

CHAPTER III

METHODOLOGY AND PROCEDURES

In the first part of this chapter, an overview of the meta-analysis methodology is provided. General procedures inherent in meta-analysis are described. The other sections describe the use of these procedures in the current study with emphasis on the selection of studies based on established criteria. Procedures for calculating effect sizes are also presented.

In this study, the overall effectiveness of CAI in technical education and training was determined by using the meta-analysis methodology to quantitatively synthesize results from the various studies that had been conducted in this area. The relation between CAI effectiveness and various features of the individual studies was also determined.

The following research questions related to higher order learning in technical education and training guided this meta-analysis.

1. What is the overall effectiveness of CAI on achievement as compared to traditional instruction?
2. What are the study features and their corresponding categories on which differences in CAI effectiveness can be investigated?
3. What differences in CAI effectiveness exist between categories in each of the identified study features?

Overview of Meta-analysis

The term “meta-analysis” was originally coined by Gene V. Glass in 1976 (Glass, 1976). In relating it to existing analyses, Glass classified research analysis into primary

analysis (primary studies), secondary analysis (secondary studies), and meta-analysis.

Primary analysis is the original analysis of raw data. Secondary analysis uses alternative analytical techniques to analyze the same data to answer the same research questions, or uses the same techniques to answer different questions from the same data. An example of primary analysis is “Pygmalion in the classroom” by Rosenthal and Jacobson while “Pygmalion revisited” by Elashoff and Snow is the corresponding secondary analysis. Glass’s reference to meta-analysis as the “analysis of analyses” (Glass, 1976, p.3) has been described by Kulik and Kulik (1988, p.1) as “the most succinct definition that has yet been proposed for this methodology.” The Kuliks further highlighted several of Glass’s characterizations of meta-analysis, as follows:

1. A meta-analysis encompasses results of studies that are already conducted.

Glass did not use the term to refer to the analysis of a planned series of investigations.

2. A meta-analysis is applied to summary statistics, not raw data. These include means, standard deviations, and results from statistical tests.

3. A meta-analysis focuses on size of treatment effects, not just statistical significance.

4. A meta-analysis focuses on study features and outcomes. The purpose of a meta-analysis is not simply to summarize a whole body of literature with a single effect size or overall significance level. It also tries to determine how study features influence effect sizes.

The Kuliks noted that some meta-analysts do not conform to the above characterizations. As an example, they cited Rosenthal (1984) as defining meta-analysis to mean the use of statistical techniques to either combine or compare either effect size

measures or probability levels from at least two studies. The Kuliks also observed that Glass's characterizations of meta-analysis seems most consistent with common usage.

Several other forms of the meta-analysis definition are enlightening. Cooper and Hedges (1994) defined it as the statistical analysis of a collection of analysis results from individual studies for integrating the findings. In comparing meta-analysis to the narrative literature review, Whitley (1997) explained that the former is a quantitative synthesis of a set of studies that integrates the results of their statistical analyses, while the latter uses qualitative techniques to integrate a body of research. One important advantage of meta-analysis over the narrative review is the explicit information provided on the decision processes used by the reviewer (Mullen & Rosenthal, 1985, cited in Hays et al., 1992).

Research Design

This study used a design introduced by Glass, McGaw and Smith (1981). The approach 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 methods to summarize overall findings and explore relationships between study features and outcomes. Other researchers who have used this approach include Cohen and Dacanay (1992), Fletcher-Flinn and Gravatt (1995) and Liao and Bright (1991).

The effectiveness of CAI in technical education and training was determined by the overall effect of the treatment. Meta-analytic procedures were applied to calculate the size of this effect. As the meta-analysis progressed, and results unfolded, studies were grouped according to their common study features, and the various categories within the features were identified. The relation of these features to effect sizes was investigated using one-way ANOVA.

Since a meta-analysis synthesizes the statistical results from many different studies, it is not necessary to define the independent variable in very specific terms. For this meta-analysis, it was sufficient to state the independent variable as the teaching method in technical education and training, which could be either computer-assisted instruction or traditional instruction. For the same reason, it was adequate to begin the meta-analysis by defining the dependent variable as student learning or achievement in learning resulting from technical education and training, as measured by some test or tests.

Study Selection

Several criteria were used to ensure that selected studies were within the focus of the meta-analysis. Criteria were also used to include only those studies from which calculations for a meta-analysis could be performed. Ensuring the methodological soundness of selected studies was also a purpose of these criteria. The criteria were:

1. The study must involve higher order learning in the field of technical education and training, which include those conducted in the civilian and military sectors. For the purpose of selecting studies that involve higher order learning in this meta-analysis, the definition of higher order learning in the definitions section in Chapter 1 was used.

2. The study in its full text form must be accessible.

3. A comparison must be made in the study between a group of students that received computer assisted instruction and another group that was taught in the traditional manner. Student learning in both groups was measured in some form.

4. Summary statistics must be reported in the study such that calculation of effect sizes can be completed. Examples include several combinations of means, standard

deviations, t-statistic, F-statistic, group sizes in samples, and ANOVA or ANCOVA summary tables.

5. Studies must be free from serious design flaws. If subjects were not randomly assigned into control and experimental groups, appropriate measures such as pretest and posttest, and statistical analysis must be used to account for preexisting differences between the groups. Any instrumentation used must be reported as having validity and reliability. Instructional content must be the same for both groups. Performance of both groups must be measured by the same test or tests. The previously mentioned criteria were applied to potential studies using the following procedure.

Identifying Studies

Two approaches were taken to meet the first criterion. First, computer searches were conducted on relevant databases: (a) National Technical Information Service (NTIS), (b) Defense Technical Information Center (DTIC), (c) Educational Resources Information Center (ERIC), and (d) Dissertation Abstracts International (DAI). The computer searches were performed using several combinations of the keywords “computer,” “effectiveness,” “technical,” “vocational,” “troubleshooting,” “training,” and “group”. The word “group” was especially important to differentiate studies that use control groups from those that do not.

Next, a Published Search was purchased from NTIS that contained the most current data available on the topic of CAI. However, the scope of the bibliography was too broad for this particular meta-analysis. Although it contains 250 citations, only 8 pertinent study abstracts were found that were not already located through the other computer searches.

Study abstracts obtained were read to eliminate studies that were not concerned with instruction in technical education and training. Among the excluded studies were those concerned with elementary education, and those not involving higher order learning.

Accessing Studies

Based on information from study abstracts, the studies were selected for further screening of their full text form. These studies were gathered from their sources through the university library and interlibrary loan services. It should be noted that a number of additional studies were identified from bibliographies of studies obtained. While retrieving studies from the civilian sources was completed without major problems, the same cannot be said of military studies. A person or a university library must be registered with the Defense Technical Information Center (DTIC) to acquire these studies. To be registered with DTIC, the applicant must be a federal government contractor. After two months were spent dealing with this “red tape,” the full texts of these military studies were finally acquired from the National Technical Information Service in microfiche format. The detailed screening resulted in 70 studies being gathered for further examination.

Examining Study Data

The 70 studies were then examined to insure sufficient data were available for the calculation of effect sizes. Studies published or conducted during the last 10 years were examined first i.e. those conducted between 1987 and 1997. Of the 70 studies, 38 were conducted during this time interval. It was determined that if further screening reduced the number of studies to fewer than 15, the interval would be expanded by another five

years. Limiting the age of studies to 10 years was also intended to ensure the inclusion of CAI studies that used relatively newer computer technology, while keeping the number of studies sufficiently large for variations in study features to be compared. This strategy has been used by Fletcher-Flinn and Gravatt (1995), whose meta-analysis was limited to studies that were published from 1987 to 1992.

The data examination process resulted in five studies being classified as deficient in the data area. One of the five studies was found to contain unsuitable data. Three of the five studies were conducted by the same organization. Achievement scores in these three studies were presented on histograms showing percentages of students receiving grades A, B, C, and D. Since no individual student score was provided, and no means and standard deviations were reported, the histograms were not useful for this meta-analysis. The contact person for this organization was contacted through electronic mail to request the missing information. A reply was received with the electronic mail address for another person from whom the data might be available. The other person was also contacted but no reply was received. The author of another of the five studies replied that she could not oblige because her study was conducted almost 10 years ago that it will be difficult for her to trace the data. Since efforts at retrieving the missing data from the five studies became futile, they were not included in subsequent analysis.

While examining the 38 studies conducted or published between 1987 and 1997, some were found unsuitable for reasons other than missing data. Three studies did not have control groups. Five studies were not within the focus of this meta-analysis. One study turned out to be a review of past evaluations rather than a primary study. Another study investigated the effects of age and GPA on learning via CAI but did not compare

performance between the CAI and control groups. Data from another dissertation was analyzed in an unusual manner that it could not be comprehended. Two dissertations initially thought available through interlibrary loan were unable to be borrowed.

Data examination described in the preceding two paragraphs resulted in 18 studies being excluded from the list of 38 studies, leaving 20. However, the study (Shute, 1991) that reviewed past evaluations provided a link to another study (Gott, 1989) that was eventually found suitable for inclusion. Three other studies (Buergermeister, 1989; Hwang, 1989; Kennison, 1990) that were unintentionally omitted from the original list of 70 studies were also found suitable. Hence, 24 studies were selected for further examination. This size was considered acceptable since (a) there is no specific rule for minimum number of studies, and (b) prominent researchers in the meta-analysis area have used between 20 to 30 studies before (Hays et al., 1992; Kulik et al., 1986).

Examining Study Design

Studies were examined for their use of sound experimental design. Factors examined included (a) assignment of subjects into groups, (b) pretesting and posttesting, (c) instrumentation, and (d) nature of treatment. For example, in the Faryniarz and Lockwood (1992) study, a quasi-experimental design with pretest and posttest on two intact groups were used. Statistical tests revealed no significant difference in initial problem solving ability between both groups. The instrument used to measure achievement was the Test of Integrated Process Skills (TIPS I & II). This instrument was shown free from reading-level bias by the good correlation of its scores with the Nelson-Denny reading test. Additionally, the pretest and posttest were reported to have construct

and content validity. Test reliability for the experimental and control groups was reported as high.

As another example, in the study conducted by Ogle et al. (1989), it was reported that content validity was obtained for the lesson plans and achievement test instruments. The test instruments used were also reported as having good K-R 20 reliability coefficients. The study conducted by Shute and Glaser (1990) was also demonstrated to have sound research design. The pretest and posttest were pilot-tested for clarity, timing, and difficulty level and were reviewed by an independent instructor for content validity in terms of completeness and accuracy. Thus, all the 24 studies were examined in the preceding manner. None of the 24 studies was considered as having serious flaws to warrant their exclusion. However, two studies did not focus on higher order learning and were excluded. These two studies focused on basic skills in reading and mathematics. Therefore, 22 studies were included in the statistical analysis. In the subsequent analysis, one study was found to have an effect size of 2.92 while the other 21 effect sizes ranged from -0.09 to 1.13. The effect size of 2.92 was considered an outlier and thus eliminated from the analysis. Therefore, 21 studies were eventually used to answer the research questions.

Statistical Analysis

In order for the studies to be compatibly synthesized, a common metric of comparison was used. This metric was termed “effect size.” Kulik and Kulik (1989, p.263) described effect size as a “general measure of the magnitude of a treatment effect on a dependent variable, expressed in such a way that the treatments in many different studies can be directly compared.”

Choice of Effect Size Indicator

Fern and Monroe (1996) classified effect size according to, among others, type, indicator, and recommended uses. The four types of effect size they listed were standardized mean difference, correlation, explained variance, and intraclass correlation coefficient. For between-group differences in experimental research, they recommended the use of standardized mean difference. Since in this meta-analysis, CAI studies were selected that investigated the difference between treatment and control groups, the most suitable effect size indicator was the standardized mean difference. Several researchers have successfully used this effect size in meta-analyses that investigated CAI effectiveness in other contexts (e.g., Kulik et al., 1986; Fletcher-Flinn & Gravatt, 1995). Thus for purposes of this study, effect size is taken to mean the standardized mean difference, and vice versa.

Standardized Mean Difference

Glass, McGaw, and Smith (1981) stated that the standardized mean difference is calculated by dividing the difference between treatment and control group means by the standard deviation of the control group. Suppose a group which received computer-based coaching on the Scholastic Aptitude Test (SAT) obtained an average score of 550 on the test, and a group that received conventional teaching averaged 500. The effect size for the computer-based treatment is 0.5, since the standard deviation on the SAT is 100 (Kulik, 1994). In other words, the computer-coached group outperformed the conventional teaching group by 0.5 standard deviation. Sixty-nine percent of the area under the standard normal curve falls below a z-score of 0.5. It can then be said that the typical student in an average computer-coached SAT class would perform at the 69th percentile

on the SAT examination, while the typical student from the conventional class would perform at the 50th percentile. Similarly in this study, the magnitude of the effect size will be an indication of CAI effectiveness compared to traditional instruction.

Overall Effect Size

In this meta-analysis, effect sizes of all selected primary studies were averaged to produce an overall effect size. Several previous results from related meta-analyses are noteworthy. In their meta-analysis of 37 studies of computer-based instruction in the health profession, Cohen and Dacanay (1992) found an overall achievement effect size of 0.41. Fletcher-Flinn and Gravatt (1995) calculated an overall effect size of 0.24 from their meta-analysis of 120 studies of computer-assisted instruction that were conducted across various educational levels ranging from elementary school to adult training.

Effect Size and Study Features

Another practice in meta-analysis is to investigate the relationship between effect size and various study features. For this meta-analysis, the selection of these study features was initially guided by previous CAI meta-analyses. Some common study features include (a) treatment duration (e.g., Kulik et al., 1986; Liao & Bright, 1991), (b) type of computer interaction -- mainframe or microcomputer (e.g., Kulik et al., 1986; Cohen & Dacanay, 1992), and (c) nature of instruction -- supplement or replacement (e.g., Cohen & Dacanay, 1992; Ryan, 1990). Study features that have been shown to be significantly related to effect sizes are treatment duration (Kulik & Kulik, 1991; Liao & Bright, 1991), and control for instructor effects (Kulik & Kulik, 1991). Based on the findings, these study features were among those investigated in terms of their relationship with effect size. Three studies were given to a second person to determine the interrater

agreement in placing study effect sizes into the various categories. Agreement was high on all study features except instructor assignment and study duration. Instructor assignment was not adequately reported in the studies. Therefore, it was not included as a feature in this meta-analysis. It was also discovered that study duration could not be used as a study feature because of the various frequencies of treatment applied. The various time arrangements reported in the studies made it difficult to group studies according to categories of study duration. To avoid subjectivity, this study feature was not examined in this meta-analysis. The choice of study features and categories within the study features is discussed further in Chapter IV. One-way analyses of variance and t-tests were performed to study the difference in effect sizes between groups of studies in the various study features.

The categories within study features were determined by the frequency of their occurrences in the various studies. The categories within a particular study feature were made mutually exclusive to avoid ambiguity. Therefore, there was no need to establish an interrater reliability coefficient.

Effect Size Calculation

As mentioned previously, the indicator chosen for effect size in this study was the standardized mean difference. The following formulae were available for the calculation of this standardized score, i.e., Cohen's d, Glass's Δ , and Hedges' g (Fern & Monroe, 1996):

$$\text{Cohen's } d = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\text{pooled}}}; \quad \text{Glass's } \Delta = \frac{\bar{X}_1 - \bar{X}_2}{S_c}; \quad \text{Hedges' } g = \frac{\bar{X}_1 - \bar{X}_2}{S_{\text{pooled}}}$$

where \bar{X}_1 = mean of the treatment group;

\bar{X}_2 = mean of the control group;

σ_{pooled} = pooled standard deviation of the treatment and control group population;

s_c = standard deviation of the control group;

s_{pooled} = pooled standard deviation of the treatment and control groups.

The use of Cohen's d is seldom possible since σ_{pooled} is rarely reported in primary studies. The choice was then between the Glass and Hedges formulae. It can be seen that the only difference between these two lies in the denominator. In order to choose a suitable formula for the calculation of effect size in this meta-analysis, suggestions from previous meta-analyses were considered. Glass, McGaw and Smith (1981) argued that experimental treatments may affect variation on the dependent variable as well on its mean value, therefore recommending the use of the standard deviation of the control group. This view was supported by Kulik and Kulik (1989). However, the Kuliks also observed that many meta-analysts had to use the pooled value because separate standard deviations for the treatment and control groups were not reported in many of the primary studies. It is interesting to note that a seemingly conciliatory finding was observed by Ryan (1990) in the meta-analysis of CAI effectiveness in elementary education. Ryan discovered that effect sizes based on Glass and Hedges formulas were exactly the same in 17% of her primary studies, and very similar on most of the rest. This result is surprising considering "the choice of the standard deviation with which to scale the differences between group means to determine effect size is crucial" (Glass et al., 1981, p.106). For studies that did not report means and standard deviations but reported summary statistics such as t and F values, Glass et al. (1981) provided procedures for calculating effect sizes. Since 40% of Ryan's primary studies fell into this category, she decided to use the

Glass formula. Other researchers who have used standard deviation of the control group in their meta-analyses included Fletcher-Flinn and Gravatt (1995), and Kulik et al. (1986), and Liao and Bright (1991).

It was anticipated that some primary studies for this meta-analysis might not report means and standard deviations. Since the Glass procedures for calculating effect size for these studies uses Glass's formula, its use became a logical choice. However, as suggested by Kulik and Kulik (1989), pooled standard deviations were used when statistical test results are not provided in the primary studies, and separate standard deviations are not reported.

Procedures for Effect Size Calculation

Fletcher-Flinn and Gravatt (1995) listed the following procedures for choosing the equation to be used for the calculation of effect size. In this meta-analysis, the choice of these equations depended on the information that was available in the primary studies.

1. Effect Size (Δ) is equal to the difference between the mean scores of the experimental group (\bar{X}_e) and the control group (\bar{X}_c) divided by the standard deviation of the control group (s_c).

$$\Delta = (\bar{X}_e - \bar{X}_c) / s_c$$

2. Effect size is equal to the difference in effect size of posttest and effect size of pretest.

$$\Delta = \Delta (\text{posttest}) - \Delta (\text{pretest})$$

3. Effect size is equal to the product of t-statistic and the square root of the sum of one divided by the sample size (n) of the experiment and control groups.

$$\Delta = t\sqrt{[(1/n_e) + (1/n_c)]}$$

4. Effect size is equal to the difference in the means of experimental and control groups divided by the square root of the mean square variance between groups (MS_b) divided by F-statistic.

$$\Delta = (\bar{X}_e - \bar{X}_c) / \sqrt{MS_w}$$

where $MS_w = MS_b / F$

As anticipated, some studies provided enough information for the application of formulae (1), (2), or (3). As recommended by Fletcher-Flinn and Gravatt, formula (2) was used instead of formula (1) or (3) in these cases, as this would give the most accurate estimate of the true treatment effect.

Calculation of Pooled Standard Deviations

The following equation was used to calculate the pooled standard deviation (Fern & Monroe, 1996; Ryan, 1990):

$$s_{\text{pooled}} = [(n_e - 1)(s_e)^2 + (n_c - 1)(s_c)^2 / (n_e + n_c - 2)]^{1/2}$$

where n_e = size of treatment group;

n_c = size of control group;

s_e = standard deviation of treatment group;

s_c = standard deviation of control group.

Sample Size Weighting

When the range in sample sizes of primary studies is too large, a bias in the overall effect size due to the influence of effect sizes from small samples is incurred. To correct for this bias, Hedges and Olkin (1985) introduced a weighting factor in the following formula:

$$d = [1 - 3 / (4n - 9)]g$$

where g is the uncorrected effect size;

n is the sample size;

d is the effect size corrected for sample bias;

$[1 - 3 / (4n - 9)]$ is the weighting factor.

Ryan (1990, p.44) compared uncorrected (g) with corrected (d) effect sizes and decided that “Since the resultant weighted effect sizes differed by 2% or less from the unweighted effect sizes, and a majority differed by less than 1%, only the unweighted effect sizes were used in this analysis.” Furthermore, Kulik and Kulik (1989) observed that many meta-analysts reported this correction had at most a trivial effect on their results. They cited their own meta-analysis (Bangert-Drowns, Kulik & Kulik, 1983) and discovered a 0.999 correlation between 27 pairs of corrected and uncorrected effect sizes, with most agreeing to 2 decimal places. They also noted that due to the small difference on bias-corrected effect sizes, many meta-analysts do not bother to use the weighting factor. Based on this evidence, this meta-analysis did not use the weighting factor to correct for sample size bias.

Effect Sizes from ANCOVA Tables

Whenever adjusted means are available, these should be used for calculating effect sizes because they provide the true effects of the treatment after partialling out the effects of the covariate(s). For this purpose, Hedges, Shymansky, and Woodworth (1989, p. 30) supplied the following equation.

$$\text{Effect Size} = (\text{adjusted treatment group mean} - \text{adjusted control group mean}) / S_{\text{pooled}}$$

where s_{pooled} is the pooled standard deviation.

Whenever ANCOVA summary tables are provided but adjusted means are not reported, Hedges et al. (1989, p. 32) provided the following formula for calculating effect sizes:

$$\text{Effect Size} = \pm [(SS_{\text{treatment}} / SS_{\text{balance}}) \cdot m \cdot (1/n_c + 1/n_t)]^{1/2} .$$

The sign of the effect size is positive if the treatment mean is greater than the control mean and negative if the treatment mean is smaller than the control mean. Hedges et al. discovered that effect sizes calculated from the above equation are biased values. They provided a correction factor of $(1 - R^2)$ to correct for this bias, where R is the multiple correlation between the criterion measure and the covariates. They further noted that since this correlation is rarely reported, it would be conservative to assume a value of .5 for R^2 , resulting in a value of .7 for $(1 - R^2)$. Their suggestion was followed in this meta-analysis. Therefore, effect sizes derived from ANCOVA tables were multiplied by .7 to obtain the corrected effect sizes.

For analysis of covariance that used one covariate, the F-statistic can be transformed into a t-statistic by the relation $t = \sqrt{F}$. Given the control and experimental group sizes, effect size can then be calculated from the corresponding formula given earlier. The correction factor of .7 must still be applied to the effect size to obtain the corrected effect size. For ANCOVA with a single covariate, this method is simpler than the one explained in the preceding paragraph.

Summary

An explanation of the meta-analysis methodology was presented as the basis for the research design in this study. Criteria and procedures for selecting, identifying, accessing, and examining studies were described including the problems encountered and how these problems were solved. Statistical analyses were explained with emphasis placed on factors to be considered when calculating effect sizes.