Chapter III

Method

3.1. Introduction

This chapter introduces the research procedures of (1) Operationalizing the concept of Meta-analysis, (2) Restatement of the problem, (3) Research design, (4) The database for selecting studies, (5) Sample, (6) Selection criterion, (7) The coding of the studies, (8) Co-coding and inter-coder reliability, and (9) The method of computing effect size.

3.1.1. Meta-analysis

Meta-analysis is a statistical analytical approach utilized for the purpose of synthesizing quantitative data to determine the degree of effect of the independent variable on the dependent variable. It uses statistical results from individual quantitative studies as the units of measurement in an overall study. These results are summarized by indices of effect size or standardized mean difference that may then be averaged to obtain an overall estimate of effect magnitude (Hedges & Becker, 1986).

3.1.2. Restatement of the problem

This study aimed to compare two meta-analytical approaches, namely, Glass and Hunter& Schmidt with single data set. This meta-analytical study synthesizes quantitatively effect sizes across the studies that had been conducted on creativity in the setting of study, Iran. It was assumed that the studies will reveal different
results regarding effects of instructional programs on creativity and its specific components. Due to this, these studies were subjected to analysis to find out how big effects actually are and more important which meta-analysis approach applied in this study can estimate effect sizes more accurate.

3.1.3. Research design

The study is mainly a descriptive analytical study wherein two approaches of meta-analysis proposed by Glass and Hunter-Schmidt were compared for their efficiency in yielding effect size. This study used a design introduced by Glass, McGaw and Smith (1981). This design is somewhat similar to Hunter – Schmidt approach (Hough, S. L. & Hall, Bruce W.1994). The design requires a reviewer to (a) use objective procedures for locating studies, (b) use quantitative techniques to describe study features and outcomes, and (c) use statistical method to summarize overall findings and explore relationships between study features and outcomes.

3.1.4. Variables of the study

Since a meta-analysis synthesizes the statistical results from many different studies, it is not necessary to define the variables in very specific terms. For this study, when comparing the two approaches, namely, Glass and Hunter-Schmidt the meta-analytical approach, was as the independent variable and effect size yielded by the two approaches was dependent variable. Nevertheless, when Comprehensive Meta-analysis Software was applied to find out hypothesized link between instructional programs and creativity in general, the instructional programs in school education, which could be either teaching methods or
educational programs in school-site level was as the independent variable and the dependent variable was creativity or achievement in creative performance resulting from instructional programs, as measured by Torrance Tests of Creative Thinking.

3.1.5. Database used for selecting data.

Considerable efforts were made to locate most comprehensive list of available experimental studies of creativity with respect to instructional program. Using the Universities Digital Libraries and Iranian Information and Documentation Center through the website search engine and entering appropriate key terms for each source, such as, “creativity” and “instructional program,” a comprehensive search for relevant and appropriate journal articles and dissertation were conducted. The Iranian Information and Documentation Center (Irandoc) was used as a main resource to reach the data included in this study because of following reasons:

i. It is the most useful database certified by research council of MSRT,

ii. It has accessed all dissertations, theses, articles and research projects produced by the universities in Iran.

iii. It has accessed high quality master degree dissertations, Ph.D. theses, journal articles, and research projects in terms of research methodology.

iv. Classified data in different subjects and categories.

In this study master degree dissertations and Ph.D. theses were used to increase the quality of study and reducing errors as much as possible. Therefore,
bachelor degree dissertations were not considered at the present study since they are not monitored by evaluation committee in the universities.

Moreover, it is supposed to include all the researches- which have been done regarding Instructional program and creativity –in a meta-analysis but to decrease some factors which may influence the results of the study like cultural factors, heterogeneity, and variation of studies, collecting data was limited to Iranian database. It is also noteworthy that not all research studies can be included in a collection for meta-analysis. Only well-designed studies which conform to the standards (e.g., Cook and Campbell, 1979) should be considered (Chong-ho Yu, 2010).

3.1.6. Inclusion of studies for meta-analysis

In the present study all the studies that were included into the database up to Dec. 2008 were considered. These studies included master degree dissertations, Ph.D. theses, journal articles, and research projects. The total number came to 204. These studies were subjected to the following criteria.

3.1.7. Criteria of Selection

One of the primary goals of the study was to utilize a large population of the studies; therefore a minimum of carefully selected limitation and delimitations was implemented. However, as it was critically important to this study to (a) evaluate the direct effect of instructional programs on creativity, and (b) maintain acceptable standards for meta-analysis research, only studies that met the criteria summarized below were included in the population base. First, the study must be
related to creativity and provided creativity measurement information. This study included school programs (e.g., Arts, music, etc.). Second, each study must be an experimental or quasi-experimental study that reported sufficient data (means, standard deviations, \(t\)-test, and \(F\) statistics) to allow the calculation of effect size. Third, the study was required to provide information about the research design (pre-post test, experimental and control group), subject’s information (e.g., sample size, age, grade, and etc.), description of the instructional program, and measurement tool used in the study. Applying these criteria, the number of studies that could be included for meta-analysis was reduced to 42.

### 3.1.8. Sufficiency of the number of studies required for meta-analysis

Meta-analytic procedures can be applied to as few as two studies but when there are very few studies the meta-analytic results will be relatively unstable. When there are very few studies available on a given research question it would be more economical of journal space and editors and reviewers time to incorporate the meta-analysis as an extension of the results section of the last in the series of a few studies. (Rosenthal & Dematteo, 2002). The number of studies in meta-analytic as Gobioff (2006) mentioned in his research is usually (\(K = 10, 20, 50\) and 100). Also Field (2004) demonstrated in his article, as a general rule, neither approach (Fixed effect model and Random effect model) is accurate when fewer than 20 studies are in the meta-analysis. Then he explain if the significance of estimates is important and there are 80 or more studies in the meta-analytic either approaches (Fixed effect model and Random effect model) will be fairly reliable.
In the present study each study included for meta-analysis had more than one hypothesis. Although there were 42 studies, the total number of hypothesis that could be derived was 142. These could be treated as individual study. Thus, for purpose of analysis in the present study, there were virtually 142 studies which could be reliable for the present study.

3.1.9. Coding of Studies

Each study was carefully read three times by researchers and then study variables were coded. The details and clarity with which these variables were reported varied dramatically among studies and any variables that were distinctly referenced were coded as existing within the study. If the existence of a variable was implied but not overtly stated it was coded as unknown. The coding scheme utilized was specifically developed for this study, and was informed by a review of the creativity literature, a review of coding schemes utilized by the previous creativity –based meta-analysis. General information about the studies included: (a) author; (b) date of publication; (c) subject’s demographic information (i.e., age, sex, grade, ses, and etc.); (d) sample size; (e) type of experimental design (e.g., pre post test, control group present or not); (f) published (journal articles) or unpublished (dissertation); (g) types of instructional program, e.g., problem solving teaching approach, any named instructional programs, other unnamed creativity instructional programs, school programs; (h) the psychological measurement tool used in the study which should be Torrance Tests of Creative
Thinking including validity and reliability; and (i) time period or length of instruction in minutes.

3.1.10. Co-Coding and Inter-Coder Reliability

From the pool of selected studies, 10 studies were randomly selected and independently coded by the primary investigator and a former Ph. D. student who graduated in the Department of Education & Psychology at Tehran University. A standardized coding form was created (Appendix II) that allowed the second coder to extract information regarding independent variables, i.e., subjects’ information including age, gender, grade, sample sizes (experimental and control group), types of instructional programs, time period or length of instruction (in minutes), and measurement tools including validity and reliability. The inter-coder agreement for each study reviewed exceeded the 85 percent criterion.

3.1.11. Computations and Analysis of Effect Sizes

A list of the individual studies was obtained and relevant statistical data were collected. These statistical data included means, standard deviations, t-test, and F statistics, the sample size of each study, and a reliability estimate from instrument used in the study to measure the dependent variable. To calibrate the studies’ results, or place them on a common scale, effect sizes were calculated. For this synthesis the investigator recorded 142 effect sizes from the 42 studies. The calculations of effect sizes involved determining the experimental group mean minus the control mean divided by the control standard deviation in Glass approach of meta-analysis or pooled standard deviation in Hunter & Schmidt
approach of meta-analysis. Formulae for calculating effect size were separately listed below for Glass and Hunter-Schmidt approaches.

**Figure (4). Formulae for Calculation Effect Size by the Approaches**

<table>
<thead>
<tr>
<th>Glass Approach</th>
<th>Hunter-Schmidt Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect size $\hat{\theta} = \frac{M_1 - M_2}{S_c}$</td>
<td>Effect size $\hat{\theta} = \frac{M_1 - M_2}{S_{\text{pooled}}}$</td>
</tr>
<tr>
<td>Where:</td>
<td>Where:</td>
</tr>
<tr>
<td>M1 = mean of the treatment group;</td>
<td>M2 = mean of the control group;</td>
</tr>
<tr>
<td>M1 = mean of the treatment group;</td>
<td>$S_{\text{pooled}} = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{(n_1 - 1) + (n_2 - 1)}$</td>
</tr>
<tr>
<td>$S_{\text{pooled}}$ = standard deviation of the control group;</td>
<td>Where:</td>
</tr>
<tr>
<td></td>
<td>$n_1$ = size of treatment group;</td>
</tr>
<tr>
<td></td>
<td>$n_2$ = size of control group;</td>
</tr>
<tr>
<td></td>
<td>$s_1$ = standard deviation of treatment group;</td>
</tr>
<tr>
<td></td>
<td>$s_2$ = standard deviation of control group.</td>
</tr>
</tbody>
</table>

### 3.1.12. Measurement error correction

Hunter and Schmidt (1990) recommend averaging the available reliability coefficients for each data set and adjusting (correction for measurement error) the overall effect size by the mean reliability coefficient. Thus, the mean reliability coefficient was used to adjust the overall mean effect size in this study. If the reliability coefficient was not published in the study, reported coefficients in similar studies with same population was used. Reliability coefficients were available for 85% of the studies. The following formula (Figure 5) was available for the calculation of this adjustment or correction for measurement error.
3.1.13. Cohen’s Classification for Effect Sizes Computed Using the Two Approaches

A number of ways of interpreting effect size have been proposed. This study utilized methods advised by Cohen (1988). Cohen (1988) developed a system that effect sizes can be classified as “small= .2”, “medium= .5”, and “large= .8”.

Using Cohen’s power table (given in Appendix I) the effect sizes computed by Glass approach and Hunter-Schmidt approach (both before and after correction for measurement error) were classified as ‘large’, ‘medium’, and ‘small’. Percentages of effect sizes classified under each group were compared descriptively for the two approaches before any correction for measurement error. Similar computations were made for percentages of effect sizes in each category before & after correction for measurement error in Hunter-Schmidt approach.

3.1.14. Calculation of mean effect sizes for 42 studies

Effect size can be considered a standardized estimate of where the experimental group stands in comparison with the control group distribution. In
the case of studies examined for this study, positive effect size indicated the treatment groups being under special instructional programs received higher scores than the control group. In each study, all of the subscales’ effect sizes were assessed (e.g., fluency, flexibility, originality, and elaboration in Torrance Thinking Creativity Test). Then, all of the subscales’ effect sizes were averaged into one single effect size index to present the effect of the study. If the study had more than one treatment group, then each treatment group would be calculated separately, and the study would have more than one effect size index to present each treatment’s effect. Thus, there were 42 mean effect sizes of the experimental studies that could be calculated based on data given in the 42 studies. These were separately available for creativity in general and each of the sub components of creativity. Such a set of 42 mean effect sizes were computed for those derived from Glass approach before measurement error. Likewise, mean effect sizes were computed for those derived from Hunter-Schmidt approach both before and after correction for measurement error.

3.1.15. Calculation of overall mean effect sizes in both approaches

For meta-analysis purpose and comparing the two approaches, overall mean effect sizes had to be computed. For mean effect sizes computed using Glass approach overall mean effect sizes were calculated for creativity in general and each sub components of creativity. These computations were limited to mean effect sizes derived from Glass approach before correction for any measurement
error. Obviously, there was no correction made for measurement error in this approach.

Similarly, for mean effect sizes using Hunter-Schmidt approach, overall mean effect sizes were calculated for creativity in general and each subcomponents of creativity. All these computations were made for mean effect sizes derived from Hunter-Schmidt approach before as well as after correction for measurement error.

3.1.16. Comparison of overall mean effect sizes computed by the two approaches

The researcher undertook the comparison to determine (a) if the use of the pooled within – group standard deviation makes any differences between the Glass and Hunter – Schmidt overall mean effect sizes in creativity and its specific components before correction for measurement error. If there were significant differences between the approaches, then it can be said that probably the pooled within – group standard deviation cause the difference, (b) otherwise any differences between the overall mean effect sizes derived from the both approaches can be attributed to correction for measurement error which has to be computed in Hunter- Schmidt approach. Paired-sample t- Test was used to compare the overall mean effect sizes of the Glass and Hunter-Schmidt. Paired sample t-Test is a statistical technique that is used to compare two population means in the case of two samples that are matched or correlated. The t-Test for correlated samples uses a different formula to take into account the correlation which exists between the two groups. Any positive correlation between the pairs
reduces the standard error of difference between the groups. The higher correlation, the greater reduction in the standard error of difference. This makes the Correlated Sample t-Test a more powerful test than the Independent Samples t-Test (Yount, 2006).

3.1.17. Tools

Analysis of the data was done via (a) hand calculations, (b) utilization of the Statistical Package for the Social Sciences (SPSS version 16), (c) Comprehensive Meta-Analysis Software version 2, and (d) MS Excel Effect Size Computation Program (Lipsey & Wilson, 2001).