet, when information about these sources of error is available, this feature of the Hunter & Schmidt approaches may recommended its use.
2.7. **Comparison of Approaches in Meta-Analysis:**

Several recent studies have compared some of the approaches that mentioned above. At the following paragraphs some of these studies which can be connected to the present study have been reviewed.

Johnson, Mullen and Salas (1995) compared Hedges-Olkin, Rosenthal-Rubin and Hunter-Schmidt meta-analytic approaches by manipulating a single data set. They concluded that the significance of the mean effect size differed substantially across the approaches: the Hunter and Schmidt approach reached more conservative estimates of significance than the other two approaches so should be used cautiously. They assumed that three general meta-analytic approaches should generally show equivalent results. Their comparison of the frameworks demonstrates some respects in which three approaches coverage but other respects in which they diverge. All three frameworks converged in their estimate of mean effect size and in variability of the effect sizes, but they diverged in terms of estimating the significance of the mean effect size and in predicting effect sizes by using a moderator. Across the various meta-analytic outcomes, the Hedges and Olkin and Rosenthal and Rubin approaches proved to be quite similar to each other, whereas the Hunter & Schmidt approach proved to be the most dissimilar from the others. This later framework consistently reached:

(a) More conservative estimates of the significance of effect size

(b) Widely variant estimates of moderators
According to findings of this study perhaps the simplest way to summarize the degree to which the three meta-analytic frameworks converged is to correlate the derivative standardized statistics that each framework produced across databases. Thus their analysis found that the assumption of equivalence of frameworks is justified in considering the Hedges and Olkin approach versus the Rosenthal and Robin approach. However, the Hunter & Schmidt approach often responses differently o variations in parameters of meta-analytic databases. These results suggest that the Hedges and Olkin and the Rosenthal and Rubin approach will yield equivalent results under most, if not all, meta-analytic circumstances but that the Hunter &Schmidt approach will frequently yield different results.

The patterns presented in this article indicated that the choice of either the Hedges & Olkin approach or the Rosenthal &Rubin approach appears to be a reasonable one. However, the divergent results produced by the Hunter &Schmidt approach suggests that , in the future, meta-analysists should be prepared to justify their choice of this approach over the other two approaches.

To sum it should be recognized that the equivalent results referred by the Hedges& Olkin and the Rosenthal &Rubin approaches are not only consistent with one another but also remarkably consistent with conventional expectations, whereas the distinct tendency of Hunter & Schmidt, framework to produce result tat violate conventional expectations suggests that it should be used with caution.
Schmidt and Hunter (1999) subsequently claimed that Johnson et al. incorrectly applied their approach and showed that, theoretically when the approach was correctly applied, their approach was comparable to that of Hedges.

Fuller and Hester (1999) compared the sample weighted approach (Hunter & Schmidt, 1990) and a newer unweighted approach (Oburn & Callender 1992) by using actual data. Their results showed that while both approaches may generate similar parameter and variance estimates in primary meta-analysis, they may lead researchers to reach different substantive conclusions in the analysis of moderators (Fuller & Hester, 1999).

Field (2001) highlighted some other concerns with Johnson et al.’s approaches and rectified these concerns in a series of simulations that compared the approaches across a variety of situations. Field found that when comparing random effect approaches, the Hunter-Schmidt approach yielded the most accurate estimates of population effect size across a variety of situations. However, neither approach controlled the type I error rate when 15 or fewer studies were included in the meta-analysis, and that the approach described by Hedges and Vevea (1998) controlled the type I error rate better than the Hunter-Schmidt approach when 20 or more studies were included.

In a more recent set of simulations, Field (2002) demonstrated that across a far-ranging set of situations both approaches produce biased estimates of the population effect size: However, the biases in the Hunter-Schmidt approach are not as large as in Hedges’ approach. Hedges approach did tend to keep tighter
control of type I error rate but with 80 or more studies in the meta-analysis, there was little to separate the two approaches. With fewer studies in the meta-analysis (20-40), Hedges approach controlled the type I error rate considerably better than Hunter and Schmidt’s approach. Aguinis, Sturman and Pierce (2008) compared the Hedges and Olkin, the Hunter and Schmidt and a refinement of the Aguinis and Pierce meta-analytic approaches for estimating moderating effects of categorical variables. They compared three meta-analytic approaches in terms of their point estimation accuracy and type I and type II error rates. Results show that the Hunter and Schmidt approach yields the most accurate estimate for the moderating effect magnitude, and, therefore, it should be used for point estimation. Also, all three approaches yield similar overall type I and type II error rates for moderating effect tests. So there are no clear advantages of using one approach over the other. In this study an attempt is made to compare approaches given by Hedges and Olkin with approaches given by Hunter and Schmidt in terms effectiveness and accuracy in the field of creativity.

Sanchez –Meca and Marin –Martines (1998) compared Hedge & Olkin(1985) with Hunter &Schmidt (1990) as weighting by variance or by sample size. In their study authors explained that, one of the most widely used effect size indexes in meta-analysis is the standardized mean difference, $d$, defined as the difference between two group means(usually experimental versus control) divided by the within-group standard deviation. The use of this index is indicated especially when the studies to be integrated are experimental or quasi-experimental. Hedges
Olkin (1985) have shown that the best produce to average a set of independent \( ds \) is a weighted average, with the inverse variance of each \( d \) as the optimal weight factor.

However, the variance of \( d \) depends on the population standardized mean difference, \( \delta \), a parameter unknown in practice. Therefore, an estimate of the optimal weight is required. Hedges & Olkin proposed to substitute the sample standardized mean difference, \( d_i \), for the \( \delta \) parameter in each single study. On the other hand, Hunter & Schmidt (1990) proposed a simpler procedure consisting of weighting by sample size of each study. Hunter & Schmidt argued that their procedure is less biased than Hedges & Olkin (1985), even in the case of non-homogenous effect sizes. In Meca and Martines (1998), a Monte Carlo simulation was carried out to assess the bias and efficiency of two estimators for averaging independent \( ds(\text{effect sizes}) \) in conditions similar to those of real meta-analysis.

Hedges & Olkin and Hunter & Schmidt proposed two alternative estimators of optimal weights. In this article, the bias and relative efficiency of both estimators had assessed via Monte Carlo simulation. Conclusions showed that Hedges & Olkin’s estimator was more efficient, although more biased, than Hunter & Schmidt estimator.

Hall, S. & Brannick, M. T. (2002) compared two approaches of random-effect meta-analysis: the Hedges & Vevea approach and the Schmidt & Hunter (1990) approach. General goals of their study were to determine whether the choice of approaches mattered and, if so, whether one approach was preferable to the other.
In service of these goals, the authors conducted a Mont -Carlo study and reanalyzed four published meta-analysis of test validation results.

*Does the approach matter?*

The answer to this question appears to be “yes, it can”. The reanalysis of the published data showed that the choice of the Schmidt &Hunter and Vevea(1998) approach could influence the primary conclusion of the study, even though the credibility intervals generated by the two approaches overlapped considerably. The MonteCarlo study showed that although both \( p \) and \( \delta_p \) appeared larger for the H-V model, credibility intervals were actually conservative. Credibility intervals for the S-H model also found to perfectly matches population values. Finally H-V approach performed poorly when the data were attenuated making it a poor choice for situations in which correction for artifacts in desired.

*Which approach is preferable?*

On the basis of this study results showed for applications tat focus on the credibility intervals, such as validity generalization, approach of Schmidt-Hunter(1990) appeared preferable, because Hedges&Vevea (1998) approach is more likely to produce credibility intervals that falsely contain zero.

In case in which the author chooses to correct the data for artifacts H-V approach will underestimate parameters and should not be used so S-H approach is better because even corrected H-V approach is more likely to produce credibility intervals that falsely contain zero.
Another advantage of using S-H approach over the H-V approach is that the S-H approach provides estimates of \( p \) and \( \delta_p \) that are in \( r \) units not \( z \) units. The choice of \( r \) rather than \( z \) makes interpretation of estimates more straightforward. In most circumstances it appears that similar results can be obtained using either approach. Although, S-H estimates tends to be more accurate. In other words, choosing to correct for artifacts and the way in which artifacts are corrected, can be expected to have a much greater impact on the outcome than the meta-analysis approach chosen. Authors and journal editors should place more emphasis on explaining and justifying correction procedures rather than justifying one’s decision to use the S-H approach instead of the H-V approach.

Aguinis & Sturman and Pierce (2007). According to what authors have written in this paper, they were following three main goals in their research.

First, meta-analytic tests of hypotheses regarding moderating effects of categorical variables are increasingly pervasive in behavioral science. Also because meta-analytic tests for categorical moderators differ from those for continuous moderators there is a specific need to investigate the procedures most widely used in the organizational science.

Second, in contrast to primary level researchers, meta-analysts do not seem to have guidelines regarding which technique to use, and under which conditions, to test hypotheses regarding moderating effects of categorical variables.

Third, although recent Mont Carlo investigations have been published regarding the performance of meta-analysis techniques, these studies have not
been comprehensive in the inclusion of meta-analytic approaches. Consequently, the purpose of this study was to compare the Hedges & Olkin and Hunter & Schmidt, and a refinement of the Aguinis and Pierce approaches for estimating effects of categorical variables meta-analytically.

This study examined these techniques under a wide range of situations typically encountered by applied psychology and management researchers and assessed the technique’s accuracy in terms of Type I and Type II error rates and their ability to estimate the population effect size. Following these goals, authors have conducted a Monte Carlo simulation to compare the relative performance of three techniques. When meta-analytically testing for the presence of a categorical moderator variable.

To facilitate the comparison, they simulated situations involving a moderator variable with two levels only (gender). As authors explained, first, they specified the simulation parameters. Second, they generated the primary –level data to be subsequently combined into sets to simulate 656,100 separate meta-analyses. Third, each set of effect sizes (each meta-analysis) was analyzed using the three approaches. Fourth, they examined the accuracy of the point estimates (mean effect size estimate vs. Population effect size specified in the first step) and the accuracy of hypothesis tests regarding Type I and Type II error rates.

They concluded that meta-analysis is just another data analytic technique and is no substitute for good theory. Much like any other data analytic technique can lead to incorrect conclusion even when all procedures are correctly implemented from
a technical standpoint. Results show that, under some conditions, the probability of detecting the moderating effect of a categorial variable can be as low as .10 (Type II error rates as high as .90). Overall, following guidelines regarding conduct of meta-analysis was provided.

First: Results show that the Hunter & Schmidt approach yields the most accurate estimate for the moderating effect magnitude, and, therefore, it should be for point estimation.

Second: regarding homogeneity tests, the Hunter & Schmidt approach provide a slight advantage regarding Type I error rates, and Aguinis and Pierce approach provide as slight advantage regarding Type II error rates. Thus, the Hunter & Schmidt approach is best for situations when theory development is at the initial stages and there are no strong theory based hypotheses to be tested. (exploring or post hoc testing). Alternatively, the Aguinis and Pierce approach is best when theory development is at more advanced stages.

Third: all three approaches yields similar overall Type I and Type II error rates for moderating effect tests. So there are no clear advantages of using one approach over the other.

Fourth: the Hunter & Schmidt procedure is the least affected by increasing levels of range restriction and measurement error regarding homogeneity test, Type I error rates, and the Aguinins and Pierce homogeneity test Type II error rates are least affected by these research design conditions (in the case of measurement error this is particularly true for effect size around 0.2).
Thus, the choice of one approach over the other needs to consider the extent to which range restriction and measurement error are research design issues present in the meta-analytic database to be analyzed.

In closing, authors of this paper hope that their recommendations will help researchers choose meta-analytic techniques based on theory and research design considerations as opposed to habit, the availability of a user-friendly computer program, or usage trends in specific topic areas.

Hess and Olejnik, (2001) investigated the efficacy of Glass’s estimator of effect size for program impact analysis. They manipulated four data condition to study the sampling characteristics of the Glass index: (1) population separation (effect size), (2) variance pattern, (3) total sample size with equal and unequal $n$, and (4) distribution shape. They concluded that when population variances are equal, the Glass effect size is recommended regardless of the distribution shape and $n$ ratio provided sample sizes and population effect sizes are not jointly small. In other words, the Glass index performs well when the total sample size is sufficient given the size of the population effect size (using Cohen’s power table). Similarly when variances are heterogeneous, the Glass index can also be recommended regardless of the distribution shape and $n$ ratio provided the sample size is sufficient given the size of the population effect size. Finally for greater precision, the Glass index performs best under large sample sizes across all condition. They concluded that if an evaluator conducts a power analysis when planning an impact analysis, he or she will find the Glass effect size to be a good
measure of treatment impact. In other words, if an evaluator anticipates a particular effect size believed to be manifested by the treatment population and selects the proper sample size in order to maintain sufficient power for detecting the effect at a particular significance level, then the Glass index may provide an accurate measure of effect size.

A comparison of the effect size estimators in meta-analysis by Yildiz, N. et al. (2009). Effect size estimators by Glass, Hedges, the Maximum Likelihood, and Shrunken Estimators of effect size were employed in their study. The study showed that Glass estimator has the smallest heterogeneity variance in all the effect size estimators. According to them Glass estimator is used in many applications. Especially, with small sample size (at least 10 subjects per group) it has better empirical properties than the effect size estimators. However, the Hedges estimator maybe preferred over the other effect size estimators when sample size is small. They concluded that the differences among the effect size estimators are negligible.

2.8. Meta-Analysis on Creativity:

In the field of creativity two meta-analysis studies were reviewed. Although none of them was carried out in the setting of the present study.

Bertrand, (2005) carried out a meta-analytic review of research on creativity training programs that evaluate efficacy and the impact of specific program features on creative ability. The meta-analysis results showed that creative ability can be moderately enhanced by training as a factor (Mean ES = .67), but the real
strength of current programs appears to be the improvement of verbal rather than figural creative ability. Bertrand reported that the analysis and conclusions drawn from this research synthesis were severely restricted by the overall weak research designs by many primary creativity studies as well as the failure to collect and or report statistical data, and the failure to adequately describe treatment variables and procedures. If the study of creativity is to advance, primary researchers must be more diligent in designing studies and reporting statistics to allow for effective research synthesis. Scope (1998) showed that overall; instruction had a positive effect on the creativity of school-aged children.

Scope, (1998) to study a relationship between instructional programs and creativity conducted a meta-analysis of research on creativity with regard to effects of instructional variables. The results derived from the meta-analysis indicated that instruction can have a positive impact on creativity. However, some of the hypothesized link in the meta-analysis such as link between time spent on instruction, structuring, reviewing, questioning and responding, and creativity scores were not supported by the data. In this meta-analysis a modest positive correlation between independent practice and creativity was reported. This meta-analysis study suggested that apart from instructional variables, motivational, social, personal, and cognitive variables which may contribute to creativity must be considered. Recently researchers have emphasized the complexity of the interactions among variables that may give rise to creativity.