The Pennsylvania State University

The Graduate School

Learning and Performance Systems

A META-ANALYTIC ASSESSMENT OF THE USE OF REHEARSAL STRATEGIES IN THE PROGRAM OF SYSTEMATIC EVALUATION

A Thesis in

Instructional Systems

by

Rose Marie Baker

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2006

The thesis of Rose Marie Baker was reviewed and approved* by the following:

Francis M. Dwyer, Jr
Professor of Instructional Systems
Thesis Advisor
Chair of Committee

Roy B. Clariana
Associate Professor of Instructional Systems

Wesley E. Donahue
Associate Professor of Management Development

Hoi K. Suen
Professor of Educational Psychology

Susan M. Land
Associate Professor of Education
Professor-in-Charge Instructional Systems Program

*Signatures are on file in the Graduate School

ABSTRACT

The goal of this research study was to synthesize the use of knowledge generation and discrimination tasks as methods of overt and covert rehearsal as used within studies in the Program of Systematic Evaluation and determine if there were interactions of these rehearsal strategies with the employment of visuals, the medium chosen for presentation, and the type of test used to assess achievement of the specific educational objectives. More than 350 published articles, conference proceedings, dissertations, and masters papers were reviewed from the Program of Systematic Evaluation; of which, 50 studies (9,620 participants) were analyzed using meta-analytic techniques and hierarchical linear modeling.

For properly positioned rehearsal strategies, the unconditional level-1 model resulted in a statistically significant estimated mean population effect size of 0.21 for paper-based and screen presentation instructional environments and the conditional level-2 model resulted in a statistically significant estimated mean population effect size of 0.34 for computer-based instructional environments. The other researched conditions were not statistically significant predictors of the estimated mean population effect size.

The employment of item analysis to determine the position for the rehearsal strategies results in savings for time and money and increased learner achievement. For a learner that would score 50% on any of the criterion measures for the specific educational objectives without interaction with any rehearsal strategy, directing the learner's attention using properly positioned knowledge generation or discrimination tasks and overt or covert rehearsal strategies, with or without visuals that are black and white or colored, results in scores between 56.0% and 60.6% for paper-based materials or screen presentations and between 52.6% and 73.1% for computer-based instructional environments.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

My thanks and appreciation are given to my committee, Dr. Francis Dwyer, Dr. Roy Clariana, Dr. Hoi Suen, and Dr. Wesley Donahue. Each has contributed to this rewarding educational experience using his strengths to guide me on my path of performing meta-analysis with hierarchical linear modeling. I have learned much about mentoring, statistical analysis, and setting priorities.

My second rater has spent hours with me in the library and has become an expert in searching for journals and finding the details of a research study. I learned from you a special bond and reliance.

A special thank you to all of you that have finished your thesis or masters paper or presented your research at a conference; your research within the Program of Systematic Evaluation has been invaluable to this research project.

My greatest gratitude is to my family and friends who spent many days being patient waiting for me to be there when they needed me most.

Chapter 1

Introduction

Rehearsal of information is an important process in human learning. Rehearsal strategies are employed for basic learning tasks and for complex learning tasks requiring the learner to differentiate between the essential information that is to be remembered and supplemental information that is not fundamental yet is advantageous to know. If the rehearsal strategies are properly designed and positioned within the instructional environment to cue rehearsal of information determined to be most difficult for the learner to acquire, they enhance the acquisition and retention of the information (Dwyer & Dwyer, 1993). The mental processing, as a result of rehearsal, facilitates the learner's interaction with the information in working memory, increasing the opportunity of activating prior knowledge stores to promote encoding of the information and transfer to long-term memory for storage.

Rehearsal strategies are employed within instruction to provide support for the individual differences of the learners (Jonassen & Grabowski, 1993). Prior knowledge levels, reading comprehension ability, and intelligence are features of individual differences that influence how the learner perceives, interacts, and profits from an instructional environment. For all types of learners, the properly designed and positioned rehearsal strategy identifies the important information that is necessary to be attended to in order to achieve the learning objectives of the instructional environment.

The Program of Systematic Evaluation is a research program dedicated to the study of effective and efficient instructional design strategy research (Dwyer, 2004). The Program of Systematic Evaluation began in 1965 with research of the effects of varied levels of detail in visuals (Dwyer, 1965). The research program has grown to four phases. The first phase was dedicated to addressing problems associated with research design of visualization research (Dwyer, 1987, p. 6). The second phase provided an integrated and comprehensive treatment of visualized instruction research (Dwyer, 1987, p. 7). In phase three, the scope of the program expanded to the strategy and tactical levels of instructional design by emphasizing independent variables and their subsequent effects on learner acquisition of information (Dwyer, 1987, p. 8). This information is critical to instructional design decisions and provides a high level of predictability of the level of achievement of specified audiences (Dwyer, 1987, p. 8). The fourth phase takes advantage of the advances in technology while continuing to explore the effects of visualization within instructional environments (Dwyer, 1987, p. 8).

The Program of Systematic Evaluation implemented rehearsal strategies as part of the research program within phase three of the program of research (Dwyer, 2004). These research studies within the Program of Systematic Evaluation employed instructional tactics and strategies as independent variables within the instructional environments to be instructionally efficient and effective and designed to facilitate learner achievement of the specific educational objectives. In the instructional environments of the Program of Systematic Evaluation, rehearsal strategies have taken many forms. A few examples are embedded questions, advance organizers, summary statements and questions, concept maps, outlines, study guides, and graphical displays

(Hagenberger, 2003; Irizarry, 1990; Lin, Kidwai, Munyofu, Swain, Ausman, & Dwyer, 2005; Wood & Dwyer, 1989; Yamashiro, 2001). The criterion measures within the Program of Systematic Evalution measure facts, concepts, rules, and principles in a hierarchy of learning to provide insight during research to determine if the learners are having difficulty with the foundational facts and concepts and/or the comprehension level of knowledge acquisition (Dwyer, 1978, p. 45). If the learner performs well on the comprehension test, a test that measures the rules and principles, then the learner should do similarly well on the tests measuring the facts, concepts, and principles (Szabo, Dwyer, & De Melo, 1981).

The design and development of instruction is expensive. The actual expense is related to the organization, learning tasks, content, rehearsal strategies used, and the media of the instructional environment. Adding rehearsal strategies to the instructional environment increases the development expense. The cost of developing rehearsal strategies ranges from the inexpensive attention gaining devices to the very expensive adaptive question, answer evaluation, and feedback systems.

Placement and choice of rehearsal strategies in an instructional environment is directed by the learning objectives and learning outcomes. If the learner must generate the information after the instructional session, the amount of rehearsal will be greater, thus the instructional environment is more expensive to develop than if the learner was required to recognize the correct response or action. In addition to the learning objectives and outcomes, the skills and abilities of the learners to interact with the instructional environment influence the placement and choice of rehearsal strategies. Provision of additional resources for those that need foundational information and supplemental resources adds to the development expense of the instructional environment.

Item analysis of the performance on the criterion measures of a cohort from the target audience for the instructional materials is completed to examine item difficulty. The items in the criterion measures are mapped to positions within the instructional materials. The locations identified as the most difficult, the items with the highest item difficulty in the item analysis, were the legitimate positions within the instruction that require supplemental resources and structures to direct the learner's attention to the important facts, concepts, rules, and/or principles within the instruction. When developing quality-based instructional environments that are both effective and efficient, prior to the process of rapid prototyping (Seels & Glasglow, 1997; Smith & Ragan, 1999), item analysis is recommended to identify these most difficult items (Dwyer, 1978, p. 51).

A learner can interact with the rehearsal strategies in a covert or an overt manner. The learner is the only one that knows the extent to which the information was attended to and practiced when employing covert rehearsal strategies (e.g., how long a visual was examined, the completeness of the response to an embedded question). Overt rehearsal is observable; a physical motion of some dimension is required. Answering a question orally, writing a response on paper, or placing the cursor on a correct response and clicking in the computer program are examples of overt rehearsal techniques.

When interacting with an instructional environment, the learner will implement personal rehearsal techniques that have been successful in prior learning experiences. These techniques can be covert (e.g., the learner repeating the information or thinking about other topics that are similar and relating the new information to prior learned information) or they can be overt (e.g., the learner taking notes). When rehearsal strategies are provided or suggested, the learner uses these strategies in place of personal techniques or in addition to the personal techniques. The prompting for rehearsal of the information within the environments facilitates the covert or overt rehearsal at levels that can be classified as knowledge generation or discrimination tasks.

Prompting for knowledge generation or prompting for discrimination tasks initiates the learner's interaction with the information at different levels of processing (i.e., levels of rehearsal: elaborative or maintenance) in working memory (Craik & Lockhart, 1972; Craik & Tulving, 1975; Wittrock, 1992). Subsequent encoding and storage in long term memory result in different levels of retrieval during assessment or when applying the information while performing a task. Prior research has indicated that learners performing knowledge generation tasks during rehearsal of the instructional materials will perform at a statistically significant higher level on knowledge generation and discrimination tasks during assessment of the information as compared to those not performing knowledge generation tasks, whereas learners interacting with discrimination tasks during rehearsal of the instructional materials will perform statistically significantly higher or equally well on discrimination tasks as compared to knowledge generation tasks (Clariana, 2003; Clariana & Lee, 2001; Moscovitch & Craik, 1976).

Hierarchy of Learning Acquisition and Retrieval in the Program of Systematic Evaluation

The instructional materials and the criterion measures have been designed for the Program of Systematic Evaluation to be instructionally congruent, the content of the instructional materials is measured by the test items (Szabo et al., 1981). The instructional materials have been organized in a learning hierarchy that presents the complex learning activities in a systematic manner beginning with basic facts to complex problem solving (Dwyer, 1978, p. 44). For example, in this hierarchical design, the instruction section for the diastolic and systolic phases of the heart was preceded by instruction of the different parts of the heart, their locations, and their interrelationships; the basic information prerequisite to understanding the diastolic and systolic phases. This design is used in the Program of Systematic Evaluation because “students need to possess competency in prerequisite skills or basic learning levels before they can be successfully introduced to ideas which are comparatively complex or abstract or which extend beyond the learner’s existing range of experiences” (Dwyer, 1978, p. 44).

The hierarchy of learning tasks of the instruction in the Program of Systematic Evaluation extended from the simple (terminology) to the complex (comprehension) (Dwyer, 1978, p. 45). The learning objectives were evaluated using criterion measures that transcend specific discipline boundaries (Dwyer, 1978, p. 45). The performance measures are the terminology test, the identification test, the drawing test, and the comprehension test. These performance measures are described as follows (Dwyer, 1978, pp. 45-46):

Terminology Test. This test consisted of 20 multiple-choice items designed to measure the student's knowledge of specific facts, terms, and definitions. The objectives measured by this type of test are appropriate to all content areas which have as a prerequisite to the more complicated types of learning a comprehensive understanding of the basic elements (terminology, facts, and definitions) indigeneous [sic] to the discipline.

Identification Test. The objective of the identification test was to evaluate student ability to identify parts or positions of an object. This multiple choice test (N=20 items) required students to identify the numbered parts on a detailed drawing of a heart. Each part of the heart which had been discussed in the presentation was numbered on the drawing and appeared in a list on the answer sheet. The objective of this test was to measure the student's ability to use visual cues to discriminate one structure of the heart from another and to associate specific parts of the heart with their proper name. Tests similar to the identification test could be used in any course in which the student is required to be able to locate and identify the various parts of objects (for example, in an automotive repair course to identify the parts of a car engine, or in a botany course to locate and name the parts of different types of plants).

Drawing Test. The objective of the drawing test was to evaluate student ability to construct and/or reproduce items in their appropriate context. For example, the drawing test (N=18 items) provided the students with a numbered list of terms, e.g., (1) superior vena cava, (2) aorta, (3) tricuspid valve, (4) pulmonary vein, etc., corresponding to the parts of the heart discussed in the instructional presentation. The students were required to draw a representative diagram of the heart (a symbol like a valentine sufficed; the quality of the drawing did not enter into the scoring) and place the number on the listed parts in their respective positions. For this test the emphasis was on the correct positioning of the verbal symbols with respect to one another and in respect to their concrete referents. The educational objectives measured by this type of test could very easily be applied to a social studies class in which the relative location of the fifty states, with respect to one another, was being taught, or to an automotive course in which students were being taught the relative locations of the various parts of a carburetor.

Comprehension Test. The comprehension tests consisted of 20 multiple-choice items. Given the location of certain parts of the heart at a particular moment of its functioning, the student then was asked to locate the position of other specified parts of the heart at the same point in time. This test required that the students have a thorough understanding of the heart, its parts, its internal functioning, and the simultaneous processes occurring during the systolic and diastolic phases. The comprehension

> test was designed to measure a type of understanding that occurs when the individual understands what is being communicated and can use the information being received to explain some other phenomenon occurring simultaneously. Tests similar to the comprehension test can be used in any discipline area in which the objective is to measure the student's understanding of complex procedures and processes. … for students to consistently select the correct alternative … they must first be familiar with the terminology used to describe the heart; they must be able to recollect the location of the various parts of the heart and be able to position the individual parts of the heart in their "minds eye" as they relate to one another. Furthermore, they must also be able to mentally simulate the functions and movements of the various parts of the heart as they would occur during both systolic and diastolic phases of the heart. Only when the students have acquired a comprehensive understanding of the content material—the description and location of various parts of the heart and their individual functions during both the systolic and diastolic phases of the heart—can they consistently respond correctly to the type of questions contained in the comprehension test.

The different types of learning objectives used in the Program of Systematic Evaluation are generalizable to other disciplines. The conditions for successful attainment of learning tasks addressing similar learning objectives (e.g., identification of parts of an object, positions of objects relative to other objects) are the same regardless of the content area; the learning process for learning the basic terminology, facts, rules and/or principles would be the same. Naïve students interacting with unfamiliar content for the first time have to learn the basic terminology and facts to be prepared to relate and combine these units to form concepts and rule. This orientation to learning implies that the "ease of subsequent learning is directly related to the quality and quantity of the prerequisite learning" (Dwyer, 1978, p. 49).

Overview of the Use of Rehearsal Strategies in the Program of Systematic Evaluation

Implementing design decisions for including learning strategies that cue the learner to rehearse the information at an elaborative rehearsal level for later knowledge generation increases the effectiveness of the learning environment. These elaborative rehearsal strategies (e.g., answering open-ended, short response questions; coloring parts of an object based on directions; choosing a correct response from multiple choices; completing a concept map) facilitate increased interaction within working memory and activate the retrieval of prior knowledge for processing, encoding, and storage in long-term memory of the new information. The processing of images, audio, and/or text increases the interaction with the visuospatial sketch pad and phonological loop, two slave systems that are within working memory, with working memory's central executive control (Baddeley, 1992). The visuospatial sketch pad manipulates visual images. The phonological loop stores and rehearses speech-based information. These interactions, of the information being rehearsed, with the slave systems of working memory are dually coded (i.e., visually and verbally), increasing the interconnectivity of visuals and words (i.e., pictures can be names and words can evoke images) (Paivio, 1976).

The Program of Systematic Evaluation has employed elaborative rehearsal strategies to determine in the individual studies ways to activate prior knowledge and stimulate increased interactions within working memory for information processing for knowledge acquisition, encoding, and storage in long-term memory. These studies have used advance organizers, note-taking strategies, questions embedded within the instruction, and concept map completion and generation as techniques to cue the learner to rehearse the information within the instructional materials. The positioning of these rehearsal strategies has been analyzed through item analysis as the most difficult facts, concepts, rules, and principles to generate a response or discriminate on the criterion measures following the instruction.

Visuals with colorization, visuals that are black and white, or no visuals within the treatments have been graphical organizers for the rehearsal strategies used as the independent variable in studies within the Program of Systematic Evaluation. Research on the use of visualization in instruction within the Program of Systematic Evaluation has indicated that the use of visuals within instruction is an effective learning strategy, yet the interaction with the redundant visuals is not enough to ensure that the learner acquires with the information at a level of proficiency (Baker & Dwyer, 2000).

The rehearsal strategies tend to cue the learner to interact with information at processing levels (rehearsal levels) consistent with discrimination tasks or with knowledge generation tasks. The discrimination tasks required the learner to recognize the information within the rehearsal strategy task. In the discrimination task research studies within the Program of Systematic Evaluation, the researchers have designed rehearsal strategies using discrimination tasks. These discrimination tasks include questions, advance organizers, study guides, and instructor generated concept maps. The discrimination task questions have been multiple choice questions that required differentiation of information rather than generation of information, the advance organizers required the learner to recognize situations of prior knowledge (e.g., a metaphor of the human heart operating in the same manner as a water pump), and the study guides with summary statements and instructor-generated concept maps were referenced during the reading of the prose text for organization, reinforcement, and discrimination of important points within the instructional materials. The knowledge generations tasks used within the Program of Systematic Evaluation required the learner to retrieve information from prior stores of knowledge to answer questions that were cued recall questions. In knowledge generation task research studies, the learners were required to know additional information or state the newly stored information, to generate concept maps or complete a study guide from the information the learner had just completed reading without referencing the text, or retrieve information from prior knowledge to be related to the information within the advance organizer.

The medium for presentation of instructional materials in the Program of Systematic Evaluation has been screen presentation, paper booklets, and computer and web-based instructional environments. The screen presentations were presented to the learner with visuals and/or verbal cues shown via projected images from slides or video projector and a taped recording of the prose text playing on the room's speakers. The paper-based prose materials were printed and bound in booklet form with and without visuals. The computer based instructional environments displayed the prose text with and without visuals and were stand alone computer based instructional units or were served over the Internet via a web browser.

As a measure of levels of learner achievement, these studies used one or more of the criterion measures to assess the effect of the instructional intervention as a means of facilitating higher levels of achievement of the specific learning objectives within each of the criterion measures. The criterion measures examine the learners' achievement of facts, concepts, rules, and principles by completing tasks that measure knowledge generation and discrimination of the information presented in the prose instructional unit in the performance assessment or objective measure, respectively.

Matching knowledge generation tasks with knowledge generation assessment tasks has resulted in statistically significant increases in achievement of these knowledge generation assessment tasks (Moscovitch & Craik, 1976) and of discrimination assessment tasks (Clariana, 2003; Clariana & Lee, 2001). The use of overt rehearsal strategies requiring the learner to practice the facts and concepts presented within the instruction have resulted in increased achievement of specific learning objectives (Dwyer & Dwyer, 1987). The goal of this research study was an examination of the use of these knowledge generation and discrimination tasks as methods of overt and covert rehearsal within instruction as used within the phase three studies in the Program of Systematic Evaluation (Dwyer, 2004).

Overview of Methodology

To determine the relationships between rehearsal strategies and learning tasks across multiple studies requires synthesis of the results through the use of meta-analytic techniques. Meta-analysis is a secondary research method with its greatest potential use in areas that are well established and have a large number of primary studies; the Program of Systematic Evaluation is one such area. Meta-analysis integrates and aggregates the multiple studies to examine general trends and consistency of effects across the studies, investigates relationships not investigated in the original primary studies, and finds trends too subtle to identify in the traditional, experimental research.

This study uses a combination of approaches including Glass' bare bones analysis techniques (Glass, 1976); Arthur, Bennett, and Huffcutt's techniques of performing meta-analysis using the Statistical Analysis System (SAS) statistical software (a modular, integrated, hardware-independent system of software for enterprise-wide information delivery) (Arthur, Bennett, & Huffcutt, 2001); and Hunter and Schmidt's techniques of correcting error and biases in meta-analysis (Hunter & Schmidt, 2004). The details of the techniques are described in Chapter 3, Methodology.

The estimate of the population effect size was calculated for the effect of rehearsal strategies for an unconditional level-1model and a conditional level-2 model (Bryk & Raudenbush, 1992). An unconditional level-1 model estimates the mean of the effect size statistics for the population. The level-1 model estimate represents the value of the mean effect size of the effect sizes for the entire population of well-designed studies examining the effect of rehearsal strategies, regardless of the type of rehearsal strategy or any additional independent or controlled variables in the studies. The level-2 model estimate of the population mean effect size includes the relationships of predictors (the independent variables and conditions in the studies, the level-2 variables) in the model found to be statistically significant predictors.

The unconditional level-1model was defined as the study effect size is equal to the mean population effect size and error. The level-1 model was analyzed to determine the mean population effect size of the use of rehearsal strategies for the entire population studies having attributes similar to the 50 studies selected from the Program of Systematic Evaluation. The level-2 model was defined as the mean population effect size is equal to the grand mean, the relationship of the statistically significant predictors, and error. The level-2 model was analyzed to determine the mean population effect size relative to the conditions of the statistically significant predictors (Bryk & Raudenbush, 1992).

Criticisms of Meta-Analysis

The most common criticisms of meta-analysis are the combining of studies that have not used a common criterion measure (i.e., comparing apples and oranges), inclusion of studies that have not controlled for competing explanations and outcomes (i.e., have not used common content or method of presentation), and the exclusion of studies that have not been published within journals due to a lack of finding significant differences within the research (i.e., the “file-drawer” syndrome—studies that are in file drawers and not available to any search of published works) (Rosenthal, 1991, 1998). Statistical analysis to determine the number of missing studies to affect the results and techniques to correct for variability, sampling, and different criterion measures have been developed to address each of these criticisms.

The many techniques for performing meta-analysis address the criticism of not using a common criterion measure, which is measuring the effect of a common construct, by converting the outcomes to a common metric (e.g., the effect size). This problem is often referred to as the comparison of apples and oranges. The effect size determines the difference in performance of the experimental group and the control group in standard deviations. Cohen (1992) describes effects sizes as the probability that the difference noted is related to something other than chance, the magnitude of the effect size is the practical significance of the difference.

At times, there truly are apples and oranges—the subsets are different from each other and require a separate meta-analysis for each subset. To calculate the common metric for comparison, the meta-analytic procedure focuses on the use of the means and the standard deviations from each study in such a way that the calculation of the effect size is performed with data including the same level-2 variable (i.e., the variables responsible for variation that exists in a population evaluated with meta-analysis after the specified corrections have been made, these are the independent variables and/or the conditions in the primary studies included in the meta-analysis). Recording other study characteristics permits the corrections due to attenuation by the study's artifacts (i.e., a study imperfection that is the result of artifactual or man-made errors and not a property of nature, for example measurement error is a man-made error) (Hunter & Schmidt, 2004). The level-2 variables are analyzed for relationships and statistical significance as a predictor of the population effect size estimate. The bare-bones technique of meta-analysis using level-2 variables results in a mean effect size for the studies in the sample that include the level-2 variable for each level-2 variable without regard to interactions, noise in the measurements due to error, or the variance of the effect sizes. Performing a hierarchical linear model with meta-analytic data determines the relationship of the level-2 variables to the population effect size estimate with corrections for error, noise, sample size, and variance.

Missing studies are believed to be a problem that leads to an availability bias and results with a biased effect size. Statistical analysis has been performed focusing on inclusion of only studies with statistically significant results within a meta-analysis (Rosenthal, 1979) and the inclusion of studies that were determined to have statistically significant effect sizes after the calculation (Hedges & Olkin, 1985). Hedges and Olkin found that 900 lost studies are necessary to affect the results enough for an error in the conclusions for each 100 studies included in a meta-analysis that had a statistically significant effect size. Using the analysis of the probability of a Type I or a Type II error and having studies that are lost within file drawers, Rosenthal and Rubin (1988) have performed calculations describing if there were studies that averaged a null hypothesis, the number of missing studies would have 19 unpublished studies with a nonsignificant result for every study included that had a significant result. Therefore, since it is unlikely that the large numbers of studies would be lost in file drawers (i.e., more than 10 studies lost for every study included in the meta-analysis), those studies available through published journals and unpublished theses would result in the best available estimate of the effect size for the independent variables on the dependent measure.

An accurate mean population effect size estimate is attainable through careful analysis and corrections for the various artifacts (i.e., study imperfections). Inclusion of research studies that have a common criterion measure, a common instructional unit for construct consistency and method of presentation, and have been thoroughly reviewed for proper research method application would also address the primary criticisms of meta-analysis. Additional criticisms (e.g., homogeneity of results, sampling error) are found to be situation-specific and are also addressed through statistical research demonstrating the unlikelihood of all the studies included having identical results or the differences between the experimental and control groups when calculating effect sizes to be only due to sampling error. Corrections for artifacts and controlling for level-2 variables also increase the accuracy of the effect size calculated and addresses the additional criticisms.

Criticism Resolution in the Program of Systematic Evaluation

The Program of Systematic Evaluation is a collection of studies that use the common core instructional module and common criteria measures. Every study utilizes an experimental design with random assignment of subjects to the treatment groups. The independent variables have been selected through literature review and placement of the strategies has been through item analysis of the criterion measures after a pilot study with a representative sample from the study population.

The criticism of meta-analysis of comparing apples and oranges has two resolutions within the Program of Systematic Evaluation and this meta-analysis. The common criterion measures ensure that the construct being measured is consistent in every study. Within this meta-analysis, examination of level-2 variables (independent variables and the conditions of the research studies) controls for real apples and oranges that influence the variability within the population effect sizes.

The criticism of including studies that have not controlled for competing explanations and outcomes is resolved in the Program of Systematic Evaluation by the strict control and approval of using the instructional materials for experimental research, random assignment of subjects, and placement of the rehearsal strategies in the common instructional unit by analysis of item difficulty in the criterion measures. There is strict control in each of the studies for construct validity, dichotomization and range variation of the dependent and independent variables, and variance due to extraneous factors that affect the relationship.

The criticism of the exclusion of studies that have not been published within journals due to a lack of finding significant differences within the research is addressed within this meta-analysis by including studies that have both statistically and non-statistically significant differences. The reliance of inclusion in this meta-analysis was not on publications within referred journals. This meta-analytic study evaluated conference presentations, masters paper, doctoral dissertations, and published experimental studies for inclusion in the study.

Delimitations

The studies to be included in this meta-analysis range from completion in 1984 to 2006. There were studies that included rehearsal strategies as early as 1971; however, the publications did not report enough statistical data to calculate the effect sizes. The studies have been completed in domestic and international environments including the United States, Panama, Korea, the Caribbean, and China. The age ranges include high school students, traditional and non-traditional college students, and adult learners. Each of the studies reports limitations of generalizability to the general population. Due to the mixture of educational settings, age ranges, and educational backgrounds for the participants and the random assignment of the experimental treatments, the results of this meta-analysis are generalizable to a larger population than any of the individual studies.

Rationale for the Study

Overt and covert rehearsal strategies have been researched to determine under which conditions, for which learning objectives, and which of the two types of rehearsal, overt or covert, are most effective (Cha, 1990; Dwyer & Dwyer, 1987). Knowledge generation and discrimination level rehearsal strategies and knowledge generation and discrimination assessment tasks in the form of performance assessments and objective measures have also been researched (Clariana, 2003; Clariana & Lee, 2001; Moscovitch & Craik, 1976).

There is a need to examine the use of overt and covert rehearsal in combination with knowledge generation and discrimination tasks and assessments to determine the most effective and efficient design decisions. Are there conditions of instructional design (e.g., the use of visuals, the media of presentation) that are more supportive to facilitate interaction in working memory, encoding, and performance on the assessments of the learning objectives if a particular type of rehearsal is used with a particular task strategy?

This study will provide generalizable evidence regarding the use of properly positioned overt and covert rehearsal strategies in instructional environments that facilitate elaboration and maintenance rehearsal levels of processing with the rehearsal tasks.

Problem Statement

A series of experimental studies using the same core instructional materials and criterion measures began in 1965; this is now known as the Program of Systematic

Evaluation (Dwyer, 1972, 2004). The focus of the Program of Systematic Evaluation is the identification of types of visual materials used singly or in combination with other independent variables to determine which is the most effective in facilitating learner achievement of specific educational objectives. Rehearsal strategies aimed at increasing learner achievement by facilitating information processing have been selected as independent variables in studies from 1971 through 2006. Each research study was a quantitative, experimental study to observe their relationship to improved effectiveness, efficiency, and utility of instructional environments and software. More than 200,000 learners have interacted with instructional materials within the Program of Systematic Evaluation (Dwyer, 2004). Each of the studies used item analysis to position the visualization and rehearsal strategies to best facilitate increased achievement by directing the learners to the information that was most difficult to generate or discriminate without the strategies in position.

An examination of these studies reveals the use of rehearsal strategies and tasks that can be classified as covert or overt rehearsal and knowledge generation or discrimination tasks. Overt rehearsal strategies are observable; they require a physical action when making the response. Covert rehearsal is rehearsal that is not observable; it is performed mentally. Knowledge generation tasks activate prior knowledge and require the learner to generate a response to a stimulus. Knowledge generation tasks include constructed response questions, fill-in-the blank questions, essays, and concept map completion (Brown, 1976; Lockhart, Craik, & Jacoby, 1976). A discrimination task requires the learner to differentiate among a series of choices. Discrimination tasks include multiple choice questions that require the learner to select the correct response and statements about the information that the learner recognizes as correct (Brown, 1976; Lockhart et al., 1976).

In phase three of the Program of Systematic Evaluation, the examination of rehearsal strategies (i.e., advance organizers, questions, note-taking, concept map completion or generation) was employed in the visualized and non-visualized learning environments (Dwyer, 1987, p. 7). By synthesizing these studies, the research question related to the effects of rehearsal strategies in an unconditional, level-1model to determine the estimated population effect size were answered. Using the effect sizes calculated for the studies chosen for inclusion in the meta-analysis, the unconditional level-1 model determines the estimated population effect size that would result if every well-designed study researching the effects of rehearsal strategies were available for analysis. Analysis using a level-2 model determines if any of the independent variables or study conditions has a relationship to the estimated mean population effect size. The relationship was modeled using hierarchical linear modeling. The result was an examination of the statistically significant predictors in the model. These predictors are the level-2 variables.

The purpose of this research was to summarize the results from studies that have employed covert and overt rehearsal strategies and have been completed as part of the Program of Systematic Evaluation. This synthesis of effect sizes calculated from the included studies was to answer the question of how the instructional effect of covert and overt rehearsal strategies was influenced by the employment of knowledge generation and discrimination tasks (i.e., the rehearsal strategies in the instructional environments used for research within the Program of Systematic Evaluation) to result in changes in student achievement in the assessments of specific educational objectives. The goal of this study is to provide guidance to instructional designers when designing and developing instructional environments to promote increased mastery of the learning objectives.

Research Questions

To combine the many studies is to employ the quantitative research method developed as meta-analysis. Many meta-analytic studies have synthesized the results of studies examining the same construct; few meta-analytic studies have combined the results of a program of evaluation. Those that have combined the same program of evaluation have been within the medical field and not the social sciences.

Within each study are controls for the rehearsal strategy and treatments that employ the rehearsal strategy as an independent variable. A rehearsal strategy can be rehearsed covertly or overtly. The task requirement of the rehearsal strategy promotes interaction in working memory at a elaborative rehearsal or at a maintenance rehearsal level of processing. The study may use a visual and be presented in an screen, paper, or computer based format. To measure the level of achievement, the criterion measures used assess at the knowledge generation or discrimination level, in the form of performance and objective measures, and target the specific learning objectives respective to facts, concepts, or rules and principles.

The calculation of the effect of rehearsal strategies is calculated to determine the individual effect sizes. In the hierarchical linear modeling technique in this meta-analysis, the level-1 analysis was completed for overall effects to determine the estimate mean population effect size in all conditions and for all independent variables. The effect of level-2 variables was evaluated on a hierarchical basis (i.e., the variables are nested within the studies and moderate the population effect sizes on a hierarchical basis) and was examined in the level-2 model in this meta-analysis.

The research questions were:

1. What is the size of the effect of rehearsal strategies used within the Program of Systematic Evaluation on student achievement of specific learning objectives?
2. How do the level-2 variables *[visualization (colorized, black and white, no visual), prompting type (knowledge generation tasks, discrimination tasks, no prompt), media presentation style (screen presentation, paper based, computer based), and criterion measure (objective measure, performance assessment)]* moderate (i.e., reduce the variance) the population effect size for rehearsal effectiveness?

Definitions and Key Terms

The following definitions are provided to ensure a common understanding of the terminology used within this study.

Term	Definition
Artifact	An artifact is a study imperfection that is the result of artifactual or man-made errors and not a property of nature (Hunter & Schmidt, 2004).
Black and White Visual	A black and white visual is a graphic that uses only black and white lines and arrows to depict the various aspects of the graphic being presented within the instruction (Dwyer, 1978, p. 54).
Color-Coded Visual	A colorized visual is a graphic that uses color-coded lines, black arrows and colorized shading of various parts of the graphic to depict the various aspects of the graphic being presented within the instruction (Dwyer, 1978, p. 55).
Computer-based Instruction	Computer-based instruction is presented wholly using computer-media as the platform for presentation. Computer-based instruction also includes web-based instruction.
Covert Rehearsal	Covert rehearsal is rehearsal that is not observable. Covert rehearsal is performed mentally (Dwyer & Dwyer, 1987).
Discrimination task / discrimination rehearsal strategy	A discrimination task or discrimination rehearsal strategy requires the learner to discriminate among a series of choices. Discrimination tasks include multiple choice questions that require the learner to select the correct response and statements about the information that the learner recognizes as correct (Brown, 1976; Lockhart et al., 1976). The term, discrimination task, is used instead of the term, recognition task, for clarity.

Term	Definition
Elaborative Rehearsal	Elaborative rehearsal involves integrated semantic processing of a to-be-remembered item resulting in the production of durable memories (Craik & Lockhart, 1972).
Knowledge Generation task / knowledge generation rehearsal strategy	A knowledge generation task or knowledge generation rehearsal strategy activates prior knowledge and requires the learner to generate a response to a stimulus. Knowledge generation tasks include constructed response questions, fill-in-the blank questions, essays, and concept map generation or completion (Brown, 1976; Lockhart et al., 1976). The term, knowledge generation, is used instead of the term, recall, for clarity.
Level-1 model	The level-1 model is an unconditional model that is estimated by the means of the effect size statistics (Bryk & Raudenbush, 1992).
Level-2 model	The level-2 model is a conditional model that estimates the true population effect size as a function of the predictor variables measured at level-2 plus error. The predictor variables are the level-2 variables (Bryk & Raudenbush, 1992).
Level-2 variable	A level-2 variable is a variable that causes nonartifactual variation from one study to the next. Level-2 variables are also called moderator variables (Hunter & Schmidt, 2004).
Maintenance Rehearsal	Maintenance rehearsal is simple rote repetition (Craik & Lockhart, 1972).
Objective Measure	An objective measure is a task that is scored objectively rather than subjectively. The responses are provided for the learner to discriminate among rather than the learner generating the response. The objective measures in the Program of Systematic Evaluation are the identification test, the comprehension test, and the terminology test. The term objective measures are used instead of recognition assessment for clarity.

Term	Definition
Overt Rehearsal	Overt rehearsal is rehearsal that is observable. Overt rehearsal requires a physical action when making the response (Dwyer & Dwyer, 1987).
Paper-based Instruction	Paper-based instruction is presented wholly using printed information on paper as the platform for presentation.
Performance Assessment	A performance assessment is a task that requires the learner to generate a response to a question or statement. The performance assessment within the Program of Systematic Evaluation is the drawing test. The term performance assessment is used instead of recall assessment for clarity.
Phonological loop	The phonological loop, part of the slave system of working memory, stores and rehearses speech based information and is necessary for the acquisition of vocabulary (Baddeley, 1992).
Prompted-Covert Rehearsal	Prompted-covert rehearsal is mental rehearsal that is prompted by a statement or question placed within the instructional materials and is performed mentally.
Question	A question is a prompt to examine the knowledge of the learner. Questions may be responded to covertly or overtly. Questions may require the learner to generate the information without any additional prompts (recall) or to recognize a response from a series of choices (Smith & Ragan, 1999, p. 168).
Rehearsal	Rehearsal is a repetition of the information received in the short-term memory that permits the information to be recognized, retained and organized for storage within long-term memory (Weinstein & Mayer, 1986). Rehearsal may be covert, prompted-covert, or overt.
Screen-based Instruction	Instruction presented wholly using audio-tape media as the platform for presentation. This instruction may be supplemented with projected images.

Term	Definition
Visuospatial sketch pad	The visuospatial sketch pad, part of the slave system of working memory, manipulates visual images in working memory (Baddeley, 1992).

Summary

The Program of Systematic Evaluation is a comprehensive research endeavor focusing on determining the most effective and efficient instructional environments and increasing the utility of instructional and training software materials. The studies within the Program of Systematic Evaluation have examined the effects of the learner's acquisition of information while interacting with instructional environments employing covert and overt rehearsal strategies and knowledge generation and discrimination tasks.

Advancements in technology are making it possible for instructional designers to insert more complex rehearsal strategies within instructional environments. These complex rehearsal strategies are employing knowledge generation tasks to rehearse the information elaboratively and encode the information for greater levels of recall after the instruction. Additional research is needed to be completed to provide instructional designers guidance on the interaction of covert and overt rehearsal strategies and knowledge generation and discrimination tasks within the rehearsal strategies. This research can be completed by using meta-analytic techniques to analyze the studies completed within the Program of Systematic Evaluation that have used rehearsal strategies.

Chapter 2

Literature Review

Rehearsal is a core component to all instructional situations. Rehearsal techniques are typically embedded within the instruction, created by the learner, or initiated by an outside source (e.g., a teacher or other learner). Regardless of the source of rehearsal, the fundamental aspect is the cognitive processing required for the learner to rehearse the information for acquisition and subsequent later recognition or recall. Researchers have been studying techniques of rehearsal in an attempt to increase the efficiency of the instructional materials and the probability of the learner recognizing or recalling the correct response after the instructional experience.

Within the Program of Systematic Evaluation (Dwyer, 2004), numerous researchers have used rehearsal techniques as independent variables in research studies to determine ways for learners to increase their achievement of specific learning objectives. These researchers have applied a variety of learning strategies to facilitate rehearsal. The goal of every strategy was to focus attention of the learner on the fact, concept, rule, and/or principle to be learned; have the learner interact with the information; encode the information; then recall or recognize the correct response on an assessment measuring the learning objectives.

The literature reviewed in this chapter will review the cognitive theories of information processing (e.g., Thorndike’s law of exercise (1913), Atkinson and Shriffen’s stage model (1968), Schneider and Graham’s connectionist model (1992)) and rehearsal for discrimination and knowledge generation. These theories have been the foundation for continued research within the Program of Systematic Evaluation. The foundational assumption of every study in the Program of Systematic Evaluation was that the learner interacted with the selected independent variables in the study treatments that had been positioned according to item analysis of the criterion measures for the study population. This interaction was intended to facilitate information processing of the targeted areas in the instruction in need of additional rehearsal in order to facilitate increased achievement of the facts, concepts, rules, and/or principles.

Information Processing

One topic that has spanned the research literature is rehearsal and the strategies necessary to practice an action, fact, concept, rule, or procedure. Early in the theory development of behaviorism, Thorndike (1913) developed the law of exercise. The law stated that the probability of a correct response to a stimulus was increased with repetition of the experience. Using the behavioral approach of stimulus and response, Thorndike’s research studies were designed to test if the learner would keep doing an action based on the reward-stimulus, and then get faster with each successful attempt. The core aspect of this rehearsal in the law of exercise is a satisfied feeling by the learner based on reward-stimulus for the repetition. The cyclical aspect meant that the learner was responding to a stimulus, then responding to a different stimulus that was internally created.

Additional research was completed by Thorndike to examine the response to a stimulus that required an action to get the reward. Thorndike was concerned with the states of mind and how they changed with experience. This work labeled Thorndike as one of the first connectionists; people that base information processing on parallel processing of sub-symbols, using statistical properties instead of logical rules to transform information (Bowers, 2002). Thorndike's work was cross-categorized in behaviorism and cognitivism since his research and reports focused on topics associated with what happens when a learner thinks about the content rather than responds to a stimulus and performing a behavioral response. Thorndike's theories of connectionism were the beginning of theories about learners applying actions to previously learned information through association.

As research shifted focus, the cognitive revolution documented how humans made meaning in their daily encounters with information (Bruner, 1990). There was another shift in research from the meanings to the processing of information within the human mind. The most widely accepted was the stage model (Atkinson & Shriffen, 1968) where information is processed and stored in three stages, the basis for the current information processing model. The levels of processing theory (Craik & Lockhart, 1972) was the second widely accepted theory of how the human mind processes information; different levels of elaboration are required to process information along a continuum of perception, attention, labeling, and meaning. A third, the parallel-distributed processing model, a connectionist model (Schneider & Graham, 1992), maintains that multiple processing occurs simultaneously in the mind. With the increase in computer technology, the models shifted to follow the computer processing metaphor; the information is received, stored, and retrieved as necessary.

Common assumptions associated with the computer processing metaphor of information processing are that processing occurs in stages that intervene between the reception of a stimulus and the production of a response and that the human system functions as a receptor of information, stores that information in memory, and retrieves the information as necessary (Schunk, 2000, p. 120). Many models of this processing are displayed within texts for reference (Gredler, 1997, p. 144; Schunk, 2000, p. 121; Smith & Ragan, 1999, p. 20).

Figure 2.1 depicts a representation of the information-processing model based on the computer processing metaphor. Information stored in memory passes through a series of transformations within sensory perception, short-term memory/working memory, and long-term memory (Atkinson & Shriffen, 1968). The executive control process regulates the flow of information within the information processing system.

The inputs to the sensory register are primarily audio and visual signals. Of the signals input into the sensory register, only a few are maintained for 0.5 to 2.0 seconds for the preliminary analysis. Even fewer of those retained by the sensory register are passed to working memory for additional processing. Additional inputs are sent to the sensory register and compete with those signals already received and passed to working memory and with those in the sensory register to be processed. Many signals will be lost unless there is stimulus to pay attention to the signals in the sensory register.
Working memory has a limited capacity, proposed to be seven plus or minus two units of information (Miller, 1956). Without rehearsal or higher levels of processing, these units

will be lost from working memory. The information within the working memory stays for a short time, approximately 20 seconds. For this reason, working memory is often called short-term memory. Working memory is divided into three subcomponents: the central executive for attention and controlling and two slave systems (a) the visuospatial sketch pad for visual images and (b) the phonological loop for speech-based information (Baddeley, 1992). Like the sensory register, some of the information is lost from working memory without being encoded into a meaningful form and transferred to long term memory for storage.

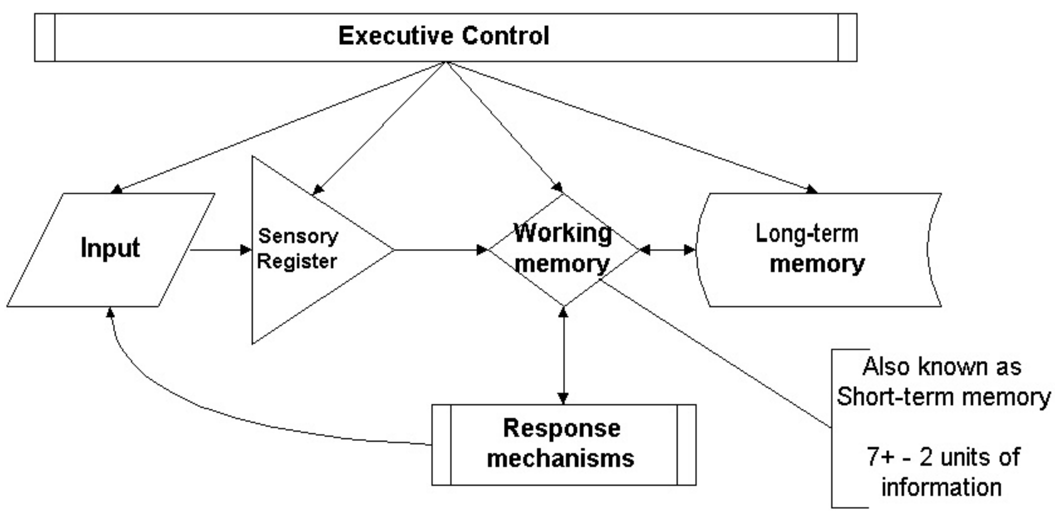

Figure 2-1: Information Processing Model, adapted from Gredler (2001, p. 144)

The information within working memory is interrelated with information from long term memory and processed for retention and recall of the information (Bentin, Moscovitch, & Nirhod, 1998). Encoding of the information can be facilitated by using chunking strategies to reduce the number of units of information being processed. Other learning strategies (e.g., advance organizers, adjunct questions, notes, concept maps) are used by instructional designers as ways to keep the information within the learner's working memory and to help the learner encode the information for storage in long term memory. These encoding strategies are rehearsal techniques.

Long-term memory is portrayed in the computer processing metaphor model as a place for permanent storage, a place where the information is retained until retrieved for recall or recognition. How the information is stored in long-term memory is proposed to take place through connections with prior knowledge (Schneider & Graham, 1992); as episodic, semantic, or procedural components (Tulving & Thomson, 1973); or as verbal or non-verbal forms in a dual-code (Paivio, 1986). The information in long-term memory, regardless of the title given, stays until needed to interact with information within the working memory, then encoded for additional storage or used to perform a task (e.g., answer a question, solve a problem).

The key components of the human information processing model are that the information is perceived, attended to, encoded, stored, and retrieved. The cognitive processes are based on perception, attention, and encoding. Without the interaction through perception, attention, and encoding, the probability of retaining the information is highly reduced and retrieval for performance on assessments to measure learner achievement is not expected. It is important to assist the learner with which components in instructional materials to perceive, attend to, and encode. To assist the learner, rehearsal is a repetition of the information received in the working memory that permits the information to be recognized, retained, and organized for storage within long-term memory. Rehearsal may be performed covertly (e.g., within the mind, thinking about information, repeating a line of a song) or overtly (e.g., writing the answer to a constructed response question, playing a game).

Rehearsal

Rehearsal of information facilitates the retention of the information in the working memory for longer periods of time, increasing the probability of encoding and storing the information in long-term memory. Rehearsal works with the executive control processes within the information processing model. The level at which the information is processed also affects the encoding and storage of the information within long-term memory (Craik & Lockhart, 1972; Craik & Tulving, 1975).

The levels of processing model proposes: the deeper the processing (i.e., the more elaborative the rehearsal) of the information, the higher the level of recall since more meaning is assigned to the information (Craik & Lockhart, 1972; Craik & Tulving, 1975). Elaborative rehearsal is the construction of integrated representations within working memory (Lehman & Schraw, 2002). Elaborative rehearsal of the information presented increases recall of information, with action encoding being the most influential on increased levels of recall (Zimmer & Engelkamp, 1999). Questions increase the time that information is in working/short-term memory, thus increasing the probability of elaborative rehearsal and encoding the information in long-term memory through chunking of information (i.e., combining smaller bits of information into a coherent group for processing) and linking to prior knowledge (Wager & Mory, 1992). Also, elaborative rehearsal reduces the amount of false recall and false recognition of information (Thapar & McDermott, 2001).

The quantity of practice is not enough for encoding, the type of processing and the type of task are also important. Encoding for a recall assessment is enhanced if the type of processing is recall (Moscovitch & Craik, 1976), meaning the learner will perform at a higher level on assessments that require the learner to generate a response if the rehearsal activities within the instruction require the learner to attend to and practice the information at a level that integrates the new information with prior knowledge.

Another process of rehearsal to facilitate attention and encoding of information is elaboration, the process of expanding new information by adding to it or linking the information to prior knowledge. Elaborative rehearsal strategies increase working memory activation and also increase the amount of retention of the learners (Harris & Qualls, 2000). Elaborative interrogation, a method of higher-order questioning, permits learners to interact with the information in such a manner as to relate the new information to prior knowledge (Willoughby & Wood, 1994). This elaborative interrogation permits the information to be manipulated longer within working memory, increasing the probability of recall or recognition.

Learning strategies are an important aspect to facilitate rehearsal, encoding, and recall or recognition of the learning objectives. Learning strategies prompt the learner to interact with the instructional materials in such a way that the information is held within working memory for a longer period of time, giving the learner a better opportunity to encode the information in a meaningful way for later recall. Rehearsal strategies used within the Program of Systematic Evaluation include advance organizers, questions, note-taking, and concept map generation.

Advance Organizers

Advance organizers are placed at the beginning of a learning unit to alert the learner to information to be presented, help link the information to prior knowledge, and scaffold the ideas by providing guidance and support required to master the skill. The learner uses the advance organizer as a channel to relate the topics within the instruction. The learner incorporates additional information as the learning progresses (Gredler, 2001). Many studies have incorrectly identified activities such as pre-reading, paragraph headings, learning objectives, and outlines as advance organizers (Corkill, 1992).

Ausubel (1968) proposed that advance organizers can be expository or comparative and are at one level more abstract than the information to be presented. Expository advance organizers include concept definitions and generalizations for helping the learner comprehend the instructional materials. The expository advance organizers are a statement that is intended to give the learner information about a difficult topic and/or explain the topic's difficult meaning. Comparative advance organizers introduce new material by presenting analogies with similar material to which the learner is familiar. An example of a comparative advance organizer used within the Program of Systematic Evaluation is the metaphor of a water pump: the human heart works like a water pump. Descriptions are given and visuals were used to provide the comparison for the learners.

Either expository or comparative advance organizers can be used to facilitate elaborative or maintenance rehearsal. The difference is that elaborative rehearsal will have the learner integrate the information with prior learning to a level that the learner can generate the information from memory whereas maintenance rehearsal will be for recollection and familiarization.

As a learning strategy, advance organizers are believed to draw attention to the concepts to be learned in the instruction, activate information from long-term memory, and help the learner keep information within the working memory for the learner to encode the information and link the information to prior knowledge. Ruthkosky and Dwyer (1996) examined the effectiveness of advance organizers within instruction in the Program of Systematic Evaluation; results indicated that all the advance organizer strategies were equally effective. The advance organizers included information to activate prior knowledge of the learners. The placement of the advance organizers were in positions determined by item analysis as locations to facilitate attention to the information included within the instruction and link that information to prior knowledge. The item analysis was performed on outcomes of the criterion measures during a pilot study examining where the study population would have the most difficulty answering the questions in the criterion measures correctly.

Questions

The use of questions within text has improved performance of learners (Osman & Hannafin, 1994). The level of the question is critical. Questions are most effective for long text passages. Placed within the text, the questions influence the processing of information of the text read prior to the question placement and the text to be read after the placement (Andre, 1979; Duchastel, 1983; Hamaker, 1986). Higher-level questions have been shown more frequently to obtain larger positive effects than factual-level question for lower-order posttest outcomes (e.g., facts and verbal information) and certain higher-order posttest outcomes (e.g., applying knowledge of concepts and principles to recognize a new example, and solving problems involving the same concepts and/or principles) (Andre, 1979, 1990; Lindner & Rickards, 1985).

The questions prepare or orient the learner to the important information to process, encode, and store in long-term memory for later recall or recognition. Prompted by the questions within instruction, the learner encodes a prior knowledge structure about the questions and answers into long-term memory in such a manner that the encoding and storage facilitates answering questions more correctly when asked at a later time (Glover, 1989; Willoughby & Wood, 1994).

The use of questions for rehearsal of information within the Program of Systematic Evaluation has shown that the use of questions does affect how learners perform on assessments of specific learning objectives (Lee & Dwyer, 1992). Item analysis was completed to ensure that the questions were placed in positions within the instruction to direct the learner's attention to the information in the text found to be most difficult for the learners in the study population to answer correctly on the criterion measures.

The outcome of the use of questions within prose learning environments (Dornisch & Sperling, 2004) and within the Program of Systematic Evaluation (Slater &

Dwyer, 1996; Vance & Dwyer, 1988) has been inconclusive in some research studies; the use of the questions in the instruction was found to have no significant statistical differences among treatments with questions placed in redundant positions and those treatments with no questions as a rehearsal strategy. This outcome may be due to a mismatch between the type of rehearsal (i.e., knowledge generation rehearsal or discrimination rehearsal) and the type of assessment measuring the specific learning objectives. Those studies that did examine the matching of rehearsal and assessment activity did result in statistically significant differences where the knowledge generation practice matched the recall assessment or discrimination practice matched the recognition assessment (Clariana, 2003; Lee & Dwyer, 1992).

Note-taking

Research on note-taking supports that note-taking facilitates learners to encode information into long-term memory and to recall the information correctly when completing an assessment on the designated information (Beecher, 1988). Crawford (1925) and Henk and Stahl (1985) concluded that the positive relationship between taking notes and subsequent grades was due to taking, reviewing, and organizing the notes. A meta-analysis of 57 studies that examined the effects of note-taking versus no note-taking found that positive interventions (e.g., pre-training of note-taking skills or techniques, providing framework notes, giving verbal instructions to employ an effective note-taking procedure) did not enhance the benefits of note-taking, visual presentation of the instructional materials interfered with the note-taking process, longer presentations did not interfere with the note-taking process, and the recall tests reflected a statistically significant higher level of encoding and retrieval of the information than did the multiple-choice objective measures (Kobayashi, 2005).

Within the Program of Systematic Evaluation, Walko and Dwyer (1990) examined the use of note-taking as a complement to visualized instruction as a means to facilitate information processing. Results indicate that the learners using their own note-taking strategy performed at a statistically higher level on the comprehension test. Walko and Dwyer used item analysis to place the note-taking strategies where the rehearsal of recording notes would be positioned to reinforce the information that was found to be the most difficult for the learners in the study population to answer correctly on the criterion measures.

Concept Map Generation

Concept maps are a schematic device used to represent concept meanings embedded in a framework of propositions (Novak & Gowin, 1984, p. 15). Placed in a two dimensional setting, the learner shows relationships and inter-relationships between selected concepts in a hierarchical arrangement (Jonassen & Grabowski, 1993). Concept maps foster the processing of information to be retained by having the learner attend to and encode the information in working memory, increasing the probability of storage in long-term memory. Concept mapping of a reading passage has been instrumental in statistically significant increases of student recall of information within the passage on free recall and multiple-choice objective measures (Wachter, 1993).

Studies within the Program of Systematic Evaluation that have used concept mapping as a learning strategy to facilitate rehearsal include Roshan and Dwyer (1998), Smith and Dwyer (1995), Taricani (2002), Wang (2003) and Yamashiro (2001). Results from these studies indicated the various types of concept maps were equally effective in terms of facilitating achievement with the exception of the learner-generated maps in the Taricani study (2002): these treatments facilitated a statistically significant increase in learner achievement for the comprehension criterion measure.

Overt and Covert Rehearsal

Each of the rehearsal learning strategies can be categorized as being an overt or a covert rehearsal strategy. Overt is open and visible in that it is observable by others; it involves cognitive processing and some type of observable action. Covert is not openly practiced in that it is not observable; it is completed cognitively by the learner. Both forms of rehearsal have been found to facilitate encoding of information and increase learner achievement.

Overtly rehearsing a fact, concept, rule, or principle is initiated by attending to the prompt to rehearse the information and completing the action required. Examples of overt rehearsal learning strategies are coloring an image, reordering sentence elements to create a sentence that is complete and correct, writing the answer to questions, playing a game, taking notes, completing a study guide, and completing or creating a concept map. Each of these strategies is a viable choice in instructional materials regardless of the media used for presentation.

With the changes in technology, instructional designers can implement learning strategy tools that will provide opportunities to learners to participate in rehearsal strategies that are more complex and require overt responses by answering constructed response, fill-in-the-blank questions, and other cued recall type activities. In a computer-based learning environment, Clariana and Lee (2001) found that learners giving an overt response to feedback to a knowledge generation question or to a discrimination question performed better than those learners that gave no overt response, those that rehearsed the information covertly or with a prompt to rehearse. In both knowledge generation and discrimination situations, overt rehearsal can increase to the learner's performance on assessments of the specific learning objectives measured in such as way that the use of the overt rehearsal strategy results in statistically significant differences (Cameron, 2004; Haag, 1995).

Covert rehearsal is an internal process that occurs within the information processing of the learner. Covert rehearsal can be prompted within the instruction using a rehearsal strategy to prompt the learner to think about the answer to a question, through review of an instructor generated concept map, or reading through and thinking about the relationships to prior knowledge introduced in an advance organizer. When a learner interacts with an instructional environment, with or without rehearsal strategies, the learner will use personal strategies to covertly rehearse the information. The extent to which the learner uses these personal strategies is based on individual differences. The learner will use these strategies while interacting with an instructional environment in addition to any treatment conditions that are included within the environment as independent variables in a research study or in a non-research study learning situation.

Overt and covert rehearsal strategies have been used within the Program of Systematic Evaluation to facilitate the learner's achievement of specific learning objectives. The rehearsal strategies that were employed as independent variables for the purpose of this study were classified as knowledge generation or discrimination rehearsal strategies. They also were classified by the use of covert or overt rehearsal strategies. Those strategies that are covert, personal learning strategies were employed in the control treatments of the primary research studies where a rehearsal strategy was not prompted. When the learner was expected to covertly rehearse within the instructional environment, this rehearsal was coded as prompted-covert rehearsal.

Summary of Rehearsal

Each of the rehearsal strategies used within phase three of the Program of Systematic Evaluation has been properly placed and positioned according to item analysis in a pilot study with a sample representing the study population. There were mixed outcomes for statistical significance in the studies using the various types of rehearsal strategies.

Knowledge Generation, Recall, Discrimination, and Recognition

The action and processing a learner must do in order to generate a response or discriminate among possible responses determines the classification, as a knowledge generation or as a discrimination task, of the action and response. The core aspect of knowledge generation is that the learner must generate a response that meets the definition of the correct response in the recall instruction. A discrimination response is the acceptance or rejection of a choice, a differentiation among possible correct responses (Brown, 1976). Recall requires both search and decision, discrimination requires only decision (Lockhart et al., 1976). From a cognition process perspective, the processing required to interact with information for knowledge generation is a more complex process requiring elaborative rehearsal than if the interaction is for recognition; recognition is much less cognitively demanding than recall (Smith & Ragan, 1999, p. 169).

Recall of information from a passage has been shown to increase for information that is hierarchically related (Meyer & McConkie, 1973), the elements of the text provide coherence through the emphasis of the main ideas within the text and deemphasis of the peripheral ideas within the text. The information is clustered together for the recall of the information through the hierarchy, the emphasis, within the text. This emphasis of the main ideas is similar to placement of a rehearsal strategy within the instruction to emphasize a fact, concept, rule, or principle being presented.

The learner cannot process all the information that is being presented within a prose instructional unit (Baddeley, 1992). By providing opportunities for the learner to interact with the information within the prose instructional unit, the learner rehearses parts of the information more elaboratively than others; the information is integrated with prior knowledge. These parts have been recalled at a statistically significant higher level after reading the prose instructional unit than the parts that were not rehearsed elaboratively (Benton, Glover, & Bruning, 1983; Frase, 1975). Millis and Just (1994) found the information that was signaled through the use of connectives (i.e., learning strategies that influence information processing) was recalled at a statistically higher level when the learner performed a cued recall assessment (e.g., Write the definition of the given word in the list). Signals within the instruction provide a structure for the learner and affect encoding of the information for improved recall for learners reading a prose instructional unit (Meyer & Poon, 2001).

In studies using rehearsal strategies to facilitate recall performance, the learner's performance has been positively influenced on assessments that include discrimination tasks, these differences are statistically significant (Clariana, 2003; Clariana & Lee, 2001). Other studies have resulted in outcomes that indicate the cues for retrieval of information vary in effectiveness as determined by the specific encoding operations performed on the input stimuli and what is required as an output; discrimination output statistically significantly exceeds recall of that same information (Shepard, 1967; Tulving & Thomson, 1973). The learners perform at a statistically significant higher rate on an objective measure exam if they use learning strategies that facilitate recall on a recall assessment than if they had used only learning strategies to facilitate discrimination of the correct response (Moscovitch & Craik, 1976). The depth of processing (the use of elaborative rehearsal and maintenance rehearsal strategies) of the information by the learner is influenced by the requirement of the assessment; the practice requirements for recall are different for different tasks (e.g., if the goal is for the recall to be verbatim or if it is to be paraphrased) (Lockhart et al., 1976; Tulving & Thomson, 1973).

Studies from cognitive, neuropsychological, and neuroimaging studies of human memory indicate that two distinct processes of memory, recollection and familiarity, compose recognition memory (Yonelinas, 2002, p. 441). Familiarity of a subject is the knowing that the learner has been introduced to the subject before yet can't recall any information other than the familiarization; the recall that something is familiar is fast and intuitive. Recollection is the ability to recall information related to the subject; the recall of recollection is slow and deliberate. Recollection and familiarity send signals to different parts of the brain (Yonelinas, 2002, p. 442). Elaborative rehearsal of the information leads to a large increase in recollection (Yonelinas, 2002, p. 457), thus increasing the level of discrimination. Those learners that interact with learning environments with elaborative rehearsal are expected to perform discrimination tasks at a higher level compared to those that did not experience the elaborative rehearsal.

The beneficial effects of cueing (i.e., giving a hint of what is important and necessary to attend to) are statistically significantly greater for integrated encoding (Moscovitch & Craik, 1976). Thus, learners that have interacted with an integrated encoding learning environment are expected to perform at higher achievement levels than if the assessment cues discrimination responses. Those learners that have interacted with learning environments with goals for recall learning strategies are expected to perform at higher achievement levels if the assessment cues recall or recognition responses.

In summary, empirical social science studies (Clariana, 2003; Clariana & Lee, 2001) and cognitive, neuropsychological, and neuroimaging studies of human memory (Yonelinas, 2002) indicate higher levels of achievement of performance that are statistically significant for learners interacting with recall/knowledge generation rehearsal strategies and elaborative rehearsal of the information being presented and answering assessment requiring discrimination tasks. Discrimination level learning strategies increase the learner's familiarity with the facts, concepts, rules, and principles; this familiarity is not enough to facilitate recall, it is only enough to facilitate recognition (Yonelinas, 2002).

Visualization Research in the Program of Systematic Evaluation

Visuals have been shown to be an effective learning strategy if they are placed in a redundant position within the instruction to be enhanced to cue the learner to the important information to be learned (Dwyer, 2003). The use of redundant visuals at various levels of realism have been studied within the Program of Systematic Evaluation to find that particular levels are appropriate for specific levels of prior knowledge of the learners: "students identified as possessing different levels of individual difference variables (e.g., intelligence, reading comprehension, and entering behavior) achieve differently from different types of visualization" (Dwyer, 1978, p. 227).

The detailed, shaded drawing, a visual style that has statistically significant positive effects on learning and has been the most effective in a representative sample of learners in high school and college (Baker & Dwyer, 2000; Dwyer, 1978, p. 227), was the same visual that was used by Lamberski and Dwyer (1981) in a study examining the effectiveness of colorized and black and white visuals to enhance the students' achievement of specific learning objectives. "In general, the simple line drawing (b&w and color) and the detailed, shaded drawings (b&w and color) were found to be consistently effective in reducing differences in achievement among students in the high, medium and low categories of individual difference variables" (Dwyer, 1978, p. 227). The studies selected from the Program of Systematic Evaluation to be included in this study that employed visuals in the instructional environment used either simple line drawings or detailed, shaded drawings in either black and white or colorized.

Media Presentation Style

The media presentation style (the format/medium used to present the instructional materials) is a decision that is made based on audience, facility, availability, cost, and content requirements. The medium is the physical means by which an instructional message is communicated. No particular medium is superior to all other types for all types of learning outcomes (Clark, 1983, 1994; Smith & Ragan, 1999, p. 286). The number of channels of communication can be a single channel, dual channel, or multimodal.

The method must be the appropriate technology for the instructional objectives. The amount of practice required for the learner to achieve the objectives and performance levels required upon completion and the learners' need for direction within the instructional unit are two of the prominent factors in the decision of medium choice. Examining the attributes of the instructional objectives and capabilities of the medium is the most effective way to determine which medium to choose for supporting the instructional materials (Smith & Ragan, 1999, p. 287). By programming the instructional environment to not advance to subsequent sections until the learner has completed or mastered particular learning objectives allows computer-based learning environments to potentially command more attention from the learner than other media presenting the same information. Screen-based instructions have been designed to have the learner observe the screen image of visuals and/or verbal cues with an audio presentation of the instructional unit. Paper-based instructional materials are the most common and least expensive to be developed.

The choice for the medium used for the research studies in the Program of Systematic Evaluation were based on the researcher's questions, the audience, the instructional objectives, the resources available, and the type of learning strategies placed within the instruction to facilitate the instructional objectives. The instructional media selected within the Program of Systematic Evaluation have been screen presentations with visual and/or verbal cues, paper-based materials in booklet form, and computer based, both online and housed on computer's hard drive.

Summary

As more cognitive theories were formulated and researched, designers adopted instructional choices to have the learner encode the information and process the facts, concepts, rules, and procedures. Rehearsal of the information is necessary for the learner to maintain the information within the working memory long enough to be encoded and stored within long-term memory. Learning strategies work within the working memory to help with visual and verbal coding in the visuospatial sketch pad and phonological loop subcomponents of working memory in an effort to dually code the information for storage.

The rehearsal strategies were designed to facilitate the interactions of the information presented in the instructional module with prior knowledge to increase the probability of retrieval of information from long-term memory. The recalled information was then encoded with the new information and/or was used to complete the discrimination and/or knowledge generation tasks within the instructional environment and the assessments of the specific learning objectives. The learning strategies were employed to increase the depth of processing and facilitate elaborative or maintenance rehearsal.

Advance organizers, questions, note-taking, and concept map generation are techniques that can be employed in screen, paper, and computer based instructional environments. Covert and overt rehearsal strategies can be employed with each learning strategy to facilitate achievement of the specific learning objectives. Covert rehearsal is performed mentally. Overt rehearsal requires a physical action when making the response.

Each rehearsal strategy can be defined and employ a knowledge generation or a discrimination task, depending upon the processing and action required by the learner. Tasks that evoke processing and generation of a response with only a cue to respond are knowledge generation tasks. An example of a knowledge generation task is: Describe what happens to the tricuspid valve when the heart is in the diastolic phase. The learner is formulating an answer for the knowledge generation task. Discrimination tasks require the acceptance or rejection of a choice. The learner must discriminate between possible responses. An example of a discrimination task is: Arrow number one (1) points to the (a) septum, (b) aorta, (c) pulmonary artery, (d) pulmonary vein.

Recollection and familiarity compose the recognition (discrimination) memory. To discriminate among responses to select a correct response, the learner interacts with information that is stored within long-term memory as recollections and familiar facts, concepts, rules, or principles. Elaborative rehearsal of both knowledge generation and discrimination learning strategies will facilitate greater retrieval of the information on knowledge generation and discrimination tasks, respectively. Knowledge generation learning strategies increase the learner's interaction with the information, increasing encoding, and increasing the performance of both recall and recognition assessment tasks.

Visuals have been used as a means for the learners to interact with the prose instructional units and the audio-presented instruction units in a second channel of communication. The visuals have been shown to be effective in facilitating increased achievement of specific learning objectives.

The medium of presentation in design decisions has been based on the researcher's questions, the audience, the instructional objectives, available resources, and the type of learning strategies placed within the instruction to facilitate the instructional objectives. Prior research for medium choice indicates that no one medium is best for all situations; the medium selected is based on designing the best instructional materials for the situation (i.e., objectives, cost, availability, resources).

This study summarizes the results of studies performed within the Program of Systematic Evaluation that have employed rehearsal strategies to facilitate knowledge generation and discrimination learning.

Chapter 3

Methodology

This chapter provides an historical overview of meta-analysis as a research method and describes the eleven steps, categorized into four aspects, to performing a meta-analytic study. The first aspect is the determination of the research question and criteria for evaluation of the criteria components within the research reports for inclusion within the meta-analysis. The second aspect is the bibliographic procedures required to identify and locate research reports that include the Program of Systematic Evaluation. The third is the coding of study characteristics for the studies determined to be included. The fourth, the statistical procedures required to analyze the data are specified and described.

Overview of Meta-Analysis

Meta-analysis, an analysis of analysis, began in 1904 with Pearson's combining of correlation coefficients (Pearson, 1904). Meta-analysis was popularized in 1976 through the work of Gene Glass examining the effect of psychotherapy (Glass, 1976, 2000). During the 1980s, more than 800 articles using and discussing meta-analysis had been published in the ERIC databases (Bangert-Drowns & Rudner, 1991). A search of the ERIC databases on December 23, 2005 shows that at least 1,884 articles have used or discussed meta-analysis. Currently, meta-analytic techniques are being adopted for research more frequently.

As a quantitative research method, meta-analysis is a statistical procedure used to integrate the findings across studies to reveal the simpler patterns of relationships that underlie research literature and to develop new theories based on the outcomes (Hunter & Schmidt, 2004, p. 17). Unlike ANOVA or ANCOVA, which compare means of experimental treatments, meta-analysis can be used to aggregate the results of multiple studies to determine an overall conclusion or summary of the combined studies. Using a common metric for all studies (e.g., effect sizes: the number of standard deviations difference between two treatment means), meta-analysis combines the results of multiple studies to calculate a single, mean population effect size estimate representing the change resulting from the independent variable employed in the experimental treatment or research environment. The primary goal of a meta-analytic study is to determine what has been learned from the results of other studies that have been conducted and discover what is yet to be learned (Rosenthal, 1998, p. 372). For examples of meta-analyses, see Kulik and Kulik (1988) and Baker and Dwyer (2000; 2005).

Meta-analysis is often described as a secondary research method, with its greatest potential use in research areas that are well established and have a large number of primary studies. Meta-analysis can integrate and aggregate these large numbers of studies, examine general trends and consistency of effects across the studies, investigate relationships not investigated within the original primary studies, and find trends too subtle to identify in traditional, experimental research.

The most popular approaches of meta-analysis are the meta-analytic procedures developed by Glass (1976), Hedges and Olkin (1985), and Hunter and Schmidt (1990; 2004). All three techniques are similar when applied to experimental studies in that each converts the results of the treatments within the primary studies to a common metric, an effect size, to be compared and analyzed. The difference in the techniques is how the statistical control for differences found within the primary studies is calculated and applied. The Glassian approach to meta-analysis computes and aggregates sample-weighted effect sizes without any statistical artifact correction. The Glassian approach is often referred to as bare-bones meta-analysis because the difference of means of the control and experimental treatments is divided by the standard deviation of the control treatment when calculating the effect sizes and there are no corrections for artifacts, only weighting of effects by sample size when calculating the population effect size estimate. Hedges and Olkin (1982) advanced the work of Glass by using a pooled deviation to correct for the experimental and the control treatments' variation within the original studies. The Hunter and Schmidt technique (1990, p. 45) (often referred to as validity generalization) corrects the summary statistics for inference of statistical artifacts (e.g., sampling error, measurement error, range variation) using each primary study's characteristics (i.e., sample sizes, means, standard deviations, and estimates of reliability).

Regardless of the approach used for performing meta-analysis, the general process for implementation includes the 11 steps (Arthur et al., 2001, p. 11) shown in Table 3-1. The 11 steps to performing meta-analysis are a systematic way to effectively and efficiently design a meta-analytic research study, determine the appropriate studies to include, and perform the necessary corrections and analysis to integrate the study outcomes, find relationships that underlie research literatures, and development new theories.

Table 3-1: Steps to be Performed within a Meta-Analysis

Step	Description
1. Topic selection—research domain definition	❑ Specify the research question of interest. ❑ Specify the experimental group and the control group variables and the postulated relationships between the variables.
2. Specification of inclusion criteria	❑ Clearly and specifically specify the criteria used to include and exclude primary studies.
3. Searching for and locating relevant studies to be included	❑ Perform electronic and manual searches for primary studies to be included.
4. Selecting the final set of studies to be included based on predetermined criteria	❑ Apply inclusion and exclusion criteria to studies located during the search.
5. Extracting the data, coding the study characteristics, calculating study effect sizes	❑ Extract the data—include independent and dependent variables, sample sizes, means, reliability of the measures, range restrictions, standard deviations, and study characteristics (i.e., author, type of study, year, etc.). ❑ Calculate the study effect sizes.

Step	Description
6. Determination of independence or non-independence of data points from the same sample and if the data points should remain separate or be aggregated	❑ Determine if effect sizes are from the same sample ❑ Determine if non-aggregation will artificially inflate the sample size. ❑ Determine if the aggregation of the effect sizes will obscure the effects of variables considered independent.
7. Testing for and detecting outliers	❑ Detect outliers within the data set since such outliers potentially affect the residual variability and a possible shift in the mean effect size. ❑ Use the sample-adjusted meta-analytic deviancy (SAMD) statistic (Huffcutt & Arthur, 1995) to detect outliers.
8. Data analysis	❑ Calculate mean effects, variability, and corrections for artifacts. ❑ Use statistics for the variation of the effect sizes, variance of sampling error and other artifacts, magnitude of the corrected effect size, chi-square test of homogeneity, and confidence intervals. ❑ Determine the homogeneity of the effect sizes.
9. Determining if searching for level-2 variables is necessary	❑ Determine if any level-2 variables exist. ❑ Identify level-2 variables.
10. Selecting and testing potential level-2 variables	❑ If any level-2 variables exist, determine the amount they moderate or alter the magnitude of the relationship.

Step	Description
11. Interpreting results and formulating conclusions	❑ Determine the estimated population effect size and its confidence interval. ❑ Report highest level interactions of level-2 variables.

Research Question Formation and Criteria for Including Studies

The completion of two prior meta-analytic studies (Baker & Dwyer, 2000, 2005) influenced the development of the research questions for this meta-analytic study. The evaluation of the effects of rehearsal strategies was selected. The research questions were formulated to examine the estimated population effect size for the effect of using rehearsal strategies in the Program of Systematic Evaluation. The experimental groups for the meta-analysis are treatments that employed rehearsal strategies as an independent variable in the primary study. The control groups for the meta-analysis are treatments within the same study as the experimental group that did not employ rehearsal strategies as an independent variable.

After formulating the research questions, determination for inclusion within the meta-analysis was initiated. The inclusion criteria were constructed from the core of the research questions, the effect of different types of rehearsal strategies. The rehearsal strategies for this meta-analysis are related to how the information is processed and rehearsed. A first strategy is overt rehearsal: the learner performs an observable action as part of the practice of the information for acquisition. A second strategy is performed covertly after being prompted to rehearse the information. The learner may read a question, think about a statement within an advance organizer as it relates to prior knowledge, or examine a study guide when rehearsing information after being prompted. A third strategy is when the instructional environment contains no prompts for rehearsal and the learner relies on a personal covert strategy. Every learner is performing a personal strategy when attending to instructional materials.

A meta-analytic assessment does not evaluate the quality of the studies to be included; it evaluates the study as meeting or not meeting the inclusion criteria. If the study meets the inclusion criteria, the data is collected from the study and effect sizes for the study are calculated. To calculate an effect size, d, the means, sample sizes, and standard deviations for the control and the experimental groups must be known. Additional information is recorded from the attributes of the studies. These attributes include reported reliabilities of the criterion measures, the use of visuals, the type of prompt used within the instruction for knowledge generation or discrimination levels of information processing for acquisition, the type of media presentation style for the instructional environment, and the names of the criterion measures used to measure the achievement of the specific learning objectives for the study. These attributes were used to correct the study effects and explain the variability of the study effects as determined within this meta-analytic study.

Corrections for error and bias in the original studies is managed through recording and correcting for sampling error, dichotomization and range variation of dependent and/or independent variables, deviation from construct validity, reporting errors, computation errors, transcription errors, and variance due to extraneous factors that affect the relationship (Hunter & Schmidt, 2004, 243). This meta-analytic study was free from many of these errors and biases since every study is part of the Program of Systematic Evaluation. The studies within the Program of Systematic Evaluation employed experimental designs with random assignment of subjects. The placement of the independent variables' strategies is determined through item analysis of the criterion measure outcomes in pilot studies in the respective study populations. There is strict control in each of the studies for construct validity, dichotomization and range variation of the dependent and independent variables, and variance due to extraneous factors that affect the relationship.

Distinguishing between variations in the population effect sizes is one of the most important tasks in a meta-analytic study. Sampling error is one component of the variance in the population effect sizes. The remaining variance is ascribed to local uniqueness of the studies with level-2 variables.

The use of level-2 variables in meta-analysis lets the researcher explore for interaction effects that cannot be seen during a single study or would require too many subjects to examine the effects and is not financially efficient. Recording the study characteristics related to inclusion and type of visuals, prompting styles for rehearsal, the media presentation style of the instructional materials, and the type of criterion measures used as dependent variables was required for analysis of the potential level-2 variables.

The criteria for including a study within this meta-analysis are:

1. The study must be from the Program of Systematic Evaluation. The instructional content and criterion measures employed in the study are those developed by Dwyer (1965).

2. The study must have included rehearsal strategies within the experimental instructional treatment and have all other independent variables held constant for the control treatment (e.g., if the experimental group in the effect size calculation was paper-based, then the control group in the calculation must also have been paper-based).
3. The study must contain at least two of the three rehearsal strategy techniques: covert, prompted-covert, and/or overt rehearsal.
4. The study must have used experimental designs with random assignment of subjects to the treatment conditions.
5. The study must report statistical data that can be used to calculate the effect size of the treatment: the means, sample sizes, and standard deviations for the control and the experimental groups.

Studies that do not meet these criteria were eliminated from inclusion in this meta-analysis. A review of research reports ($N > 350$) indicated that 50 independent studies met the criteria for inclusion within this meta-analysis. These 50 studies included 9,620 individual participants. An analysis of the data obtained from these studies yielded 640 independent effect sizes.

Bibliographic Procedures

Locating research reports that include the Program of Systematic Evaluation involved both computer and manual searches. All studies that are from the Program of Systematic Evaluation include references to Francis M. Dwyer's work beginning in 1965

(Dwyer, 1965). The initial search was to find all research reports (experimental research that is published and studies reported) that employ the instructional content, instructional tools, and criterion measures from the Program of Systematic Evaluation and/or cite related work done by Francis M. Dwyer. These research reports include masters papers, doctoral theses, conference proceedings, and published research articles.

Computer Search

Computer searching for published research reports related to the Program of Systematic Evaluation includes searching with the term “program of systematic evaluation” and the author/cited reference, Dwyer. The libraries of Penn State offer several electronic searching tools. The most productive and inclusive of the databases for research reports related to the Program of Systematic Evaluation are the Web of Science and the Social Science Abstracts.

The Web of Science is a citation database that includes bibliographies or reference lists to indexed articles. The article access includes the Science Citation Index expanded (1900 to present), the Social Sciences Citation Index (1956 to present), and the Arts & Humanities Citation Index (1975 to present). Information within the Web of Science database can be retrieved by author, subject term, journal title, or cited reference. The Social Science Abstracts are included in two databases: articles published from January 2001 through September 2005 and February 1983 through December 2000.

The suggested databases to search for the Program of Systematic Evaluation within WebSPIRS, the Silverplatter.com search tool, which includes the Social Science

Abstract databases, are Education Full Text 6/83-12/99, Information Science & Technology Abstracts Plus 1966-2004/12, The ERIC ISA Subset 1966-2004/12, SPORT Discus 1830-1996, Applied Science & Technology Abstracts 1/96-12/00, Applied Science & Technology Abstracts 10/83-12/96, Biological Abstracts 1999, Biological Abstracts 1983-1984, Biological Abstracts 1980-1982, and Education Full Text 1/00-3/05.

The databases suggested for the author and cited references that would contain “Dwyer” are Biological & Agricultural Index 7/83-12/00, Social Sciences Abstracts 2/83-12/00, Education Full Text 6/83-12/99, Humanities Abstracts 2/84-9/05, EconLit 1969-2002, SPORT Discus 1830-1996, Biological Abstracts 1987-1988, Biological & Agricultural Index Plus 1/01-9/05, Health and Psychosocial Instruments 1985-2005/09, Applied Science &Technology Abstracts 10/83-12/95, Biological Abstracts 1996, Biological Abstracts 1980-1982, Biological Abstracts 1998, Art Abstracts 9/84-9/05, Biological Abstracts 1985-1986, Biological Abstracts 1975-1979, Biological Abstracts 2001, FSTA 1990-2006/01, Biological Abstracts 2000, Biological Abstracts 2002, Biological Abstracts 1994, Biological Abstracts 1995, Biological Abstracts 1969-1974, Biological Abstracts 1999, Biological Abstracts 1992, Biological Abstracts 1983-1984, Biological Abstracts 2003, Biological Abstracts 1989, The Philosopher's Index 1940-2005/09, Biological Abstracts 2004, Biological Abstracts 1993, Biological Abstracts 1997, Biological Abstracts 2005/01-2005/10, Biological Abstracts 1991, SPORT Discus 1997-2005/10, EconLit 2003-2005/11, Applied Science &Technology Abstracts 1/01-9/05, Biological Abstracts 1990, Information Science & Technology Abstracts Plus 1966-2004/12, Social Sciences Abstracts 1/01-9/05, Education Full Text 1/00-3/05, The ERIC

ISA Subset 1966-2004/12, FSTA Retrospective 1969-1989, Applied Science & Technology Abstracts 1/96-12/00, Treasures from the Film Archives, Art Index Retrospective 1929-1984, Index Film Periodicals 1972-3/2005, Education Full Text 4/05-9/05, Index TV Periodicals 1979-3/2005, Mental Measurements Yearbook, and Fiaf Members Publications.

Dissertation Abstracts provides an additional searchable resource for dissertations that have been completed since 1861. The full-text document is available for dissertations completed since 1997. Searchable fields within include keywords, authors, title, school, subject, abstract, adviser, and degree. There are several research partnerships within the Program of Systematic Evaluation that have generated research reports and dissertations in other universities. Searching by adviser was most beneficial to find dissertations that have employed the Program of Systematic Evaluation.

Analysis to determine the appropriateness, correct author, and correct cited reference requires further searching within the results. An initial search to determine the suggested databases yielded 4,114 articles that contain “Dwyer” as an author or cited reference. Within the Social Sciences Abstracts are 228 articles that have “Dwyer” as an author or as a cited reference. Specification with additional criteria was necessary to focus the search to the appropriate articles for evaluation for inclusion. Not all articles that have used the Program of Systematic Evaluation are included within the databases. Articles submitted to journals that are not indexed will not be included. Research reports that are dissertations, masters papers, or conference proceedings required manual search techniques to be located.

Manual Search

In addition to finding dissertations, masters papers, and conference proceedings through manual searches, analysis of cited references within journal articles identified ERIC citations and conference presentations that were not found through computer search techniques. For example, several articles have been published over the years to discuss research related to the Program of Systematic Evaluation (Canelos, 1987; Dwyer, 1987; Dwyer, 2003). These articles identify the structure of research that has been performed and cite many of the research reports that have been completed using the Program of Systematic Evaluation. Communication with Dr. Francis Dwyer and review of his records were additional manual sources for names of researchers that have completed research using the instructional and evaluation materials inherent in the Program of Systematic Evalution.

Computer and manual searches were meant to be an exhaustive approach to find every research report related to the Program of Systematic Evaluation. An overview of the Program of Systematic Evaluation indicates that more than 250 scholarly publications, more than 350 scholarly presentations, and more than 100 quantitative doctoral dissertations have been completed using the experimental materials utilized in the Program of Systematic Evalution (Dwyer, 2004). In the time since the web site was last updated and accessed for reference, additional studies have been completed, presented, and published in the Program of Systematic Evaluation.

Study Coding

Defining the attributes of the studies to be included within the meta-analysis assisted in selecting studies that were appropriate for inclusion, provided relevant data to answer the research questions, and produced a population estimate that was generalizable to the general population rather than the populations studied within the Program of Systematic Evaluation. The major variables of interest were the type of rehearsal strategy and the attributes of the study related to visuals, prompting style, media presentation type, performance on the dependent measures, and type of criterion measures.

A single study must have contained at least two of the three designated rehearsal strategy techniques: covert, prompted-covert, and/or overt rehearsal. For calculation of effect sizes, either the covert or prompted-covert was the control group, with either prompted-covert or overt rehearsal as the experimental group. Table 3-2 depicts the experimental and control groups for inclusion in the numerator of the formula for calculating the effect sizes for the studies. These take the position of the M_E (mean of the experimental group) and M_C (mean of the control group) in the calculation of the study effect size.

Table 3-2: Comparisons for Effect Size Calculations

Experimental Group	Control Group
Overt rehearsal	Covert rehearsal (personal strategy)
Prompted-covert rehearsal	Covert rehearsal (personal strategy)
Overt rehearsal	Prompted-covert rehearsal

The means, standard deviation, and sample size of the experimental group and control group were entered into an Excel spreadsheet. The following columns in the spreadsheet contained the coding for characteristics for the experimental group and control group rehearsal strategies (overt, prompted-covert, or covert) and descriptions for the expected level-2 variables. This coding categorized the expected level-2 variables for visual type, prompting style for rehearsal of the information for acquisition, media presentation style, and criterion measure.

An outside agent was used as the second person to code studies. The results were used to verify the inclusion and coding decisions. The second rater was presented the inclusion criteria and the coding sheet for the studies. Strict adherence to the criteria for inclusion and communication of the coding scheme were most important. Disagreements were resolved by reviewing the definitions for the rehearsal techniques and expected level-2 variables then coming to consensus of placement and coding. These definitions are described in the next section, Conceptual Background for Coding the Studies, for the level-2 variables.

Conceptual Background for Coding the Studies

Rehearsal strategies within instruction are designed to facilitate the learner's knowledge acquisition through some type of rehearsal relating the information (i.e., facts, concepts, rules, and principles) to previously learned and retained information (prior knowledge) or by repeating the new information for the stimulus to move from short-term memory into long-term memory for later retrieval. Following the information processing model (figure 2-1), information to be learned activates the sensory register. The information that is perceived is passed through to the short-term memory. If rehearsal of the information takes place, this information is held for a longer while in the short-term memory. The longer the information is acted upon within short-term memory, the higher the probability that the information will be encoded, passed, and stored within long-term memory where it will be retained until needed and/or activated for interaction with additional information for recall, recognition, or additional learning.

Maintenance rehearsal works with the information processing model by the learner keeping the information within short-term memory with repetition of the information. Covert rehearsal strategies (personal strategies) and prompted-covert rehearsal strategies use a maintenance rehearsal technique by having additional repetition of the facts, concepts, rules, and principles to be learned. Overt rehearsal strategies are more involved and stimulate long-term memory and other processes to activate. This elaborative rehearsal strategy works with the information processing model by keeping the information within the short-term memory for even a longer time and with increased interaction with the content material by the learner creating and re-creating links to prior knowledge and storing the information in new or novel formats. The time factor and the interaction with prior knowledge increase the probability that the information will be linked to prior learning, a physical activity, or retained as new information.

Each of the identified, potential level-2 variables has a positive relationship with rehearsing of information. Instructional strategies that are placed in locations to provide redundancy to the facts, concepts, rules, and principles being presented facilitate an opportunity for rehearsal of the designated content that increases the likelihood that additional achievement and recall will occur. Strategies placed where the learner has already mastered the information, or where the information is easily rehearsed and passed to long-term memory, often result in lower achievement scores (Baker & Dwyer, 2000; Dwyer, 1984, 2003; Elliott & Dwyer, 1995).

Level-2 Variable Coding

The identified, potential level-2 variables for this meta-analysis included visual type, prompting style of rehearsal of the information for acquisition, media presentation style, and criterion measures. These variables were selected upon the completion of two pilot meta-analytic studies examining the effects of various types of visuals used within instruction on student achievement (Baker & Dwyer, 2000) and the effect of instructional strategies and individual differences on student achievement (Baker & Dwyer, 2005).

The categories for the types of feedback were examined in the study of the effect of instructional strategies and individual differences on student achievement (Baker & Dwyer, 2005). Within the regression calculations, feedback and prompting style variables produced nearly the same regression line, thereby they were collinear. This situation is known as multicollinearity. The calculations that included feedback as a level-2 variable resulted in multicollinearity and halted the progress of the hierarchical linear modeling analysis. In a multicollineaity situation, one of the variables may be removed from the equation to allow the calculations to identify the coefficients for the remaining variables without the additional variability due to the multicollinearity.

The coding of feedback (e.g., knowledge of correct response, knowledge of response) was removed from the list of level-2 variables due to multicollinearity with prompting style for rehearsal of the information for acquisition in the prior study. Feedback within the instructional environment in each of the studies did not occur unless there was a prompt to rehearse the information; feedback was placed in the instruction only after the learner was prompted to rehearse the information.

Categorization of the other level-2 variables (visual type, prompting style for rehearsal, media presentation style, criterion measure) was determined from the written descriptions within the methods sections of the research reports and from viewing the actual study treatments where available. Strict adherence was made to the definitions. Two raters reviewed the coding scheme for consensus on the categorization.

Visual Type

The initial studies completed within the Program of Systematic Evaluation determined the use of simple line drawings (b&w and colored) and shaded, detailed drawings (b&w and colored) were found to be consistently effective in reducing achievement differences among students of high, medium, and low categories of individual difference variables (Dwyer, 1978, p. 227). The detailed, shaded drawing of images (i.e., detailed black and white shaded and detailed color-coded images) were used within the studies completed by Lamberski (1972; 1980; Lamberski & Dwyer, 1981). The detailed, shaded drawings were found to be the most effective style of visual used for a combined population of high school students and college students in a prior meta-analysis by Baker and Dwyer (2000). Many of the subsequent studies within the Program of Systematic Evaluation used these color-coded and black and white detailed, shaded drawings and their placements within the instructional environments. The coding of visuals used within the studies for this meta-analysis include if the images are black and white, color-coded, or if no visuals were used in the treatments.

Prompting for Rehearsal of Information for Acquisition Style

Interaction with any instructional environment by a learner results in rehearsal of the information for acquisition. The type of rehearsal taking place is dependent upon the type of tasks the prompt for rehearsal within the instructional environment is prompting or it a prompt is employed. If no prompt is employed, the learner relies on a personal covert strategy to interact with the information in the instructional environment. The prompts for rehearsal can be labeled as prompts for a knowledge generation task or prompts for a discrimination task. The learner uses the rehearsal task within the instruction as a strategy to encode the information at a level that will facilitate later knowledge generation or later discrimination on the performance assessment or the objective measures.

Either technique of rehearsal of the information stimulates the learner to retrieve information from long-term memory. For the purposes of this study, if the information is retrieved and an answer is generated, then the task is categorized as a knowledge generation task. For the purposes of this study, if the information is retrieved for the learner to distinguishing among possible correct responses what is correct, the task is categorized as a discrimination task.

Media Presentation Style

The medium is the physical means by which an instructional message is communicated. The number of channels of communication can be a single channel, dual channel, or multimodal. The method must be the appropriate technology for the instructional objectives. The amount of practice required for the learner to achieve the objectives and performance levels required upon completion and the learners' need for direction within the instructional unit are two of the prominent factors in the decision of medium choice.

Considerable research has been completed for instructional development in various mediums including screen formats (e.g., slide projections with an audio presentation, television programming), paper and text based publications (e.g., textbooks, training booklets), and computer presentations (e.g., web based interactive environments, computer-based training environments) (Clark, 1983, 1994; Cobb, 1997). No particular medium is superior to all other types for all types of learning outcomes (Smith & Ragan, 1999, p. 286); each has its more beneficial situations where the media is more capable of efficiently delivering and facilitating the different learning outcomes to the target audience (Smith & Ragan, 1999, p. 287).

The coding for the media presentation style used within the studies included in this meta-analysis were categorized as screen presentation (slides with audio presentations, video based presentation), paper based materials (booklets), and computer based (web based and stand-alone computer based instructional environments).

Criterion Measures

The criterion measures within the Program of Systematic Evaluation are designed to evaluate learning related to facts, concepts, rules, principles, and comprehension. Each measure consists of 20 questions. These criterion measures are the drawing test, terminology test, identification test, and comprehension test. The average reliability coefficient of the criterion measures using the Kuder-Richardson Formula 20 was reported from a random sampling of studies from 1965 to 1978 as terminology test 0.83, identification test 0.81, drawing test 0.83, and comprehension test 0.77 (Dwyer, 1978, p. 47)

The learning objectives within the criterion measures represent a hierarchy of learning tasks. The instructional materials are designed to first teach the basic terminology associated with the content; identify the basic parts, structures, and relative positions; and display comprehension of the interrelated functions during phases. The criterion measures evaluate different levels of learning, facilitate different cognitive structures, and assess specific educational objectives (Dwyer, 1978, p. 45-46).

The terminology test evaluates the learner's knowledge of specific facts and definitions and is presented as a 20-question multiple choice test that has the learner distinguish the correct response from the distracters for the stem in each question. The identification test evaluates the learner's ability to identify parts or positions of an object and is presented as a 20-question multiple choice test that has the learner distinguish the correct response from the distracters for the stem in each question. The drawing test evaluates the learner's ability to reproduce the objects in their appropriate context and is presented as a 20-question cued recall test that has the learner draw a picture and label the components using the 20 names in the list as a cue as to what to identify in the picture. The comprehension test evaluates the learner's understanding of the complex procedures and/or processes and is presented as a 20-question multiple choice test that has the learner retrieve information about the rules and principles of the process after reading the stem of the question before selecting the correct response from the distracters.

The classification of the criterion measures for analysis in this study is dependent upon the type of test construction for the assessment. The drawing test is a cued recall test and is scored subjectively. The drawing test is classified as a performance assessment for this analysis. The terminology, identification, and comprehension tests are 20-question multiple choice tests. Each of these multiple choice tests are scored objectively. The terminology test, identification test, and the comprehension test are classified as objective measures for this analysis.

While the Program of Systematic Evaluation uses an instructional unit based upon the parts and functions of the human heart, other research using physics and science concepts have yielded the similar results for visualization use in instruction and employment of various rehearsal strategies (Dwyer, 2003, 2004; Mayer, 2001). .

Statistical Analysis

After identifying and coding the means, standard deviations, sample sizes, independent research attributes for the experimental treatments, and the outcomes of the dependent measures, the data was entered within an Excel spreadsheet, the study effect sizes were calculated, and level-2 variables were analyzed.

The common metric used for this meta-analysis is an effect size, *d*. The effect size metric is employed the most for research domains that have an experimental or quasi-experimental research design containing identifiable dependent measures for the experimental and the control groups. Evaluation of the effect sizes within a meta-analysis can disentangle the differences of the effects due to sampling error and those due to real level-2 variables.

This section presents the technique for performing meta-analysis with the effect size metric (Hunter & Schmidt, 2004, p. 277-279) and a random-effects model (Hunter & Schmidt, 2004, p. 201-203). By using a random effects model, the results of the meta-analytic study are more accurate than if a fixed-effect model was used and violated. Using a fixed-effects model inappropriately results in a confidence interval that is too narrow and all significance tests have Type I error biases.

A fixed-effects model assumes that the same population effect size underlies all the study effects. The variance of the population effect sizes is assumed to be zero. The fixed-effect model is a special case of the random effects model. With the random-effects model, there is also the assumption that there is real variability across the effect sizes. The variance of the population effect sizes is assumed to be equal to or greater than zero. This variability can be estimated. The corrections for artifacts are made to reduce this variance. After correcting for artifacts in a random-effects model, the examination of level-2 variables is completed to explain the variance not due to sampling error.

Calculating Effect Sizes

After collecting the relevant studies for the research questions, the effect size metric is calculated from the recorded means, sample sizes, and standard deviations. For post-test only experimental designs, the *d* statistic is calculated by finding the difference between the mean of the experimental group and the mean of the control group within the research study and dividing the difference by a measure of variation (formula 3.1) (Glass, 1976; Hunter & Schmidt, 1990, 2004). Hunter and Schmidt recommend a pooled, with-in group standard deviation (formula 3.2) since it has only about half of the sampling error as the standard deviation of the control group.

$$d = \frac{M_E - M_C}{S_W} \tag{3.1}$$

where:

d = effect size

M_E = mean of the experimental group

M_C = mean of the control group

S_W = pooled, within group standard deviation of measurement

To calculate a pooled, within group standard deviation of measurement:

$$S_W = \sqrt{\frac{(N_E - 1)S_E^2 + (N_C - 1)S_C^2}{N_E + N_C - 2}} \tag{3.2}$$

where:

S_W = pooled, within group standard deviation of measurement

N_E = number of participants in the experimental group

S_E = standard deviation of the experimental group

N_C = number of participants in the control group

S_C = standard deviation of the control group

The experimental design of this meta-analysis is the independent group design. This is the most common meta-analysis approach. The studies included within the meta-analysis use different subjects randomly assigned to experimental and control groups within each independent study.

Since each of the experimental and control groups are independent of the other groups within the study and independent of other study groups, the calculated effect sizes are independent of each other and will not be aggregated within the studies, only aggregated within the meta-analysis so as to maintain the identification for the level-2 variables. The independent variables within each study are examined to control for consistency and recording of potential level-2 variable categories.

All studies include the use of one or more of the criterion measures as dependent variables. A particular study could have the same research participants accounted for within as many as four effect sizes. Even though the same research subjects are recorded with the criterion measures and their respective effect sizes, these effects size calculations are independent due to the measurement of a different dependent variable. These effects must be separate for the level-2 variable analysis.

Calculation of the Population Effect Size

When calculating the sample-weighted mean effect size, $\bar{d}$, and its variance, *Var(d),* the studies are assigned weights based upon the sample size. The larger samples sizes are given a larger weight based upon the concept that studies with larger sample sizes have less sample error in the calculations than studies with smaller sample sizes. Therefore, the weighted mean effect becomes an accurate estimate of the true effect size of the population. The formulas are a summation of the individual effect sizes multiplied by the assigned weight divided by the summation of effect sizes (Hunter & Schmidt, 1990, 285; 2004, 287)

$$\bar{d} = \frac{\sum \omega_i d_i}{\sum \omega_i} \qquad (3.3)$$

where:

ω_i = the assigned weight

d_i = individual effect size

The variance accounted for by sampling error is calculated using $\bar{d}$, *d*, and the assigned weight. In a bare bones meta-analysis, the only artifact controlled for is sampling error. Variation due to other artifacts (e.g., error of measurement, strength of the treatment) has the effect of lowering the observed effect size. The effect size calculated in a bare bones meta-analysis is most likely an underestimate of the true effect size (Hunter & Schmidt, 1990, 291; 2004, 287). The estimated population effect size calculated in a bare-bones meta-analysis is a summary of the effect sizes that are being combined. The bare-bones estimated effect size is laden with errors to due noise and variance (standard error) in the study effect sizes.

$$Var(d) = \frac{\sum \omega_i \left[d_i - \bar{d} \right]^2}{\sum \omega_i} \tag{3.4}$$

where:

ω_i = the assigned weight

d_i = individual effect size

$\bar{d}$ = average effect size

This meta-analysis used the sample size and variance as the weight. In a bare-bones meat-analysis, weighting by sample size produces more accurate estimates of the population variance. Using hierarchical linear modeling weights by the sample size and the variance and produces more accurate estimates of the population variance and the estimated population effect size then the bare-bones technique. The accuracy of the population variance is critical to meta-analysis (Hunter & Schmidt, 2004, p. 202).

Corrections for outlier errors in the study effect sizes were made prior to the calculations for the population effect size, population variance, sampling error, and confidence intervals. The weighted mean effect sizes were calculated using PROC MEANS in SAS to demonstrate the differences of the results using the different techniques. The weighted mean effect size in hierarchical linear modeling was completed within the calculations for the level-1 model.

Correcting for Errors

Errors associated with artifacts within a meta-analysis include sampling error, error of measurement, dichotomization and range variation of dependent and/or independent variables, deviation from construct validity, reporting errors, computation errors, transcription errors, and variance due to extraneous factors that affect the relationship.

Dichotomization and range variation and deviation from construct validity were not errors addressed within this meta-analysis since each study was using a consistent core instructional unit and the same criterion measures. The variance due to extraneous factors error was managed through the randomization of the subjects into treatment groups.

Error of measurement of the dependent variables is corrected for at the study level by using the calculated reliabilities reported for the individual dependent variable measures in each study. Care must be taken with corrections for error of measurement due to an increase in sampling error variance. For this meta-analysis, corrections for error of measurement were not made because the correction for reliabilities would increase the amount of sampling error more than the correction would correct the error of measurement.

Transcription, reporting, and typing errors were managed through outlier analysis. Sampling error was controlled for at the population level by calculating the variance of the population effect size and calculating confidence intervals.

Correction for Transcription, Reporting, and Typing Errors

There is not a mathematical way to correct for errors associated with transcription, reporting, or typing within the reported data sets or even those that occur when entering of data into the computer system for analysis within the meta-analysis. A technique for finding potential errors related to these errors is the detection of outliers; the calculations related to transcription, reporting, or typing may result in an effect size that is an outlier to the other data. Even if these errors do not exist, detection of outliers within the data set will reduce residual variability and possible shifts in the population effect size calculations due to the outlier values.

One way to test and address these concerns is the sample-adjusted meta-analytic deviancy (SAMD) statistic, developed by Huffcutt and Arthur (1995) to account for sample size and outlying effect sizes. By using the SAMD statistic, the actual meta-analytic data set is corrected for the percentage of variance associated with any removed outlying study coefficients before the population effect size, its variance, and confidence intervals are calculated.

The SAMD technique takes into account the number of participants in each study to reduce the errors associated with analyzing for outliers by observing only the calculated effect sizes. The SAMD technique compares a study's effect size to the mean sample-weighted population effect size without the study included in the calculation then adjusts the value for the sample size of the study. Each study is assigned a SAMD statistic.

The formulas are as follows:

$$Var(i) = \frac{\left[4(N_i - 1)\left(1 + \frac{\bar{d}^2_{w/ostudyi}}{8}\right)\right]}{N_i(N_i - 3)} \tag{3.5}$$

where:

i = ith study in the meta-analysis

$\bar{d}_{w/ostudyi}$ = mean of all the studies in the meta-analysis without the value for the ith study

N_i = sample size for the ith study in the meta-analysis

$$Var(\overline{d}) = \frac{\left[4\left(\overline{N}-1\right)\left(\frac{1+\overline{d}^{2}_{w/ostudyi}}{8}\right)\right]}{\overline{N}\left(\overline{N}-3\right)K} \qquad (3.6)$$

where

$\overline{d}_{w/ostudyi}$ = mean of all the studies in the meta-analysis without the value for the *i*th study

$\overline{N}$ = average sample size of the studies in the meta-analysis

K = the number of studies used to compute the mean coefficient

$$SAMD(i) = \frac{d_{studyi} - \overline{d}_{w/ostudyi}}{\sqrt{Var(i) + Var(\overline{d})}} \qquad (3.7)$$

where:

i = *i*th study in the meta-analysis

d_{studyi} = effect of the *i*th study in the meta-analysis

$\overline{d}_{w/ostudyi}$ = mean of all the studies in the meta-analysis without the value for the *i*th study

After calculating the SAMD statistics, a scree plot of the values in rank order from highest to lowest is created. This plot is evaluated for outliers. Those values that rise above the flat gradual slope should be evaluated as potential outliers. If there are no values that are away from the slope to be evaluated, since the distribution of the SAMD statistics is approximately a *t* distribution, values greater than 3.0 could be flagged for analysis as potential outliers. Assessment of these potential outliers will reveal if the values are due to an error or due to some unusual study characteristic and if they should be removed from the meta-analysis.

Sampling Error Corrections

After calculating all of the effect sizes, *d*, for the experimental and control treatments, the variance of the population effect sizes are calculated to control for sampling error, confidence intervals around the population effect sizes are determined, and the chi square test of homogeneity of the effect sizes are evaluated. Corrections for artifacts are also performed to ensure that the estimate of the population effect size is more accurate.

Correcting for sampling error does not take place on an individual study basis; the correction is made on the study population effect sizes. This meta-analysis used a random effects model; it was assumed the sampling error variance of the effect sizes was randomly distributed. To calculate confidence intervals around the estimated population effect size, the variance of the effect sizes and the sampling error variance was determined.

Calculation of the variance of the effects is based on the number of participants in all the samples, the effect size for all studies, and the mean of the effect sizes. Formula 3.8 is the technique to calculate the variance of the effect sizes.

$$Var(d) = \frac{\sum\left(N_i * \left(d_i - \bar{d}\right)^2 \right)}{\sum N_i} \tag{3.8}$$

where:

N_i = number of participants in study *i*

d = effect size for study *i*

$\bar{d}$ = estimate for the population effect size

Sampling error variance is calculated to determine the confidence interval around the estimated population effect size (formula 3.9). For large samples, the first term, $\left[\frac{(N-1)}{(N-3)}\right]$, may be eliminated.

$$Var(e) = \left[\frac{(N-1)}{(N-3)}\right]\left[\left(\frac{4}{N}\right)\left(\frac{1+\bar{d}^2}{8}\right)\right] \tag{3.9}$$

where:

N = number of participants in all studies

$\bar{d}$ = estimate for the population effect size

The observed variance of the effect sizes is corrected for sampling error by subtracting the sampling error variance (formula 3.9) from the observed variance (formula 3.8). This result is the variance of the population effect sizes.

$$Var(\delta) = Var(d) - Var(e) \tag{3.10}$$

where:

$Var(\delta)$ = variance of the population effect size

$Var(d)$ = variance of the effect sizes

$Var(e)$ = variance of the sampling error

The square root of the variance of the population effect sizes yields the standard deviation of the study population effect sizes (formula 3.11).

$$SD_{\delta} = \sqrt{Var(\delta)} \tag{3.11}$$

where:

SD_{δ} = standard deviation of the study population effect sizes

$Var(\delta)$ = variance of the population effect size

If the population effect sizes are normally distributed, confidence intervals can be calculated. A 95% confidence interval for population effect size uses the standard deviation of the study population effects and the population effect size.

$$\delta - 1.96SD_{\delta} < \delta < \delta + 1.96SD_{\delta} \tag{3.12}$$

where:

SD_{δ} = standard deviation of the study population effect sizes

δ = population effect size estimate

Homogeneity and Level-2 Variables

Testing the effect sizes for homogeneity is a way to determine if the variation in the effect sizes is due to sampling error or a level-2 variable. If the chi square tests of homogeneity of the population effect sizes are not statistically significant, then the population effect sizes are accepted without further searching for level-2 variables. Because the chi square has low power to detect differences beyond sampling error, it is recommended to search for level-2 variables regardless of the outcome of the tests (Hedges & Pigott, 2001). After completing the tests for homogeneity for the population effect sizes, a search for level-2 variables was completed, even if the tests were not statistically significant.

For this meta-analysis, there were potential level-2 variables that had been identified based on prior research (Baker & Dwyer, 2005). The most important aspect of level-2 variables is that every study must record the condition (e.g., if types of visuals are a level-2 variable, every study must have data on the type of visuals used recorded with their respective data, not just one data set with combined information).

These level-2 variables were hierarchically assessed. For this meta-analysis, the level-2 variables are analyzed within hierarchical linear modeling (HLM 5.04) as level-2 variables in the level-2 conditional model. Determining the presence and effect of level-2 variables fulfills the meta-analyst's role of sorting out the variation across studies. The data structure for meta-analysis is not one-dimensional, thus HLM can be applied to handle the multi-level data analysis.

Hierarchical Linear Modeling

The outcomes of meta-analysis are measured through population effect sizes. The variance of the effect sizes is explained through correction of artifacts and through the analysis of level-2 variables. Using hierarchical linear modeling (HLM) to analyze the level-2 variables, the amount of the effect is distributed to the variable in a prediction equation (Bryk & Raudenbush, 1992; Raudenbush, Bryk, Cheong, & Congdon, 2000). A coefficient is calculated for each potential level-2 variable to determine its relationship to the population effect size being analyzed and the variability of the population effect size. In the calculation of the estimated population effect size, each effect size is weighted by its sample size and its variance (the correction is completed with the individual standard error term entered into HLM with the effect size and attribute coding).

HLM is an extension of simple linear regression in that it handles the estimation of variables at two or more levels. In the case of this meta-analysis, each population effect size is to be investigated by individual effect sizes (in the level-1 models) and by level-2 variable (in the level-2 models) because the effects are nested within the level-2 variables within each of the studies.

Employing the HLM approach, meta-analysis can be performed under the variance-known model (Bryk & Raudenbush, 1992, p. 155-174). This model assumes normality of the sample distribution of statistics (the study effects are normally distributed about the true population effect size) and sampling variance is known if there are enough studies. With the variance-known assumption, the Level–1 model is formula 3.13 for studies $j = 1, 2, \ldots, J$, where e_j is the sampling error associated with d_j as an estimate of δ_j and for which we assume $e_j \sim N(0, V_j)$

$$d_j = \delta_j + e_j \tag{3.13}$$

where:

d_j = the effect size for study j

δ_j = the estimated population effect size

e_j = error associated with study j

The Level–2 model, the true population effect size, δ_j, depends on study characteristics (level-2 variables) and a Level-2 random error. The true population effect size has the noise removed and accounts for sample size and variance in the individual effect sizes being analyzed.

$$\delta_j = \gamma_0 + \gamma_1[BW] + \gamma_2[COLOR] + \gamma_3[KNOWGEN]$$

$$+ \gamma_4[DISCRIM] + \gamma_5[PAPER] + \gamma_6[COMPT] + \gamma_7[OBJMSR] + u_j \quad (3.14)$$

where:

[BW] = black and white visual type predicting the effect sizes

[COLOR] = color visual type predicting the effect sizes

[KNOWGEN] = knowledge generation prompting style for rehearsal predicting the effect sizes

[DISCRIM] = discrimination prompting style for rehearsal predicting the effect sizes

[PAPER] = paper presentation style predicting the effect sizes

[COMPT] = computer presentation style predicting the effect sizes

[OBJMSR] = object measure type predicting the effect sizes

$\gamma_0, \ldots, \gamma_7$ = regression coefficients

u_j = Level-2 random error for which it is assumed $u_j \sim N(0,\tau)$

Putting the level-1 and level 2 models into a single mathematical formula (formula 3.15) creates the combined model. In the combined model, there are two sources of error from the level-1 and level-2 models: $e_j \sim N(0, V_j)$ and $u_j \sim N(0,\tau)$. Thus the single model for the observed d_j is:

$$d_j = \gamma_0 + \Sigma\, \gamma_S W_{Sj} + u_j + e_j \qquad (3.15)$$

$$Var(d_j) = \tau + V_j \qquad (3.16)$$

where:

d_j = the effect size for study j

γ_0 = regression coefficient

$\Sigma\, \gamma_S W_{Sj}$ = summation of level-2 variables predicting the effect sizes

u_j = Level-2 random error for which it is assumed $u_j \sim N(0,\tau)$

e_j = error associated with study j

The estimated true variance is the sum of the variance from the level-1 and level-2 models. Since the model assumes that V_j is known, τ is the value to be estimated. The reduction or non-reduction of the τ value for the population effect size being analyzed in HLM determines when the resulting equation is the best predictor of the population effect size.

Care must be given when using hierarchical linear modeling with meta-analytic data to determine if multicollinearity is a result. At times, level-2 variables may be so closely related to each other that multiple collinear lines are found during the HLM analysis. To determine the effect of the collinear variable, the analysis is run with one, then two, increasing the number of level-2 variables, in all combinations, used until collinearity is determined. Analysis of the correlations among the variables will also indicate the potential collinear problems (Arthur et al., 2001; Mertler & Vannatta, 2002; Tabachnick & Fidell, 2001).

Interpreting the Results

It is important in meta-analysis that interpretation of the single level-2 variables may be erroneous if there is an interaction between two or more level-2 variables. The rule for analysis of variance, if there is an interaction between two or more of the factors in the design, then interpretation of lower order main effects or interactions may be erroneous, applies to meta-analysis results also.

It is most important in the reporting and interpretation of the hierarchical presentation of meta-analysis data to consider large level-2 variables together and to focus the main interpretation on the final breakdown of the data with the level-2 interactions, the highest level of interaction, than on the individual confounded averages for each level-2 variable.

For the level-2 variables to be considered singly is to assume that the level-2 variables are independent and additive in their effects. Usually, a meta-analysis does not have enough studies to analyze the interactions of the level-2 variables. Using the selected contrasts, the 50 studies included within this meta-analysis had enough studies in each cell to analyze the interactions of the level-2 variables.

Summary

The analysis performed in this study was completed on research performed within the Program of Systematic Evaluation and employed the use of rehearsal strategies. Strict rules for inclusion had been determined based upon the research questions. There were 50 studies included within this study.

The 640 effect sizes calculated from the 50 studies were corrected for sampling error and outliers. The other corrections either add additional error to the analysis (i.e., error of measurement corrections) or were not relevant. Each of the studies employed the same criterion measures and the same core instructional unit. The only differences in the studies were the use of the identified potential level-2 variables as independent variables.

The analysis was performed using Excel 2003, MINITAB Release 14, SAS 9.1, SPSS 13, and HLM 5.04. Each program performed parts of the analysis where the functionality was most applicable. Initial effect sizes were calculated within Excel. MINITAB was used for graphical displays. SAS' PROC MEANS and SAMD technique determines the corrections for sampling error and outliers. SPSS was used to generate the *.dat files required for analysis of level-2 variables in HLM and to analyze the contrasts. HLM and SAS were used to analyze the level-2 variables.

Chapter 4

Results

This chapter reports the results of the meta-analysis of selected studies from the Program of Systematic Evaluation. The data collected from the 50 studies that met the criteria for inclusion in the meta-analysis generated 640 effects sizes and included 9,260 participants. All the studies were entered in the Microsoft Excel 2003, Minitab Release 14, SPSS version 13.0, HLM 5.04, and SAS version 9.1 for their respective analysis. Aspects of each of the most popular approaches were used for the analysis as necessary. The three most popular approaches are Glass (1976), Hedges and Olkin (1985), and Hunter and Schmidt (2004).

The chapter is organized according the to the completion of the following eleven steps to be performed within a meta-analytic study (Arthur et al., 2001, p. 11):

1. Select the topic and define the research domain;
2. Specify the inclusion criteria for the studies;
3. Search for and locate relevant studies to be included;
4. Select the final set of studies to be included based on the criteria;
5. Extract the data, code the study characteristics, and calculate study effects sizes;
6. Determine the independence and non-independence of data points;
7. Test for and detect outliers;

8. Calculate mean effect size, variability, and confidence intervals of effect sizes;
9. Analyze residual variability for need to search for level-2 variables;
10. Select and test potential level-2 variables;
11. Interpret results and formulate conclusions.

Topic Selection

The topic selected in this study was the relationship of knowledge generation and discrimination rehearsal strategy tasks and assessments as examined in studies where the learning environments employed visuals/no visuals in a medium that was selected for the study. The media presentation styles selected for each of the 50 studies, included in the meta-analysis, were one of the following: screen presentation, paper base, or computer (web) based. All the studies were part of the Program of Systematic Evaluation.

The studies within the Program of Systematic Evaluation have used the human information processing model to explain the interaction and processing happening cognitively for the learner to interact with the instructional environment for knowledge acquisition. The acquisition of knowledge is the eliciting of information from the instructional environment, including rehearsal strategies (e.g., visualization, embedded questions). By rehearsing the information through a personal rehearsal strategy or cued by a prompt to rehearse the information in a particular manner, the learner's sensory register is stimulated by the perception, the working memory facilitates the interaction of information from the perception and information that activated and retrieved from prior knowledge. The processing in working memory brings forth interaction with the information, encoding, and integration the information with the prior knowledge if rehearsed at knowledge generation level of processing or recollection if rehearsed at a discrimination level.

The research questions answered in this study were:

1. What is the size of the effect of rehearsal strategies used within the Program of Systematic Evaluation on student achievement of specific learning objectives?
2. How do the level-2 variables *[visualization (colorized, black and white, no visual), prompting type (knowledge generation tasks, discrimination tasks, no prompt), media presentation style (screen presentation, paper based, computer based), and criterion measure (objective measure, performance assessment)]* moderate (i.e., reduce the variance) the population effect size for rehearsal effectiveness?

Inclusion Criteria Specification

Before locating and evaluating studies, the following criteria for inclusion of a study in this analysis were specified.

1. The study must be from the Program of Systematic Evaluation. The instructional content and criterion measures employed in the study are those developed by Dwyer (1978).

2. The study must have included rehearsal strategies within the experimental instructional treatment and have all other independent variables held constant for the control treatment (e.g., if the experimental treatment included the use of a visual within the instruction, the control must also include a visual within the instruction).
3. The study must contain at least two of the three rehearsal strategy techniques: covert, prompted-covert, and/or overt rehearsal.
4. The study must have used experimental designs with random assignment of subjects to the treatment conditions.
5. The study must report statistical data that can be used to calculate the effect size of the treatment: the means, sample sizes, and standard deviations for the control and the experimental treatment groups.

Studies that did not meet these criteria were excluded from this study.

Relevant Study Search and Location

Examination of computer sources where “Dwyer” was entered as an author or cited entry, yielded 4,114 articles. Within the Social Sciences Abstracts were 228 articles that have “Dwyer” as an author or as a cited reference. An overview of the Program of Systematic Evaluation indicates that more than 250 scholarly publications, more than 350 scholarly presentations, and more than 100 quantitative doctoral dissertations have been completed using the Program of Systematic Evalution (Dwyer, 2004).

Application of the criteria for inclusion of rehearsal strategies was applied during the initial search. The search yielded 128 studies to be examined more closely for inclusion in this meta-analysis.

Studies Meeting the Inclusion Criteria

Sixty studies included rehearsal strategies as required in the criteria. Ten of these studies were not included due to missing data related to the necessary statistics for calculations in the analysis phase of the meta-analysis or they had treatments that did not include consistent independent variables for the study treatments selected as the experimental group and the control group. An example of this situation where there were not consistent independent variables is a study that had used a summary page for rehearsal in the treatments that included a visual and did not include a visual in the control treatment; there is not a way to discern if the differences were due to the visual in the treatments or the other components of the summary page in this meta-analysis. The final inclusion decisions yielded 50 independent research studies that met the criteria for inclusion in this meta-analysis. The number of study reports and publications exceeds 50 since many of the research studies for dissertations resulted in published articles in referred journals.

Data Extraction, Coding, and Study Effect Size Calculations

The data extraction was completed for the four categories of information about the study. These categories are Article Analysis, Design and Method Analysis, Descriptive Data of the Subjects and Study, and General Background Markers. A study coding sheet is displayed in Appendix A.

The coding of studies was completed by the researcher and a second rater. The researcher and the rater discussed any differences to the coding of the studies to come to consensus. No inter-rater reliability was measured because any differences on the 50 studies were discussed and agreement was reached. A table of the studies meeting the criteria for inclusion and their respective coding is shown in Appendix B.

The effect size calculations were completed using formulas 3-1 and 3-2. The effect sizes ranged from -1.87 to 3.32. An effect size indicates the number of standard deviations between the experimental group and the control group in the effect size calculation. The total number of effect sizes calculated from the 50 included studies was 640. The categorization and counts for each category are shown in Table 4-1.

Independence and Non-independence

The independence or non-independence of the studies was assessed. The dependent variables included in this meta-analysis are all from post-test measurement condition in experimental research design (Campbell, Stanley, & Gage, 1966). These dependent variables are independent within the studies. Each study indicated that the participants recruited had not participated in a previous study using the Program of

Systematic Evaluation materials. The participants are also independent. All effect sizes calculated were retained as individual effect sizes for analysis. The effect sizes were not measures of the same point. Each was a measure of a unique sample in the primary study.

Table 4-1: Numbers of Study Effect Sizes by Rehearsal Type, Media, Prompting for Rehearsal Type, and Visual Type

	Prompting for Rehearsal Type and Visual Type									
	Knowledge Generation			Discrimination			Hybrid[b]			Total
Rehearsal Type[a] and Media	b&w[f]	color[g]	none[h]	b&w	color	none	b&w	color	none	
I. Comprehension Test										
Overt – Covert[c]										
screen presentation	0	2	0	0	0	0	0	0	0	2
computer	7	0	0	7	9	0	0	0	0	23
paper	47	0	17	7	0	0	0	0	0	71
Overt – Prompted[d]										
screen presentation	0	2	0	0	0	0	0	0	0	2
computer	2	0	0	5	2	0	2	0	0	11
paper	12	0	1	2	0	0	13	0	6	34
Prompted – Covert[e]										
screen presentation	0	2	0	0	0	0	0	0	0	2
computer	1	0	0	6	0	0	0	0	0	7
paper	13	0	1	12	0	5	0	0	0	31
Comprehension Test Total	82	6	19	39	11	5	15	0	6	183
II. Drawing Test										
Overt – Covert										
screen presentation	1	2	0	0	0	0	0	0	0	3
computer	4	0	0	3	3	0	0	0	0	10
paper	34	0	9	5	0	0	0	0	0	48
Overt – Prompted										
screen presentation	0	2	0	0	0	0	1	0	0	3
computer	4	0	0	1	0	0	2	0	0	7
paper	9	0	1	0	0	0	5	0	4	19
Prompted – Covert										
screen presentation	0	2	0	1	0	0	0	0	0	3
computer	1	0	0	4	0	0	0	0	0	5
paper	8	0	2	4	0	4	0	0	0	18
Drawing Test Total	61	6	12	18	3	4	8	0	4	116

Rehearsal Type[a] and Media	Prompting for Rehearsal Type and Visual Type: Knowledge Generation			Discrimination			Hybrid[b]			Total
	b&w[f]	color[g]	none[h]	b&w	color	none	b&w	color	none	
III. Identification Test										
Overt – Covert										
screen presentation	1	2	0	0	0	0	0	0	0	3
computer	7	0	0	3	9	0	0	0	0	19
paper	44	0	18	7	0	0	0	0	0	69
Overt – Prompted										
screen presentation	0	2	0	0	0	0	1	0	0	3
computer	4	0	0	1	2	0	2	0	0	9
paper	11	0	1	2	0	0	13	0	4	31
Prompted – Covert										
screen presentation	0	2	0	1	0	0	0	0	0	3
computer	1	0	0	4	0	0	0	0	0	19
paper	12	0	2	12	0	1	0	0	0	27
Identification Test Total	80	6	21	30	11	1	16	0	4	169
IV. Terminology Test										
Overt – Covert										
screen presentation	1	2	0	0	0	0	0	0	0	3
computer	7	0	0	3	9	0	0	0	0	19
paper	44	0	17	7	0	0	0	0	0	68
Overt – Prompted										
screen presentation	0	2	0	0	0	0	1	0	0	3
computer	4	0	0	1	2	0	2	0	0	9
paper	11	0	1	2	0	0	13	0	6	33
Prompted – Covert										
screen presentation	0	2	0	1	0	0	0	0	0	3
computer	1	0	0	4	0	0	0	0	0	5
paper	11	0	1	12	0	5	0	0	0	29
Terminology Test Total	79	6	19	30	11	5	16	0	6	172
							Total Effects			640

[a]Rehearsal Type displays the types of rehearsal used for the experimental group and the control group, respectively, when calculating the study effect sizes.
[b]Hybrid is defined as the resulting study effects sizes when an overt rehearsal strategy used a knowledge generation strategy and the prompt for covert rehearsal used a discrimination strategy.
[c]Overt – Covert is defined as study effect sizes that were calculated where the experimental group used an overt rehearsal strategy and the control group used a personal covert rehearsal strategy.
[d]Overt – Prompted is defined as study effect sizes that were calculated where the experimental group used an overt rehearsal strategy and the control group used a prompted-covert rehearsal strategy.
[e]Prompted – Covert is defined as study effect sizes that were calculated where the experimental group used an prompted-covert rehearsal strategy and the control group used a personal covert rehearsal strategy.
[f]b&w is defined as visuals used within the instructional environment that were only black and white.
[g]color is defined as visuals used within the instructional environment that were colorized.
[h]none is defined as instructional environments that used no visuals within the instructional materials.

Outlier Testing and Detection

The sample-adjusted meta-analytic deviancy (SAMD) technique compares a study's effect size to the mean sample-weighted population effect size without the study included in the calculation then adjusts the value for the sample size of the study. Each study was assigned a SAMD statistic. The maximum SAMD value was 18.32. The minimum SAMD value was 0.00. Ninety-nine SAMD values were calculated to be higher than 3.00. Review of the studies for accuracy was completed. All data values were entered correctly. These 99 effect sizes were retained in the study to facilitate the analysis for level-2 variables. The SAMD values are shown in figure 4-1.

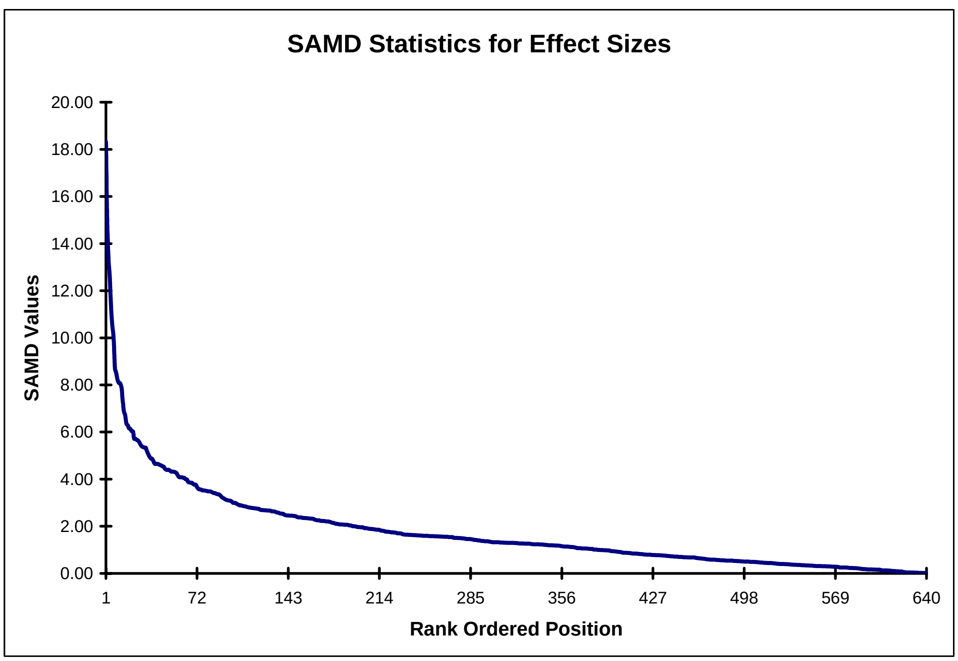

Figure 4-1: SAMD Statistics for Effect Sizes

Data Analysis of Effect Sizes and Level-2 Variables

The data analysis for the level-1 model using hierarchical linear modeling is to estimate the mean population effect size of the studies included within the meta-analysis (fixed effect model) and the variance of the effect sizes (random effect model). This level-1 model is the unconditional model. The resulting estimate for the mean population effect size is statistically inferred to estimate the parameter. Prior to calculating the fixed and random effects for the level-1 model, the hierarchical linear model program was used to generate the Empirical Bayes residuals. Adding the Empirical Bayes residuals to the study effect sizes calculated using formulas 3-1 and 3-2 results in stabilized effect sizes.

The estimates for the population effect size, γ_0 (fixed effect), and the residual variance, τ (random effect), are presented in Table 4-2 and Table 4-3. The null hypothesis that the population effect size is equal to zero was rejected (see Table 4-2). Examination of the residual variance of the population effect sizes in the unconditional level-1model indicated that the variance to not be statistically significant from zero (see Table 4-3).

The formula for the level-1 model, as shown as formula 3-13, is

$$d_j = \delta_j + e_j$$

where:

d_j = the effect size for study j

δ_j = the estimated population effect size

e_j = error associated with study *j*

Table 4-2: Results of HLM Analysis of Unconditional Model (fixed effect)

	Coefficient	Standard Error	*t*-ratio	*p*-value
Population effect size, δ_j[a]				
Stabilized effect size data	0.21	0.03	7.73	< 0.01

[a]δ_j is defined as the population effect size in the level-1 model and is equal to $\gamma_0 + u_j$.

Table 4-3: Results of HLM Analysis of Unconditional Model (random effect)

	Standard Deviation	Variance Component	*df*	χ^2	*p*-value
Population effect size, δ_j					
Stabilized effect size data	0.04	0.002	639	468.32	> 0.50

[a]δ_j is defined as the population effect size in the level-1 model and is equal to $\gamma_0 + u_j$.

Even though the null hypothesis that the stabilized effect sizes have variability equal to zero is not rejected, the conditional analysis was completed. The chi square test of homogeneity of the population effect size was not statistically significant (χ^2 = 486.32, degrees of freedom = 639, p > 0.50). Because the chi square has low power to detect differences beyond sampling error, it is recommended to search for level-2 variables regardless of the outcome of the test (Hedges & Pigott, 2001).

The data analysis for the level-2 model using hierarchical linear modeling was to formulate a model to predict the effect size with all the level-2 variables included in the model (fixed effect model) and to estimate the residual variance across studies (random effect model). There were nine level-2 variables entered into the level-2 hierarchical linear model. These 9 variables were coded into 14 dummy-coded variables using 1 and 0.

The coding scheme is as follows:

Variable	Code
Visuals	
Black and White	1 if in category; 0 otherwise
Color	1 if in category; 0 otherwise
No visual	Reference category
Prompt for Rehearsal	
Knowledge Generation	1 if in category; 0 otherwise
Discrimination	1 if in category; 0 otherwise
Hybrid	Reference category
Rehearsal types for Effect Size Calculation	
Overt – Covert	1 if in category; 0 otherwise
Overt – Prompted	1 if in category; 0 otherwise
Prompted – Covert	Reference category
Media Presentation Style	
Computer base	1 if in category; 0 otherwise
Paper base	1 if in category; 0 otherwise
Screen presentation	Reference category
Grouped Criterion Measures	
Objective Measures	1 if in category; 0 otherwise
Performance Assessment	Reference category

The first analysis with all nine level-2 variables included resulted in multicollinearity. Subsequent runs determined this to be due to the Rehearsal types for Effect Size Calculation category variables. These level-2 variables were dropped from the analysis. The results of the conditional analysis are shown in Table 4-4 and Table 4-5.

The level–2 model, the true population effect size, δ_j, depends on study characteristics (level-2 variables) and a level-2 random error. The formula for the level-2 model, as shown as formula 3-14, is

$$\delta_j = \gamma_0 + \gamma_1[BW] + \gamma_2[COLOR] + \gamma_3[KNOWGEN]$$
$$+ \gamma_4[DISCRIM] + \gamma_5[PAPER] + \gamma_6[COMPT] + \gamma_7[OBJMSR] + u_j$$

where:

[BW] = black and white visual type predicting the effect sizes

[COLOR] = color visual type predicting the effect sizes

[KNOWGEN] = knowledge generation prompting style for rehearsal predicting the effect sizes

[DISCRIM] = discrimination prompting style for rehearsal predicting the effect sizes

[PAPER] = paper presentation style predicting the effect sizes

[COMPT] = computer presentation style predicting the effect sizes

[OBJMSR] = object measure type predicting the effect sizes

$\gamma_0, \ldots, \gamma_7$ = regression coefficients

u_j = Level-2 random error for which it is assumed $u_j \sim N(0,\tau)$

Table 4-4: Results of HLM Analysis of Conditional Model (fixed effect)

	Coefficient	Standard Error	*t*-ratio	*p*-value
Intercept	-0.28	0.19	-1.48	0.14
Visuals				
Black and White	0.14	0.08	1.80	0.07
Color	0.16	0.14	1.11	0.27
No visual	rc[a]			
Prompt for Rehearsal				
Knowledge Generation	0.10	0.09	1.16	0.25
Discrimination	0.01	0.10	0.12	0.91
Hybrid	rc			
Media Presentation Style				
Computer base	0.38	0.15	2.47	0.01*
Paper base	0.22	0.16	1.39	0.17
Screen presentation	rc			
Grouped Criterion Measures				
Objective Measures	0.07	0.07	0.93	0.35
Performance Assessment	rc			

[a]rc is defined as the reference category.
* The p-value with * indicates that the variable is statistically significant ($p<.05$).

Table 4-5: Results of HLM Analysis of Conditional Model (random effect)

	Standard Deviation	Variance Component	*df*	χ^2	*p*-value
True Effect Size, δ_j	0.03	7.7×10^{-4}	632	451.88	> 0.50

The computer (web) base variable was statistically significant. Another level-2 model was analyzed with only the Media Presentation Style level-2 variables. This analysis was completed to determine if the residual variance in the level-2 (random effect model) would be reduced by including only the statistically significant variable and the other variables in the dummy coding scheme. Tables 4-6 and 4-7 show the fixed and random effect models.

Table 4-6: Results of HLM Analysis of Conditional Model (fixed effect) with Media Presentation Style Level-2 Variables only

	Coefficient	Standard Error	*t*-ratio	*p*-value
Intercept	2.1×10^{-3}	0.13	0.02	0.99
Media Presentation Mode				
Computer base	0.34	0.14	2.47	0.01*
Paper base	0.17	0.13	1.32	0.19
Screen presentation	rc[a]			

[a]rc is defined as the reference category.
* The p-value with * indicates that the variable is statistically significant (p<.05).

Table 4-7: Results of HLM Analysis of Conditional Model (random effect) with Media Presentation Style Level-2 Variables only

	Standard Deviation	Variance Component	*df*	χ^2	*p*-value
True Effect Size, δ_j	0.03	8.7×10^{-4}	637	458.58	> 0.50

The amount of variance in the hierarchical linear model with the Media Presentation Style level-2 variables only has a higher value of the residual variance. Therefore, the conditional model depicted in Table 4-4 and Table 4-5 is the best predictor of effect sizes for all situations other than those that contain the computer base presentation.

Results Interpretation and Conclusions

In summary of the results, the findings in this chapter are used to answer the two research questions.

The first research question: *What is the size of the effect of rehearsal strategies used within the Program of Systematic Evaluation on student achievement of specific learning objectives?* Analysis of the level-1 model to estimate the population effect size, δ_j, resulted in a rejection of the null hypothesis that the population effect size was equal to zero ($p < 0.01$). The estimate the population effect size, δ_j, in the level-1 analysis is 0.21.

The second research question: *How do the level-2 variables [visualization (colorized, black and white, no visual), prompting type (knowledge generation tasks, discrimination tasks, no prompt), media presentation style (screen presentation, paper-based, computer-based), and criterion measure (objective measure, performance assessment)] moderate the population effect size (i.e., reduce the variance for the population effect size) for rehearsal effectiveness?* The analysis of the level-2 variables resulted in the level-2 model with only the computer base indicating a statistically significant *t*-ratio (*p* =. 01). The subsequent level-2 model with computer base and paper base (both included due to dummy coding of the variable to maintain the proper reference category) indicated that the variable for computer base was still statistically significant

(p = .01). Therefore, the second level-2 model analyzed is the equation for predicting the best estimate of the population effect size for instructional environments employing the computer base medium for presentation. All other conditions are best predicted using the level-1 model. The results of the level-1 model are shown in Tables 4-1 and 4-2. The results of the level-2 models are shown in Table 4-4, Table 4-5, Table 4-6, and Table 4-7. The estimate the population effect size, δ_j, in the level-2 analysis is 0.34 when the instructional environment employs a computer-based medium for presentation.

In Chapter 5, the results presented in this chapter are discussed.

Chapter 5

Discussion

Rehearsal strategies are employed for basic learning tasks and for complex learning tasks requiring the learner to differentiate between the essential information that is to be remembered and supplemental information that is not fundamental yet is advantageous to know. If the rehearsal strategies are properly designed and positioned within the instructional environment to cue rehearsal of information determined to be most difficult for the learner to acquire, they enhance the acquisition and retention of the information (Dwyer & Dwyer, 1993). Research studies within the Program of Systematic Evaluation have employed instructional tactics and strategies as independent variables within the instructional environments to be instructionally efficient and effective and designed to facilitate learner achievement of the specific educational objectives.

Research was needed to examine the use of overt and covert rehearsal in combination with knowledge generation and discrimination tasks, instructional environment media, the use of visuals, and assessments of specific learning objectives to determine the most effective and efficient design decisions. This research employed meta-analytic techniques to analyze the studies completed within the Program of Systematic Evaluation that have used rehearsal strategies. The primary goal of this meta-analytic study was to determine what has been learned from the results of other studies that have been conducted and discover what is yet to be learned.

Statement of the Problem

The Program of Systematic Evaluation began in 1965 as a research series to examine the effects of visualization within instruction and to determine the most effective and efficient use of instructional strategies for knowledge acquisition of specific educational objectives (Dwyer, 1972, 1978, 2003, 2004). Rehearsal strategies designed to facilitate learner achievement were incorporated as an instructional strategy in studies within the Program of Systematic Evaluation from 1971 to 2006. The placement of the rehearsal strategies was identified through item analysis as the positions within the text that were most difficult for a representative group from the target population for the research. These positions were the legitimate positions for the placement of instructional strategies; these positions required supplemental resources and structures to direct the learner's attention to the important facts, concepts, rules, and/or principles within the instruction.

The techniques of interacting with a rehearsal strategy were covert (mentally) or overt (require a physical, observable action). Covert rehearsal strategies used within the studies (N=31 studies) selected for this study included embedded questions, advanced organizers, concept maps for review, study guides for review, and metaphors. Instructional strategies requiring the learner to write responses to questions, generate and complete concept maps, take notes, select the correct response for multiple choice questions, or color parts of the heart were overt rehearsal strategies used within the studies (N=40 studies) included in this study. Studies that examined the effects of the rehearsal strategy as an independent variable contained treatments that used no prompt for rehearsal (N=41 studies). Each of the covert and overt rehearsal strategies used either a discrimination (N=20 studies) or knowledge generation type (N=38 studies) task within the rehearsal strategy to cue the learner to the information to be acquired. The instructional environments of the studies were screen presentations (N=2 studies), paper-based booklets (N=33 studies), or computer or web-based (N=15 studies) environments. Visuals used were black and white (N=41 studies) or colorized (N=4 studies). Eight of the studies that used no visuals within the treatments (N=13 studies) also used black and white visuals within other treatments.

The goal of this research study was to examine the use of these knowledge generation and discrimination tasks as methods of overt and covert rehearsal as used within studies in the Program of Systematic Evaluation and determine if there were interactions of these rehearsal strategies with the employment of visuals, the medium chosen for presentation, and the type of test used to assess achievement of the specific educational objectives. The research questions were:

1. What is the size of the effect of rehearsal strategies used within the Program of Systematic Evaluation on student achievement of specific learning objectives?
2. How do the level-2 variables *[visualization (colorized, black and white, no visual), prompting type (knowledge generation tasks, discrimination tasks, no prompt), media presentation style (screen presentation, paper based, computer based), and criterion measure (objective measure, performance assessment)]* moderate (i.e., reduce the variance) the population effect size for rehearsal effectiveness?

Review of Methodology

Meta-analysis integrates and aggregates the multiple studies to examine general trends and consistency of effects across the studies, investigates relationships not investigated in the primary studies, and finds trends too subtle to identify in the traditional, experimental research. As a quantitative research method, meta-analysis is a statistical procedure used to synthesize the findings across studies to reveal the simpler patterns of relationships that underlie research literature and to develop new theories based on the outcomes (Hunter & Schmidt, 2004, p. 17).

This study employed hierarchical linear modeling as the technique to integrate and aggregate the multiple studies to examine the general trends and consistency of the 640 effect sizes from the 50 studies included and to investigate relationships that were not studied in the primary studies. The 640 effect sizes were first stabilized using Empirical Bayes residuals. The unconditional level-1 model (formula 3.13) weighted the stabilized effect sizes by their standard error, a weighting including both the sample size and the variance in the effect size. The conditional level-2 model (formula 3.14) evaluated the relationship of the predictors (level-2 variables) to the estimated mean population effect size.

The resulting estimate of the mean population effect size for a hierarchical linear model is a more accurate estimate of the true population effect size than using a bare-bones analysis. The bare-bones meta-analytic result is most likely an underestimate of the true population effect size and is a summary of the data within the study. The fixed effect model outcome in hierarchical linear model was generalizable to the population and the random effects model was used to analyze the variance of the residuals. If only a fixed effect model was used without the random effects model, the results would indicate a confidence interval that was too narrow and all significance tests would have Type I error biases.

Summary of Results

The general trend found within the integration and aggregation of the 640 effect sizes from the 50 studies examined from the Program of Systematic Evaluation is that the use of rehearsal strategies in all environmental conditions results in an estimated mean population effect size of 0.21. The consistency of the effect sizes is such that the variance of the residuals was not statistically significant from zero. The chi-square test of homogeneity of variance did not reject the null hypothesis that the effect sizes are homogeneous ($\chi^2 = 486.32$, $df = 639$, $p > .50$).

The examination of the relationships not investigated in the primary studies was performed using the conditional level-2 model in the hierarchical linear modeling analysis. The level-2 model was analyzed as recommended by Hedges and Pigott (2001) even though the level-1 model failed to reject the null hypothesis of homogeneous effect sizes. The relationships not investigated in the primary studies were the interactions of the type of rehearsal strategy used within the calculations of the effect sizes (covert rehearsal with no prompt, covert rehearsal with a prompt, overt rehearsal), the type of rehearsal task used (knowledge generation task or discrimination task), the medium of presentation of the instructional environment (screen presentation, paper-based materials, or computer (web) based materials), the use and type of visuals (black and white visuals, colorized visuals, or no visuals), and the type of test used to measure the achievement of the specific educational objectives (terminology test, identification test, comprehension test, or drawing test).

The first analysis using the conditional level-2 model resulted in multicollinearity. Multicollinearity means the categories for analysis are predicting the same phenomenon; the variables are redundant and correlated to each other. Within the analysis, removal of the variable or a re-categorization of the elements is recommended. Analysis of the type of rehearsal strategy used in the calculations of the effect sizes resulted in multicollinearity within the hierarchical linear model. This variable was removed from the level-2 model due to the relationship of the categorization of the rehearsal strategy type (covert, overt, prompted-covert) with the technique used to calculate the effect sizes; they were the same. Analysis of the type of criterion measure used to assess the achievement of the specific educational objectives (identification, terminology, comprehension, drawing) also resulted in multicollinearity. This variable was re-categorized by the type of test structure. The new categorization for the analysis for the criterion measure was categorized into two categories: objective measure or performance assessment. This re-categorization did not result in multicollinearity in the subsequent analysis.

The conditional level-2 model was analyzed with dummy coded variables for visuals, type of rehearsal task that was prompted, media presentation style, and test structure category for the criterion measures. The results of the conditional level-2 model reduced the residual variance of the unconditional level-1 model by 61.5% (from 0.002 in the level-1 random effect model to 7.7 X 10^{-4} in the level-2 random effect model). Of the dummy coded variables in the model, only the media presentation style, computer-based instructional environment, was statistically significant as a predictor of the estimated mean population effect size ($p = .01$). The model was reduced to include the dummy coded variables for media presentation style only as predictors of the estimated mean population effect size. The reduced conditional level-2 model included the paper-based and computer-based dummy coded variables to maintain the proper reference category of screen presentation. The computer-based variable remained statistically significant as a predictor of the estimated mean population effect size ($p = .01$) with an estimated mean population effect size of 0.34.

Discussion and Implications of the Analysis and Results

The multicollinearity of the rehearsal types and the test names and the lack of statistical significance of the rehearsal strategy task, the visuals, and the test types, when categorized as objective measures and performance assessment, were not expected. The level-2 variables were chosen based on prior meta-analytic research (Baker & Dwyer, 2005) and the research outcomes of the research included in the meta-analysis from the Program of Systematic Evaluation (the studies included in the meta-analysis are listed in the Bibliography of Studies Included in the Meta-Analysis).

The multicollinearity of the types of rehearsal strategy resulted from the relationship that existed between the technique for calculating the effect sizes from the studies and the labels given the categories. The effect sizes that used the mean of the treatment using the overt rehearsal strategy as the experimental group and the mean of the treatment using no prompt (the covert category) for the control group were categorized as overt–covert. The effect sizes that used the mean of the treatment using the covert rehearsal strategy that had a prompt to rehearse as the experimental group and the mean of the treatment using no prompt (the covert category) for the control group were categorized as promoted-covert–covert. The effect sizes that used the mean of the treatment using the overt rehearsal strategy as the experimental group and the mean of the treatment using the covert rehearsal strategy that had a prompt for the control group were categorized as overt–prompted-covert. The three categories were predicting the same linear equation and were collinear. Analysis was completed with the dummy coding for the three rehearsal strategy types with no other predictors. The multicollinearity of the analysis with only the rehearsal strategy type as the predictor confirmed the collinearity of the types with the outcomes. The statistical rule is to remove the variable from the analysis as a predictor and re-run the analysis with all other predictors.

The Program of Systematic Evaluation was designed to have a hierarchical structure of learning tasks from the simple (terminology) to the complex (comprehension) (Dwyer, 1978, p. 45). The beginning basic facts are presented prior to instruction of complex problem solving. The learning hierarchy is also measured in the criterion measures for the Program of Systematic Evaluation. Based on the learning hierarchy design, the leaner would need to possess competency in prerequisite or basic learning facts before being successfully introduced to the complex or abstract ideas (Dwyer, 1978, p. 44). Consequently, if a learner performs well on the comprehension test (measures complex understanding of functions and movements of the heart), the learner would, most likely, perform well on the terminology and identification tests (the prerequisite basic knowledge). The multicollinearity of the analysis resulted in the tests predicting the hierarchical linear model in a collinear manner. The structure of the tests was a second possibility for categorization. The tests were categorized as objective measures for the multiple choice tests (identification, terminology, and comprehension) and as performance assessment (Smith & Ragan, 1999, p. 99) for the drawing test due to the subjective measure of scoring and the open-ended task required. This analysis resulted in the tests not being statistically significant predictors with the objective measure and performance measure categorization.

The lack of statistical significance as predictors for the estimated mean population effect size by the types of rehearsal strategies, the use or non-use of visuals, and the type of test indicates that the variations in the effect sizes calculated from the studies are not related to the choice of independent and dependent variables in the primary studies. These independent and dependent variables were not included in the final model to predict the outcome of the estimated mean population effect size.

Interpretation of the Results

One of the aspects of meta-analysis is its ability to find trends too subtle to identify in the traditional, experimental research. The use of item analysis was not studied in any of the primary research studies as an independent variable; item analysis is a study condition. The results of this study indicated that the various types of rehearsal tasks, the type of rehearsal (overt or covert), the use and type of visual, and the type of assessment used to measure achievement are not the key components of the statistically significant increase the achievement scores of learners interacting with rehearsal strategies within the instructional environment. The key component was the common study condition of placement of the rehearsal strategies in locations that have been identified through item analysis for item difficulty as those facts, concepts, rules, and/or principles that were most difficult for the learner's to acquire and respond to on the criterion measures. Neither the type of rehearsal strategy or task used, nor the test type, influenced the positive outcome as much as the proper position of the rehearsal strategy to direct the learner's attention to the information that has been found to be the most difficult to acquire.

The unconditional level-1 model resulted in an estimated mean population effect size of 0.21. The unconditional model verified that the use of properly positioned rehearsal strategies results in a positive increase in performance in all conditions of the level-2 variables (the type of rehearsal task, use and type of visuals, media presentation style, and type of criterion measure). A learner interacting with a rehearsal strategy that is properly positioned can expect to increase performance on assessments measuring specific educational objectives by 8.3%. A learner scoring 50% without the use of any rehearsal strategies would improve, with 95% confidence, to a score in the range of 56.0% to 60.6% after interacting with properly positioned rehearsal strategies.

The statistically significant predictor in the conditional level-2 model, computer-based medium presentation style, is interpreted as instructional environments that employ a computer-based presentation medium and properly placed rehearsal strategies expect a 0.34 effect change on learner achievement of assessments measuring specific educational objectives. This 0.34 effect change is the same as increasing the learner's performance on an assessment measuring specific educational objects by 13.3%. The 95% confidence interval for the learner's performance on the assessment has a lower bound of 2.6% and an upper bound of 23.1%. Hence, for a learner that would score 50% on the assessment, interaction with properly positioned rehearsal strategies in a computer-based learning environment would increase the learner's score, with 95% confidence, to a score between 52.6% and 73.1%.

Discussion of Expectations and Relationships

Based on the outcomes of prior research, the expectations of this research were that there would be relationships found among the level-2 variables. Relationships to the changes in magnitude of the effect sizes were expected to be related to change in the categorization for the level-2 variables: the knowledge generation rehearsal task and discrimination tasks used within the rehearsal strategy, the overt and covert prompting of rehearsal, the use and type of visuals within instruction, the medium used for presentation, and the type of test used to measure the specific educational objectives.

Prior research of knowledge generation and discrimination tasks has led to statistically significant higher levels of achievement when used as rehearsal strategies and assessed with multiple-choice tests (Clariana, 2003; Clariana & Lee, 2001). The expectation was that knowledge generation tasks would be positively related to changes in effect sizes for the performance assessment (drawing test), which is a knowledge generation activity. The relationship of knowledge generation and discrimination tasks to the change in effect sizes for the performance assessment and objective measures was the same as a relationship due to chance. The outcome of the study was that properly placed knowledge generation and discrimination tasks will increase learner achievement equally well.

Overt rehearsal has resulted in statistically significant positive differences for learners as compared to no prompt for rehearsal or to prompted covert rehearsal (Dwyer & Dwyer, 1987, 1993; Taricani, 2002). Prompting for covert rehearsal has resulted in statistically significant higher achievement compared to no prompt for rehearsal (Chen, 2002; Irizarry, 1990). The change in effect sizes was expected to be related to the use of overt rehearsal strategies and the use of prompts in such as manner that the use of overt rehearsal would predict the higher effect sizes; however, that was not the case. The employment of overt or prompted covert rehearsal within the treatment as an independent variable in the study did not predict the effect size.

The use of visuals has been shown to be an effective learning strategy when placed in redundant positions within the instructional environment resulting in statistically significant positive effects on knowledge acquisition (Dwyer, 1978, p. 227). Color and black and white visuals have been found to be equally effective in reducing differences of individual difference variables (Dwyer, 1978, p. 227) and visuals have been found to enhance learner’s achievement of specific learning objectives compared to no visuals (Baker & Dwyer, 2000). The expectation was that instructional environments with visuals would supplement the use of rehearsal strategies to direct the learner’s attention to the important aspects of the instructional module. The outcome of the study was that the properly positioned rehearsal strategies facilitate the direction of attention of the learner equally well in the presence of visuals, regardless of black and white or color, as with no visuals within the instructional environment.

The medium used for presentation has been researched to find that no definitive outcomes and generalizations can be made for superiority of one medium as compared to another (Clark, 1983, 1994). The type of medium that is most appropriate is based on objectives, cost, availability, facility requirements, and the audience. The most effective and efficient medium is based on cognitive efficiency for the learner, how well the learner may interact with the information and acquire the key facts, concepts, rules, and principles being presented (Cobb, 1997). Due to cognitive efficiency, learners interacting with the information in the computer-based environment may have found the opportunity to interact with the properly positioned rehearsal strategies in a manner that facilitated an increase in achievement on all the criterion measures as compared to no rehearsal if using overt or prompted covert rehearsal or as compared to prompted covert rehearsal if using overt rehearsal. The cognitive efficiency was enhanced for the properly positioned rehearsal strategies in the computer-based medium.

In summary, properly positioned rehearsal strategies facilitated an increase in student achievement of specific learning objectives in all conditions and for all types of assessments examined in this study. The increase in achievement was statistically significant with a magnitude of 0.21 for all conditions, and 0.34 for computer-based instructional environments.

Comparison of Meta-Analytic Techniques

A bare bones analysis is limited to evaluating the results that are included in the study. The data is laden with noise and error. The result is generalizable only to the population being studied. A bare bones analysis most likely underestimates the true population effect size. The mean population effect size estimated in a hierarchical linear modeling analysis is generalizable to the population of well designed studies and their populations that have been and could be used to examine the effects of rehearsal strategies within an instructional environment.

In this study, the estimated mean population effect size for all conditions was 0.21 when using hierarchical linear modeling with a confidence interval from 0.15 to 0.27. Using bare bones analysis techniques, the overall mean effect size was 0.18. This value was an underestimate of the estimated population effect size found with hierarchical linear modeling. The 95% confidence interval for the barebones analysis had a lower limit of -0.71 and an upper limit of 1.06. The confidence interval for the bare bones analysis was wider than the confidence interval for the result of the hierarchical linear modeling analysis.

Results of a bare bones analysis for the level-2 variables in the study are shown in Table 5-1. The statistical significance of the chi-square tests for homogeneity indicated that the null hypothesis was rejected and the alternative hypothesis, the effect sizes were not homogeneous, was accepted. The lack of homogeneity implied that another factor was influencing the results and acceptance of the sample weighted mean effect size for each of the variables was errant. These sample weighted mean effect sizes contain noise due to sampling error and variance. Any comparisons of differences made with these outcomes in the bare bones analysis are not statistically significant.

The hierarchical linear modeling using the conditional level-2 model indicated that the only statistically significant predictor was the computer-based media presentation style. The estimated mean population effect size for computer-based instructional environments was estimated to be 0.34 with a lower confident limit of 0.07 and an upper limit of 0.61 for the 95% confidence interval. The bare bones analysis of the sample weighted effect size for computer-based instruction resulted in a chi-square test of homogeneity that was statistically significant indicating the 129 effect sizes used to calculate the sample weighted mean effect size for computer base media presentation style were not homogeneous. To accept the sample weighted mean effect size and 95% confidence intervals is to most likely make a Type I error.

Table 5-1: Results of Glassian Meta-Analysis where Errors are not Corrected

Variable	Total number of data points (k)	Sampled weighted mean (d)	Corrected SD	% variance due to sampling error	95% Confidence Interval L	95% Confidence Interval U	χ^2
Overall	640	0.18	0.45	20.97	-0.71	1.06	3051.64*
Visuals							
Black and White	474	0.19	0.48	18.54	-0.75	1.13	2556.52*
Color	60	0.22	0	139.48[f]	0.22	0.22	43.01
No visual	106	0.08	0.42	24.48	-0.75	0.91	432.94
All visual types	534	0.19	0.46	20.55	-0.70	1.09	2598.21*
Prompt for Rehearsal[a]							
Knowledge Generation	397	0.21	0.47	19.62	-0.70	1.13	2028.13*
Discrimination	168	0.12	0.39	25.57	-0.65	0.90	657.07
Hybrid[b]	75	0.08	0.49	21.87	-0.88	1.04	342.98
Media Presentation Style							
Computer base	129	0.32	0.57	11.43	-0.81	1.45	1128.35*
Paper base	478	0.13	0.39	27.90	-0.62	0.89	1713.24*
Screen presentation	33	-0.06	0.43	31.21	-0.91	0.78	105.73
Rehearsal types for Effect Size Calculation							
Overt – Covert[c]	338	0.19	0.47	18.39	-0.72	1.11	1838.24*
Overt – Prompted[d]	164	0.14	0.50	19.38	-0.83	1.11	846.33
Prompted – Covert[e]	138	0.17	0.33	38.18	-0.47	0.81	361.48
Criterion Measures							
Drawing	116	0.11	0.32	33.77	-0.51	0.74	343.49
Identification	169	0.17	0.56	14.80	-0.92	1.26	1141.69*
Terminology	172	0.19	0.44	21.80	-0.68	1.06	788.98*
Comprehension	183	0.21	0.42	23.99	-0.62	1.04	762.88*
Grouped Criterion Measures							
Objective Measures	524	0.19	0.47	19.44	-0.75	1.13	2695.18*
Performance Assessment	116	0.11	0.32	33.77	-0.51	0.74	343.49

[a]Rehearsal Type displays the types of rehearsal used for the experimental group and the control group, respectively, when calculating the study effect sizes.
[b]Hybrid is defined as the resulting study effects sizes when an overt rehearsal strategy used a knowledge generation strategy and the prompt for covert rehearsal used a discrimination strategy.
[c]Overt – Covert is defined as study effect sizes that were calculated where the experimental group used an overt rehearsal strategy and the control group used a personal covert rehearsal strategy.
[d]Overt – Prompted is defined as study effect sizes that were calculated where the experimental group used an overt rehearsal strategy and the control group used a prompted-covert rehearsal strategy.
[e]Prompted - Covert is defined as study effect sizes that were calculated where the experimental group used an prompted-covert rehearsal strategy and the control group used a personal covert rehearsal strategy.
[f]This value is the correct output. The corrected standard deviation for the effect size is reported as 0 if the variance is calculated as less than one. The calculations for the percent variance due to sampling uses negative variance within the calculation resulting in greater than 100% of the variance due to sampling error.
* A χ^2 value with a * indicates that the variable is statistically significant for non-homogeneity of the effect sizes.

Recommendations for Instructional Designers

The goal of this study was to provide guidance to instructional designers when designing and developing instructional environments to promote increased mastery of the learning objectives. Decisions for the design of the instructional environment are based on cost, availability, resources, facility requirements, objectives, and the audience. The design of instruction is expensive. Adding rehearsal strategies increases the development expense. The expense ranges from inexpensive attention gaining devices to the very expensive adaptive question, answer evaluation, and feedback systems. If two techniques are equally effective, the choice is generally based on cost; choose the one of the techniques which is least expensive.

The systematic design of instruction begins with an analysis to answer the questions of “Where are we going?”, “How do we get there?”, and “How will we know when we arrived?” (Smith & Ragan, 1999, p. 7). The analysis has three parts: learners, learning context, and learning task. When determining the instructional goals and objectives, the instructional designer is developing way to assess the goals and objectives to answer the question of “How do we know when we arrived?”. Using this assessment as a tool to identify the locations within the instruction that the learners have the most difficulty allows the instructional designer to place the relatively expensive rehearsal strategies in positions where they will be most efficient and effective. By placing rehearsal strategies in only these positions, there will not be any rehearsal strategies in locations where the learners can interact and acquire the information using their own personal strategies.

Item analysis is completed by examining the item difficulty. The proportion of the number of respondents that answered the question correctly is examined. Those items that have the lowest score for a valid test are the more difficult items. Within the Program of Systematic Evalution, the researchers placed the instructional strategies within the instructional materials in the positions where the objective, tested by the difficult item, was presented.

In a cost saving measure, placing the rehearsal strategies in positions that are identified as the most difficult through item analysis, the instructional designer saves both time and money for development and for the learner. The designer's decision of location to place the rehearsal strategy is determined by the analysis; this saves time for making the decision. Needless rehearsal strategies are not used nor are they placed in positions where the learners will not require their use or will find them to be a distraction.

The properly positioned rehearsal strategy, regardless of type of task or type of rehearsal, results in an 8.3% increase for all types of educational objectives in paper-based or screen presentations (lecture style presentations). Properly positioned rehearsal strategies within a computer-based environment yield an increase of 13.3% in achievement for all types of educational objectives. Performance of item analysis to properly place the rehearsal strategy reduces development expense and increases the probability of learners acquiring the information within the learning objectives.

Limitations of the Study

This study was a synthesis of studies completed in the Program of Systematic Evaluation. These studies have used the same technique of item analysis to place the treatments within the instructional environments as required by the pilot tested members of the target audience. The target audiences have ranged from high school students, college undergraduates, and adult learners in the United States and several international settings. Each study reports within the limitations of the study that the results are generalizable to the target population. This study is generalizable beyond those target populations as a result of employing hierarchical linear modeling as the analysis technique for the 50 studies included in the meta-analysis.

Additional limitations exist to the unequal representation of each of the 108 combinations of the level-2 variables. Many cells had low numbers of effect sizes or none at all. For example, this meta-analysis included no paper based instructional materials that used color images. All images used in the paper base instructional materials were black and white. The hybrid category was only available for the condition of an experimental group with an overt rehearsal strategy employing a knowledge generation rehearsal task and a control group with a prompt for covert rehearsal that employed a discrimination rehearsal strategy. Table 4-1 shows the counts of effect sizes in each category. The lack of representation in the cells causes the outcomes of the hierarchical linear modeling to not calculate the predictor as a significant predictor.

The data available is secondary data. There is no record of the data that was analyzed within the studies. Access to the primary data would permit analysis of more advanced relationships within the outcomes of the dependent measures. The means, standard deviations, and subject counts do not facilitate the advanced levels of analysis for the advanced relationships. There is no statistical correction available for errors associated with reporting or transcription. Without the primary data, there is no technique to confirm the accuracy of the reported data in the research reports and publications.

Suggestions for Future Research

Based on the implications and limitations of this study, two recommendations for future research are made to investigate the outcomes of the level-2 variables on achievement of specific educational objectives when rehearsal strategies are employed.

The first recommendation is to extend the meta-analysis beyond the Program of Systematic Evaluation to examine the effect of rehearsal strategies within other instructional environments and within other techniques of placing the rehearsal strategies. This extension of the meta-analysis beyond the Program of Systematic Evaluation would have a greater representation in the various cells indicated in Table 4-1. This meta-analysis included no studies that had used a paper base media presentation style with color visuals. No conclusions can be determined from this study relative to the effect size of color visuals in a paper-based presentation. The analysis of color visuals in paper base instructional materials compared to black and white as a partner to the rehearsal strategy should be investigated.

A second recommendation is to implement the advancements in technology to facilitate the use of knowledge generation rehearsal tasks and achievement of specific educational objectives. An example of these advancements is latent semantic analysis to analyze the results of constructed response questions in a computer-based instructional environment (Wolfe & Goldman, 2003). This new interactive electronic technology has the potential for increased utilization of rehearsal strategies leading to more creative and effective instructional environments. Used in an instructional environment, latent semantic analysis (the technology that is used by search engines in the Internet), evaluates the learner's responses for correctness, predicts the learner's reasoning for a subject based on analysis of the words and groups of words within the response , and links the learner to additional resources. Developed to meet the needs of the learner and to address the individual differences of the learners, additional resources are built within the instructional materials and are accessed based on the scoring of the learner's responses given after prompts to rehearse the information by responding to a constructed response question in a computer based instructional environment (Wolfe & Goldman, 2003). These supplements in the instructional environment can be more challenging activities for those that have mastered the instruction or have high prior knowledge. The added information may provide foundational information for increased understanding of the facts, concepts, rules, and principles for those learners that need additional information due to low prior knowledge or lack of experience with the instructional objectives. The outcomes of the latent semantic analysis individualize the instructional experience for the learner. Future research exploring the effects and viability as an instructional strategy should be explored.

Summary

Textbook design decisions that have questions in the margins, review questions at the end of the module, practice questions at the end of the chapter, and higher-order questions at the end of the unit have all been influenced by research results (Heinich, Molenda, & Russell, 1989). The use of visualization as an instructional tool has been widely researched in environments including newspapers, educational materials, multimedia, web-based environments, and even virtual reality (Griffin, Gibbs, & Weigmann, 1999). The Program of Systematic Evaluation has more than 700 scholarly publications and research reports examining the effects of instructional strategies with visualization (Dwyer, 2004). This study is another step in these series of research.

The results of this study indicate that the use of properly placed visuals and knowledge generation and discrimination rehearsal tasks within paper base and screen presentation media presentation formats that the expected increase in achievement of 0.21σ or 8.3% increase in achievement. If these same instructional strategies are employed in a computer base media presentation, the expected increase in achievement is $0.34\ \sigma$ or 13.3% increase in achievement.

Instructional designers are recommended to perform item analysis to identify the locations to place rehearsal strategies within the instructional environments to facilitate the direction of the learner’s attention to the objectives that are most difficult to acquire. The resulting outcomes for instructional materials for a learner that would perform at the 50% level without properly positioned rehearsal strategies would be between 56.0% and

60.6% for paper-based materials or screen presentations (lecture formats) and between 52.6% and 73.1% for computer-based learning environments.

Bibliography

Andre, T. (1979). Does answering higher-level questions while reading facilitate productive learning? *Review of Educational Research, 49*, 280-318.

Andre, T. (1990). Type of inserted question and the study-posttest delay. *Journal of Experimental Education, 58*(2), 77-86.

Arthur, W., Bennett, W., & Huffcutt, A. I. (2001). *Conducting meta-analysis using SAS*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Atkinson, R., & Shriffen, R. (1968). Human memory: A proposed system and its control processes. In K. Spence & J. Spence (Eds.), *The psychology of learning and motivation: Advanced research and theory* (Vol. 2). New York: Academic Press.

Ausubel, D. P. (1968). *Educational psychology: A cognitive review*. New York: Holt, Rinehart & Winston.

Baddeley, A. D. (1992). Working memory. *Science, 255*(January), 556-559.

Baker, R. M., & Dwyer, F. M. (2000). A meta-analysis assessment of the effect of visualized instruction. *International Journal of Instructional Media, 27*(4), 417-425.

Baker, R. M., & Dwyer, F. M. (2005). Effect of instructional strategies and individual differences: A meta-analytic assessment. *International Journal of Instructional Media, 32*(1), 69-84.

Bangert-Drowns, R. L., & Rudner, L. M. (1991). *Meta-analysis in educational research* (No. ED339748). Washington D.C.: ERIC Clearinghouse on Tests Measurement and Evaluation.

Beecher, J. (1988, December 12). *Note-taking: What do we know about the benefits?* Retrieved January, 6, 2005, from http://reading.indiana.edu/ieo/digests/d37.html

Bentin, S., Moscovitch, M., & Nirhod, O. (1998). Levels of processing and selective attention effects on encoding in memory. *Acta Psychologia, 98*, 311-341.

Benton, S. L., Glover, J. A., & Bruning, R. H. (1983). The effect of number decisions on prose recall. *Journal of Educational Psychology, 75*(3), 382-390.

Bowers, J. S. (2002). Challenging the widespread assumption that connectionism and distributed representations go hand-in-hand. *Cognitive Psychology, 45*, 413-445.

Brown, J. (1976). An analysis of recognition and recall and of problems in their comparison. In J. Brown (Ed.), *Recall and recognition* (pp. 1-35). New York: John Wiley & Sons.

Bruner, J. S. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models : applications and data analysis methods* (Vol. 1). Newbury Park: Sage Publications.

Cameron, B. (2004). *The effect of gaming, cognitive style and feedback type in facilitating delayed achievement of different learning objectives.* Unpublished dissertation, The Pennsylvania State University, University Park, PA.

Campbell, D. T., Stanley, J. C., & Gage, N. L. (1966). *Experimental and quasi-experimental designs for research*. Chicago: R. McNally.

Canelos, J. (1987). Research justification and implementation of the Program for Systematic Evaluation (PSE). In F. M. Dwyer (Ed.), *Enhancing visualized instruction: Recommendations for practitioners* (pp. 1-3). State College, PA: Learning Services.

Cha, D. C. (1990). *Effect of visual and embedded rehearsal strategies in facilitating student achievement on different educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Chen, W.-F. (2002). *The effect of varied elaboration strategies in complementing visualized instruction and prior knowledge in a hypertext environment on student achievement of different educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Clariana, R. B. (2003). The effectiveness of constructed-response and multiple-choice study tasks in computer aided learning. *Journal of Educational Computing Research, 28*(4), 395-406.

Clariana, R. B., & Lee, D. (2001). The effects of recognition and recall study tasks with feedback in a computer-based vocabulary lesson. *Educational Technology, Research and Development, 49*(3), 23-36.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445-459.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology, Research and Development, 42*(2), 21-29.

Cobb, T. (1997). Cognitive efficiency: Toward a revised theory of media. *Educational Technology, Research and Development, 45*(4), 21-35.

Corkill, A. J. (1992). Advance organizers: Facilitators of recall. *Educational Psychology Review, 4*(1), 33-68.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268-294.

Crawford, C. C. (1925). Some experimental studies of the results of college note-taking. *Journal of Educational Research, 12*, 379-386.

Dornisch, M., & Sperling, R. (2004). Elaborative questions in web-based text materials. *International Journal of Instructional Media, 31*(1), 49-59.

Duchastel, P. C. (1983). Improving adjunct question research: Processes and ecological validity. *Human Learning, 2*, 1-5.

Dwyer, C. A. (1987). An overview of the Program of Systematic Evaluation. In F. M. Dwyer (Ed.), *Enhancing visualized instruction; Recommendations for practitioners* (pp. 4-8). State College, PA: Learning Services.

Dwyer, C. A., & Dwyer, F. M. (1987). Effect of depth of information processing on students' ability to acquire and retrieve information related to different instructional objectives. *Programmed Learning and Educational Technology (Great Britain), 13*, 265-279.

Dwyer, C. A., & Dwyer, F. M. (1993). Effects of covert and overt rehearsal strategies used to complement visualization possessing differing amounts of realistic detail. *International Journal of Instructional Media, 20*(1), 53-69.

Dwyer, F. M. (1965). *An experimental evaluation of the relative effectiveness of selected visual illustrations in teaching science concepts to college freshmen.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Dwyer, F. M. (1972). *A guide for improving visualized instruction*. State College, Pa.: Learning Services.

Dwyer, F. M. (1978). *Strategies for improving visual learning: A handbook for the effective selection, design, and use of visualized materials*. State College, Pa.: Learning Services.

Dwyer, F. M. (1984). *Enhancing visualized instruction: Recommendations for practitioners*. State College, PA: Learning Services.

Dwyer, F. M. (2003). Assessing strategies for developing effective and efficient text for distance education: Traditional and electronic. *International Journal of Instructional Media, 30*(1), 11-23.

Dwyer, F. M. (2004). *The Program of Systematic Evaluation*. Retrieved May 10, 2005, from http://www.ed.psu.edu/dwyer/index.htm

Elliott, C. E., & Dwyer, F. M. (1995). Effect of prior knowledge and varied rehearsal strategies on student achievement of different educational objectives. *International Journal of Instructional Media*, *22*(4), 341-351.

Frase, L. T. (1975). Prose Processing. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 1-47). New York: Academic.

Glass, G. V. (1976). Primary, secondary and meta-analysis research. *Educational Researcher, 5*, 3-8.

Glass, G. V. (2000). *Meta-analysis at 25*. Retrieved January 27, 2001, from http://glass.ed.asu.edu/gene/papers/meta25.html

Glover, J. A. (1989). Improving readers' estimates of learning from text: The role of inserted questions. *Reading Research and Instruction*, *28*(3), 68-75.

Gredler, M. E. (1997). *Learning and instruction : theory into practice* (3rd ed.). Upper Saddle River, N.J.: Merrill.

Gredler, M. E. (2001). *Learning and instruction: Theory into practice* (4th ed.). Upper Saddle River, N.J.: Merrill.

Griffin, R. E., Gibbs, W. J., & Weigmann, B. (Eds.). (1999). *Visual literacy in an information age*. University Park, PA: International Visual Literacy Association.

Haag, B. B. (1995). *The effects of visual manipulation strategies within computer-based instruction on various types of learning objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Hagenberger, J. M. (2003). *Effect of Field Dependence-Independence and Elaborative Levels of Job Aids in Facilitating Learner Achievement of Different Educational Objectives.* Unpublished dissertation, The Pennsylvania State University, University Park, PA.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56*, 212-242.

Harris, J. I., & Qualls, C. D. (2000). The association of elaborative or maintenance rehearsal with age, reading, comprehension, and verbal working memory performance. *Aphasiology, 14*(5/6), 515-526.

Hedges, L. V. (1982). Fitting continuous models to effect size data. *Journal of Educational Statistics, 7*, 245-270.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. In. Orlando, FL: Academic.

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods, 6*, 203-217.

Heinich, R., Molenda, M., & Russell, J. D. (1989). *Instructional media and the new technologies of instruction*. New York: Macmillan Publishing Company.

Henk, W. A., & Stahl, N. A. (1985). *A meta-analysis of the effect of notetaking on learning from lecture.* Paper presented at the 34th Annual Meeting of the National Reading Conference.

Huffcutt, A. J., & Arthur, W. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology, 80*(2), 327-334.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis : correcting error and bias in research findings*. Newbury Park, Calif.: Sage Publications.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Irizarry, D. (1990). *The effects of embedded analogies and question generation strategies in facilitating student achievement of specific educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Jonassen, D. H., & Grabowski, B. L. H. (1993). *Handbook of individual differences, learning, and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Kobayashi, K. (2005). What limits the encoding effect of note-taking? A meta-analytic examination. *Contemporary Educational Psychology, 30*, 242-262.

Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*(1), 79-97.

Lamberski, R. J. (1972). *An exploratory investigation of the instructional effect of color and black and white cueing on immediate and delayed retention.* Unpublished masters, Pennsylvania State University, University Park, PA.

Lamberski, R. J. (1980). *The effect of a verbal and visual color code on self-paced instruction and testing for retention on different tasks.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Lamberski, R. J., & Dwyer, F. M. (1981). The instructional effect of color in intermediate and delated reaction. *Journal of Instructional Psychology, 8*(4), 122-131.

Lee, M. G., & Dwyer, F. M. (1992). Developmental analysis of the effects of adjunct questions and visualization on student learning. *International Journal of Instructional Media, 19*(2), 157-168.

Lehman, S., & Schraw, G. (2002). Effects of coherence and relevance on shallow and deep text processing. *Journal of Educational Psychology, 94*(4), 738-750.

Lin, H., Kidwai, K., Munyofu, M., Swain, J., Ausman, B., & Dwyer, F. M. (2005). The effect of verbal advance organizers in complementing animated instruction. *Journal of Visual Literacy, 25*(2), 237-248.

Lindner, R. W., & Rickards, J. P. (1985). Questions inserted in text: Issues and implictions. In D. H. Jonassen (Ed.), *The technology of text: Principles for structuring, designing, and displaying text* (pp. 131-157). Englewood Cliffs, NJ: Educational Technology.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. (1976). Depth of processing, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 75-102). New York: John Wiley & Sons.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mertler, C. A., & Vannatta, R. A. (2002). *Advanced and multivariate statistical methods : practical application and interpretation* (2nd ed.). Los Angeles: Pyrczak.

Meyer, B. J. F., & McConkie, G. W. (1973). What is recalled after hearing a passage? *Journal of Educational Psychology, 65*(1), 109-117.

Meyer, B. J. F., & Poon, L. W. (2001). Effects of structure strategy training and signalling on recall of text. *Journal of Educational Psychology, 93*(1), 141-159.

Miller, G. A. (1956). the magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language, 33*(128-147).

Moscovitch, M., & Craik, F. I. M. (1976). Depth of processing, retrieval cues, and uniquesness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior, 15,* 447-458.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. New York: Cambridge University Press.

Osman, M. E., & Hannafin, M. J. (1994). Effects of advance questioning and prior knowledge on science learning. *Journal of Educational Research, 88*(1), 5-13.

Paivio, A. (1976). Imagery in recall and recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 103-129). New York: John Wiley & Sons.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.

Pearson, K. (1904). Report on certain enteric fever innoculation statistics. *British Medical Journal*, 1243-1246.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2000). *HLM 5: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.

Rosenthal, R. (1979). The "file-drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638-641.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed. Vol. v. 6). Newbury Park [CA]: Sage Publications.

Rosenthal, R. (1998). Meta-analysis: Concepts, corollaries and controversies. In J. G. Adair, D. Belanger & K. L. Dion (Eds.), *Advances in psychological science (Volume I): Social, personal, and cultural aspects*. East Sussex: Psychological Press.

Rosenthal, R., & Rubin, D. B. (1988). Comment: Assumptions and procedures in the file drawer problem. *Statistical Science, 3*, 120-125.

Roshan, V., & Dwyer, F. M. (1998). Effect of embedded graphic mapping strategies in complementing verbal instruction. *International Journal of Instructional Media, 25*(4), 389-398.

Ruthkosky, K. O. N., & Dwyer, F. M. (1996). The effect of adding visualization and rehearsal strategies to advance organizers in facilitating long term retention. *International Journal of Instructional Media, 23*(1), 31-40.

Schneider, W., & Graham, D. J. (1992). Introduction to connectionist modeling in education. *Educational Psychologist, 27*(4), 513-530.

Schunk, D. H. (2000). *Learning theories: An educational perspective* (Vol. 3). Uper Saddle River, NJ: Prentice-Hall.

Seels, B., & Glasglow, Z. (1997). *Making instructional design decisions* (2nd ed.). Englewood, NJ: Prentice Hall.

Shepard, R. N. (1967). Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*, 156-163.

Slater, R. B., & Dwyer, F. M. (1996). The effect of varied interactive questioning strategies in complementing visualized instruction. *International Journal of Instructional Media, 23*(3), 273-280.

Smith, K. M., & Dwyer, F. M. (1995). The effect of concept mapping strategies in facilitating student achievement. *International Journal of Instructional Media, 22*(1), 25-31.

Smith, P. L., & Ragan, T. J. (1999). *Instructional design* (2nd ed.). New York: John Wiley & Sons Inc.

Szabo, M., Dwyer, F. M., & De Melo, H. (1981). Visual testing - Visual literacy's second dimension. *educational Communications and Technology Journal, 29*(3), 177-187.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn and Bacon.

Taricani, E. M. (2002). *Effects of the level of generativity in concept mapping with knowledge of correct response feedback on learning.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects on retention interval and levels of processing. *Memory & Cognition, 29*(3), 424-432.

Thorndike, E. L. (1913). *Educational psychology: Volume II. The psychology of learning*. New York: Teacher's College Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352-373.

Vance, L. E., & Dwyer, F. M. (1988). The effect of question response mode in facilitating verbal instruction. *International Journal of Instructional Media, 15*(3), 272-281.

Wachter, L. N. (1993). *An investigation of the effects of hierarchical concept mapping as a prefatory organizer on fourth-grade students' comprehension and retention on expository prose.* Unpublished doctoral dissertation, Pennsylvania State University, University Park, Pennsylvania.

Wager, W., & Mory, E. H. (1992). Feedback, questions and information processing - Pulling it all together. *ERIC Document*, #348038.

Walko, J. K., & Dwyer, F. M. (1990). Effect of varied note taking strategies in complementing visualized instruction. *International Journal of Instructional Media, 17*(4), 293-301.

Wang, C. X. (2003). *The instructional effects of prior knowledge and three concept mapping strategies in facilitation achievement of different educational objectives.* Unpublished doctoral dissertation, The Pennsylvania State University, University Park, PA.

Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook of Research in Teaching*. New York: Macmillan.

Willoughby, T., & Wood, E. (1994). Elaborative interrogation examined at encoding and retrieval. *Learning and Instruction, 4*, 139-149.

Wittrock, M. C. (1992). Generative learning processes of the brain. *Educational Psychologist, 24*(4), 345-376.

Wolfe, M. B., & Goldman, S. (2003). Use of latent semantic analysis for predicting psychological phenomena: Two issues and proposed solutions. *Behavior Research Methods, Instruments, & Computers, 35*(1), 22-31.

Wood, G. M., & Dwyer, F. M. (1989). The additive effect of visualization and overt response/visual feedback in facilitating student achievement of different educational objectives. *International Journal of Instructional Media*, *16*(1), 57-68.

Yamashiro, K. A. C. (2001). *The effect of semantic maps and different adjunct processing strategies on student achievement of different types of learning outcomes.* Unpublished doctoral dissertation, Pennsylvania State University, University Park, PA.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517.

Zimmer, H. D., & Engelkamp, J. (1999). Levels of processing effects subject-performed tasks. *Memory & Cognition, 27*(5), 907-914.

Bibliography of Studies Included in the Meta-Analysis

Akambi, M. (1986). *The effects of inductive and deductive instructional strategies on students' ability to profit from visualized instruction.* Unpublished dissertation, The Pennsylvania State University, University Park, PA.

Al-Saai, A. J. (1993). *The effect of different visualized treatments on field-dependent and field-independent students at the University of Qatar.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Bennett, L. T. (1991). *The effectiveness of varied interactive visuals on students' ability to achieve on tests measuring different educational objectives.* Unpublished dissertation, The Pennsylvania State University, University Park, PA.

Buckley, K. A., & Dwyer, F. M. (1984). *The instructional effects of level of locus of control and type or rehearsal strategy on student achievement of different educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Cameron, B. (2004). *The effect of gaming, cognitive style and feedback type in facilitating delayed achievement of different learning objectives.* Unpublished dissertation, The Pennsylvania State University, University Park, PA.

Cha, D. C. (1990). *Effect of visual and embedded rehearsal strategies in facilitating student achievement on different educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Chen, W.-F. (2002). *The effect of varied elaboration strategies in complementing visualized instruction and prior knowledge in a hypertext environment on student achievement of different educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Chezik, M. A. G. (1984). *The effects of review and practice activities on learning from prose text.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Clark, J. M. (1986). *An investigation of the effects of three levels of adjunct postquestions on different levels of repeated and nonrepeated item posttest scores of college students.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Couch, R. (1990). *The effects of imagery rehearsal strategy and cognitive style on the learning of different levels of instructional objectives.* Unpublished dissertation, Virginia Technical Institute, Blacksburg, VA.

Dwyer, C. A. (1984). *The effect of varied rehearsal strategies in facilitating achievement of different educational objectives as measured by verbal and visual testing modes.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Dwyer, F. M. (1985). The effect of visualized instruction and varied rehearsal and evaluation strategies (verbal and visual) in facilitating students' long-term retention on tests measuring different instructional objectives. *Journal of Visual/Verbal Languaging, 5*, 15-27.

Elliott, C. E. (1993). *Effects of prior knowledge and various rehearsal strategies on student achievement of different educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Folsom, A. C. (1995). *The contribution of locus of control to the effect of mentoring tactics in computer-based instruction on college students' achievement.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Gulick, E. J. (1992). *The effect of selected testing sequences and gaming methods on delayed retention, using self-paced programmed instruction, and evaluation for different instructional objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Haag, B. B. (1995). *The effects of visual manipulation strategies within computer-based instruction on various types of learning objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Hatch, C. A. (1998). *The effect of varied advance organizer strategies in facilitating learner achievement of specific educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Heeter, S. V. (2002). *The effects of judgements of learning and locus of control on achievement of various educational objectives and allocated study time.* Unpublished dissertation, The Pennsylvania State University, University Park, PA.

Hodes, C. L. (1990). *The induction, use, and effectiveness of mental imagery as an instructional variable.* Unpublished dissertation, The Pennsylvania State University, University Park, PA.

Hsu, P.-S., & Dwyer, F. M. (2004). Effect of level of adjunct questions on achievement of field independent/field dependent learners. *International Journal of Instructional Media, 31*(1), 99-106.

Irizarry, D. (1990). *The effects of embedded analogies and question generation strategies in facilitating student achievement of specific educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Jackson, R. G. (1993). *The effect of elaboration theory in facilitating achievement of varied educational objectives in print/text materials.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Joseph, J. H. (1978). *The instructional effectiveness of integrating abstract and realistic visualization.* Unpublished dissertation, The Pennsylvania State University, University Park, PA.

Kim, I. W. (1992). *The effects of different kinds of instructional cueing strategies in facilitating student achievement of different educational objectives on immediate and delayed tests.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Labant, J. C. (2000). *The effects of varied verbal rehearsal strategies on immediate and delayed retention of learned objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Lee, M. G. (1991). *A multivariate analysis of developmental effects of adjunct questions and visualization on self-paced prose learning.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Lin, C.-L. (2001). *The effect of varied enhancements to animated instruction on tests measuring different educational objectives.* Unpublished dissertation, The Pennsylvania State University, University Park, PA.

Lin, H. (2006). *The effect of questions and feedback used to complement static and animated visualization on students' achievement of different learning outcomes in a computer-based instructional environment.* Unpublished doctoral dissertation, The Pennsylvania State University, University Park, PA.

Lin, H., Kidwai, K., Munyofu, M., Swain, J., Ausman, B., & Dwyer, F. M. (2005). The effect of verbal advance organizers in complementing animated instruction. *Journal of Visual Literacy, 25*(2), 237-248.

Litzinger, M. E. (1994). *The effect of three instructional prescriptions on the achievement of different educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

McCrae, M. (2004). *Effects of Three Visualized Strategies and College Students' Self-Reporting of Metacognitive Self-Regulation on Physiology Learning Outcomes.* Unpublished dissertation, The Pennsylvania State University, University Park, PA.

McNeal, G. (1986). *The effect of agreement or disagreement of student learning style and method of instruction.* Unpublished dissertation, University of Washington, Olympia, WA.

McNeal, J. M. (1994). *Effect of rehearsal strategies and testing format on student achievement of different educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Partee, G. L. (1984). *The instructional effect of varied layout and design formats in facilitating student achievement of different educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Pineda De Romero, L. (2002). *The effects of visuals complemented with rehearsal strategies in facilitating the achievement of learners with different levels of prior knowledge.* Unpublished dissertation, The Pennsylvania State University, University Park, PA.

Poindexter, M. (2003). *The effect of cognitive processing strategies and level of prior knowledge on student achievement of different*

educational objectives. Unpublished dissertation, The Pennsylvania State University, University Park, PA.

Pollack, L. E. (1987). *The use of interactive video to examine the interactive effect of cognitive style and structural organizers on learning.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Richards, D. R. (1987). *An experimental assessment of the relative effectiveness of varied types of computer-generated feedback strategies in facilitating achievement of different educational objectives as measured by verbal and visual tests.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Roshan, V. (1997). *The effect of varied structural knowledge learning strategies in complementing verbal instruction.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Ruthkosky, K. O. N. (1993). *The effects of adding visualization and rehearsal strategies to advance organizers on facilitating long term retention.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Slater, R. B. (1996). *The effect of varied interactive questioning strategies with feedback in complementing visualized instruction.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Sulaiman, J. (2000). *The effect of varied instructional text design strategies on the achievement of different educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Taricani, E. M. (2002). *Effects of the level of generativity in concept mapping with knowledge of correct response feedback on learning.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Vance, L. E. (1986). *The effect of question response mode in complementing self-paced verbal instruction.* Unpublished dissertation, The Pennsylvania State University, University Park, PA.

Walko, J. K. (1989). *The effect of varied note-taking strategies on students' ability to profit from visualized instruction as measured by tests assessing different educational objectives.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Wang, C. X. (2003). *The instructional effects of prior knowledge and three concept mapping strategies in facilitation achievement of different educational objectives.* Unpublished doctoral dissertation, The Pennsylvania State University, University Park, PA.

Williams, B. O. (1995). *The instructional effects of a mixed-chunk, elaborative feedback strategy on comprehension tasks in CAI.* Unpublished dissertation, Pennsylvania State University, University Park, PA.

Wood, G. M. (1986). *The additive effect of visualization and overt rehearsal/visual reinforcement in facilitating student achievement of different educational objectives.* Unpublished masters, Pennsylvania State University, University Park, PA.

Yamashiro, K. A. C. (2001). *The effect of semantic maps and different adjunct processing strategies on student achievement of different types of learning outcomes.* Unpublished doctoral dissertation, Pennsylvania State University, University Park, PA.

Zhu, L. (2005). *Question and feedback: Effects of direct manipulation and animation in facilitating student achievement on tests measuring different educational objectives.* Unpublished dissertation, The Pennsylvania State University, University Park, PA.

Appendix A

Coding Form for Study Characteristics

Article Analysis

1. Complete citation in APA format
2. Names of authors
3. Frequency in Social Science Index

Design and Method Analysis

1. Most frequently cited author(s) in references (other than author of article/paper)
2. Research question(s)
3. Hypothesis directionality

Descriptive Data of the Subjects and Study

1. Psychometric data
    a. Age range of subjects, location
    b. Academic measures for separation into categories
    c. Pre-test information
        i. Type of test
        ii. Source of test
        iii. Reliability
        iv. Validity

v. Separate into classes? (yes, no, how if used)

2. Type of research design
3. Statistical analysis
4. Criterion test scores
5. Validity and reliability of dependent measures
    a. Type of reliability measure
    b. Scores
6. What controls the pacing of instruction?
    a. Independent, self-paced
    b. Externally paced
7. Type of instructional design strategies used
    a. Specific name of strategy
    b. Category of strategy
        i. Visual (b&w, color, none)
        ii. Learning style
        iii. Feedback
        iv. Prompt for rehearsal (recall, recognition, none)
        v. Overt/covert rehearsal (overt, prompted-covert, covert)
        vi. Medium (screen, paper, computer)
        vii. Item analysis for placement (yes/no)
8. Stated results
9. Concept terms used in introduction/sample description

General Background Markers

1. Year of the study

    a. Month/year study began

    b. Month/year study completed

    c. Not indicated

2. Any exclusions of subjects from data? Why?

3. How experimental and control subjects differ?

4. How experimental and control conditions differ?

5. Internal validity

6. External validitiy

Appendix B

Table of Studies Included

Each study was coded according to the coding sheet included in Appendix A. Table B-1 depicts the reference for each study included, the number of participants within the treatments used to calculate the effect sizes, the number of participants in the complete study, the rehearsal style for the included treatments, the types of visuals included (if used in the treatment), the type of task the rehearsal strategy presented for the participant to perform, the medium of the presentation, the type of pacing for the instructional materials, and the criterion measures used to assessment the participants' achievement. The names of the treatments are indicated in the table as titled by the study author. All references are included in the bibliography of the studies used within the meta-analysis.

Table B-1: Research Studies Included within the Meta-analysis: Reference, Participant Counts, and Coding

Study	N included in meta analysis	Number in study	Rehearsal strategies (covert, prompted covert, overt)	Visual types (b&w, cc, none)	Prompting for Rehearsal (knowledge generation, discrimination, none)	Media Presentation Type (computer, paper, screen presentation)	Pacing (self, external)	Tests
(Akambi, 1986)	128	206	Covert (ta,tb,tc,td) Pcovert (te,tf,tg,th,ti,tj)	B&W (ta,tb,tc,td,te,ti,tj) None (tf,tg,th)	None (ta,tb,tc,td) KnowGen (te,tf,tg,th,ti,,tj)	Computer	Self	C
(Al-Saai, 1993)	150	150	Pcovert (t2) Overt (t3)	B&W	KnowGen (t3)	Paper	Self	D,I,T,C
(Bennett, 1991)	178	178	Covert (t1) Overt (t2, t3, t4)	B&W (t1, t4) B&W (as tool – t2, t3)	Discrim (t2) KnowGen (t3,t4)	Paper	Self	D,I,T,C
(Buckley & Dwyer, 1984)	100	100	Pcovert (t2) Overt (t3)	B&W	KnowGen (t2,t3)	Paper	Self	D,C
(Cameron, 2004)	422	422	Covert (t1) Pcovert (t2) Overt (t3,t4)	B&W	KnowGen	Computer	Self	D,I,T,C
(Cha, 1990)	468	468	Covert (t1,t2,t4) Pcovert (t3) Overt (t5,t6)	B&W	None (t1,t2) KnowGen (t3,t4,t5,t6)	Paper	Self	D,I,T,C
(Chen, 2002)	82	125	Covert (t1) Pcovert (t3)	B&W	Discrim (t3- summaries) None (t1)	Computer	Self	D,I,T,C
(Chezik, 1984)	161	161	Covert (t1) Pcovert (t2) Overt (t3,t4,t5)	B&W	None (t1,t2) KnowGen (t3,t4)	Paper	Self	D,I,T,C
(Clark, 1986)	100	100	Covert (control) Pcovert (comp,id,term)	B&W	None (control) Discrim (comp,id,term)	Paper	Self	I,T,C
(Couch, 1990)	113	113	Overt (self-gen, recreated) Covert (instruct only)	B&W (recreated) None (self gen, instruct only)	KnowGen (recreated, self-gen) None (instruct only)	Screen	External	D,I,T,C
(Dwyer, 1984)	136	136	Covert (t1) Pcovert (t2) Overt (t3,t4)	B&W	Discrim (t2,t4) KnowGen (t3)	Paper	Self	D,I,T,C
(Dwyer, 1985)	160	200	Covert(t2) Pcovert(t3) Overt(t4,t5)	B&W	KnowGen(t4,t5) Discrim(t3)	Paper	Self	I,T,C
(Elliott, 1993)	132	132	Covert (t1,t3) Pcovert (t2,t4)	B&W	KnowGen (t2,t4)	Paper	Self	D,I,T,C
(Folsom, 1995)	133	133	Overt (t2, t3, t4) Pcovert (t3,t4)	B&W	None (t1) KnowGen (t2, t3, t4)	Computer	Self	D,I,T,C
(Gulick, 1992)	96	96	Pcovert (t2,t4) Overt (t1,t3)	B&W	KnowGen (t1,t2,t3,t4)	Paper	Self	D,I,T,C
(Haag, 1995)	184	184	Covert (control) Pcovert (comp-manip) Overt (visual sum w/manip, learner manip)	B&W	None (control) KnowGen (vis sum w/manip) Discrim (comp manip, learner manip)	Computer	Self	D,I,T,C
(Hatch, 1998)	161	161	Overt (tb,tc,td) Covert (ta)	B&W	KnowGen (tb,tc,td) None (ta)	Paper	Self	D,I,T,C
(Heeter, 2002)	143	143	Covert (t1) Pcovert (t2) Overt (t3,t4)	color	Discrim (t2,t3,t4)	Computer	Self	I,T,C

Study	N included in meta analysis	Number in study	Rehearsal strategies (covert, prompted covert, overt)	Visual types (b&w, cc, none)	Prompting for Rehearsal (knowledge generation, discrimination, none)	Media Presentation Type (computer, paper, screen presentation)	Pacing (self, external)	Tests
(Hodes, 1990)	56	112	Covert (t1) Pcovert (t2)	None	KnowGen (t2)	Paper	Self	D,I,T,C
(Hsu & Dwyer, 2004)	132	132	Covert (t1) Overt (t2,t3)	B&W	Discrim (t2,t3)	Computer	Self	C
(Irizarry, 1990)	160	160	Covert (control, analogies) Overt (question gener, combination)	B&W	None (control) Discrim (analogies) KnowGen (question gener, combination)	Paper	Self	I,T,C
(Jackson, 1993)	144	144	Covert (t1,t3) Pcovert(t2,t4)	None	None (t1,t3) Discrim (t2,t4)	Paper	Self	D,T,C
(Joseph, 1978)	36	54	Covert (t1, t2) Overt (t3)	None (t1) B&W (t2,t3)	None (t1,t2) KnowGen (t3)	Computer	Self	D,I,T,C
(Kim, 1992)	117	157	Covert (1,t2) Pcovert (t3,t4)	None (t1) B&W (t2,t3,t4)	KnowGen(t2,t3,t4)	Paper	Self	D,I,T,C
(Labant, 2000)	248	248	Covert (control) Pcovert (summary statement) Overt (fillin statement, summary writing)	None	None (control) Discrim (sum statement) KnowGen (fillin, summary writing)	Paper	Self	D,I,T,C
(Lee, 1991)	243	243	Covert (w/o elab, w/vis) Overt (w/ adj quest, w/ vis & adj quest(	B&W (visual treatments only) None	KnowGen (w/ adj ques) None (others)	Paper	Self	D,I,T,C
(Lin et al., 2005)	89	89	Covert(t1) Pcovert(t2,t3)	B&W	Discrim (t3)	Computer	Self	D,I,T,C
(Lin, 2001)	71	93	Covert (t2) Pcovert (t3) Overt (t4)	B&W	KnowGen (t4)	Computer	Self	D,I,T,C
(Lin, 2006)	582	582	Covert(no strategy) Overt(question, questions+feedback)	B&W	None(no strategy) Discrim(question, question+feedback)	Computer	Self	D,I,T,C
(Litzinger, 1994)	160	160	Covert (control) Pcovert (CDT) Overt (behavioral, gagne)	B&W	KnowGen (gagne, behavioral) None (control, cdt)	Paper	Self	D,I,T,C
(McCrae, 2004)	189	189	Covert (t1) Pcovert (t2, t3)	B&W	None	Paper	Self	I,T,C
(McNeal, 1986)	155	155	Covert (tc,td) Overt (ta, tb)	B&W (ta, tc) None (tb, td)	None (tc,td) KnowGen (ta, tb)	Paper	Self	D,I,T,C
(McNeal, 1994)	352	352	Overt (t2,t4) Covert (t1,t3)	B&W (t3,t4) None (t1,t2)	KnowGen (t2,t4)	Paper	Self	I,T,C
(Partee, 1984)	135	168	Pcovert (43,t4) Covert (t1,t2)	None(t1,t3) B&W (t2,t4)	KnowGen	Paper	Self	I,T,C
(Pineda De Romero, 2002)	449	449	Covert (t1) Overt (t2, t3, t4)	B&W	KnowGen (t2,t3,t4)	Paper	Self	D,I,T,C
(Poindexter, 2003)	146	146	Covert (t1) Overt (t2,t3)	B&W	KnowGen (t2) None(t1) Discrim(t3)	Paper	Self	T,I,C

Study	N included in meta analysis	Number in study	Rehearsal strategies (covert, prompted covert, overt)	Visual types (b&w, cc, none)	Prompting for Rehearsal (knowledge generation, discrimination, none)	Media Presentation Type (computer, paper, screen presentation)	Pacing (self, external)	Tests
(Pollack, 1987)	118	118	Covert (control) Pcovert(study) Overt (mastery)	Color	None (control) KnowGen (study,mastery)	Screen (video)	External	D,I,T,C
(Richards, 1987)	102	126	Pcovert (t1,t2) Overt (t3)	B&W (t1,t2,t3)	KnowGen (t1,t2,t3)	Computer	Self	D,I,T,C
(Roshan, 1997)	248	248	Covert (t1) Overt (t2,t3,t4)	B&W	None (t1) KnowGen (t2,t3,t4)	Paper	Self	D,I,T,C
(Ruthkosky, 1993)	160	200	Pcovert (t2,t4) Overt (t3,t5)	B&W (t4,t5)	KnowGen (t3,t5)	Paper	Self	D,I,T,C
(Slater, 1996)	388	388	Covert (t1) Overt (t2, t3, t4)	B&W	None (t1) KnowGen (t2,t3,t4)	Paper	Self	D,I,T,C
(Sulaiman, 2000)	123	123	Covert (t1) Pcovert (t3,t4) Overt (t2)	B&W(t1,t2,t3,t4)	KnowGen	Paper	Self	D,I,T,C
(Taricani, 2002)	120	150	Pcovert (pf,pn) Overt (tn,tf)	none	KnowGen (tf,tn) Discrim (pn,pf)	paper	Self	T,C
(Vance, 1986)	75	75	Overt (t2, t3, t4) Covert (t1)	None	None (t1) Discrim (t2) KnowGen (t3, t4)	Paper	Self	D,I,T,C
(Walko, 1989)	140	140	Overt (t2,t3,t4) Covert (t1)	B&W	None (c) KnowGen (2,3,4)	Paper	Self	D,I,T,C
(Wang, 2003)	182	182	Covert (t1 c) Overt (t2, t3, t4)	None	KnowGen (t2,t3, t4)	Computer	Self	I,T,C
(Williams, 1995)	105	105	Pcovert (t1) Overt (t2,t3)	B&W	Discrim (t1,t2,t3)	Computer	Self	C
(Wood, 1986)	75	110	Covert (t2) Overt (t3)	B&W	Discrim (t3)	Paper	Self	D,I,T,C
(Yamashiro, 2001)	270	270	Pcovert (t1) Overt (t2,t3,t4)	B&W	KnowGen (t2,t3,t4) Discrim (t1,t2,t3,t4)	Paper	Self	I,T,C
(Zhu, 2005)	184	184	Covert(t1) Overt(t2,t3,t4,t5)	Color	None (t1) Discrim(t2,t3,t4,t5)	Computer	Self	D,I,T,C

VITA

Rose Marie Baker

Education

Ph. D., Instructional Systems, Pennsylvania State University, May 2006
Emphasis: Training, Technology, and Systems Design and Development

M. Ed., Adult Education, Pennsylvania State University, August 1998

B. A., Mathematics/ Chemistry, Washington & Jefferson College, May 1987

Professional Experience

Faculty Member, Project Manager, Instructional Designer
Management Development Programs and Services, University Park, Pennsylvania

Web and Print Based Researcher, Instructional Designer
PricewaterhouseCoopers, Florham Park, New Jersey

Web-based Course Evaluator
Educational Technology Integration Program
World Campus, Pennsylvania State University, University Park, Pennsylvania

Training, Technology and Systems Development, Instructional Designer
Controls, Robotics, and Welding, General Motors University, Warren, Michigan

Mathematics, Chemistry, Computer Applications Teacher
Southern Huntingdon County School District, Three Springs, Pennsylvania

Office, Sales and Marketing Manager, Accountant
Trew's Distributing, Waynesburg, Pennsylvania

Selected Publications

Passmore, D. L. & Baker, R. M. (2005). Sampling strategies and power analysis. In R. A.Swanson & E. Holten (Eds.), *Research in organizations: Foundations and methods of inquiry*. San Francisco: Berrett-Koehler.

Baker, R.M. & Dwyer, F.M. (2005). A meta-analytic assessment of the effects of instructional strategies and individual differences in reducing achievement differences in between FI/FD learners. *International Journal of Instructional Media, 32*, 69-84.

Baker, R. M. & Dwyer, F. M. (2000). The meta-analytic effects of using visualization in instruction, *International Journal of Instructional Media, 27*, 417-426.