

4-28-2010

Evaluation of the Flicker Effect as a Generative Strategy in Enhancing ComputerBased Instruction (CBI) of Visual Recognition and Classification

Ping Luo
University of South Florida

Follow this and additional works at: <https://scholarcommons.usf.edu/etd>



Part of the [American Studies Commons](#)

Scholar Commons Citation

Luo, Ping, "Evaluation of the Flicker Effect as a Generative Strategy in Enhancing ComputerBased Instruction (CBI) of Visual Recognition and Classification" (2010). *Graduate Theses and Dissertations*.
<https://scholarcommons.usf.edu/etd/1702>

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Evaluation of the Flicker Effect as a Generative Strategy in Enhancing Computer-
Based Instruction (CBI) of Visual Recognition and Classification

by

Ping Luo

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Instructional Technology
College of Education
University of South Florida

Major Professor: James A. White, Ph.D.
Ann E. Barron, Ed.D.
John M. Ferron, Ph.D.
Thomas A. Sanocki, Ph.D.

Date of Approval:
April 28, 2010

Keywords: technology affordances, learning, pedagogy, interactive, images

© Copyright 2010, Ping Luo

Dedication

I would like to thank my major professor and committee for encouraging me in my studies. With their guidance and passion in research, I could have overcome many difficulties and made this project a completion. They smiled less and less day after day and finally they gave me hugs after my defense.

I am thankful to all of the other professors of mine who have helped me learn in instructional technology, educational psychology, vision, cognition and the other areas.

My thanks also go to my family and friends for their support in time of need. Year in and year out, they have kept asking me when I can graduate and they sounded exhausted when they got the answer they wanted to have.

Please also allow me to express my gratitude to all of the institutions that make it possible for me to pursue my graduate studies.

Thank God for His blessings.

Table of Contents

List of Tables	iv
List of Figures	vii
Abstract	ix
Chapter 1 Introduction	1
Statement of the Problem	7
Rationale	8
Purpose of the Study	12
Research Questions	12
Significance of the Study	14
Definitions of Terms	15
Limitations and Delimitations	17
Chapter 2 Review of Literature	19
Introduction	19
Problem Solving, Similarity-Based Reasoning, Information Processing, and Meaning Construction	20
Expert Studies	25
General Studies	25
The Nature of Radiological Expertise	27
Perception Models, Visual Literacy, and Visual Concept Learning	36
Interactivity	41
Interactivity and Types or Dimensions of Interaction from Different Perspectives	41
The Importance of Interactivity to Learners and Active and Meaningful Learning	42
Levels of Interactivity and Technology Affordance	44
Effectiveness of Technology-Based Instruction	45
Instructional Methods in Radiology Education	49
Existing Methods and Desired Ones	49
Innovative Instructional Methods	52
Generative Strategies	55
Generative Learning: the Theory and Evidence	56
Generation Effect	59
The Theory, Evidence, and Interpretation	59
Flicker Effect	66
The Flicker Enhances the Cognitive Process in Studying Images Attention	68

Internal and External Representations	73
Comparison and Contrast, Internal and External Images	77
Generative Encoding and Passive Encoding	80
Discrimination and Generalization across and within Categories	82
Forming and Testing Hypotheses	83
Facilitating Perceptual, Conceptual, and the Interactions of These Processes	84
Scanning Images with the Flicker Treatment	85
The System Enhances the Metacognitive Processes in Studying images	86
The Flicker Enhances the Affective Respect in Studying Images	88
The Flicker Paradigm and the Other Treatments: the Curriculum and Participants	89
Conclusions	91
 Chapter 3 Research Methods	 98
Introduction	98
Research Design of the Study	101
Recruitment	105
Participants and Sample Size	105
Instrumentation	106
Development Processes	106
The General Structure and Activities of the Study Materials	109
Three Independent Treatment Programs	118
Instruments of Dependent Measures	121
Validating Instruments	122
Results of Instrument Validation	124
Methods and Results of the Instrument Reliability	134
Procedures	135
Ethical Considerations	136
Statistical Analysis Procedures	137
Pilot Studies	141
The First Pilot Study	141
Usability Test	146
Instrument Reliability	147
The Second Pilot Study	163
Conclusions	167
 Chapter 4 Results of the Study	 168
Introduction	168
Sample Size and Demographic Information of the Participants	169
Analysis of the Relationship between the Pretest and the Posttests	169
Descriptive Statistics	170
Assessing Group Differences in Outcome Measures	176

Analysis of the Relationship between the Recognition Test and the Classification Test	176
Effectiveness Null Hypotheses Testing	177
More Covariate Analysis: Necessary or Not	190
Conclusion	193
Chapter 5 Discussion	195
Introduction	195
Findings of the Experimental Study	195
The Participants in the Three Groups Learned	198
No Significant Effect in the Recognition Test	210
Significant Flicker Effect in the Classification Test	211
Significant Differences in the Other Outcome Measures	212
Theoretical Implications	215
Implications for Practitioners	219
Limitations of the Study	225
Recommendations for Future Research	227
Conclusions	231
List of References	232
Appendices	256
Appendix A IRB Approval	257
Appendix B A Screenshot of the Comparison Treatment	258
Appendix C Screenshots of the Flicker Treatment	259
Appendix D Screenshots of the No-Flicker Treatment	262
Appendix E A Screenshot of a Test Item in the Pretest	264
Appendix F A Screenshot of a Test Item in the Recognition Test	265
Appendix G A Screenshot of a Test Item in the Classification Test	266
Appendix H A Recruitment Flyer	267
Appendix I Evaluation Instruments	268
Appendix J A Demographic Survey	272
About the Author	End Page

List of Tables

Table 3.1 Response Results by Treatment Group for the Usability Test Survey Items in the First Pilot Study	129
Table 3.2 Results of ANOVAs for the Three Groups' Responses to Each Item of the Usability Test Survey in the First Pilot Study	130
Table 3.3 Learning Objectives and Content Validity of the Study	132
Table 3.4 Test Reliability Coefficients of the Posttest Instruments	147
Table 3.5 Pearson Correlation Values Indicating the Relationship between the Pretest and Posttests in the First Pilot Study (n=76)	148
Table 3.6 Mean, Standard Deviation, Sample Size and Other Descriptive Statistics Results of the Three Tests by Treatment Group in the First Pilot Study	149
Table 3.7 Results of Multivariate Analysis of Covariance in the First Pilot Study	153
Table 3.8 ANCOVA Results of the Group Recognition Scores in the First Pilot Study	154
Table 3.9 ANCOVA Results of the Group Classification Scores in the First Pilot Study	155
Table 3.10 Mean, Standard Deviation, Sample Size, and Other Statistics of the Display Rates Selection Frequency in the First Pilot Study	157
Table 3.11 Mean, Standard Deviation, Sample Size, and Other Statistics of the Number of Incorrect Responses by Treatment Group in the First Pilot Study	157
Table 3.12 Results of an Analysis of Variance of the Number of Incorrect Responses in the First Pilot Study	158
Table 3.13 Mean, Standard Deviation, Sample Size, and Other Statistics of the Number of Trials by Treatment Group in the First Pilot Study	158

Table 3.14 Results of an Analysis of Variance of the Number of Trials in the First Pilot Study	159
Table 3.15 Results of Item Analysis of Recognition Test Images with Item Difficulty and Discrimination Indices	160
Table 3.16 Results of Item Analysis of Classification Test Images with Item Difficulty and Discrimination Indices	161
Table 3.17 Mean, Standard Deviation, Sample Size, and Other Descriptive Statistics of Number of Trials for Each Case in the Study Sessions of the Three Groups in the First Pilot Study	162
Table 3.18 Mean, Standard Deviation, Sample Size, and the Other Descriptive Statistics by Test and Treatment Group in the Second Pilot Study	164
Table 3.19 Duration Results by Treatment Group in the Second Pilot Study	165
Table 3.20 ANOVA Results of Group Differences in Duration in the Second Pilot Study	165
Table 3.21 Means, Standard Deviations and the Other Descriptive Statistics of the Number of Incorrect Responses by Treatment Group in the Second Pilot Study	166
Table 3.22 ANOVA Results of the Number of Incorrect Responses in the Second Pilot Study	166
Table 3.23 ANOVA Results of the Number of Trials in the Second Pilot Study	166
Table 4.1 Pearson Correlation Values of the Relationships between the Pretest Scores and the Posttest Scores of Recognition and Classification Tests in the Experiment (n=228, All Items)	170
Table 4.2 Mean, Standard Deviation, Sample Size and Other Descriptive Statistics Results by Treatment Group and Dependent Variable in the Study (n=228, All Items)	171
Table 4.3 Pearson Correlation Values Indicating the Relationship between the Posttests in the Experiment (n=228, All Items)	177
Table 4.4 Results of Multivariate Analysis of Covariance of the Overall Group Differences in the Study (n=228, All Items)	179
Table 4.5 ANCOVA Results of the Group Recognition Scores in the Experiment (n=228)	180

Table 4.6 ANCOVA Results of the Group Classification Scores in the Experiment (n=228)	182
Table 4.7 Results of Group Contrast of Adjusted Means of Posttest Scores with the Pretest Scores as a Covariate (n=228)	184
Table 4.8 ANOVA Results of Group Differences in Duration (n=228)	185
Table 4.9 Results of Multiple Comparisons of Group Duration with Tukey HSD (n=228)	186
Table 4.10 The ANOVA Results of Group Differences in the Number of Incorrect Responses	187
Table 4.11 The ANOVA Results of Group Differences in the Number of Trials (n=228)	187
Table 4.12 Results of Multiple Comparisons of the Number of Incorrect Responses with Tukey HSD (n=228)	188
Table 4.13 Results of Multiple Comparisons of Group Differences in the Number of Trials with Tukey HSD (n=228)	189
Table 4.14 Results of Selection Frequency of Display Rates in the Flicker and No-Flicker Groups	190
Table 4.15 Correlation Coefficients of Duration, Number of Incorrect Responses, and Number of Trials with the Posttest Scores	191
Table 4.16 Results of the Analysis of the Instructional Strategy Effects upon Learning without the Pretest Scores as the Covariate with MANOVA (n=228)	192
Table I1 The evaluation instrument for the subject matter expert	268
Table I2 The evaluation instrument for the instructional technology expert	269
Table I3 Usability Test Survey	271

List of Figures

Figure 3.1 An overview of variables manipulateded and observed in the study	102
Figure 3.2 Diagram of research design of the pretest-posttest control group experimental study	104
Figure 3.3 Flow chart illustrating the flow and structure of the three parallel versions of the program	111
Figure 3.4 An analysis of the construct pattern recognition	133
Figure 3.5 The construct pattern recognition and measures of the construct	133
Figure 3.6 Diagram of an overview of statistical analysis procedures	140
Figure 3.7 Participants gender distributions in the first pilot study	143
Figure 3.8 Participants age distributions in the first pilot study	144
Figure 3.9 Participants ethnicity distribution in the first pilot study	145
Figure 3.10 Participants program distributions in the first pilot study	146
Figure 3.11 Pretest performance by group in the first pilot study shown with box plots	150
Figure 3.12 Recognition test performance by group in the first pilot study shown with box plots	151
Figure 3.13 Classification test performance by group in the first pilot study shown with box plots	152
Figure 4.1 Pretest performance by group in the study shown with box plots	173
Figure 4.2 Recognition test performance by group in the study shown with box plots	174
Figure 4.3 Classification test performance by group in the study shown with box plots	175

Figure B1 An Instructional Screen of the Comparison Treatment	258
Figure C1 The First Screen of a Case Study in the Flicker Treatment	259
Figure C2 The Second Screen of a Case Study in the Flicker Treatment	260
Figure C3 The Third Screen of a Case Study in the Flicker Treatment	261
Figure D1 The First Screen of a Case Study in the No-Flicker Treatment	262
Figure D2 The Second Screen of a Case Study in the No-Flicker Treatment	263
Figure E1 A Screenshot of a Test Item in the Pretest	264
Figure F1 A Screenshot of a Test Item in the Recognition Test	265
Figure G1 A Screenshot of a Test Item in the Classification Test	266

Evaluation of the Flicker Effect as a Generative Strategy in Enhancing Computer-Based Instruction (CBI) of Visual Recognition and Classification

Ping Luo

Abstract

Few studies address the question of the technology-based instructional methods of visual patterns, so the overarching purpose of this study was to investigate the effects of three treatments on pattern recognition. Specifically, with a pretest-posttest control group experimental study, the effectiveness of three instructional strategies, a flicker treatment, a no-flicker treatment, and a comparison treatment, (groups respectively analyzing sequential displays of two similar images with and without a blank screen in between and simultaneous displays of two images), was compared based on recognition (memory) and classification (transfer) test scores. The group differences in learning effectiveness and efficiency were also measured by study duration, the number of incorrect responses and the number of trials. Pretest scores were taken as a covariate to equate the groups' prior knowledge.

College students (n=228) recruited from the liberal arts, science, and engineering programs in a Southeast university of the United States were randomly assigned to one of the three treatments. Their immediate learning was assessed with validated tests of recognition and classification, and their study time and response accuracy was tracked. All of the three groups learned and gained approximately an 80% accuracy rate in both

posttests. An overall statistically significant difference was identified among the groups. In the classification test, both the flicker and comparison groups performed significantly better than the no-flicker group with small effect sizes. However, there were no significant differences among the groups in the recognition test. Moreover, the three groups demonstrated statistically significant differences in duration, number of incorrect responses, and number of trials.

The study results are consistent with generative learning and related theories and evidence. Outcome measures inform practitioners of potential effective methods and a validated instructional system while effect sizes indicate relatively small advantages at relatively high cost.

Chapter 1 Introduction

Instructional designers can more or less see potential and possibilities but face uncertainties and ambiguities in the design of computer-based or Web-based learning environments. Maybe an obvious problem that they have been experiencing is the debate on the existence, significance, and effectiveness of computer and the Internet use in education (e.g., Cuban, 2001; Clark, 1994; Kozma, 1994; Hannum, 2007). To address these arguments, it is necessary to consider integrating instructional technology (IT) into classrooms through effective instructional methods. However, there is a lack of prescriptive instructional methods in existing human learning theories and instructional design models, principles, and heuristics (Alessi & Trollip, 2001; Gagne, Wager, Golas, & Keller, 2005; Jonassen, 1999, 2004; Mayer, 2001), suggesting the urgency to investigate the evidence of instructional methods, especially in such an area as visual category learning with little empirical pedagogical information (Sharples, 1991). Therefore, this study examined how different instructional strategies impact complex image study in technology-based instruction.

In order to help learners improve their performance and learning outcomes with IT, instructional methods ought to satisfy the needs of technology-based learning and apply technology affordances. For example, online learning environments are characteristic of voluntary participation and independent learning (Davidson-Shivers, 2002; Gagne, Wager, Golas, & Keller, 2005; Mayer, 2001). Therefore, instruction in

these environments needs to contain strategies to engage learners and enhance mental participation. However, much information without active learning activities (Davidson-Shivers, 2002; Gagne, Wager, Golas, & Keller, 2005; Mayer, 2001) is a widely existing trap in online learning. Here active learning refers to cognitive participation indicated in Mayer's active learning assumption (2001). On the other hand, technology offers possibilities of interactive instruction to engage learners in thinking and other cognitive activities. Hence, enhanced interactivity through technology can be integrated into instructional strategies to foster learning.

One of the neglected areas of pedagogical inquiry for technology-based instruction is visual category instruction (Sharples, 1991; Kim & Astion, 2000) or pattern recognition instruction. Sharples (1991) defined a visual concept as “a named mental construct associated with a set of visual images” (1991, p. 124). Kim and Astion (2000) further explained that “a visual concept lies at the intersection of what we see in an object (perception) and what we know about the object (meaning)” (p. 350). The visual concept these researchers referred to is equivalent to the concrete concept defined in the classic intellectual skill hierarchy of instructional design (Gagne, Wager, Golas, & Keller, 2005). Therefore, learning a visual category means learning the individual representatives of the category (recognition) and classifying new instances into categories with rules (classification) (Fleming, 1993; Bruning, Schraw, Norby, & Ronning, 2004). The recognition and classification of visual or concrete categories is regarded as pattern recognition (Bruning, Schraw, Norby, & Ronning, 2004; Norman, Coblenz, Brooks, & Babcook, 1992; Wood, 1999) in this study. In visual category/pattern recognition instruction, previous studies focused on how to present images and offered scant theory-

based and empirically supported information on effective technology-based instruction. As Sharples identified (1991), there was little research of visual category instructional methods for computer-based instruction (CBI).

However, visual category learning or pattern recognition, especially that in technology environments, is important for education because of broad application, usefulness, and complexity of visual patterns and images. First, images are widely used in many academic and professional areas, such as math, biology, architecture, medicine, and radiology (e. g., Braden, 1996; Sharples, 1991). Digital images have become a main modality in such an area as radiology while computerized images have frequently been applied in online education of math, biology, architecture, and some other areas. Second, images can demonstrate different perceptual dimensions of objects, including shape, size, texture, contrast, brightness, and other features. Images can illustrate spatial relationship and processes. For examples, radiographic images can show locations of glands and tissues and changes in organs. Images can also represent basic concepts in an area, such as geometrical shapes in math, cell structures in biology, architectural styles in architecture, and anatomical structures in radiology. Therefore, perceptual recognition and conceptual understanding of images are important in these areas. Third, visual concepts can be so complex that it usually takes years of training for novices to become experts in such a professional area as radiology (e. g., Gibson, 1969; Lesgold, Rubinson, Feltovitch, Glaser, Klopfer, & Wang, 1988; Norman, Coblenz, Brooks, & Babcook, 1992). Thus, benefits and difficulties in learning visual concepts demand effective instructional methods.

For concept learning, generative strategies were proposed as one general type of instructional strategies for concept learning (Smith & Ragan, 1993). “Generative strategies (Wittrock, 1974) are those approaches in which learners encounter the content in such a way that they are encouraged or allowed to construct their own idiosyncratic meanings from the instruction by generating their own educational goals, organization, elaborations, sequencing and emphasis of content, monitoring of understanding, and transfer to other contexts” (Smith & Ragan, p.151-152, 1993). Based on a constructivist view of learning, generative strategies drive learners to be active and responsible for constructing meanings in learning.

Furthermore, generative learning and generation effect theories and studies offer theoretical and empirical evidence for effectiveness of generative strategies. In particular, studying generation effect with pictures, researchers found that recall and recognition were increased when learners generated solutions to problems by themselves rather than received pictures and/or solutions directly from experimenters or any other sources (Carlin, Soraci, & Strawbridge, 2005; Kinjo & Snodgrass, 2000; Peynircioglu, 1989; Wills, Soraci, Chechile, & Taylor, 2000).

In one of the studies (Carlin, Soraci, & Strawbridge, 2005), researchers examined the flicker task as a generative strategy. In the flicker task, two images were flashed alternatively with a blank screen in between and learners were asked to identify the change(s) in the image. The effect of the flicker task was compared with that of a no flicker task, in which two images were flashed alternatively without any screen in between. They found flicker effect on participants’ recall memory. However, this study and the other few generation studies with pictures did not examine generation effect on

learners' transfer learning – classification of images or image patterns that have not been viewed in study. In fact, comprehension is one of the most outstanding outcomes from generative learning strategies proposed by Wittrock (1974, 1990, 1991, 1992). Thus, it is reasonable examining comprehension or classification of image patterns.

In addition, the pictures applied in these studies are those of everyday objects and scenes but not from any academic or professional domains. Complex images in science have rarely been studied in generation effect studies for technology-based instruction. Therefore, this study examined generative strategies with computer-based radiographic image learning.

As in many other areas, educators in radiology have recognized the benefits of instructional technology (IT) (Gunderman, Kang, Fraley, & Williamson, 2001). With the advent of digital radiographic images, residents tend to rely on computers to view and interpret images and make reports in clinical training. Another phenomenon in radiology education is the increasing development and use of technology-based education, including Websites, online teaching files, and educational software. However, training methods of computer-based instruction in this area were understudied (Sharples, 1995; Luo, Eikman, Kealy, & Qian, 2006; Luo, Szabunio, & White, 2008). Traditional instructional methods in radiology education include conferences, lectures, teaching files, and self-study (Chew, 2001; Collins, 2000, 2006). One of the case conference methods commented as engaging is side-by-side comparison and contrast (Roberts & Chew, 2003), but effectiveness of this method has seldom been investigated in previous studies, especially with radiographic images.

In addition, instructional design ought to be grounded in an understanding of how learning occurs (e.g., Bransford, Brown, and Cocking, 1999; Jonassen, 2004). Therefore, this study was based on previous cognitive studies of expertise characteristics and development (Ericsson & Charness, 1997; Myles-Worsley, Johnston & Simons, 1988; Lesgold, Rubinson, Feltovitch, Glaser, Klopfer, & Wang, 1988; Alexander, 2003). Existing knowledge in this respect can justify learning processes and goals.

In summary, it is important to address the relationship between visual concepts and instructional strategies based on the knowledge of human cognition and learning. First, pedagogical research is critical for improving computer-based instruction (CBI) or Web-based training (WBT). Delay in this line of research may otherwise hamper effective incorporation of IT into complex image learning. Second, the study can have theoretical implications for generative theories because it extends existing studies from everyday images to complex scientific images. Evidence can be derived from this study, validating the existing hypotheses in generative learning and generation theories. Third, in practice, it can increase instructional designers' skills and confidence in solving instructional design problems in CBI or WBT instruction. Starting with fundamental media-embedded instructional research and using this knowledge to enhance media and harness technology, possibilities of effective and efficient instructional design may become a reachable goal. Fourth, visual concept instruction is an essential curriculum component in many academic and professional areas. In particular, instructional practice of effective instructional strategies with radiographic images may improve learners' recognizing and classifying image patterns and facilitate them in pattern recognition and

concept formation. This improvement may lay a foundation for them to develop higher level of thinking and solve difficult diagnostic problems afterwards.

Statement of the Problem

Relevant research problems and gaps were identified and presented as follows: First, few studies were conducted on what potential CBI and/or WBT methods can be used to promote visual category learning and how different instructional methods affect visual category learning. IT has become a trend in education, but the methods of applying IT in visual pattern learning are limited. Although CBI and/or WBT is rich in visual applications and visual categories have been widely learned on computers, it was rarely studied what instructional strategies can be designed to promote visual learning. Specifically, few researchers had ever compared the effect of the flicker task as a generative searching strategy with that of the no-flicker task as a direct searching strategy and the conventional comparison strategy for instructional design of visual categories. Second, the study examined the effect of generative strategies on a new type of learning, visual category learning. Generative strategies were examined in science, reading, and other academic and professional areas, but little was known in the effects of generative strategies on complex image categories. Third, for the purpose of assessing visual category learning, the study designed and developed new criterion measures of recognition and categorization, on which few investigators had pursued evidence by comparing the effect of these strategies. Fourth, classification performance had not been assessed in generation effect studies, but it was regarded as the major assessment approach to testing concept/category learning to indicate transfer of learning (Gagne, Wager, Golas, & Keller, 2005; Smith & Ragan, 1993). Thus, it was proposed as one of

the criterion measures in this study. Fifth, there is a lack of theoretical frameworks for visual category learning and visual literacy (Braden, 1996) and this study can serve as an effort of experimenting with new approaches in psychology by identifying, redesigning, and assessing their effects through CBI and/or WBT design, development, and an experimental evaluation. Briefly, it was imperative to conduct this study to fill in these existing research gaps.

Rationale

Without studies in how to apply IT in CBI and/or WBT, its benefits would be questioned, challenged, and compromised. Technology-based teaching materials are emerging and increasing, but there is a lack of instructional design practice and research support for these projects. Mostly, these materials consist of online teaching files, tutorials, and other forms of information transmission, duplicating textbooks and atlases (Cook, 2005; Friedman, 1996). They were developed with limited consideration of how people think and learn (Bransford, Brown, & Cocking, 1999). Therefore, it is necessary to examine theory-informed and learner-centered instructional methods to enable learners to engage them in processing information and making sense of what they study rather than merely receive information as observers (Jonassen, 1999; Mayer, 2001, 2005; Morrison et al., 1994).

One of the areas that deserve attention is radiographic image instruction. On one hand, computer technology has been widely used in radiology education because of the increasing application of digital images in radiology. On the other hand, few studies of instructional activities have been conducted in visual concept instruction, especially in radiographic image instruction (Kim & Astion, 2000; Sharples, 1991). Although general

guidelines are available for presenting and sequencing visual concept instruction (Sharples, 1991), there is little evidence of effective instructional strategies in this area.

Furthermore, the existing challenges in radiology education demand IT research. Radiology education is traditionally teacher-centered, and this model needs to be replaced by a learner-directed model (Chan & Gunderman, 2005), which means that learners are supposed to have more autonomy and independence in their learning processes than before. One of the reasons for the urgency of learner-centered learning is the shortage of academic radiologists in teaching (Gunderman, Heitkamp, Kipfer, Frank, Jackson, & Williamson, 2003). Radiologists are usually overloaded with clinic work and conferences. When they play multiple roles of physician, faculty researcher, and educator, they may probably have to prioritize these tasks with clinical work on the top of the task list largely because of clinic reading volumes and institutional responsibilities. Therefore, they do not have adequate time for designing instructional programs. As a result, instruction in radiology may probably become an ad hoc apprenticeship (Azevedo, 1998), demanding standard and detailed curriculum (e. g., Collins, 2000, 2006; Gunderman, Heitkamp, Kipfer, Frank, Jackson, & Williamson, 2003). Furthermore, technologies, such as the Internet and Picture Archiving and Communication Systems (PACS), provide storage, retrieval, delivery, and presentation vehicles and platforms but have few pedagogical and cognitive tools to engage learners in learning and practice. This leaves the learner-directed model questionable in radiology education. Therefore, it is necessary to investigate effective approaches to CBI and/or WBT for learner-directed learning.

There are other difficulties in clinical teaching: one issue is the random and discrete cases in clinics, resulting in difficulties for learners to relate their prior knowledge to new cases. Therefore, structuring knowledge has become one of the most difficult tasks in radiologists' professional life. Another problem with clinical cases is that residents may have insufficient immersion in patterns because screening cases mostly comprise of the cases that residents go over during their rotations and these cases are basically normal. Furthermore, in comparing previous images with the current ones to look for changes over time, viewers have to go between computer-based images and the films hang at view-boxes. The cross-media comparisons may lead to information overload and inconvenience for observers. Besides, resident teaching is short of self-assessment schemes, which may limit the opportunities for residents to reflect on their learning and get to know their own learning curves, knowledge gaps, and skills and abilities. They may not realize what they need to make up for further progress. Therefore, radiologist educators have been searching for solutions to address these issues.

In addition, the motivation among radiology residents was reported compromised in studying mammogram evaluation. Bassett and his colleagues (2003) surveyed 201 residents at 211 accredited radiology residencies. They found that 87% of residents regarded mammography interpretation more stressful than reading other images. Although 65% of them valued sub-specialists in this area, 64% of them were reported unwilling to take breast imaging in their fellowship. Furthermore, 63% refused to spend 25% or more of their clinical practice time in interpreting mammograms. They also identified the reasons for these phenomena, including comparatively low interest, high

stresses, and possibilities of lawsuits. The researchers concluded with the lack of willingness to do mammography among residents for fellowship and future practices.

These problems may be reflected in the performance differences among radiologists (Barlow et al., 2004), reflecting the performance gap that needs to be improved through training (Alessi & Trollip, 2001; Smith & Ragan, 1993). Observer detection accuracy in radiology is usually measured with sensitivity and specificity. Sensitivity means an observer's ability to discriminate the targeted stimulus from noises and recognize it while specificity refers to an observer's ability to indicate there is no targeted stimulus found when such a stimulus does not actually exist. Newstead (2003), an associate professor of radiology at the University of Chicago, reported in *Diagnostic Imaging Online* the sensitivities for year-one to year-four residents, namely 33%, 48%, 38%, and 54%, with an average specificity of 72% found for residents. She compared the residents' average sensitivity with that of the radiologists and experts, respectively 46%, 72%, and 82%. Although the statistics reported in this study need further studies to generalize to the other populations, they can reflect an existing phenomenon of inadequate performance. On the other hand, qualified radiologists are necessary because of the large reading volume in clinics. Therefore, improving radiologists' performance deserves IT educators and researchers' attention. Improvement of performance should also consequently affect mammography's status among residents.

Educators in this area have detected the existence of instructional design models and the importance of understanding human learning (Collins, 2000; Williamson, Gunderman, Cohen, & Frank, 2004). However, understanding is one thing but applying this knowledge is another and semantic knowledge differs from procedural knowledge.

Furthermore, instructional design models and human learning principles tend to be limited to descriptive principles in instructional design and studies are necessary for improving types of learning in specific areas. It has to be admitted that theories are usually general rather than specific, but instruction does need models that involve detailed prescriptions. Therefore, there is a demand of theory-based and evidence-supported instructional strategies that these educators can employ in design and instruction. That is, it is necessary to construct effective prescriptive methods that can be more directly applied in instructional design than general principles.

Purpose of the Study

To address the effect of generative learning and generative strategies upon visual concept instruction with technology, the study examined whether the flicker method as a generative instructional strategy in CBI can better increase visual category learning than the no-flicker method as a direct strategy and the traditional comparison method. More specifically, the effects of the flicker activity in comparison to the no-flicker task and comparison strategy were examined on two criterion measures - recognition memory and classification. In addition to the comparison of the three CBI methods upon visual category learning, the other factors will be compared across groups, including duration, frequency of incorrect responses, and frequency of trials.

Research Questions

Specifically, this researcher was interested in examining the following research questions:

1. Did the participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison

method demonstrate any statistically significant differences in their overall performance as measured by recognition and classification posttest instruments?

2. Did the participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method demonstrate any statistically significant differences in their recognition performance as measured by the recognition posttest instrument?

3. Did the participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method demonstrate any statistically significant differences in their classification performance as measured by the classification instrument?

4. Where were there any statistically significant differences in their performance as measured by posttest instruments between the students who studied visual patterns in computer-based instruction with the flicker method of instruction and the no-flicker method of instruction, those studying with the flicker method and the comparison method, and/or those studying with the no-flicker method and the comparison method?

5. Were there any statistically significant group differences in their on-task duration among the participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method?

6. Were there any statistically significant differences in the number of incorrect responses and number of trials they made in their study among the

participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method?

In addition, three post-hoc research questions were raised:

1. If any significant differences in duration were identified among groups, between which groups were the significant differences detected?

2. If any significant differences in number of incorrect responses and number of trials were identified, between which groups were the significant differences detected?

3. Without the pretest score as covariate, did the participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method demonstrate any statistically significant differences in their overall performance as measured by recognition and classification posttest instruments?

Significance of the Study

This study has both theoretical and practical implications for theory development and validation and instructional design practice. More specifically, this study has potential to extend generative learning theories and visual concept instructional models. The results from the study can enhance knowledge of generative learning, generative strategies, and instructional strategies for visual concept learning. Furthermore, the study can inform practitioners of complex image instruction in related academic and professional areas of CBI, WBT, or face-to-face instruction. In addition, effective instructional strategies may improve students' understanding of medical images and prepare them for future learning.

Definitions of Terms

1. Accommodation – A constructivist view of how learning, especially conceptual changes, is achieved. It means that a person learns through creating and/or reorganizing his or her cognitive structures.
2. Affordance – The features and functions offered by the environment, here by instructional technology.
3. Assimilation - A constructivist view of how learning, especially conceptual changes, is achieved. It suggests that a person learns through relating new information to or building it into his or her existing cognitive structures.
4. Comparison method - Learners are asked to identify the difference between two juxtaposed images and they are told that the difference indicates the pattern they are supposed to learn from their image study.
5. Computer-Based Instruction (CBI) – Based on instructional design, instructional and human learning theories and principles, instruction is designed, developed, implemented, and delivered by using computer software and hardware while learners learn through interacting with computers.
6. Classification test – An assessment of categorizing newly-encountered visual patterns according to what one knows about the categories and what one views of the patterns. In other words, the performance in classification is measured with the number of right categorical decisions made with the novel images that have not been observed in the study session.
7. Flicker method – Learners are asked to identify the change between two images when the images are flashed in alternation with a blank screen in between and

they are told that the change indicates the pattern or is an instance of a category that they are supposed to learn from their image study.

8. Generative learning – Learning that assumes that learners are engaged in generative processing of information by connecting their prior knowledge and experience with what they learn. It refers to the learning in which the individual learner actively engages his or her motivation, attention, thinking, and metacognitive resources in learning to enhance encoding, understanding, and problem solving.
9. Generation effect – The effect that the stimuli learners generate can be better recalled and recognized than those provided by experimenters or other sources.
10. Generative strategies – The instructional strategies that engender generative learning processes, facilitating attention, motivation, generation, and metacognition.
11. No-flicker task – Learners are asked to identify the change between two images when two images are flashed alternately without any screen in between and they are told that the change indicates the pattern/category that they are supposed to learn from their image study.
12. Recognition test – An assessment of learners’ memory or ability of identifying the visual patterns/categories on the images that they have previously studied. The performance in recognition is measured with the number of the right decisions made on studied images and image patterns.

13. Sensory memory – The memory structure where incoming visual information is perceived and may be passed to short-term visual memory and long-term visual memory for further processing.
14. Visual short-term memory (vstm) – The memory structure where visual information is perceived, recognized, assigned meanings, and stored temporarily, and may be passed to long-term visual memory for further processing, storage, and retrieval.
15. Web-Based Training (WBT) – Based on instructional design, instructional, and human learning theories and principles, instruction is designed, programmed, delivered, and accessed with Web technologies. It is either regarded as an equivalent to or a subordinate of Computer-Based Instruction.

Delimitations and Limitations

The validity of the experimental study was considered and implemented in research design with randomization, equivalent instructional content and design, and meaningful learning materials. This study intended to reach a balance of internal validity and external validity (Ross & Morrison, 2004) by controlling extraneous variables on one hand and keeping the study meaningful for real-life practice on the other hand.

Considering variables, such as prior knowledge and experience, motivation and interest in learning, and intellectual capacities, might confound the results of the study, random assignment of participants was applied in procedures to rule out the influence of these variables. By randomly assigning participants to the three treatments of the study, the results drawn from the study were caused by the treatments but not by the other factors. Another important approach to extraneous variable control in this study was making all

the design factors and content components in the three treatments of the study equivalent except instructional strategy, the investigated independent variable of the study.

Nevertheless, controlling variables in this study did not compromise the meaningfulness of the study. That is, the study sessions for treatments were meaningful for learners because learning problems were based on clinical cases and instructional strategies were also practicable in real-life CBI and/or WBT. Therefore, the study balanced internal and external validity by randomization, equivalent instructional content and design, and meaningful learning materials.

However, caution is necessary in generalizing experimental results from one sample to the population and from one population to the other populations. Future studies may be conducted to examine the proposed methods in this study with the other samples of the population or the other populations. In addition, changing the selected type of images or the difficulty levels of the images in this study may lead to different learning outcomes. Future studies can examine the questions and hypotheses with different types of images at different difficulty levels.

Another limitation of the study was that gender differences in treatments might lead to an uncontrollable issue to affect internal validity. That is, compared with male participants, female participants were presumed to have higher interest and more prior knowledge in mammogram images and thus more female participants might attend the study. It turned out that similar numbers of male and female students participated in the study. Hence, this is not a limitation of the study any more.

Chapter 2 Review of the Literature

Introduction

In this chapter, related literature is analyzed and critiqued, setting the stage of a theoretical framework for the proposed methodology to address the previously mentioned research problems. The searched and retrieved literature includes journal articles and books in both print and electronic format. Database of different subject areas are included, such as educational database ERIC, psychological database PsychInfo, and medical database Medline, because of the interdisciplinary nature of the study and IT research. In literature filtering and integration, primary sources were regarded as more important than secondary sources. In addition, the quality and authority of studies and journals were also considered.

As a result, this chapter consists of the following interrelated themes: First, this chapter introduces and assesses literature in information processing, medical image diagnosis as pattern recognition, attention, and nature of expertise, exploring cognitive sciences and setting up the large picture and groundwork for this study and review. Second, it describes and analyzes perception models, visual memory, visual literacy, and visual category learning. Third, it discusses the significance, types, and levels of interactivity in technology-enhanced instruction, indicating the importance of learners' interaction with the instructor, computer and the other parts of an instructional system. Four, the chapter evaluates research and comments on the philosophy and methodology

of IT. Fifth, from the stance of instructional design, the chapter conducts an analysis of the instructional methods used in the subject area. Sixth, the chapter analyzes generative models in both education and psychology, reveals the connections of theories and constructs the essence of the theoretical framework for this study. Seventh, it continues to assess the proposed methods, especially the flicker method, in promoting generative processing of visual patterns.

*Problem Solving, Similarity-Based Reasoning, Information Processing,
and Meaning Construction*

Problem solving is regarded as a higher-level intellectual skill, defined in a widely accepted hierarchy in instructional design, with the other lower-level skills in the order of complexity, including discrimination, concepts, and rules (Gagne, Wager, Golas, & Keller, 2005). Cognitive scientists and educators are interested in problem solving because it widely exists in almost every domain of learning and real life (e.g., Bruning, Schraw, Norby, & Ronning, 2004; Jonassen, 2004). Humans actually solve many problems every day no matter whether these problems are math, science, reading, writing, or just everyday routines.

Medical doctors solve diagnostic problems in a great many of areas, ranging from physical examination, internal medicine, to radiology (Norman, Coblenz, Brooks, & Babcock, 1992). However, researchers in cognitive sciences noted that the problems in such an area as radiology differ from those in some other medical areas because similarity-based reasoning is essential in solving radiology problems (Norman, Coblenz, Brooks, & Babcock, 1992; Wood, 1999). That is, diagnosis is established on the basis of pattern recognition and the diagnostic decisions of previous cases. Therefore, diagnostic

problems in radiology are usually solved by correctly detecting visual patterns on radiographic images rather than by collecting various patient symptoms, analyzing them, making and testing hypotheses. This implies that visual features and concepts play an essential role in solving radiology problems. Hence, these features and concepts are the entry points of learning objectives in radiographic image education, considering that they are basic-level intellectual skills in comparison to further interpretation and reasoning processes in radiology diagnosis (Azevedo, 1998; Lesgold, Feltovich, Glaser, & Wang, 1981; Rogers, 1992).

Problem solving includes the presentation of the problem, the original state of the problem, and the goal state of the problem (Bruning, Schraw, Norby, & Ronning, 2004; Jonassen, 2004). To solve a problem is to find some routes to go from the original state to the goal state. For example, in solving a radiographic image problem, the original state is patient data and images whereas the goal state is to interpret observations and make diagnostic decisions although the patient data are usually recommended to be examined after initial detection and diagnosis.

Considering the paths and solutions to problems, researchers usually distinguished between well-defined and ill-defined problems (Bruning, Schraw, Norby, & Ronning, 2004; Jonassen, 2004). For ill-defined problems, there are no absolute steps or solutions. For example, radiography interpretation is a kind of ill-defined problems because readers may take different procedures in viewing and interpreting images. General rules, called heuristics, are followed to pursue the detection and diagnostic goal – radiologists need to figure out how to identify and make decisions on case problems. In addition, differential diagnosis may be given instead of definitive solutions to these problems.

Maybe because of the high frequency of problem solving in academic, professional, and daily life, cognitive scientists claimed the existence of some general problem solving strategies decades ago (e. g., Newell & Simon, 1972). The information-processing model (e.g., Newell & Simon, 1972; Mayer, 2001) may be a product of this assumption. In probing human problem solving, cognitive scientists described models of human information processes (e.g., Atkinson & Shiffrin, 1968; Newell and Simon, 1972). The information processing approach, one of the most essential frameworks in cognitive science, assumes that the human mind works as a computer although current connectionists (e.g., McClelland, McNaughton, & O'Reilly, 1995; Rumelhart & Todd, 1993) have revised this linear model into a networking paradigm. According to the information processing model, information coming in, the sensory system attends, perceives, and detects the information. Then information is processed in working memory and integrated into long-term memory. This stage is called the organization of information. The organized information is encoded and held in the long-term memory for retrieval in the future.

As the entrance of information processing, sensory memory is critical for learners to initiate and activate their minds. According to Goldstein (2002), the perceptual process starts with focused attention to an environmental stimulus when the observer directly looks at the stimulus, forming an image of the stimulus on the observer's receptors. The light coming in the eyes is then transformed into electrical signals in the receptors and these signals are processed and flow in networks of neurons, leading to "conscious sensory experience" (p. 6) called perception. The next step is recognition, which is explained as the ability "to place an object in a category, such as 'tiger', that gives it

meaning” (p. 6). The step of recognition here is similar to classification or a certain type of “pattern recognition” and “assignment of meanings” (p. 18) defined by Bruning, Schraw, Norby, and Ronning (2004, p. 18), who explained “pattern recognition” as “associate perceptual information with a recognizable pattern” (p. 18) and “assignment of meaning” as making decisions about the meaning of sensory information.

Furthermore, the information processing model indicates that meanings are constructed mostly in short-term memory and integrated into long-term memory (Bruning, Schraw, Norby, & Ronning, 2004). Information processing does not simply transmit and translate physical stimuli to mental representations, but essentially through information processing, meanings are constructed and reconstructed based on one’s prior knowledge and learning contexts. This meaning-making process can help one comprehend and retain information. As Craik and Lockhart (1972) identified, memory counts on depth of processing because deep processing concentrates on meanings while shallow processing focuses on superficial respects of materials. Furthermore, meaning making promotes transfer of learning (Bransford et. al., 1983; Mandler & Orlich, 1993), suggesting that learners can use what they learn to solve new problems. As a result, incoming information will be encoded, related to prior knowledge, understood, and transferred if deep perceptual and cognitive analyses are conducted. Otherwise, shallow analyses may probably lead to little learning.

The information-processing model also suggests that novices need to learn basic skills to allocate attention to higher-level skills and tasks. Automaticity (e.g., Chandler & Sweller, 1990; Sweller, 1999) of basic knowledge and skills can prepare novices for their future learning. For example, when learners are fluent in pattern recognition in

radiographic image study, they can then save their attention resources for making decisions.

Regarding visual memory, both the ability and latency to retain visual information are seriously restricted. Sperling (1960) found that the capacity to hold visual stimuli is limited and only about four items can be recalled after an exposure of letters for about 0.5 *second*. He also found that sensory memory decays quickly and visual information can only be retained for 500 *milliseconds (msec)* after the information disappears. Later on, Phillips (1974) distinguished visual short-term memory from the sensory storage, identifying that visual short-term memory has lower capacity than sensory memory and can last from 600 *msec* to a few seconds. These findings suggest that radiographic image learning can become engaging and stimulating by challenging and activating learners' visual memory.

The other important aspect about the sensory system is that prior knowledge and contexts impact upon perception, pattern recognition, and meaning assignment in perceptual processes (Adam, 1990; Anderson, 1984; Bruning, Schraw, Norby, & Ronning, 2004; Goldstein, 2002). For example, researchers (Carmichael, Hogan, & Walters, 1932) found that subjects tended to draw images according to the given verbal labels when they were provided with the same ambiguous pictures with one of two different labels. The study suggests that prior knowledge influences how a person perceives, recognizes, and makes sense of visual information.

Expertise Studies

General Studies

A large body of literature exists in expertise studies, including medical expertise. In many empirical studies, researchers found that experts are different from novices in their knowledge structure and task performance. Compared with novices, experts across domains share some general characteristics (Ericsson & Charness, 1997; Myles-Worsley, Johnston & Simons, 1988; Lesgold, Rubinson, Feltovitch, Glaser, Klopfer, & Wang, 1988; Norman, Coblenz, Brooks, & Babcook, 1992; VanDeventer & White, 2002). For example, experts are typically more accurate, automatic, and adaptive to new situations. Experts actually have “generative knowledge” (Mathews, Roussel, & Cochran, 2001). They recognize meaningful patterns based on an organized knowledge base, make fewer errors, have superior memory recall, and can solve complex problems. Experts tend to see what the novices cannot see. Furthermore, effortful explicit learning in rules and features is important for novices while experts solve problems in a more holistic and automatic way. Considering these novice-expert differences, learning activities for novices in radiographic image reading are supposed to help learners developmental representations and models (Jonassen & Henning, 1999) and improve their automaticity, accuracy, and flexibility in feature differentiation and pattern recognition. It is noticeable that this may lead to some activities directly teaching rules and features with semantic descriptions, which is a method widely used in text books and lectures. However, these existing expository activities have departed from learners’ prior knowledge and experience. Without concrete experience with images, direct instruction may hamper learners from forming representations that need to be constructed based on their concrete experience. In

brief, this body of literature informs what to teach and how to teach different learners although it does not contain any specific instructional methods.

While examining the characteristics of experts or the differences between novices and experts, this body of literature also reveals the processes from novices to experts. Researchers showed that at least three things change on the trajectory from novices to experts, including knowledge, strategies, and interest (e.g., Alexander, 2003). Students develop their expertise in an academic domain from acclimation, through proficiency, to expertise. Both quantitative and qualitative changes take place in the students' knowledge, strategies, and interest. The initial stage of expertise is featured as fragmented knowledge, surface-level strategies, and the reliance on situational interest. Moving onto competence and proficiency, novices change in these respects: their body of knowledge turns to be more integrated, their strategies tend to be more deep-processing, and their interest becomes more self-reliant. The interaction of knowledge, strategies, and interest was regarded as essential.

Considering the developmental processes from novices to experts, learning activities designed for novices will need to motivate learners to invest attentional and other cognitive resources in deep learning. With increased attention allocated to learning, learners may be engaged in such activities as seeing the environmental stimuli in their minds' eye, abstracting concrete experience, and making connections of patterns in the environment and with their prior knowledge, which may help increase organization of external information. These activities are supposed to be deep processing strategies rather than directly reading solutions to problems.

The Nature of Radiological Expertise

The general information of novice-expert differences may not be able to satisfy the needs of instructional design and practice in different subject areas. Needless to say, it is helpful for educators to get to know that experts across domains share some characteristics and education researchers have recognized the value of this literature (e.g., Alexander, 2003; Bransford, Brown, Cocking, 1999) for instructional design and practice. However, it may be more worthwhile for researchers and educators to have detailed knowledge of expertise and its development in a certain instructional area they study. Although a general knowledge of expertise can inform researchers and educators of some differences between novices and experts, this knowledge cannot help develop detailed curricula and design individualized instructional methods for specific learners in a particular area. The specific knowledge of expert and novice performance may imply what learners need to do and aim at, what instructional strategies can enhance learning, and how to assess learning. As indicated in her classic “The Nature of Expertise” (1988), Chi summed up from a collection of expertise studies and maintained that expertise studies are not just limited to one area, but in multiple domains, ranging from the academic domain of physics to professional domain of chess and typing. Therefore, understanding radiological performance may become a foundation for making decisions on instructional methods in radiology.

Based on different theories and evidence, prior researchers studied radiological expertise from different perspectives (Lesgold, Feltovich, Glaser, & Wang, 1981). Generally, these studies can fall into the following three areas: (1) visual detection studies, (2) search studies, and (3) cognitive studies. The former two lines of studies

focus on detection and detection processes in diagnosis. The latter line mainly examines the interpretation of images in diagnosis. A brief review of this body of literature may help justify the instructional methods proposed in this study.

The detection studies were grounded in the signal detection theory (e.g., Goldstein, 2002; Norman, Coblenz, Brooks, & Babcook, 1992). According to this theory, signal detection depends on two factors: one is the sensory system or the observer's sensitivity and the other is the criterion the observer uses in making decisions. The influence of these two factors upon one's performance in perception can be illustrated in signal detection experiments. The experiment tends to contain two essential concepts: signal and noise. The signal refers to the stimulus while the noise means the other stimuli beyond the presented stimuli in the environment. In signal detection experiments, a noise will always be present in every trial with a signal present or absent. There are different performance outcomes in identifying the signal, explained in details as follows, including types of performance outcomes in radiological diagnosis.

The concepts of the signal and noise and the performance outcomes are reflected in radiographic image complexity and diagnostic difficulties. First, it is difficult to detect abnormalities on these images for the complexity of the images, which may result from some physical dimensions, contexts, and anatomical structures of the images. The complexity of radiographs was demonstrated to be originated from the physical characteristics of these images, including sizes, contrast, and edge sharpness (Kundel, 1981; cited in Lesgold, Feltovich, Glaser, & Wang, 1981). The unclear appearances of these physical features may increase the noise in detection. Another interesting finding is that the observer made poor detection because the observer's view was limited to the

abnormal area and the contexts of the image were ignored (Carmody, Nodine, & Kundel, 1980; Swensson, Hessel, & Herman, 1978). This implies that the review of the contexts is important for better detection performance. Moreover, the anatomical structures on images may result in low visibility because they may interact with the abnormal features, hiding them or forming normal appearances. Apart from these difficulties of images, some other perceptual challenges were found, such as thresholds for reporting detection, criteria in making detection, and memory for experiences and patterns (cited in Lesgold, Feltovich, Glaser, & Wang, 1981). These perceptual factors may lead to complexities and difficulties in radiological diagnosis and insufficient performance.

The image complexities and challenges indicated by these studies can have implications for instruction. The complexity of the background and anatomical structures suggest that learners need to study image signals in the contexts of these features. Through interacting with varieties of figures and grounds and anatomical features, learners may form and revise their mental representations and schemata.

Next to the detection research, researchers also investigated the search behaviors and patterns of radiologists. In these studies, radiologists' eye movements in diagnosis were recorded. It was found that how they scan images is varied from image to image and from person to person. Otherwise, they tend to show inferior performance if they use uniformed scanning patterns (Tuddenham & Calvert, 1961). That is, their search patterns are "neither random nor stereotyped" (Kundel, Nodine, & Carmody, 1978). In their search, radiologists were found to fixate and refixate for constructing meanings and meaningful representations (Thomas & Lansdown, 1963). Based on these studies, researchers (e.g., Lesgold, Feltovich, Glaser, & Wang, 1981) summarized some factors

that influence image diagnosis, including initial perception of images, clinical information, prior knowledge of the characteristics of images, and memory and interpretive experiences.

With eye-tracking methods, some current search studies focused on the study of search patterns and time (e. g., Krupinski, 1996; Kundel, Nodine, Conant, & Weinstein, 2007). They seemed to extend the former search studies into comparing the search behaviors among professionals across expertise levels, including radiologists, residents, and technologists. Their findings are consistent on the faster search and more accurate outcomes for radiologists and slower search and less accurate results in less experienced professionals. Interestingly, a current study noted the development of search patterns of radiologists from slow search-to-find patterns to fast global searching (Kundel, Nodine, Conant, & Weinstein, 2007). They also found that the less experienced spend more time searching and go over more image areas than the more experienced (Krupinski, 1996). The other finding they made through their search studies is that the more experienced radiologists have higher abilities to discriminate and classify features. Interestingly, they found that lack of perceptual learning experience in mammography training is a major reason for performance differences in residents. It was explained that their limited perceptual experience confined their skills in object recognition and resulted in difficulties in determining differences of malignant, benign, and normal image patterns (Nodine, Kundel, Mello-Thomas, Weinstein, Orel, Sullivan, Conant, 1999).

These findings from search studies suggest the importance of searching in training because searching patterns are developed through searching. If learners could experience sufficient searching activities, they might have opportunities to experience the perceptual

organization of information (Goldstein, 2002). They may also have experience in eye-movements, getting familiar with the image patterns through fixations. Furthermore, searching is a meaning seeking process, which is critical for categorization of features and diseases. Scanning an image to make sense of it can become a valuable activity in helping learners improve their engagement and deep learning because understanding meanings of images rather than remembering discrete facts was recommended as crucial in radiology education (Lesgold, Feltovich, Glaser, & Wang, 1981).

Different from these two lines of studies in radiology expertise, the other one emphasizes the cognitive processes of radiographic diagnosis. Clearly, the former two types of studies focus on observers' perception and recognition and accuracy of recognition. They provide evidence on the perceptual nature of radiology expertise. However, how observers' perception extends to diagnostic decisions is unclear. Based on the information-processing model, several researchers have conducted studies examining the perceptual and cognitive processes and their interactions in radiology expertise.

Lesgold and his colleagues (1981, 1988) seemed to be the pioneers to investigate the expert problem solving process of radiological diagnosis. In their earlier study (1981), the radiological diagnosis process was explained as "an interaction between the information content of the specific film and the knowledge base of the radiologist" (p. 100). Radiologists' knowledge structure is composed of "schemas for constructing mental representations of anatomy, for recognizing abnormal film features, and for classifying and understanding the implications of disease conditions of patients" (p. 100). It seems that they started to develop a cognitive model of radiological problem solving and identified characteristics of radiological diagnosis.

In their latter study (1988), their research methods were naturalistic observation and think-aloud protocol studies although they called their protocol studies experiments. They clearly presented the participants, materials, procedures, and findings from their second experiment that they designed and developed based on their observation and the first experiment. Different from the former detection and search studies, they collected data of how residents and radiologists thought in their problem solving rather than information about images and eye-movements. Another difference is that they used cases difficult enough to “produce a substantially amount of variability in diagnoses” (p. 315). This selection of difficult cases added more weight to cognitive reasoning (Norman, 1992). In their data analysis, they selected three difficult cases among the cases they used, finding some differences of reasoning chains and clusters among their participants. The experts were found to have “longer reasoning chains, bigger clusters, more clusters, and a greater number of their findings connected to at least one other finding” (p. 317). On the contrary to the “coherent model of the patient” that experts developed, novices tended to manifest “more superficial, fragmental, and piecemeal” representations in their protocols. It seems that experts demonstrated more organized knowledge and understanding rather than discrete facts in their problem solving.

Beyond their quantitative results, the researchers mainly demonstrated some interesting qualitative findings from their protocol analysis. Major findings and some of the implications that can be derived are as follows: (1) “Experts build mental representations of patient anatomy” (p. 320). Experts used their knowledge of anatomy as a map and bound the film features and assigned features to normal anatomy schemata to identify abnormality and localize it. (2) “Experts exhibit flexibility and tuning of

schemata” (p. 323). Novices were found to be limited to some obvious responses but did not consider “remote possibilities” in their diagnosis. Importantly, they reasoned that this could result from some inefficient subprocesses, which consumed their processing capacities. Therefore, the more efficient thinking processes in some lower level thinking can lead to more efficiency in working memory for higher-level thinking. They also justified that “novices may fail because they have not yet developed the fine-tuned visual acuity needed for feature discrimination that is seen in their more experienced colleagues” (p. 324). In contrast, “experts had more refined schemata that allowed them to make finer discriminations” (p. 326). It seems essential for novices to immerse themselves in cases and use some instructional interventions to develop their discrimination at the early stage of their education. Otherwise, the inefficiency in their perception may prevent them from developing their higher-level thinking later on. (3) Experts saw image features differently from novices. (4) Experts were capable of using newly incoming data, demonstrating the opportunism even in diagnosis. (5) “The balance of recognition and inference in diagnosis seems to vary with experience” (p. 336). Finally, the researchers concluded that “the acquisition of expertise consists in ever more refined version of schemata developing through a cognitively deep form of generalization and discrimination” (p. 340). This conclusion implies that perceptual generalization and discrimination are the bases and starting point of this type of learning. Another important point the authors communicated is that they adopted developmental theories and valued the development processes in expertise. They provided evidence of sub-processes and intermediate performance, as mentioned above, suggesting that their study supports the

importance of perceptual training for novices to develop their schemata, generalization, and discrimination.

Based on their studies, two dissertation studies were conducted in the area of artificial intelligence and cognitive sciences (Rogers, 1992; Azevedo, 1998). They used similar approaches to examine radiological diagnosis processes and radiologists' knowledge base although their research purposes were not limited to find the novice-expert differences. They derived similar findings about the processes of problem solving in diagnosing radiographs even if they used different types of difficult cases. In the processes proposed by Lesgold and his colleagues (1981, 1988), there are multiple steps, starting with the perceptual process and followed by cognitive processes. The cognitive process is triggered by the perceptual decision and may lead to more searching and other perceptual activities. Rogers (1992) proposed a model consisting of a perceptual process, a visual interaction process, and a problem solving process, which are all connected to working memory and long-term memory. Azevedo (1998) developed a seven step cognitive model, including visually inspecting mammograms, identifying and characterizing image findings, and providing a definitive or differential diagnosis. The models they proposed in their studies unanimously suggest that the perceptual process is critical because this process initiates the diagnosis, triggers higher levels of thinking, and provides both schemata and evidence for perceptual and cognitive decision making. Furthermore, they found that radiologists' knowledge base also contains a substantial perceptual component. The perception related knowledge includes various image features, anatomical structures, and image categories. Therefore, it is important for learners to develop the mental representations of anatomical structures and their

variations in different cases (Lesgold, Feltovich, Glaser, & Wang, 1981), including image patterns and disease categories.

In addition, some typical errors, including search errors, detection errors, and interpretation errors, also indicate the perceptual nature of radiographic image reading (Azevedo, 1998; Kundel, Nodine, & Carmody, 1978; Tourassi, 1999). According to the generic Analysis, Design, Development, Implementation, and Evaluation (ADDIE) model, it is critical to identify the performance problems for the design and development of instructional strategies, media elements, and other approaches (Alessi & Trollip, 2001; Gagne, Wager, Golas, & Keller, 2005). Previous studies indicate that the major performance problem in radiography performance lies in the limited attention and insufficient perceptual and conceptual knowledge and skills (Myles-Worseley, Johnston, & Simons, 1988; Sowden, Davies, Roling, 2000). Specifically, some typical errors include the following items (Tourassi, 1999):

1. *Some key features are missed because of the lack of attention;*
2. *Some features are missed because of misinterpretation of features;*
3. *Some features are missed because of the problem in searching.*

To solve performance problems, the proposed training methods need to engage learners in devoting their attention to detecting and discriminating patterns in radiographic images, constructing the meanings from practice, and becoming diligent searchers of features. This type of training methods can then cultivate deliberate practice and improve construction (Lesgold, Feltovich, Glaser, & Wang, 1981) and retention of knowledge. It can also provide a problem-solving learning environment for learning how to solve these image diagnostic problems and decrease errors in problem-solving.

Perception Models, Visual Literacy, and Visual Concept Learning

Just as the foundation of a house, perception supports further conceptualization and problem solving. Goldstone and his colleagues (1997) presented the traditional view of perceptual learning as the foundation of the other types of learning. They said, “In building models of cognition, it is customary to commence construction on the foundations laid by perception. Presumably, perception is to provide us with an initial source of information operated upon by subsequent cognitive processes. As with the foundation of a house, a premium is on stability and solidity. Stable edifices require stable support structures.” (p. 2). They maintained that traditional views of the stable structure of perception overlooked the flexibility property that perceptual systems may embrace. They suggested that perception functions as a bridge connecting the outside world with conceptualization of the world. Perception is flexible rather than rigid. Hence, instruction in mammogram reading is to construct the flexibility of perception to support problem-solving processes.

For the nature of perception in radiography interpretation, the studies in visual perception can guide this study. During the past two decades, researchers and scientists in psychology have developed explicit models and experimental designs on how neurons, neural circuits, and pathways work together and how human brains attend to stimuli, separate and integrate visual information, and solve perceptual problems (e.g., Biederman, 1987; Goldstein, 2002; Sanocki, 1991, 1993, 1998, 1999; Treisman, 2006). Importantly, they proposed diverse insights and evidence for us to understand how different perceptual and cognitive processes in time courses may influence the organization and segmentation of visual information. Their explanations using the

concepts of geon, structure description, salient information, and parallel and serial search indicated that human perceptual system is robust in abstracting incoming information, connecting with cognitive systems, and using adaptive search strategies. Instruction needs to provide sufficient activity spaces for learners to apply their natural abilities to learn how to solve domain specific problems.

The researchers also provided perspectives on the interaction between the world and human visual brain. They proposed that the internal representations of objects are important for object recognition. The representations are constructed through combined communicative efforts of many neurons and neuron networks. Furthermore, salient features of objects are related to representations and essential for solving object recognition problems. Researchers have different views about how global interpretations are computed from local fields or how pieces of information are grouped in human brains, but they have gradually found the soundness of an interaction model: separate brain regions need to communicate with each other for perception. The model implies that both bottom-up and top-down mental processes are important for perception. These models and studies can help understand the processes and tasks of mammogram reading. Moreover, they provide guidance for what goals a good instructional strategy needs to reach. For example, the strategy is supposed to activate learners in viewing across cases, selectively attending to salient features, making guesses about patterns, constructing internal representations, and continuously testing hypotheses between and across cases.

Furthermore, the visual literacy studies (Braden, 1996) are related to this study. Visual literacy is the competencies to read and write images and it is related to visual thinking, the ability to think in imagery (Braden, 1996; Wileman, 1993). The researchers

in this area emphasize the importance of visuals as cognitive and affective aids. They also stress the importance of teaching and learning how to read and write visual information. Decoding and encoding are two proposed approaches to improving visual skills (Heinich, Molenda, Russell, & Smaldino, 1999). This body of literature usually informs instructional designers of how to deal with instructional message or media design.

Researchers reported a couple of studies in how to present visual information in medical education (Kim & Astion, 2003) and if presentations might imply certain interactivity. Kim and Astion (2003) found that learners gained better scores by interacting with and comparing across images than just viewing images in a computer-based urine lab. Besides, they found that presentation mode of anchored images significantly increased learning than successive single image presentation mode and simultaneous double image presentation mode. As for successive or simultaneous presentations of visual concepts, inconsistent findings existed (e.g. Whiteside, 1987; Kim & Astion, 2003).

More specifically, Kim and Astion (2003) did a study examining how different types of presentations influenced learning. The major purpose of the study was to look for the statistical significance among three different kinds of image displays in computer-based instruction in affecting medical concepts: respectively images were presented in a single mode, side-by-side pair mode, or an anchored multiple mode. They tracked how learners used these different modes and found that the anchored multiple image mode was mostly used. Furthermore, the students who used this mode performed the best in their post-test, compared with those in the other two modes. After obtaining data, they also analyzed the performance differences between students who used the comparison

and contrast feature and those who did not. They found that those used the comparison and contrast approach did better in their post-test than those who did not use this feature. They then concluded that this feature could bring up statistical significance in learning outcomes no matter what kinds of presentations were used. However, the assessment of the study did not distinguish retention and transfer, so it is unclear whether the method is significant for learning transfer.

On the other hand, theorists of concept learning informed the necessity to improve the learners' ability to weigh the probabilities whether the sum of evidence matches the criteria in memory (e.g., Wattenmaker, Dewey, Murphy, & Medin, 1986). There are three types of concept learning theories, including rule-based theory (e.g., Bruner, Goodnow, & Austin, 1956), prototype theories or exemplar theory (e.g., Rosch & Mervis, 1975), and probabilistic theories (e.g., Wattenmaker, Dewey, Murphy, & Medin, 1986). They respectively emphasized learning rules, family resemblance, and sufficient attributes presented. In medical education, the commonly used teaching methods are teaching rules through instruction and teaching exemplars and features through case-based learning. However, there is no extant evidence or theory-based instructional design approaches to integrating these knowledge and skills. Teaching rules and exemplars is common in research and teaching, but how exemplars and rules can be constructed internally through learning tasks is unclear in former instructional theories and practice.

Sharples (1991) noted the existence of visual concepts in a broad range of domains and scarce research information on visual concept instructional methods for CBI. He extracted from the existing related studies (e.g., Stones, 1979; Tennyson & Park,

1980) guidelines that can be adopted in designing this study's materials, some of which are listed as follows:

Ascertain students' prior knowledge;

Explain the terms to be used in labeling the concepts and their attributes;

Start by showing a series of simplified exemplar images, with few and obvious attributes, to emphasize the critical attributes;

Provide a sequence of matched pairs of exemplar and non-exemplar images;

Provide feedback to the learner for each discrimination;

Provide suitable cuing to ensure that learners gradually become independent in their ability to identify novel exemplars of the concepts.

(cited in Sharples, 1991, p. 124)

Sharples explained and commented on the last principle, suggesting that “images with similar critical attributes are grouped together and there are explicit links between matched or related items”. He also evaluated the guidelines as “fairly clear and consistent” (p. 124). However, these studies have not offered evidence-based instructional activities for learners to become active participants, knowledge builders, and deep learning seekers. The guidelines for what to teach and how to present information cannot replace the evidence of what and how learners think and process information and construct knowledge. Therefore, it is necessary to further seek for theoretical and empirical evidence for effective instructional strategies that can enable learners to make good use of their cognitive, metacognitive, and affective resources.

Interactivity

Interactivity and Types or Dimensions of Interactivity from Different Perspectives

After a review of literature of the types of learning in radiographic images and related visual cognition and instructional research, it is necessary to examine interactivity and technology affordances for interactivity to enhance learning. Although different definitions of interactivity exist, interactivity in the context of technology-based learning can be defined as the “technological capability for establishing connections from point-to-point” (Wagner, 1994). The “point-to-point” in computer-based instruction can be explained as the interplay between the computer and the learner, learner and learner, the learner himself or herself, and the learner and the instructor. This interpretation broadens the scope of Jonassen’s interactive teaching (1985) and highlights the two-way nature of interactivity. It reflects the communication circles Moore described (1989) although they have different emphasis in terms of computer or content. Interestingly, the computer is not the content and vice versa. The computer actually needs to do more than merely present the content with the appropriate use of interactivity.

Researchers in different areas classified interactivity in different ways. According to Proske, Narciss, and Korndle (2007), multimedia interactivity has three facets: a technical dimension, a social dimension, and a mental dimension. Technically, multimedia interactivity refers to all of the features allowing learners “to search, locate, select, access, manipulate, document and save information” (p. 511). The social respect of interactivity provides learners opportunities to communicate with their instructors and the other learners. Importantly, they identified a mental dimension of interactivity, allowing learners to “process the learning materials constructively, engage in learning

activities actively and take control of their learning processes” (p. 512). These three dimensions of interactivity may provide designers a good tool for checking what type of interactivity they would like to adopt for their specific purposes.

A comprehensive review by Chou (2003) provided a big picture of different types of interactivity. Based on the basic types of interactivity in Moore (1989) and other researchers’ work, he created tables of 9 dimensions of interactivity, including choice, non-sequential access of choice, responsiveness to learner, monitoring information use, personal-choice helper, adaptability, playfulness, facilitation of interpersonal communication, and ease of adding information. These dimensions detailed the above-mentioned functions in the three dimensions and they can be sorted into the previous three groups.

Furthermore, another classification of types of interactivity for elearning objects (IEEE 1484.12.1-2002) may be interesting: the interactivity may be active, passive, or mixed. Active learning and expository or passive learning are characteristic of the former two types of interactivity. This definition may somewhat overlap the topic of levels of interactivity that will be covered later.

The Importance of Interactivity to Learners and Active and Meaningful Learning

Interactivity is one of the important design factors and constructive pedagogy approaches, regarded more important than content in impacting learning and learners (Draves, 2000). Interactivity can activate learners’ minds with engaging inquiries, feedback, reflections (Berge, 2002) and other strategies. These interactions can extend beyond trivia interactivity (such as clicking a menu) to manipulating objects, generating products, constructing understanding, and solving problems. All these possibilities may

probably lead to more engaged learners and their better performance in learning. For these reasons, Buckley and his colleagues (1999) explored interactivity as an instructional feature and maintained that interactivity fosters active learning. Interactivity is thus regarded as one of the key factors in designing constructive learning environments.

Interactivity contributes to learners' motivation, cognitive engagement, self-regulated learning, memory, and performance (Chung & Zhao, 2004; Matthews et al., 2007; Selcer, 1993). Learners tend to prefer the contents with interactivity to those without any interactive exchanges. While learners manage to respond to questions, manipulate objects, interpret data, and create their own representations, they use their prior knowledge and generate new knowledge and/or thoughts. Their thinking processes are actually activated through interacting with computers (Ridley, 2007). Based on their classification, Proske, Narciss, and Korndle (2007) described how they used these interactive elements in a Web-based learning environment called "Studierplatz". They found that not all of the students were serious about using interactive features in learning. The researchers found that using interactivity functions promoted achievements and was related to better learning. They also discussed how self-regulated multimedia learning, with interactivity as a major component, can be applied in higher education. However, a recent study (Kennewell, Tanner, Jones, Beauchamp, 2008) found limited achievements with interactive instruction. This case study of technology-based activities demonstrated that students had confusion about learning goals and objectives when they independently studied in such an environment. This confusion resulted in distractions in learning and decreases in performance. Although mixed results were found about the interactivity as a causal factor in significantly increasing performance, researchers have concurred that

interactivity is highly related to active learning and enhanced performance (Matthews et al., 2007). For these reasons, interactivity has also become an important criterion in evaluating educational computer courseware (Comer & Geissler, 1998; Laurentiis, 1993).

Levels of Interactivity and Technology Affordances

A pragmatic definition of levels of interactivity was provided by the Department of Defense (1996). Accompanied with the levels, engagement strategies and contexts were provided for the design of interactivity levels. The four levels of interactivity are passive, limited interaction, complex interaction, and real-time interaction. E-learning designers and developers tend to use the third level of interactivity, but they also apply the first two levels when appropriate. The purpose of providing a description of these levels of interactivity is to help organizations develop cost-effective programs because higher levels of interactivity imply higher demands in time, budget, and expertise.

Furthermore, researchers and practitioners provided various points of views on this issue. For example, in multimedia design, a wide range of visualization methods can make learning interactive, ranging from simple animation to visualization with input and zooming to learner generated visualization (Saddik, 2001). Another example is that cognitive interactivity was emphasized and regarded as more important than just clickable objects and other behavioral or functional interactions (Kennedy, 2004).

Existing authoring tools provide possibilities for these levels of interactivity. Chou (2003) in his review article gave some examples of achievable interactivity with computer-assisted instruction, communication technology, distance learning, and the Web. It seems that different technologies can be superior in some aspects but may be limited in the other respects. For example, communication technology may provide

complex functions in conferences but has limited capacity for developing simulated real-life experience with interactive learning contexts, objects, and tools.

In spite of the substantial literature in interactivity, researchers have seldom studied the pedagogical design of interactivity in radiographic image reading, but researchers reported the lack of computer-based training (CBT) methods in this area (Sharples, 1991; Twitchell, 2001). In practice, online courses and materials in medical education often adopt an information transmission model due to its pedagogical tradition of didactics (Gunderman & Chan, 2003). In radiology education, instructional technology is mostly regarded as a vehicle for delivering information rather than constructing knowledge (Gunderman, Kang, Fraley, & Williamson, 2001). Although some multimedia methods exist, such as tutorials, simulations, and games (Dee, 2002; Luo, Eikman, Kealy, & Qian, 2006; Roubidoux, 2005), they are still at the initial stage of development and validation of instructional strategies.

Effectiveness of Technology-Based Instruction

Instructional designers face serious uncertainties and ambiguities in their work. There have been arguments for and against significance and efficiency of computer and the Internet use in education, resulting in continuous discussions on significance studies and meta-analysis studies of the effects of computer-based instruction (CBI) or Web-based training (WBT) (e.g., Cuban, 2001; Clark, 1994; Kulik & Kulik, 1986, 1991; Kozma, 1994; Hannum, 2007). The other uncertain aspect for instructional designers is that few prescriptive methods exist in learning theories (Reigeluth, 1999) and instructional design models (Alessi & Trollip, 2001). Certainly, design-oriented theories, principles, and heuristics (Jonassen, 2008; Mayer, 2001; Reigeluth, 1999) have

complemented with the descriptive theories and offered guidelines for instructional design, including studies of interface and spatial representation design (e.g., Hilbelink, 2007; Grace, 2005). Furthermore, researchers have examined diverse multimedia methodologies, such as tutorials, hypermedia, simulations, and educational games (Alessi & Trollip, 2001; Javidi, 2004; Jonassen, 2004). However, these studies inform instructional designers of generally applicable rules of thumb and multimedia methods, they contain scant theory-informed micro-level empirical information of pedagogical effectiveness for CBI and WBT.

To study e-learning pedagogy, it is worthwhile to look back upon existing values and studies of instructional technology. The advocates of computer use in education proposed many advantages that technology may bring about to education (e.g., Alessi & Trollip, 2001). The advent of the Internet promotes the accesses to information and information evaluation is commented as crucial to learners. For this reason, computers and the Internet are suggested as the approaches to bringing up critical thinkers and problem solvers. Computers have been further valued as cognitive tools or partners in learning (e.g., Lajoie & Azevedo; Liu & Bera, 2005). Importantly, the proponents maintain that computers can improve students' learning achievements than traditional instruction (e.g., Alessi & Trollip, 2001).

However, concerns about these technology innovations were shown in Cuban's arguments (1986). Cuban argued that technology use in education has put much pressure on teachers and schools. They have to deal with hardware and software issues, including their complexity, incompatibility, and development. The researcher recognized a variety of challenges that instructors may come across in applying and integrating technology,

which initially indicates that more efforts in instructional design and its research are necessary than ever before in the era of technology.

While the arguments of values of IT suggest further research, existing significance studies have yet to start addressing the challenges of instructional design. In no-significance reports throughout these years, researchers and educators tried to compare computer-based instruction and traditional teaching to see outcome changes, resulting in many mixed results or no significance findings. In the meta-analysis of these findings in previous research, researchers reported small effect sizes from computer-based instruction as 0.32 (Hattie, 2004), 0.26 (Kulik & Kulik, 1986), and a varying range of 0.22 to 0.57 (Kulik, 2003). Although the effect sizes from these studies did not demonstrate the promising respects of IT, but it suggested that potential do exist and deserve attention for further research.

Seeing the uncertainties and possibilities of the values and learning outcomes of e-learning programs, researchers will get interested in instructional design research to help solve problems. However, the challenges are increased because of the characters of e-learners and e-learning environments. Technology-based learning environments are characteristic of voluntary participation and independent self (Davidson-Shivers, 2002; Gagne, Wager, Golas, & Keller, 2005; Mayer, 2001). Without active engagement, learners can go to online courses without paying any attention to what is learned, wandering around, losing interest, and abandoning their studies. Although they may have tests that force them to study more, they may still easily lose their mental participation in the sea of information. Nowadays, with the development of the Internet and authoring tools such as the learning management system Blackboard, huge amount of information

is poured into online course shells. Since students have computers of largely increased memory capacity, it is easy for them to cache the bulky materials and download the online materials before they go through them (Mayer, 2005; Young, 2003). Large amount of information without mental participation may lead to rote memory and discrete information but not knowledge, understanding, and problem solving skills (e.g., Jonassen, 2004; Bruning, Schraw, Norby, Ronning, 2004).

To help increase learning, it may be helpful to look at how traditional instruction addresses the too much information issue in didactics (Gagne, Wager, Golas, & Keller, 2005; Jonassen, 1999; Mayer, 2001). In traditional learning environments, the instructor may continuously use learning tasks to activate the students' minds and students may answer different types of questions from the instructor. This type of interactions may stimulate the students' minds and they become engaged in learning. In addition to the questions, the instructor may use many other strategies to engage learners. Some other approaches include: to ask students to explain a phenomenon, to critically comment on a situation, to integrate what is learned, and to question some confusing points. Teachers seem to have many ways to activate students and they may use these approaches in high frequency in teaching.

However, in computer-mediated mammogram learning environments, such activities are far from sufficient (Gagne, Wager, Golas, & Keller, 2005; Jonassen, 1999). Existing online programs use such authoring tools as Dreamweaver and Powerpoint to provide lists of bulleted point information and images to learners and the key points may help learners obtain the major points in their readings. These programs also pay attention to the use of graphics and the other media elements to attract learners' attention.

However, the engagement level of this type of devices is usually unsatisfying, especially when the information volume is enormous. Hence, learners also need other types of approaches to keeping their minds on what they learn and achieving their learning goals.

Given the features of online learning and digitizing process in education and the characteristics of e-learning and e-learners, instructional design and technology researchers need to conduct studies in pedagogical effectiveness of online learning to examine instructional strategies (Jonassen, 2004). However, a widely existing misconception is that online learning automatically makes learning effective. This is why, in the past, many online materials were developed without considering the information processing processes of online learning.

Specifically, health sciences instructors have developed their own teaching methods, such as case-based learning (Kim, et al., 2006; Luo, Eikman, Kealy, & Qian, 2006) and problem-based learning (Norman & Schmidt, 2000; Visschers-Pleijers et al., 2006). In technology-based instruction of radiographic images, such instructional methods as tutorials, simulations, and games have started to be used in technology-based programs (Luo, Eikman, Kealy, & Qian, 2006). However, these teaching methods are still at their initial developing stages in terms of their instructional strategies and corresponding research.

Instructional Methods in Radiology Education

Existing Methods and Desired Ones

Radiologists introduced that the traditional teaching methods in radiology education include conferences (in the formats of lectures and case presentations), one-on-one teaching, small group instruction, and self-study (such as textbook reading, teaching

files, and educational software programs) (Collins, Blankenbaker, Albanese, Stack, Heiserman, Primack, & Kazerooni, 1999). Explicit instruction, such as didactic conferences and presentations, is one of the main approaches to all subspecialties of radiology (Roberts and Chew, 2003). Radiology educators encourage students to adopt the formats of self-study, which can save faculty time, be more flexible for students, and be closer to what students need to do in their professional life (Collins, Blankenbaker, Albanese, Stack, Heiserman, Primack, & Kazerooni, 1999). The authors also suggested the importance of cases in resident education and cited that the Accreditation Council for Graduate Medical Education (ACGME) emphasized the availability of various teaching file cases to students (cited in Collins, Blankenbaker, Albanese, Stack, Heiserman, Primack, & Kazerooni, 1999). In spite of the existing methods, evidence is lacking about their effectiveness for IT-based learning. The authors did not mention what methods are effective and how these methods are implemented and evaluated in technology-based instruction.

Other researchers also described and commented on the instructional methods used in radiology resident education. In one of the studies, researchers listed four types of methods and highly recommended preview activities (Deitte, 2006). According to Deitte (2006), when residents preview images before conferences with the other residents and radiologists, active learning occurs. On the other hand, when radiologists lead conferences without any preview activity for students, learning was called passive. Therefore, active and passive learning exist in radiology education because of different instructional methods. In commenting on traditional methods of teaching, researchers have pointed out the necessity to change the existing “passive viewer” syndrome

indicated in Jameson, O'Hanlon, Buckton and Hobsley's article (cited in Tachakra & Dutton, 2000).

Although these authors did not have evidence for their opinions, they did give thoughtful suggestions for designers and researchers. First, learners' mental participation in image observation is critical for learning efficiency. In the above-mentioned methods used in radiology, researchers claimed that learners' mental processes may decrease when they merely receive instructors' or the other learners' findings (Deitte, 2006; Tachakra & Dutton, 2000). From appearance, instructors take their responsibilities of teaching but their expository teaching methods may leave students less efforts and less active participation in learning. The other extreme maybe totally leave students alone with little guidance and feedback, in which students may also decrease their participation in learning because feedback and guidance was found to influence motivation and achievements (Terrell & Rendulic, 1996; Mory, 2004). Second, too much information without instructional values may result in decreasing participation. Learning is a process and expertise is developed through participation and guidance, especially at the initial status of expertise development. Therefore, online resources might become somewhat overloads for learners with few engaging methods to increase mental participation. Third, the discussion of existing teaching methods in radiology provides a framework for understanding instruction but do not have been evaluated in technology-based instruction. Hence, instructional methods need to be studied for evidence and future applications in technology-based instruction.

Innovating Instructional Methods

Academic radiologists are innovating traditional didactic teaching methods and developing engaging instructional strategies to let individual learners solve problems and increase their mental participation. Chew (2001) proposed a revised teaching method: conference with previewing cases and filling in answer sheets. The author pointed out the existing problems in the case conference, sometimes in the form of a hot-seat conference. The unknown case study in the hot-seat conference was regarded as the main traditional method in radiology resident education, but this method may cause problems. For example, the discussants may merely stare at the image with little thinking when they look at the image and talk about it because of the unknown nature of the image. The reason may be that they do not even have time to perceive and analyze the image before the presentation. Therefore, the author suggested that every attendee of the conference preview the case, make one's own diagnosis, look at it the second time, and complete one's answer sheet. Five conferences of this new format were evaluated through surveys. The evaluations indicated that 98% of the attendees preferred the new approach to the former ones and 99% of the respondents desired more of such conferences. The results from the study imply the importance for the individual learner to preview cases and solve problems by oneself before presentations and explanations. However, there has been no learning outcome evidence for this method.

The other researchers also maintained that instructional methods in radiology resident education could be improved. Deitte (2006) pointed out instructional problems that are worth further studies. For example, the author noted the lack of study efforts for the weaknesses of Picture and Archiving Communication Systems (PACS). Although

multiple researchers found that PACS can improve educational efficiency, there is a shortage of studies examining its effectiveness in impacting actual learning. The author cited Redfern, Lowe, and Kundel's study (Deitte, 2006) that reported the decrease of residents' "autonomous participation" "in image interpretation from 38% to 17%" while "the workload increased by 33%". (p. 530). Furthermore, the author presented his observations of instruction in his department: Two of the problems are "increased passive learning due to the impact of group reading" and "decreased feedback secondary to 'remote' reading" (p. 531). He then claimed the role change of radiology residents from active to passive due to the transition from a film to a digital image department. He explained that the "preview-review-dictate" model used in the film age was thought as promoting active learning because the preview activity involves active learning and feedback is provided through radiologists' follow-up interpretation. He also defined a list of methods that are used in resident education with PACS, including preview-review-dictate, review-dictate, group reading, parallel reading, and remote reading. The author pointed out that the group readout sessions with radiologists leading reading might decrease radiologists' time in instruction, but may result in passive learning. Therefore, the author suggested continuing to enhance learning with the method of preview-review-dictate, encouraging students to view images in dictation, increasing feedback, and encouraging self-directed learning.

Roberts and Chew (2003) reviewed the teaching methods commonly used in resident education. The reviewed methods are case conferences, didactic conferences, self-teaching files, textbooks and journals, clinical teaching and preparation for call, and residents as teachers. The case conference was defined as "a group teaching method in

which the moderator of the conference presents a case to a discussant. The discussant performs the traditional radiological thinking process by identifying the modality and technique, identifying the relevant positive and negative findings, listing a differential diagnosis, narrowing the differential diagnosis, and giving a best diagnosis, if possible. The educational value of the traditional case conference is highly variable; in the worst circumstances, the discussants find it too stressful to perform, the moderator becomes frustrated, and the audience grows uncomfortable and learns nothing. To ameliorate these problems, the case conference may be modified in a number of ways” (S97). They gave examples of different types of case conferences. They admitted that teaching techniques can improve residents’ confidence and competence in spite of small changes in the techniques.

Different from the other authors, they described about 5 variations of case conference in details. One type of case conference allows residents to preview images, similar to what Chew (2001) and Deitte (2006) proposed. Residents are able to view images, make their own diagnosis, and examine topics in depth. Interestingly, in the other type of the variations of case conference, students were provided with two cases simultaneously presented and asked to compare and contrast the two images. Each is allowed to make one comment upon the case about the similarities and differences of the case until information exhausted. The authors maintained that “all residents participate and are engaged with each case” (S98). In spite of this method preference shown among residents, it seems that few researchers have ever managed empirical studies on this instructional method in radiology.

However, the strategy of comparison is a recommended clinical problem solving strategy and the evidence of its effectiveness was found in this context in a recent study. Roelofs and his colleagues (2007) did a study examining the influence of prior mammograms upon performance of screening mammograms. In their study, experienced radiologists read mammograms in two different reading conditions, with the prior mammograms provided in one session and without these images available in another. In addition, the researchers also combined these two reading sessions to compute the performance when images were only available by request. They found that performance was significantly better in the reading session when prior mammograms were available, followed by the session when prior images were provided when asked for. The performance in the reading session without prior mammograms available was found significantly lower than the other two conditions. It seems that comparison can give confirmation to the recognition and interpretation. Therefore, the comparison method can be a beneficial strategy in improving performance in clinics. The use of this strategy as an instructional strategy may also be helpful for identifying patterns more accurately.

Generative Strategies

The previous sections of the literature review demonstrate the necessity and urgency of learner-centered learning activities for radiographic image study. The activities need to enhance cognitive participation and knowledge construction through constructive learning, efforts made, assimilation and accommodation, focused attention and increased interactivity. With the prescribed affordances, the activities can foster self-directed learning in CBI and WBT.

Generative strategies (Grabowski, 2004; Mayer, 2005; Smith & Ragan, 1993; Wittrock, 1990, 1991) can be powerful methods that satisfy these learning needs and offer theoretical and empirical evidence for this argument. Generative strategies were found effective in studies grounded in both generative learning theory and generation effect theory. In the remaining parts of the literature review, there will be a close examination of these studies and theories.

Generative Learning: the Theory and Evidence

Constructivists suggested the importance of constructing knowledge from experience and prior knowledge by learners rather than transmitting knowledge by instructors (Dewey, 1902; Jacoby, 1978; Jonassen, Strobel, & Gottdenker, 2005; Knowles, 1998; Mayer, 2001; Piaget, 1970; Vygotsky, 1986; Wittrock, 1974, 1990, 1992). Grounded in his constructive view and findings in neuropsychology and empirical studies (Grabowski, 2004), Wittrock (1974, 1990, 1992, 1995) proposed and tested generative learning theory and corresponding activity-based instructional strategies. During more than twenty years, Wittrock and colleagues have found substantial evidence of the effectiveness of these strategies in different subject areas, including reading, science, and economics.

As a functional model, generative learning theory and its corresponding generative teaching model help instructors design and develop meaningful learning activities and satisfy the needs sought for in this literature review. Generative learning activities are the learning activities that engage learners in comprehending learning materials with deep understanding as an outstanding learning outcome. Wittrock (1990) summed up two types of generative learning activities: some generative activities can

help construct relationship between the information in environment and other information in environment, including titles, questions, concept maps, graphs, scripts, main ideas, summaries, outlines, and so on and so forth; the other activities can help generate relationship between information in environment and prior knowledge and experience, such as examples, predictions, applications, metaphors, inferences, interpretation, and analogies.

Generative strategies can promote deep learning and generative learning is learner-centered and learning-centered. Learners are presumed as active participants rather than passive receivers in generative learning theory. To help learners make sense of experience and respond to what is perceived, generative strategies engage learners in four generative learning processes, including motivational processes, learning processes (such as attention), knowledge creation processes (such as preconceptions, concepts, and metacognition), and generation processes (Wittrock, 1990, 1992). Among these four processes, generation processes are crucial for generating relationship between information in environment as well as between information in environment and prior knowledge and experience. The purposes of the generated relationship are elaboration, reconceptualization, organization, and reorganization, which lead to comprehension. Therefore, the former three processes seem to be the basis of the generation process while the generation process is built upon the former processes, essentially reaching the learning goal of comprehension.

Generative strategies enable conceptual change in learners. It was found that learning can occur when learners actively participate in generative activities because these activities can activate the above-mentioned thinking processes and enhance

understanding. The activities can help learners selectively attend to what is learned and actively construct meanings and build mental models (Grabowski, 2004; Wittrock, 1990, 1991). Learners can be motivated to encode and organize their new knowledge as well as create meanings between their prior knowledge and newly learned knowledge (Wittrock, 1990, 1992). This emphasis on generating relationships and meanings is congruent with the most current neuropsychological findings in the interactions among different parts of the brain (Goldstein, 2002). It is also consistent with the fundamental theory of constructing knowledge through assimilation and accommodation (Piaget, 1968; Winn, 2004). The strategies are coherent with and applied in instructional theories and principles, such as conditions of learning (Gagne, Wager, Golas, & Keller, 2005) and instructional strategies for concept learning (Smith & Ragan, 1993).

Wittrock and his colleagues (1974, 1990, 1992, 1993) found substantial evidence of the effectiveness of these strategies in different subject areas, including reading, science, and economics. These experimental studies showed that generative learning activities can significantly enhance learning. These studies had power because of their large sample sizes, levels of significance, and effect sizes. Large sample sizes help improve the probability of rejecting the null hypothesis when it is false and decrease type II errors (Glass & Hopkins, 1996). In these studies, the level of significance (α value) was usually set at .01 or .001 and considerable percentage gains in tests were identified in these studies. Computing the effect sizes of these studies with Cohen's approach ($d = \frac{M_c - M_c}{SD}$), the researcher found that their effect sizes, the magnitude of differences, were large (>0.8).

However, generative strategies, such as paraphrasing, explaining, outlining, summarizing, and creating main ideas mainly deal with declarative learning or text reading. To enhance visual rich type of learning, these generative learning activities for texts need to be expanded. Admittedly, the generative strategies for learning texts, such as inferences, predictions, and examples, may be appropriate for image study. For example, examples are widely included in textbooks and cases are the main themes of clinical studies in radiology. As for inferences and predictions, clinical studies may contain similar activities to them because they are close to thinking processes in radiology detection and diagnosis. Therefore, they may be used in learning images. However, they may be insufficient to engage novice learners in mental participation because inferences and prediction activities seem to be somewhat difficult for those learners who have little prior knowledge. Furthermore, understanding texts and recognizing visual patterns are different types of learning outcomes. According to instructional design theorists (Gagne, Wager, Golas, & Keller, 1992, 2005; Smith & Ragan, 1993), instructional methods need to align with instructional goals and outcomes. Generative learning strategies seem to be close to the learning objectives of meaning seeking for recognizing visual patterns in studying radiographic images, but such specific tasks as summary and outline are typical text rather than image comprehension activities. Hence, it is necessary to develop new generative strategies for image study.

Generation Effect

The Theory, Evidence, and Interpretation

Generative learning theory is closely related to or includes another evidence-based theory, called generation effect theory (Slamecka & Graf, 1978). Generation effect

is an evidence-supported hypothesis that learner-generated stimuli can be better retained than experimenter-provided stimuli. The theory shares the active versus passive learning assumption with generative learning theory. Specifically, both theories emphasize the role of learners in learning as participants and the process or approach of generation to increase learning. One of the major differences are that the two theories emphasize different thinking processes and types of learning, with generative learning theory and activities stressing comprehension of texts and generation effect theory and tasks focusing on encoding of words and pictures. In the process of creation, validation, and extension of the generation theory throughout decades, it has been found robust with continuing empirical data to support and revise the theory but keeping its original flavor.

Slamecka and Graf (1978) first observed that learners remember words better in generating the verbal responses than merely reading word pairs. In a series of studies of generation effect, five experiments were conducted to examine the possible influence of generation versus reading method and other factors. The other independent variables beyond generation versus read they tested include the timed versus self-paced presentation rate, different generation rules, informed versus uninformed about a test, and the stimulus versus responses study conditions. The dependent variables they examined are recognition and recall test scores. A general procedure of these experiments with generation versus read variable was that subjects were provided with tens word pairs with or without responses. For the generation treatment, only the stimuli were given and subjects needed to produce the responses themselves based on the rules they were provided. In the reading treatment, both the stimuli and responses were provided to the subjects and the subjects were asked to read them. For example, one of the word pairs in

the generation treatment was rapid-f while it was fully spelled out as rapid-fast in the reading condition. They found significant differences between the generation condition and the reading condition in both recognition and recall tests. They also ruled out the possible influence of the other variables mentioned previously. Therefore, their experiments basically established the effect of learner-generated verbal materials upon recall and recognition.

The initial efforts of this study were obviously significant in identifying, analyzing, and testing this memory phenomenon. The researchers left a legacy of generation effect theory and delineation of experimental approaches to generation effect. Furthermore, they proposed interpretations of this effect. They explained that generation implies deeper or more elaborate processing that leads to better performance because deep processing focuses on meaning and leads to memory (Craik & Lock, 1972). In addition to these two explanations, they also confidently suggested an encoding distinctiveness of the relationship between the stimulus and the response. Distinctiveness of encoding means that distinctiveness of information makes it memorable (Jacoby & Craik, 1979), implying that learning materials requiring decisions in encoding result in recall of the material (Jacoby, Craik, & Begg, 1979). Besides, they argued that the initial recall in generation might substantiate better recall results in tests and they noted this justification as the least possible reason for the effect. In addition, they recommended remaining questions to be solved that may influence the deep processing explanation. One question was why this depth of processing explanation worked with the response rather than the stimulus, which was reflected in experiment 3. The other question was why the rhyme rule was not singled out as one of the significant methods although it

seemed to produce a shallow level of processing in the generation condition. Another question is that the mental act of generating might probably contain a higher level of processing than the act of reading. However, they admitted that no existing theories could support this speculation.

The contemporary of the above two researchers (Jacoby, 1978) also reported their finding that solving a problem improves retention compared with being provided with the solution and remembering it. In this study, two experiments were conducted to examine the phenomenon of generation effect. The method of experiments was that subjects were asked to complete a crossword-like puzzle (e.g., foot s _ _ e) or just read the word pair. As the previous experiments by Slamecka and Graf (1978), significant findings were also reported for the generation group when comparing with the reading group. However, only recall tests were used for criterion measure in this study. It seems that this study was less complex and analytical than the previous one, but it has its own features. In comparing the generated and immediately provided solutions, the study tested spacing effect and the factor of difficulty level of the problem. The construction group was found performing better in recall tests than the reading group. In the first experiment with spacing versus immediate variable, the spaced construction condition made the highest gain among the six conditions. In the second experiment, even the easy problem condition resulted in significantly better recall scores than the corresponding reading condition, implying that generation effect is robust even for easy problems. Therefore, no matter how easy the problem could be, it seems that construction processes are likely to increase encoding performance than just remembering the solutions directly.

On the basis of these two original studies in generation effect, researchers in psychology replicated and generalized the results to other populations, learning areas and materials, generation tasks, and memory tests. Some new learning areas are math (Crutcher & Healy, 1989; Gardiner & Rowley, 1984), non-words (Johns, & Swanson, 1988), and pictures (Carlin, Soraci, & Strawbridge, 2005; Kinjo & Snodgrass, 2000; Peynircioglu, 1989). No matter what rationale researchers used to explain generation effect with verbal materials, the causal effect of generation does exist in those contexts. Interestingly, researchers hesitated to examine this effect with pictures, maybe because of the more random features of pictures or the existence of picture superiority studies (Paivio, 1990; Reiber, 1994). If words were generated and retained because of semantic meaning connections, pictures seem to relatively lack in these connections. Furthermore, if pictures were superior in helping memory, learners may not need to make effort to retain them. However, one of the main explanations on generation effect is that the learner may exert more effort in generative learning, so they can retrieve stimuli better. This conflict may somewhat explain the delay of investigation of generation theory in learning images.

In spite of much less studies evaluating generation effect with pictures, recent literature in psychology did provide some evidence of generation effects with pictures (Carlin, Soraci, & Strawbridge, 2005; Kinjo & Snodgrass, 2000; Peynircioglu, 1989). Peynircioglu (1989) seemed to conduct the first study to evaluate the hypotheses of generation effects with line drawings of common objects and scenes in the first two experiments and nonsense pictures in the latter two experiments. In the first experiment, the subjects in the experimental treatment were given a name or description of a picture

and were asked to draw the picture according to the name or description provided. Those in the control condition were given a drawing and its name or description and asked to rate the artistic value of the picture. In the second experiment, a copy condition was added to the draw and rate conditions. With nonsense figures, the third and fourth experiments tested both generation and semantic activation hypotheses by comparing copying and drawing conditions as well as tracing and drawing conditions respectively. It was found that drawing according to description caused significantly higher recall scores than copying or looking at pictures. The initial validation of generation effects in this study led to a couple studies with pictures. Kinjo & Snodgrass (2000) did two experiments with two treatments of naming complete pictures in the name condition and naming fragmented pictures in the generation condition. They found the effect in three outcome measures, including free recall, yes/no recognition, and a source-monitoring task. More related to the proposed study, Carlin, Soraci, & Strawbridge (2005) used the flicker method as a generative strategy and compared the effect of generative search for scene changes and passive search upon memory. They found a significant difference of recall in generative search, and they reasoned that guesses generated in generative search for changes can function as retrieval cues. They proposed that the flicker method can be promising for computer-based learning environments. Therefore, these researchers developed a new approach to testing generative effect with the flicker task in learning pictures. A detailed review of the flicker task will be conducted in the next section.

The reasons for generation effect were explained with such theories as semantic coding, cognitive effort, multiple factors (Kinjo & Snodgrass, 2000), multiple cues (Soraci, Carlin, Chechile, Franks, Wills, & Watanabe, 1999), distinctiveness of the

solution (Begg, Snider, Foley, & Goddard, 1989), transfer appropriate processing (e.g., Bruning, Schraw, Norby, & Ronning, 2004), and “aha” effect (Auble, Franks, & Soraci, 1979; Wills, Soraci, Chechile, & Taylor, 2000). Some newly tested factors include source memory and implicit memory. The tests in these studies include free recall, cued recall, and recognition tests, but transfer tests have never been considered. The analysis of generation effect was extended from the memory of responses to that of cues and some other context factors. Generation effect was identified with these generation tasks in these different areas and most of the tests.

Generally, these studies of generation effect with pictures were well controlled and clearly defined. First, they all designed or replicated the generation rules for generation to happen. Second, the samples of the experiments were randomly selected and/or assigned to decrease bias and they were laboratory and well-controlled studies. Nevertheless, these studies focused on one image rather than a series of related images and the images were everyday objects and scenes but not complex images in any professional and academic domain. In addition, the learning outcomes measured in these studies remained the focus on memory but did not include problem solving and conceptual learning. Of course, memory plays an important role in learning, being the foundation of all types of learning, especially in such an area as medicine where similarity-based decision making is essential for diagnosis. Furthermore, memory is critical for pattern recognition and conceptualization in radiology education because schemata or mental models are formed through interacting of memorized or internally represented images. In addition, deep processing resulting from generation can facilitate

learning to transfer what they learn to new contexts (Toth, Reingold, & Jacoby, 1994). Thus, generation effect theory and evidence can be useful for improving image learning.

However, it is unclear whether the treatments used in generation effect can be used in radiograph reading to enhance image learning. Compared with the pictures of everyday objects, radiographs are more complex. The figure and ground of radiographic images are difficult to be segmented, and image features may be hidden, overlap with other anatomical features, and have low contrast information. If instructional methods could sharpen their eyes and let learners see more of these features in their minds' eyes, learners might have better visual memory of the images. With deep processing of meaning through generation, transfer of knowledge to new contexts may occur.

Flicker Effect

The flicker paradigm as a generative strategy (Carlin, Soraci, & Strawbridge, 2005) was originally used to test the role of attention in change detection. Rensink and his colleagues (1997, 2000) did a series of experiments with the flicker task, in which the original view of an image and the modified one of that image were flashed and alternated with a blank screen in between. The change on the modified image can be any type of these changes: a color changes, an object disappears, the location of an object changes, or any other object or dimension of features changes. Researchers found it difficult for subjects to detect the changes because of the lack of attention, and called this phenomenon change blindness (Simon & Levin, 1997). Furthermore, it was explained that it may be easy to recognize objects in a scene, but memory for the objects and the scene is transient and vulnerable (Simon & Levin, 1997). The results from the studies imply that visual memory is limited and decays in a brief time, and what is temporarily

held in visual memory will be gone with the object that disappears. The results also somewhat explain a former proposition of the illusion of the unending availability of the outside storage of the visual world (O'Regan, 1992). The illusion suggests that the visual system assumes that the stimulus in the environment will remain available so it is unnecessary to attend to and retain that information. As a result, it is natural for the visual system to rely on the outside world and look at things without consciousness.

The flicker task was first applied as a generative strategy in a study testing how different encoding methods led to the differences in recall and recognition of scene changes among groups of subjects varying in age and intelligence (Carlin, Soraci, & Strawbridge, 2005). The two treatments in this study were the flicker task for generative encoding and no flicker task for receptive encoding. The only difference between the flicker treatment and the no flicker treatment was the omission of the interruption of the blank screen in the no flicker task. It was found that all groups did better in free recall with the flicker treatment than no flicker treatment. The reasons for this significant difference were attributed to multiple guesses/solutions, the distinctiveness of the final answer, and the transfer specificity with the flicker task.

The results from these studies imply that the flicker task can be a robust generative learning strategy for improving radiographic image reading. The flicker task can optimize the internal processes in image reading because it helps draw learners' attention to images, form internal representations, involve learners in comparing the internal representations with the external representations, and continuously encoding in comparing images. All these cognitive processes satisfy the needs of studying radiographic images. Furthermore, metacognitively, the brief self-assessment and

feedback after the task can provide a moment for thinking about thinking and reflection, which can enhance deep learning. In addition to these cognitive and metacognitive processes that the flicker task can stimulate in learners, the task also implies challenges and discoveries for learners. The problem in the task can motivate and engage adult learners because they like problem solving (Knowle, 1990, 1998). In this specific problem solving situation with the flicker task, the hit of the right solution by selecting among multiple guesses may bring an “aha” moment of internal cheers for discoveries (Carlin, Soraci, & Strawbridge, 2005; Wills, Soraci, Chechile, & Taylor, 2000).

In the following sections, a detailed analysis and evaluation of the flicker paradigm for image study will be presented with theoretical and empirical evidence. The flicker strategy will be compared with the no flicker strategy and compare and contrast strategy in terms of stimulating and engaging learners in the cognitive, metacognitive, and affective processes of studying images. Meanwhile, there will also be explanations of how a certain learning process may be achieved through the flicker task. In addition, some arguments and evidence support why a certain process is important to learning images.

The Flicker Enhances the Cognitive Processes in Studying Images

Attention

Attention is critical for learning. It is remarkable that attention is listed as the first event followed up by other events in one of the classic instructional design principles – nine instructional events (Gagne, Wager, Golas, & Keller, 2005). To draw learners’ attention, instructional designers tend to use techniques, such as animation, humor, eye-catching pictures, and audio. However, it was found that these media elements may

somewhat distract learners from approaching and delving into the real learning goals because they are irrelevant or too intense perceptually (e. g., Mayer, 2001; Rieber, 1994). For example, animation or music may not reflect learning content, misleading attention to something else rather than what is learnt. On the other hand, media elements may be so strong that learners' attention may be exploited at the beginning of learning.

Attention plays an important role in perception and recognition. According to information processing theory, attention works like a bottleneck, which confines the amount of input information that is processed (e. g., Friedman, Polson, & Dafoe, 1988; Spear & Riccio, 1994). This implies the importance of guiding limited attention to learning goals. Specifically, the perception process that Goldstein depicted (2002) informs that attending to the outside world stimuli goes before perception and recognition, so attention is essential to image learning. Other psychologists, through experimental studies, found that attention is the key to perception. One of the first studies in attention could be the cocktail party study. Besides, other studies on inattention blindness (Mack & Rock, 1998), attentional blink (Shapiro, Arnell, & Raymond, 1997), and change blindness (Levin & Simon, 1997) indicate the role of attention in processing visual information. The findings of failure in attention and detection reported in these studies reflected the demand for focused attention in perceptual tasks as well as imply that different tasks require different amount of attention. Interestingly, psychologists (Treisman, 2006) provided evidence through experiments about the phenomena of both the limitations and robustness of attention. The researchers then tried to coordinate these findings and explained that the attention window can be flexible in observing the outside world and it can be focused or wide, adapting to the task demands and other conditions.

Particularly, attention plays an important role in radiographic image reading. Reading radiographic images is to detect and recognize the abnormalities among different organ features. The imaging method leaves difficulties to readers. There is much overlapping among tissues, so some features may be hidden and occluded by the other features. Furthermore, some other features may look like each other. Then it is important for image learners to overcome these difficulties. If through diligently working on images with vision, the reader can pay attention to these parts that may miss with little attention, readers' visual systems then may become more acute to identify image varieties, differences, and ambiguous visual information.

Interestingly, according to previous studies on the scanning patterns of experienced and less experienced radiologists, it was found that less experienced readers tend to more actively compare the side-by-side left and right organs than experienced observers (Azevedo, 1998; Lesgold, Rubinson, Feltovich, Glaser, Klopfer, & Wang, 1988). The reason for this difference was attributed to the fact that the less experienced readers cannot discriminate the findings from the other features and the background. That is, novices need to build up their knowledge and skills in discriminating features. In the flicker task, overt attention is invested on different anatomic objects in different parts of an image. With this task, readers can actually experience comparisons across different image features so that they can get familiar with different features, tell differences among patterns, and improve their discrimination skills.

Therefore, not all visual tasks can engage the same amount of attention in perception. For example, parallel search need less attention than serial search. According to Treisman's Feature Integration Theory (FIT) (cited in Goldstein, 2002), parallel search

is a typical pre-attentive task due to the pop-out of the searched objects while the serial search is more attentive search with one-by-one fixation upon objects. Therefore, more attention is needed in serial search than in parallel search.

The differences between parallel search and serial search are represented in the flicker and no flicker task. The flicker task is a typical serial search task because visual memory is limited and tends to start to decay 0.5 *seconds* after the image disappears. With the blank screen in between the two images for 0.5 *seconds* or longer, viewers cannot access to what they see before. Therefore, viewers need to search for the change item by item and use serial search. However, in the no flicker treatment, the change will pop out, somewhat like a simulation of parallel search, in which the target directly pops out. In the compare and contrast treatment, learners first need to do serial search to identify similarities and differences between the two images but the difference may pop out as the eyes go between the two images, which may result in parallel search. Therefore, more attention needs to be invested in the flicker treatment than in the other two treatments. Thus, the flicker task can draw more of learners' attention to images than the other two treatments.

Therefore, here are some of the possibilities or methods that can improve attention to the patterns of images and the flicker task is such a task. First, searching serially can make the objects directly observed. Focused attention is to put the observed target directly onto fovea rather than using peripheral vision. This is why overt attention and covert attention are distinguished according to fovea vision and peripheral vision. Overt attention is the attention gained with more awareness and direct fovea vision while the covert attention is the attention through indirectly looking at the target and less

awareness of the target. Serial search can result in overt attention to the object. The flicker treatment can lead to serial search and overt attention. Second, when sensitivity to the target is important and when noise may disturb vision, overt attention and fovea vision are important. For radiograph novices, overt attention to possible targets is important because this helps them improve their sensitivity – their ability in distinguishing signals from noise. Overt attention can help them compare across different possibilities and also compare the possibilities to the rest of the structure of the organ to construct meanings. Third, gaining learners’ attention is important in instruction and image learning may need more attention than learning the other subjects. On the top of the nine events of instruction lists “attract learners’ attention” (Gagne, Wager, Golas, & Keller, 2005). They also gave some brief suggestions about how to attract learners’ attention. However, the usual attention attraction devices in instructional design depend on visual display, for example, animation and attractive visuals are recommended to attract learners’ attention. Actually instructional strategies can help gain and regain learners’ attention at the beginning and in the process of learning. This function of instructional strategies seems to be achieved through some tasks and problem-solving situations. For example, one instructional strategy is to give students a few minutes in the middle of the course and ask them to write down a summary of what they have learned till this moment. When the task engages learners’ mental participation of generating the summary, the learners then devote their attention to what they are learning and try to recollect their minds in this situation. So, instructional strategies can help learners gather their attention and focus on the learning goals. Furthermore, the flicker treatment does not only draw attention at the beginning of the task, but continuously direct learners’

attention to problem solving by letting them compare the two images to locate the change, present their finding(s) by selecting the right target and going back or forward to another similar task.

The flicker paradigm can be more effective in guiding learners' attention to learning images than the no flicker method (Carlin, Soraci, & Strawbridge, 2005) and the conventional compare and contrast method whereas the conventional method can better enhance learners' attention than the no flicker method. First, the flicker treatment can draw learners' attention to the learning goals – recognizing image patterns and assigning meanings to possible image objects. That is, it can engage learners' attention in systematically searching for and identifying possible image patterns on images rather than limited to one final solution. The flicker task requires awareness and attention for solving problems demanded by the task. Comparatively, in the no flicker treatment, the learners are directly provided with final answers one way or the other, so their attention may mostly be caught and limited to this one answer. Or worse than this, novices may just glance at the answer without any attention because of its availability (O'Regan, 1992). With the compare and contrast method, learners may allocate their attention to images but may merely focus on the change between the two images but have insufficient attention to the contexts of the change. When the two images are juxtaposed, the difference between the images may stand out and be easily identified.

Internal and External Representations

Internal representations are also called mental models, which influence concept formation and conceptual change. Summarizing previous studies, Jonassen (2005) argued that conceptual change is a typical type of meaningful learning. Regarding how to make

conceptual changes in learners, he maintained that learners can obtain these changes when they build external models of what they learn with technology.

Furthermore, Jonassen argued that different tools were available for different types of reasoning. For example, database and concept maps are best in supporting comparison-contrast reasoning while expert systems can scaffold causal reasoning. He then suggested more research in comparing these different conceptual models systematically. As for how to assess conceptual changes, Jonassen admitted that model building is a good approach to testing these changes. Furthermore, he proposed more research to validate the models built by learners and providing rubrics in assessment.

In arguing for the effectiveness of model building for conceptual change, Jonassen presented previous researchers' arguments and his own ideas: First, modeling is regarded as an important means to understanding phenomena among science educators. He also defined modeling and the relationship between modeling and conceptual changes. He elaborated on the types of phenomena that can be modeled, including domain knowledge, problems, systems, experiences, and thinking. Afterwards, he summed up limitations of modeling. Although he did not include any instructional strategies that can be used to build up mental models, he pointed out some tools that are helpful. Besides, he maintained the necessity of studying these mental model-building tools.

It seems that what Jonassen proposed about constructing external models has relationships with what Carlin proposed about generative encoding. The difference is that in the former one the learner constructs something but in the latter one the learner search for different possibilities and construct external models in a different way. The reason

why Carlin's approach is more appropriate here is for the consideration of learners. Novices may have little knowledge of the varieties of instances of the concepts, so it is difficult for them to literally construct external models at this stage. This may easily cause misconceptions because of their little prior knowledge of what abnormalities are. Therefore, it is more reasonable to give them images and let students search for patterns than ask them to create external models from scratch.

The internal representations of patterns are the key to mental models. This is why computer-based instruction studies examined different display methods that can enhance internal representations and mental models. For example, in Hilbelink's study (2007), she compared the 2-D and 3-D displays and found that the 3-D display method can better help form mental models.

In learning radiographic images, it is crucial to form mental models and visual concepts. Two complexities in radiographic images make it necessary to provide tasks to help learners form mental models. One complexity is that radiographic images greatly vary. This complexity puts readers in a new reading setting whenever reading a new set of images. Therefore, the representations of patterns of anatomical structures that the reader forms internally can facilitate the reader to identify the abnormalities in the setting and segment from the setting. The other complexity of the radiographic image is that even the properties of the same type of abnormalities may vary largely from each other in terms of the size, shape, contrast, brightness, texture, configuration, and other dimensions. Therefore, it is necessary to construct mental models through working with instances, creating and modifying models rather than directly learning abstract descriptions or sketched prototypes of these instances. The other noticeable reason is that,

through the flicker treatment, learners can improve internal representations by extensive cases and the features in these images. The construction of internal representations is not constrained to one case but extended to the other cases, which can be regarded as a process of concept change.

Furthermore, limited observation may hurt understanding image features. False internal representations may be formed with little observation. In the flicker treatment, the change detection requirement in the alternations of images and the repetition of the images may give opportunities to keep building up and revising internal representations. Therefore, the flicker treatment may enable a constructive process of internal representations of images and their patterns that are studied.

Moreover, in the flicker treatment, learners will be active in forming internal representations of what they see. For identifying the change in the fast going images blocked by a blank image, the learners will guess at the meanings of possible abnormalities and spontaneously construct mental imageries of these features. With the internal representations, the learner can then make comparisons between this representation and the follow-up external representation. Therefore, the task goal of change detection and the task constraint of the blank image in between the alternated images make it necessary for the learner to develop representations internally.

However, in the other treatments, learners do not have the necessity to work out internal imageries. For the no flicker task, the difference between the two images, that is, and the change, pops out, so the learner does not have to form any internal representation to solve problems. With the quickly found solution, learners may put the instant findings at test and go to the answer sheet to submit their responses. Therefore, the underlying

requirement of the task in the no flicker treatment is different from that in the flicker task. Furthermore, the compare and contrast treatment may or may not demand internal representations because of the simultaneous view of two images demonstrates the change.

Comparison and Contrast, Internal and External Images

Studies (Schwartz & Bransford, 1998; Schwartz, Martin, & Nasir, 2005) indicated that the method of contrast improved transfer in concept learning. Researchers did experiments examining how comparing data and reading data can influence learning texts afterwards. They found that the group comparing data achieved higher scores in post-tests, especially in transfer tests. They explained that the comparison of data enabled learners to form mental models to get them ready for learning the texts.

The use of contrasting cases is empirically supported by Schwartz and his colleagues' studies (1998, 2005). They continuously developed studies on using contrasting cases to support knowledge evolution. In their 1998 studies, they did three experiments to study how contrasting cases were used in teaching psychology concepts. When students did differentiation study before they listened to a lecture or studied a text, they would end up with significant differences in prediction tests. Although the recognition tests did not show the same significant result, contrasting cases methods can still be an efficient method in teaching concepts and problem solving. Throughout these three experiments, Schwartz found that there is a point to tell knowledge to learners. Although telling is regarded as a non-constructivism method, it is actually an important part of knowledge construction. What we need to do is to prepare students for this telling process so that students can very easily map this telling part into their construction part. Telling then can become an effective constructivist method. Later on, Schwartz and his

colleagues did some other studies to test the contrasting cases method with groups of students learning statistics. Similar results were found that contrasting case group did significantly better in prediction tasks.

Mayer (2001) in his *Multimedia Learning* summed up possible approaches to organizing knowledge and comparison and contrast was regarded as an important means of organization of knowledge. Therefore, radiologist novices can use this knowledge organization tool with some tasks like the flicker task in their study endeavor, even though there are other ways to help organization, for example, presenting cases in database . This is why in this study related images are placed in a cluster rather than jumbled.

The comparison itself actually is an important skill in radiographic image reading. Azevedo (1998) found that data comparison was a problem-solving operator in mammography interpretation. In his study, experts used comparison much more than novices. This implies that comparison is actually a skill that novices need to learn and this skill is supposed to be an objective included in the curriculum. Furthermore, discrimination in mammography interpretation is the ability to distinguish abnormalities from normal features. Constant comparisons among different possibilities can help improve learners' knowledge and thinking about the similarities and differences between different features. With the feedback after comparison, the viewer can then further reflect on these possibilities and develop their awareness of different image patterns.

To search for the change in the flicker task, the viewer needs to fix upon a part of the images each time and compare it to the part of the next image blocked by the blank screen. This time span, no matter how short or long it is, always challenges visual short memory. Visual short-term memory can briefly retain a few objects every time and start to decay 0.5 seconds after the object disappears (Sperling, 1960), so the viewer needs to hold what he or she can temporarily store and make comparisons with the other incoming information. This is somewhat like flipping pages of a book continuously to compare two images on different pages.

However, when learners go to no flicker treatment, the comparison is none or little, so the relationships between the possible findings are not constructed. What learners lose in these tasks may be the comparison skill itself or maybe the relationships among patterns.

Furthermore, the flicker treatment also facilitates learners to compare their internal representations and external representations. In comparing these representations, the awareness of images is developed and strengthened. First, the internal representations become necessary because of the flash of images. To have more details in internal representations, more attention needs to be engaged in representing the image. Second, the internal representation of a concept may be modified by the comparisons across cases. When learners study a case that stands for abnormality, they will use their knowledge and compare their internal models with the images they see. Afterwards, they may revise their previous models with new instances. In addition, comparison can include both holistic and point-by-point comparison. Both of the comparison approaches can increase understanding of images as well as the awareness of details of images. Comparison and

contrast in the flicker treatment can help improve discrimination skills and sensitivity performance.

It is helpful for readers to learn with the comparison and contrast treatment in radiographic image reading. However, it may take a while to cultivate the ability and habit of making comparisons. Instead of directly asking learners to search for changes by comparing and contrasting images, the flicker treatment can activate and motivate novice image readers to make comparisons across features, between potential abnormal and normal features, and potential abnormal features. If directly asking learners to compare and contrast two images, they might not be engaged in the process.

Generative Encoding and Passive Encoding

Without telling image learners the differences with demonstrations as it is with the no flicker treatment, the flicker treatment can provide more opportunities for learners to study images and search for pattern changes, thus actively encoding image patterns. On the opposite, in the no flicker treatment, learners may passively encode responses provided without mentally engaging in any inquiry. In the comparison and contrast treatment, learners may derive the differences and similarities between two images without making wide search as what learners in the flicker treatment do.

About encoding, the explanations that Carlin and colleagues (2005) developed to explain the effectiveness of the flicker task are multiple cues, distinctiveness of the response, and transfer specificity. All these three are important strategies for deep processing of information and encoding. Deep processing actually can improve meaning making. Therefore, generative encoding can help learners construct meanings and make

the meanings retained. On the other hand, passive encoding will result in shallow processing and little comprehension of what are learned.

Specifically, the flicker treatment can enable novice radiographic learners to attend to different cues besides the changed object so that they can have more clues for the recall and recognition tests afterwards. Multiple cues can be a good reason why the flicker task can result in better learning. By assigning meanings to different objects, these objects then become more related, which may provide cues for the pattern. Furthermore, more comparisons of these possibilities will be carried out in the flicker treatment. However, in the no flicker treatment, the final answer is provided so learners do not need to attend to the other possibilities. Therefore, they do not have cues as those learners in the flicker treatment. For the learners in the compare and contrast treatment, they are more active than those in the no flicker treatment, but as they work with more images, similarities and differences may pop out, and the other possible answers will be less likely to be noticed. Therefore, this treatment may work less well than the flicker treatment but better than the no-flicker treatment.

Moreover, the distinctiveness of encoding is helpful for learning. When mental effort is made in learning, the responses gained through effort will become distinct. Compared with the other two treatments, the flicker task engages learners in more responsibilities and effort to search for responses, make guesses of the possible answers, and select the right one through filtering information. Through the cognitive effort, the response will become impressive and more meaningful. Furthermore, the existence of different options in the flicker treatment may also make the final answer standing out

because the distinctiveness is further developed through comparison among different possibilities.

The other advantage of the flicker treatment is that learners do the same tasks in their assessment as they do in the flicker task but not in the other two treatments. In assessment, they have to search for the patterns and choose right responses among potential ones, which is what they do in the flicker task. In the no flicker task, the answer of change is just there, so they do not need to make any selection and search. In the comparison and contrast treatment, the solution to the question of similarity and difference can be identified but not as difficult as in the flicker treatment. According to previous research, similar requirement and environment in learning and tests can enhance memory (Tulving & Osler, 1968). It was found that when learners study in a certain condition and tested in the same condition, learning results were better than the students learning in a condition but tested in a different environment.

Discrimination and Generalization Across and Within Categories

Discrimination and generalization of cases are important for category perception. As Keller and Schoenfeld (1950, p. 155) proposed, “Generalization within classes and discrimination between classes – this is the essence of concepts.” Visual category learning can be achieved by obtaining both similarities among cases of the same category and differences between cases of different categories from experience, but not merely one of them (Gibson, 1969). These commonalities and distinctions can help learners create associations for future retrieval and activation (Rumelhart & Todd, 1993).

The original study of the flicker task includes series of scene images, but the purpose of the study is not examining instructional strategies for improving image

perception and recognition. In clinics, cases are in random order and recorded in patients' names, so similar cases are not clustered. In this study, similar cases were clustered to make it easier for learners to differentiate different instances of the same concept or differentiate different concepts, as well as generalize across similar cases. This immersion in clusters of cases can help learners continuously construct and revise their mental models as well as improve learners' generalization across cases. Therefore, with these cases, the flicker paradigm can improve discrimination of different potential targets and the collection of the task with different cases can improve generalization. Altogether, the flicker treatment can promote both discrimination and generalization, resulting in recognition. In the other two treatments, clusters of similar and different case may not work as well as in the flicker treatment because the solutions are more easily available to learners.

Forming and Testing Hypotheses

The flicker treatment enables learners to create hypotheses of abnormal and normal features when they are told to watch the changes of these features from one to the other. The conjectures that learners make can then be tested with the facts they collect through their systematic searching. After they eventually locate the change, they will then assure themselves whether their hypotheses are right or wrong. This process of forming and testing hypotheses is usually regarded as constructive learning, which is consistent with learning theories about conceptual change through learners' experience (e. g., Dewyer, 1902; Piaget, 1968; Vygotsky, 1986). For example, classical constructivists Piaget (1968) and Vygotsky (1986) suggested that children learn through constructive processes, such as assimilation and adaptation and social interactions. Assimilation and adaptation

are two procedures that children use to create, test, and revise their knowledge while social interaction is another means of knowledge construction. Through working with data - developing theories by hypothesizing from and testing with data, knowledge can become flexible and transferable. Otherwise, information might be inert and useless if it were merely delivered from other sources (e.g., Cobb, 1999; Resnick, 1987).

Facilitating Perceptual, Conceptual, and the Interactions of These Processes

Underestimating instructional strategies in teaching visual concepts may result from the lack of awareness of the complex processes of perception, conceptualization, and the overlook of the interaction between perceptual and conceptual processes.

According to Lesgold and his colleagues (1981) and Rogers (1992), radiological diagnosis includes a process of interactions of perception and conception. Therefore, the designers need to adopt an approach that facilitates the interactions of these cognitive processes.

Previous instructional and cognitive research in this area seemed to use an isolated method, separating perceptual learning and conceptual learning. It was claimed that visual concepts could be taught by pointing out the objects and features. Researchers in instructional design also studied guidelines in teaching visual concepts, mostly focusing on the presentation of these concepts. It seems that active learning strategies dealing with both perceptual and conceptual learning have yet to be initiated in research.

Studying sets of cases with the flicker treatment seems to be an effective instructional decision for novice learners to study images. Through working with a series of cases, learners can make their guesses about the concept after perceptual activities and these conjectures will then be tested in solving the other problems. The interactions of

perceptual and conceptual processes can thus be realized through these activities.

Therefore, the proposed flicker method probably helps learners integrate their perception and conceptualization, and develop the interactions among them. The underlying reason for this conjecture is that the interactions will be strengthened through the internal problem requirement of figuring out the changes, not through demonstrating the changes in such methods as in the no-flicker method.

Scanning Images with the Flicker Treatment

Maybe scanning images can be one of the important skills that radiologists have. Practicing scanning can help learners make fast, systematic, and block-by-block eye movements when they look at images. It is important to make fast eye movements while getting meanings of image patterns. It was found that experts can scan images faster and linger on the findings in a shorter time than novices (Lesgold, Rubinson, Feltovitch, Glaser, Klopfer, & Wang, 1988). That is, the fast eye movement through the objects is important. On the other hand, the systematic movements are also important for radiological diagnosis. It was found that there may be not a standard for looking from the bottom to the top of the image or vice versa, but radiologists usually scan in a systematic way (Krupinski, 1996; Rogers, 1992; Lesgold, Rubinson, Feltovitch, Glaser, Klopfer, & Wang, 1988).

The flicker treatment actually can enable learners make fast eye movements because of the fast alternated images. The flicker task also requires systematic search because systematic search seems to be the most efficient approach to detect the change in the flicker treatment. In addition, the fast changing images with a blank screen in between in the flicker treatment may enable learners to separate the image into blocks with

attention on a small block every time when they get a chance to look at the images. However, the other two treatments do not imply the necessity of fast eye movement in making sense of patterns. Therefore, the flicker treatment can be a more efficient than them in helping learners scanning images systematically.

The System Enhances the Metacognitive Processes in Studying Images

Self-assessment in the shared system of the three treatments is an important metacognitive strategy for adult learners. It can help learners check if they have understood what they have learned. Therefore, it is a process of taking the meanings out of learners and making learners negotiate what they detect and diagnose in images.

Furthermore, feedback in the shared system can increase learners' metacognition. First, feedback is an important instructional strategy because it may provide guidance to students about where they are in their learning, what their strengths are in their study, and what they need to rethink of. Without feedback, students will stay puzzled about their strengths and weaknesses in their thinking and may gradually lose their interest and motivation in learning. Second, feedback is important in radiology teaching because accurate detection and diagnosis is crucial in this learning situation. However, the lack of feedback has been identified in literature (Azevedo, 1998; Deitte, 2006). Third, feedback is important for radiology novices to engage in self-directed learning. Self-directed learning does not mean that feedback is unnecessary but even more important than face-to-face teaching. Feedback is critical to cognitive apprenticeship because knowledge and thinking is constructed with conceptualization and its revision (Collins, Brown, & Newman, 1989). Misconceptions will influence learners in their development. Without

feedback, self-directed learners may get lost because they need guidance for improving their thinking.

With this system of guidance, the flicker treatment not only activates learners' cognitive processes, but also enhances their metacognitive processes and the improved metacognition may enable learners to become more active and independent in learning. In the flicker treatment, after the internal representation was compared with the external representation, the viewer thought about the difference between the two images and chose the identified change on the original image and feedback of right or wrong was given. By doing so, the internal representation is tested, confirmed or denied, providing an opportunity for the viewer to think over the differences between the two images, again comparing the differences between the changed object and the change, abnormal and normal features. Throughout cases, learners are led to make constant comparisons and reflections upon these differences.

In the other two treatments, self-assessment and feedback are also provided. However, they might not result in improved metacognition and reflection upon problems, solutions, and revision of existing models. The reason is because learners may get right answers instantly and easily, and no reflection is necessary. For example, the function of feedback in no flicker treatment probably is reinforcement.

However, the flicker task somewhat constrains the details of feedback. For example, it is limited to confirmation of the responses, but not explaining the reason why it is right or wrong. In instructional design, high-level feedback is regarded as advisable for guiding the learner in understanding what they are learning (cited in Mory, 2004). However, considering novices' situation, this level of feedback is appropriate. Indicated

in literature, confirming the recognition of abnormality is frequently used in resident education in clinics (e.g., Chew, 2001; Deitte, 2006). It was not mentioned the feedback also includes why it is right or wrong. Therefore, this level of feedback is supposed to be all right for novice learners.

The Flicker Enhances the Affective Respect in Studying Images

One of the emphases of the flicker paradigm is to offer challenges to learners about what are there in images and what are the changed. Considering the difficulties of the task, the flicker treatment imposes bigger challenges than the other two treatments. According to Vygotsky (1986), challenges are a key to learning. Therefore, the flicker task is supposed to engage learners in learning.

The challenge and responsibility of fixations upon details of the images are important for novice learners. The flicker paradigm can challenge viewers and let them take the responsibility of actively placing different parts of the images under surveillance. Even when learners get to know but are not sure what may be the abnormal part of the image, they will still have to place different guesses in their fovea and compare across the guesses, attending to these parts rather than looking at the pop-out answer or quickly see the change in the other two treatments.

The other advantage of the flicker paradigm is that the uncertainty of findings may engage learners to make continuous effort to find the path to problem solving and reach the final decision. This may have some common points with the “aha” effect (Wills, Soraci, Chechile, & Taylor, 2000) that was also found with pictures. It was found that when learners connected dots of images, significant differences were found in their recall and recognition than provided with images to read the images or provided with

lines and just repeat the contours of the images. They explained that the differences were caused by sudden insights that learners achieved from working out the problem by themselves.

Another possibility is that the flicker treatment may stimulate awareness and enhance metacognition (Grabowski, 2004) in viewers by making their discovery. Otherwise, the availability of differences in the other treatments may decrease viewers' interests to fixate on the change and the images. Furthermore, the pop-out change in the no flicker treatment may even decrease attention and awareness level because of the direct answer offered to learners.

The Flicker Paradigm and the Other Treatments: The Curriculum and Participants

In comparing two images in clinics and diagnosing changes in clinics and conferences, two images are put side by side for examination. In some studies of image displays, it was found that simultaneous presentation is better than successive presentation (cited in Kim & Astion, 20031). Some other studies found contradictory results. However, none of these studies used the flicker paradigm.

Side-by-side displays have been conventionally adopted in clinics. In a study of radiologist working station, the side-by-side display was proposed because this display can benefit image readers to make comparisons across images (Armato, Doshi, Engelmann, Croteau, & MacMahon, 2006). This would make sense considering the viewers are radiologists in diagnosing cases because of their expertise, tasks, and purposes of viewing the images. Their expertise enables them to quickly see the abnormalities and make right decisions for diagnostic purposes. However, the situation for novices will be another story for the differences in their expertise and purposes of

studying images. They have fewer ideas of the various image patterns, so they need to take time to immerse in these patterns and external representations. If this immersion could stimulate their interest and engage them in connecting their minds with images, learning may probably occur. With the flicker treatment, they have to attend to and study each possible image finding. In doing so, they focus their fovea on these findings directly and test whether their hypotheses are right by discriminating them from the other image patterns. The comparison of the images happens between the image they see and the image in the mind, so the external images formerly seen are supposed to be internalized.

From different aspects of cognitive processing, the flicker paradigm seems to promote all of them, including attention, short-term memory, and long-term memory. First, Rensink's study has indirectly support the use of flicker paradigm as an attention enhancement method. Second, the flicker paradigm can force to unitize the features on the image and make sense of them. With the meanings and interpretation in mind, learners may rehearse these representations and meanings internally in eye movements. Short-term memory will benefit from this meaning making and rehearsal processes. In going through a series of cases of the flicker paradigm, the meanings and visual concepts will be constructed, which will be incorporated into long-term memory.

Compared with the flicker treatment, the other two treatments may result in less mental participation, responsibility, and self-direction in image study. The no-flicker treatment makes the change pop up and learners merely need to parallel process images for finding the change. O'Regan (1992) pointed out the illusion caused by the richness and availability of the outside world, so the simultaneous presentation of the images may give viewers this feeling of lasting presence. Nevertheless, the flicker treatment can

create a situation of the disappearance of the image and stimulate the interest and urgency to see what is there rather than just being inert in perception. The going away images in the flicker treatment attract learners to catch the running train and detect the changed objects and patterns. The illusion of presence seems gone in the flicker situation.

Conclusions

In conclusion, the above-mentioned literature suggests that, in problem solving, novices differ from experts in terms of their knowledge, interest, and strategies. To solve complex image learning problem in technology-based instruction, proposed training methods need to draw learners' attention to detecting and discriminating image patterns, engage them to actively construct meanings, and help them become diligent searchers and knowledge builders of image patterns.

Moreover, technology itself does not guarantee that learning will occur (Gagne, Wager, Golas, & Keller, 2005; Mayer, 2001). Use of technology in instruction has raised a debate of technology-centered versus learner-centered applications (Mayer, 2001, 2005; Reed, 2006). The technology-based applications usually emphasize the use of cutting-edge technology to improve the delivery of information and knowledge transmission. Such technologies as radios, televisions, and computers were strongly claimed to influence learning and expected to replace teachers in the past (Cuban, 1986; Cuban & Usdan, 2003). However, technology has failed to effectively influence education for the lack of effective instructional applications, as Cuban (1986) identified and suggested. Therefore, other factors need to be considered and instructional methods are one of them.

With uncertain effective instructional methods in CBI and WBT defined in the first section of this chapter, pedagogical research in visual concept instruction is a timely

and useful project for the following reasons: First, it was identified that visual concept instruction methods for CBI have been neglected (Sharples, 1991). Although previous researchers have developed general guidelines (Sharples, 1991; Kim & Astion, 2000), there has been little information of theory-based empirical evidence of specific instructional strategies in this area. Second, researchers in visual perception have identified psychophysical characteristics and processes of image perception and recognition (Goldstein, 2002). Models of perception increase understanding in the importance of attention and perceptual organization rules of visual information. However, these researchers did not investigate instructional methods for image learning. The study of the visual learning methods may extend the perspectives of these researchers. Third, visual literacy studies focused on message presentations and offered little theoretical basis and instructional method information (Braden, 1996; Rieber, 1994). In studying presentation modes, there have been debates of simultaneous view of images in comparison to the successive mode (Kim & Astion, 2000; Whiteside, 1987). Nevertheless, presentation modes tended to be isolated from instructional methods in previous studies. Furthermore, visual literacy was found difficult to be further examined because of lack of theoretical support (Braden, 1996). Without theories, these studies tend to be piecemeal and insufficient in depth. The identification of theories in this study may help image instruction researchers to progress. In brief, this study can expand knowledge of visual concept instruction because there has been little pedagogical information for this type of learning in previous cognitive and educational research.

In addition to lack of evidence of effective instructional methods in visual concept instruction, previous studies have suggested the necessity and possibility of engaging and

active learning methods for CBI and WBT (e.g., Chou, 2003; Davidson-Shivers, 2002; Draves, 2000; Jonassen, 2004; Matthews et al., 2007; Mayer, 2001; Moore, 1989). On one hand, CBI and WBT require instructional methods that can increase learners' mental participation. Compared with face-to-face instruction, CBI and WBT are characteristic of voluntary, independent, and active engagement (Davidson-Shivers, 2002; Gagne, Wager, Golas, & Keller, 2005; Mayer, 2001). With little face-to-face interaction with instructors and peers, learners in CBI and WBT may lose their attention to what they study and cannot make sense of information and construct knowledge. Thus, CBI and WBT need to engage learners in processing information and participating in activities (Mayer, 2001; Proske, Narciss, and Korndle, 2007). On the other hand, instructional technology (IT) offers interactive features and functions to enable levels of interactivity (Chou, 2003; Draves, 2000; Jonassen, 2004; Matthews et al., 2007). Interactive activities can enhance learners' motivation, cognitive engagement, memory, and performance (Chung & Zhao, 2004; Matthews et al., 2007; Selcer, 1993). Thus, it is necessary and possible for instructional designers to engage learners in actively processing and organizing information in CBI and WBT. If researchers could address uncertain instructional methods previously mentioned, instructional designers might better satisfy requirements and utilize the affordance of IT to achieve instructional effectiveness and foster learning.

Considering these needs and possibilities, generative strategies (Carlin, Soraci, & Strawbridge, 2005; Smith & Ragan, 1993; Wittrock, 1974, 1990, 1992) may probably address pedagogical effectiveness in technology-based visual concept instruction. The generative strategy is one of the two major types of instructional strategies for concept instruction (Smith & Ragan, 1993). Generative strategies can engage learners in learning

as active participants rather than observers on the side and learners can become responsible for their learning activities in generative learning (Morrison, 1994; Smith & Ragan, 1993; Wittrock, 1990, 1995). More importantly, effective outcomes from generative learning were attributed to deep levels of information processing and cognitive efforts (e.g., Slamecka & Graf, 1978; Kinjo & Snodgrass, 2000; Carlin, Soraci, & Strawbridge, 2005). Therefore, what generative strategies are promising in addressing learners' cognitive participation.

Founded in a constructivist view of learning, generative strategies can enhance learners' ability in attending to, organizing, encoding, elaborating, and integrating information (Jonassen, 1988; Wittrock, 1990, 1992, 1995). Constructivists believe that learners can achieve learning through assimilation and accommodation, making sense of new information and changing existing cognitive structures responding to new information, according to Piaget's theory (cited in Siegler & Alibali, 2004). By generating meanings through generative strategies, learners can enhance attention, improve cognitive participation, construct mental models, and improve problem-solving abilities, matching what constructivists as indicated important for increasing learning (Dewey, 1902; Jacoby, 1978; Jonassen, Strobel, & Gottdenker, 2005; Knowles, 1998; Mayer, 2001; Piaget, 1970; Vygotsky, 1986; Wittrock, 1974, 1990, 1992). Therefore, generative strategies can develop in learners match the effective learning conditions defined by constructivists, involving learners in constructing knowledge and solving problems.

Specifically, generative strategies are grounded in two generative models, generative learning theory and generation effect theory. The common hypothesis in these

two models is that learner-generated stimulus and meanings increase learning in comparison to those provided by experimenters or instructors. Based on an active learning assumption, generative learning theory (Wittrock, 1974, 1990, 1992, 1995) proposed that generative strategies increase memory and comprehension through engendering four learning processes, including motivation, attention, prior knowledge, and generation with generative strategies, high learning gains were continuously established in previous studies (Wittrock, 1974, 1990, 1992, 1995). Furthermore, both theoretical and empirical evidence has been found in generation, reported in generation effect studies (e.g., Slamecka & Graf, 1978; Carlin, Soraci, & Strawbridge, 2005).

Generative strategies vary in types of learning. Generative learning strategies, such as summary, main idea, analogy, and explanation, tend to aim at enhancing comprehension of passages rather than pictures. Somewhat complementing with these text-oriented generative strategies, other researchers studied the generation effect on both words and pictures (e.g., Slamecka & Graf, 1978; Kinjo & Snodgrass, 2000). They found that learner-generated stimuli (words or pictures) could significantly improve the encoding of these stimuli. Compared with experimenter-provided texts or pictures, better recall and recognition outcomes were derived with learner-generated stimuli. The explanations for generation effect include deep processing of information, more effort in generating, and transfer specificity.

In particular, the flicker task as a generative strategy was found to result in improved learning with pictures, compared with no flicker task (Carlin, Soraci, & Strawbridge, 2005). The flicker paradigm was originally developed as a method to test attention in visual changes (Rensink, 1997). In a recent computer-based experimental

study (Carlin, Soraci, & Strawbridge, 2005), the flicker treatment was used as a generative strategy because it can enable learners to actively search for objects rather than receiving them. Pictures of objects and scenes were used as learning materials in the study. The flicker treatment was found to increase recall and recognition more effectively than the no flicker treatment and significant difference was found in recall. However, this has appeared to be the only study where the flicker treatment was examined as a generative strategy. It has also been one of a few studies related to image learning in studies of generative learning and generation effect.

As a generative strategy, the flicker treatment can better enhance novices' cognitive, metacognitive, and affective respects in studying images, compared with the no flicker and compare and contrast treatment. The flicker task was found more effective than the no flicker treatment in facilitating visual recall and recognition memory but the flicker effect has not been tested in radiograph study. In this study, the flicker treatment can draw learners' overt attention to image patterns, and engage them in forming internal representations and comparing their internal representations with external representations. The flicker treatment can also help learners encode what they learn actively rather than passively. Through the proposed flicker tasks, learners can study similar and diverse cases and generalize and discriminate across and within categories, leading to conceptual change for solving new problems. Through the flicker tasks, learners can also form and test their hypotheses and practice image scanning. Furthermore, self-assessment and feedback can promote metacognition while challenges and discovery in learning can enhance learners' interests with "aha".

An example of complex visual concept instruction is radiographic images. Researchers in this area found that learners' participation and instructors' guidance are important for promoting medical practice and expertise. However, the extant instructional methods and technology integration need improving.

With generative learning, novices may generate possible patterns and solutions, serving as multiple cues and highlighting the selected pattern so that they can retrieve better in performance. They may also develop relationships among their prior knowledge, experience, and current information through constructing, testing, and revising their mental models. Computer-based generative learning can be achieved through computer-enabled interactivity and may have potential to develop radiology novice learners' visual thinking and problem solving in radiographic image study.

Generative strategies were identified as effective in promoting learning in the instruction of science, reading, and other academic areas (Grabowski, 2004; Mayer, 2005; Smith & Ragan, 1993; Wittrock, 1974). However, little has been investigated about the generative strategies for studying visual patterns.

Chapter 3 Research Methods

Introduction

To examine the effectiveness of proposed new and existing instructional strategies upon pattern recognition as measured with recognition and classification instruments and scrutinize group differences in other factors, including duration (on-task time/study time), the number of incorrect responses, and the number of trials in study, this chapter provides details of research design, recruitment, participants, instrumentation and the validation of instruments, procedures of the study, ethical considerations, including an approval letter from the Internal Review Board (Appendix A), methods of statistical analysis, and pilot studies to respond to the corresponding research questions. Before unfolding these parts of the methodology, it is necessary to have a review of the research questions, considered to be more appropriate than the hypotheses in the original proposal because this study is an exploratory study without established evidence of directionality of the hypotheses.

The following are the major research questions this study addressed:

1. Did the participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method demonstrate any statistically significant differences in their overall performance as measured by recognition and classification posttest instruments?

2. Did the participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method demonstrate any statistically significant differences in their recognition performance as measured by the recognition posttest instrument?
3. Did the participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method demonstrate any statistically significant differences in their classification performance as measured by the classification instrument?
4. Were there any statistically significant differences in their performance as measured by posttest instruments between students who studied visual patterns in computer-based instruction with the flicker method of instruction and the no-flicker method of instruction, those studying with the flicker method and the comparison method, and/or those studying with the no-flicker method and the comparison method?
5. Were there any statistically significant group differences in their on-task duration among the participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method?
6. Were there any statistically significant differences in the number of incorrect responses and number of trials they made in their study among the participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method?

In addition, three post-hoc research questions were raised as follow-up ones depending on the results of the previous inquiry:

1. If any significant differences in duration were identified among groups, between which groups were the significant differences detected?
2. If any significant differences in number of incorrect responses and number of trials were identified, between which groups were the significant differences detected?
3. Without the pretest score as covariate, did the participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method demonstrate any statistically significant differences in their overall performance as measured by recognition and classification posttest instruments?

The following sections depict the research design, which, in order to reply to the above research questions, evolved into a pretest-posttest control group experimental study with instructional strategy as the independent variable and recognition and classification test scores, on-task duration, the number of incorrect responses, and the number of trials in study as the dependent variables. To decrease the measurement errors of effectiveness of instructional methods that might be caused by the different levels of prior knowledge of radiographic images among participants, a pretest instrument was complemented and the scores from the pretest were taken into considerations to adjust means and other statistics. Thus, this study had the pretest score as a covariate.

Furthermore, this chapter also describes the recruitment of participants with participants' consent of voluntary participation, the number of participants recruited for studies, the

sample size for a certain level of power and effect size, the study materials, instruments, and validation of these instruments. With these important sections presented, the chapter introduces the study procedures of randomization and administration of instruments. It also indicates ethical considerations and describes the proposed statistical analysis approaches to evaluate statistical null hypotheses to respond to the research questions. In addition, a report of pilot studies is provided to demonstrate the observations of instruments, study procedures, the group differences in duration and the number of incorrect responses, evaluation results of images in both study and tests, and other related respects.

Research Design of the Study

This study examined the effects of three instructional strategies on visual category learning represented by radiographic images, and corresponding duration, the number of incorrect responses, and the number of trials in the context of instructional design of CBI and/or WBT. In particular, instructional strategy was the independent variable while recognition scores, classification scores, duration, the number of incorrect responses, and the number of trials were the dependent variables, and the pretest score was the covariate (Figure 3.1). More specifically, the study investigated the effects of three instructional strategies, namely the comparison treatment (Appendix B) as a conventional strategy, the flicker treatment (Appendix C) as a generative strategy, and the no-flicker treatment (Appendix D) as a receptive strategy, and, with a pretest (Appendix E), on complex image recognition and classification performance as measured with a recognition posttest (Appendix F) and a classification posttest (Appendix G). The effectiveness of the comparison treatment upon learning in this area had not been examined in previous

literature, but it was regarded as a conventional method because this method is widely applied in clinical instruction and study. More details of the three types of strategies would be further explained in the instrumentation section of this chapter. The group differences with these three methods would also be examined in terms of duration, the number of incorrect responses, and the number of trials.

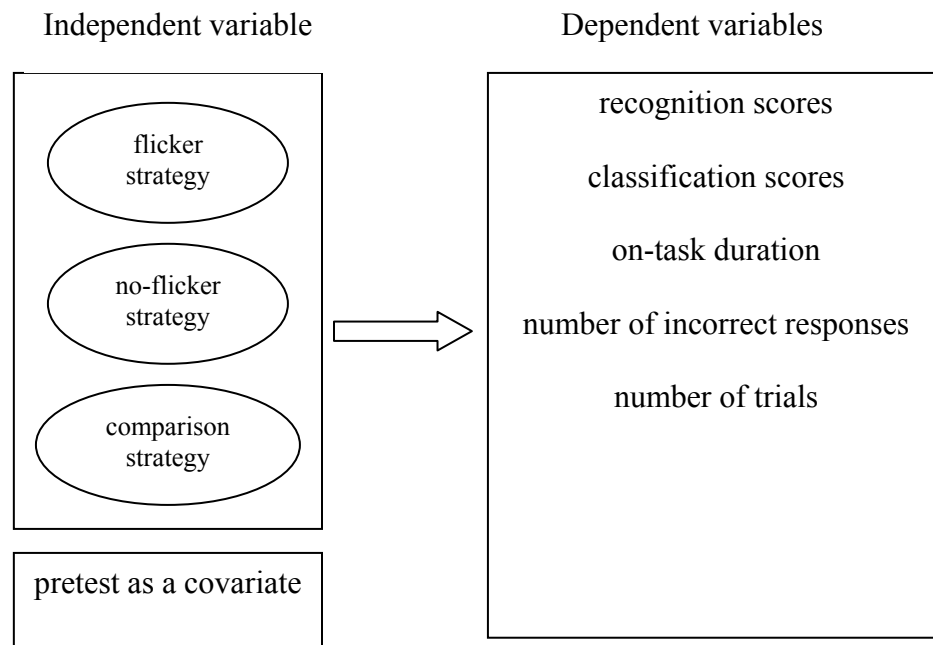


Figure 3.1. An overview of variables manipulated and observed in the study

The experimental study adopted a pretest-posttest control-group design illustrated in the diagram of research design (Figure 3.2). The participants were randomly assigned to the three groups: the participants studying with the comparison method belonged to Group 1, the participants studying with the flicker treatment were Group 2, and those studying with the no-flicker method were Group 3. Four phases were applied, including random assignment of the participants to experimental and control groups, administration of the pretest instrument, administration of three independent treatments, and

administration of two posttest instruments (Gall, Gall, & Borg, 2003). Thus, the design presumed that the general procedures of the study were random assignment of the participants to groups, a pretest of items, a study session of cases with one of the three methods of instruction, and two posttests of recognition and classification.

The pretest-posttest control group design was used to compare the effectiveness of three different instructional strategies upon the five criterion measures: recognition scores, classification scores, duration, the number of incorrect responses, and the number of trials. The reasons why selecting this method as the research design included: First, to increase the internal validity of the study, the pretest was used to decrease the influence of different levels of prior knowledge upon performance to statistically control the variable prior knowledge and rule out its influence and equate the initial points of the study among participants. Of course, in recruitment, the factor of prior knowledge was considered when potential participants were recruited with the criterion of little knowledge of radiographic images. However, the pilot studies, described in a following section, informed that the prior knowledge and skills were difficult to be evaluated without a pretest of the knowledge and skills. One of the reasons could be that the potential participants might employ criteria different from what were set by the researcher. In this case, they might mean differently from what the researcher meant by little knowledge and skills in radiographic images. The other reason can be that the visual and predicting skills could vary from person to person. For these reasons but not limited to these reasons, learners might show different learning abilities in starting to figure out patterns in viewing different cases and recognizing patterns. Therefore, their different levels of prior knowledge and skills needed to be measured and ruled out. Second, the

posttests were useful in measuring the effect of three treatments upon pattern recognition. The recognition test could examine how well the participants recognized the patterns in the images they viewed in image studies and the classification test examined how well participants categorized the patterns in the images that they did not view in the study sessions. Third, the control group served as a baseline and provided a foundation for comparison of performance across groups.

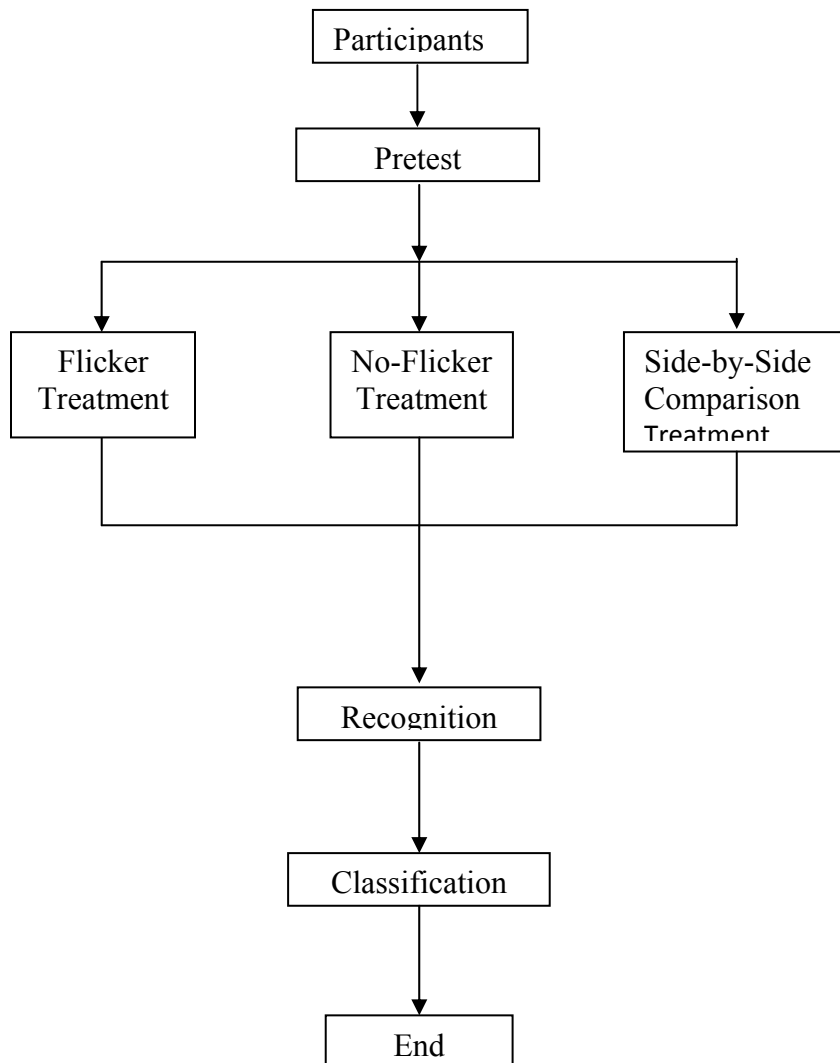


Figure 3.2. Diagram of research design of the pretest-posttest control group experimental study

Recruitment

To recruit participants and invite collaborations from the potential college student participants, two major recruitment strategies were used, namely posting recruitment flyers and talking with instructors and potential participants to invite participation. The researcher sent recruitment flyers (Appendix H) to instructors and potential participants, informing of participation criteria, the general purpose of integrating technology into higher education, the major content of the study, the duration of the study, voluntary and anonymous participation, the length of the study, and the benefits that participants would receive. More specifically, it was introduced that students who had little knowledge of radiographic images were invited to participate in the study. It was also noted that the study consisted of a pretest, a study session, and two posttests of the content area. The voluntary and anonymous participation in the study suggested that the participants volunteered to take part in the study and they were not asked to provide their names in the study. The participants were informed that through the study they could learn knowledge and skills of mammograms and they would also receive a certain amount of compensations.

Participants and Sample Size

The proposed sample size of the study was more than 150 participants and it turned out that 247 college students were recruited for the formal study in a Southeast university in the United States. These subjects were naïve learners who had little knowledge in radiographic images. According to literature (Steven, p. 247, 2002), a sample size of 75 students with 25 subjects in each group of the three groups satisfies the need of a large effect size and sufficient power (.70) at the alpha level of .05 (type I error)

in multivariate analysis of variance (MANOVA) studies and a similar sample size can reach the same purposes in the studies with the statistical method of multivariate analysis of covariance (MANCOVA). However, considering possible attrition of subjects, data that may be compromised in analysis in terms of outliers or other statistical considerations, and somewhat different/unstable effect sizes demonstrated in generative learning and generation effect studies, the proposed sample size in this study was increased to 150 participants in order to warrant the power of the study (Cohen, 1988). As a result, 247 participants were recruited for the study with 228 participants' data complete and usable for research analyses.

Instrumentation

The researcher developed the three programs, including both the independent and dependent instruments. The three programs or three parallel versions of the program consisted of the same content embedded in the same interface, including a pretest of ten items, a study session of twenty cases with forty images/ twenty sets of images and each set containing an abnormal image and corresponding normal image (edited), two posttests consisting of a recognition test and a classification test, with 10 items in each test, followed by a brief demographic survey. The only difference among the three versions of the program was the instructional strategy used. The content and interface of the study materials were explained in this section whereas the pretest and posttest instruments were introduced in the next section.

Development Processes

With her major professor's guidance, the researcher took the initiative of the instrumentation, developing the codes, interfaces, graphics, videos, and the other

elements of the instruments. Authorware™ 7.0 (Macromedia, 2004) was selected as the major authoring tool and the technical manual contained in the software package was consulted throughout the development. Although Authorware has very limited supporting resources for developers, it was selected as the authoring tool because of its capabilities. With the technology affordances of the authoring program, it is possible to create planned interactivity, including learner-computer and learner-content interactions as well as learner performance tracking. With Authorware, it is also possible to integrate multimedia into instruction, such as Flash format videos, static graphics, and texts.

In developing the instruments for this study, Authorware was specifically used to promote the major functions of the instruments: interactive instruction of quizzes and feedback; learners' selection options of study paces; experimenter's tracking learners' study process behaviors and performance information; and other functions of the planned instruments. The tracked learner information contained all of their scores of the pretest, the recognition test, and the classification test; the study time the participants spent after they finished the pretest and before they started with the posttests; the number of incorrect responses they made in the case study; the number of alternations it took them to reach correct responses; and the frequency they selected options to display the images among different modes of paces.

Coding and recoding with ongoing evaluation were guided engineering and research processes. Three basic principles used in the development were: development of the whole program in one time is impossible and may lead to overloads; development with ongoing evaluation leads to timely improvement of the program; and documenting small steps of development facilitates completion of development. Corresponding to the

principles, three approaches to development were used for constructing and improving instruments: breaking down tasks into smaller ones; evaluating the functions and design elements in the process and for finalizing the products; and documenting related procedures and variables.

On one hand, with his expertise in programming, the major professor guided the development of the instrumentation through evaluation, identifying problems in programming and pointing out the necessities and possibilities to increase the robustness of the prototypes and improve the products. Both formative and summative evaluation approaches were applied for developing the instrumentation because of the complexity of the instruments and the multiple aspects of design and development criteria. In formative evaluation, problems were identified and workable algorithms were confirmed for further development. In summative evaluation, the instruments were implemented and problems were recognized for further improvement and correction. Summative evaluation after the pilot studies enabled the finalization of the instrument.

On the other hand, the major professor gave advice on project management and asked the researcher to work on reachable goals. In prototyping, the researcher broke major tasks into smaller chunks of tasks, worked on a small chunk of functions each time, frequently implemented and evaluated the prototypes, not necessarily in a linear manner, and gradually attained the robustness of all of the codes for the instrumentation. The itemization of tasks enabled the researcher to have doable goals and complete manageable subtasks with step-by-step approaches, preventing from overwhelming mistakes. Tasks could overwhelm the developer if they were not separated into easily handled ones. For example, it was easier to separate the task of interface design from the

task of coding variables and functions and then these two tasks could still continue to be analyzed into and implemented in smaller trunks of tasks. Thinking over and experimenting with the background color(s) was one of the tasks that was more easily handled than the whole interface design and development in one time. The other method used in prototyping was coding the intended functions in natural languages and interpreted the languages into AuthorWare codes by using embedded properties, functions and variables. This method facilitated the researcher to figure out workable scripts and related setting definitions. In addition, development procedures, variables, and codes were documented in the process. Documentation improved the consistency of different portions of the instruments. In the process of coding, evaluation, and documentation, guidance, critical thinking, and searching for insights were essential.

In sum, development of the instruments for this study was filled with much work of analysis, coding, recoding, and evaluation. It was a step-by-step, bit-by-bit, and reiterative procedure although the steps were flexible and the bit was not definitely defined. Evaluation, critical thinking, thinking in natural languages and documentation were essential to the instrumentation.

The General Structure and Activities of the Study Materials

In the image study section illustrated in the flow chart of the learning section (Figure 3.3), the participants were asked to identify pattern changes in radiographic images. Each study case of the image learning materials consisted of two sets of images (image pairs) representing an abnormal category and the corresponding set of the same images but with abnormal features edited to represent normal features, with a total of forty images for the study cases. Generally, the images were sequenced from easy to

difficult, following the principle of elaboration theory (Reigeluth, 1999) and the other guidelines (Sharples, 1991). That is, each participant learned a total of forty images of the same sequence with three different instructional strategies. Each abnormal image and its corresponding normal image were studied with a certain strategy and corresponding tasks, according to the definition of the individual instructional strategy.

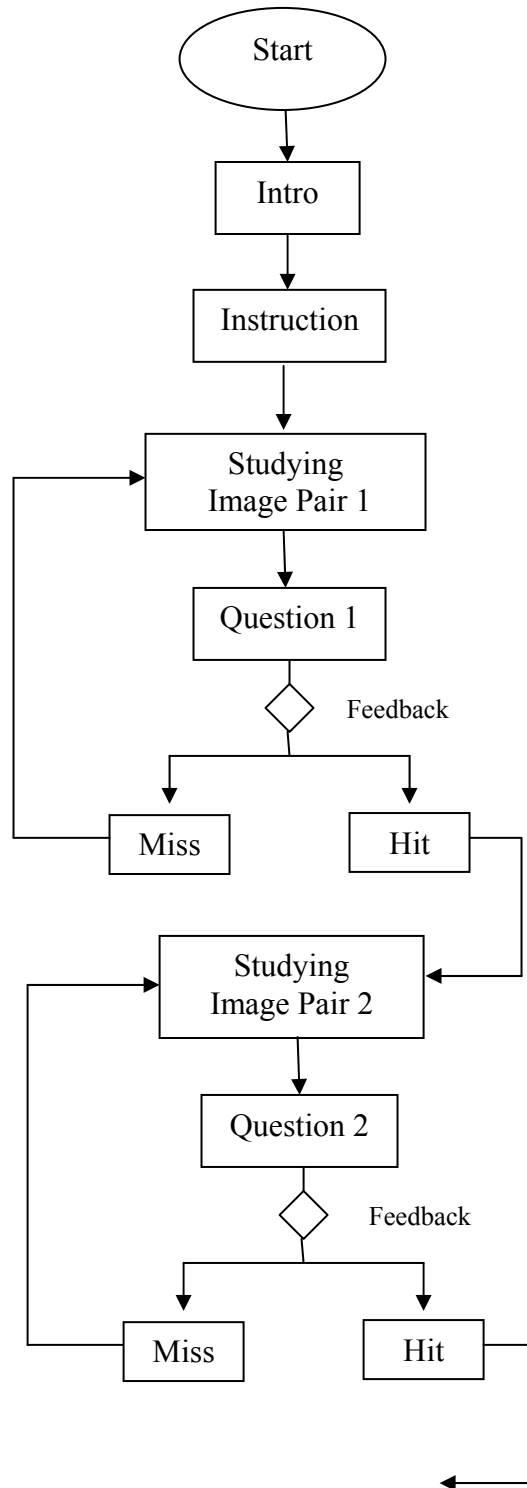
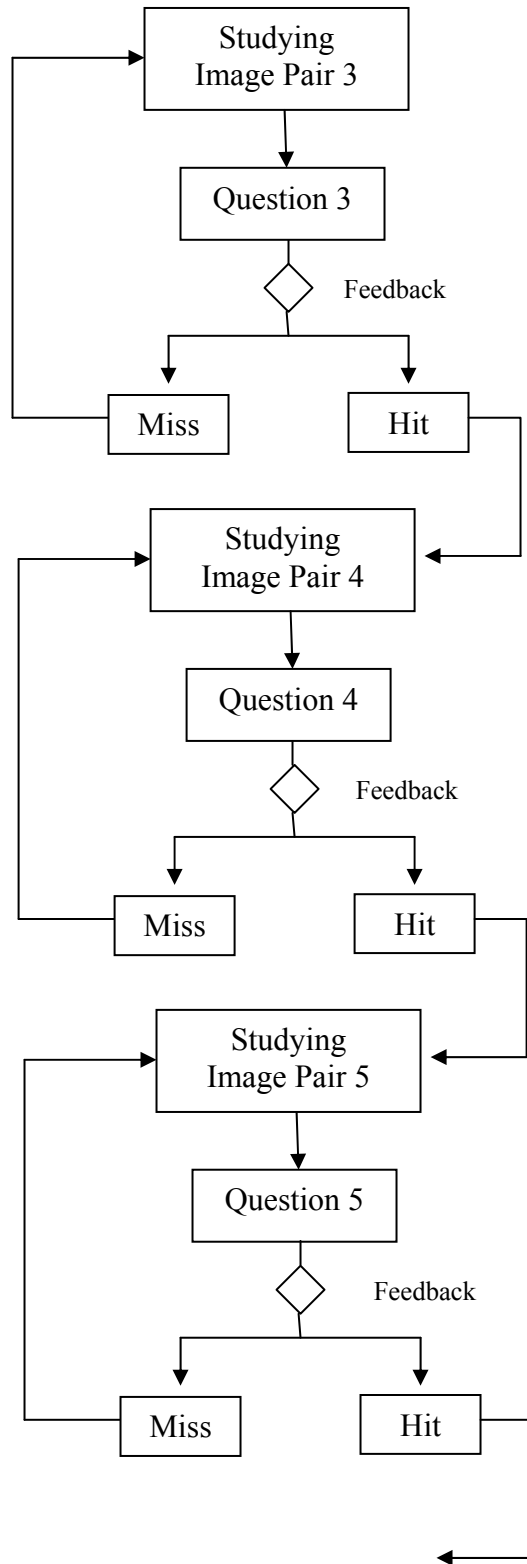
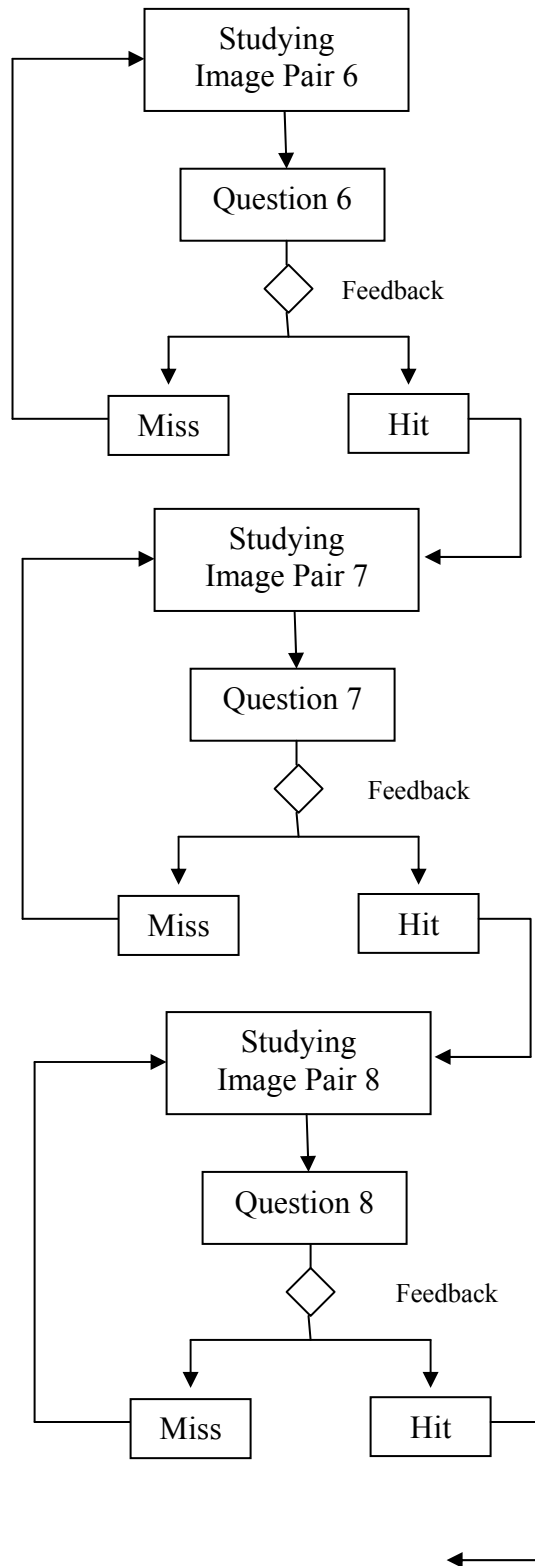
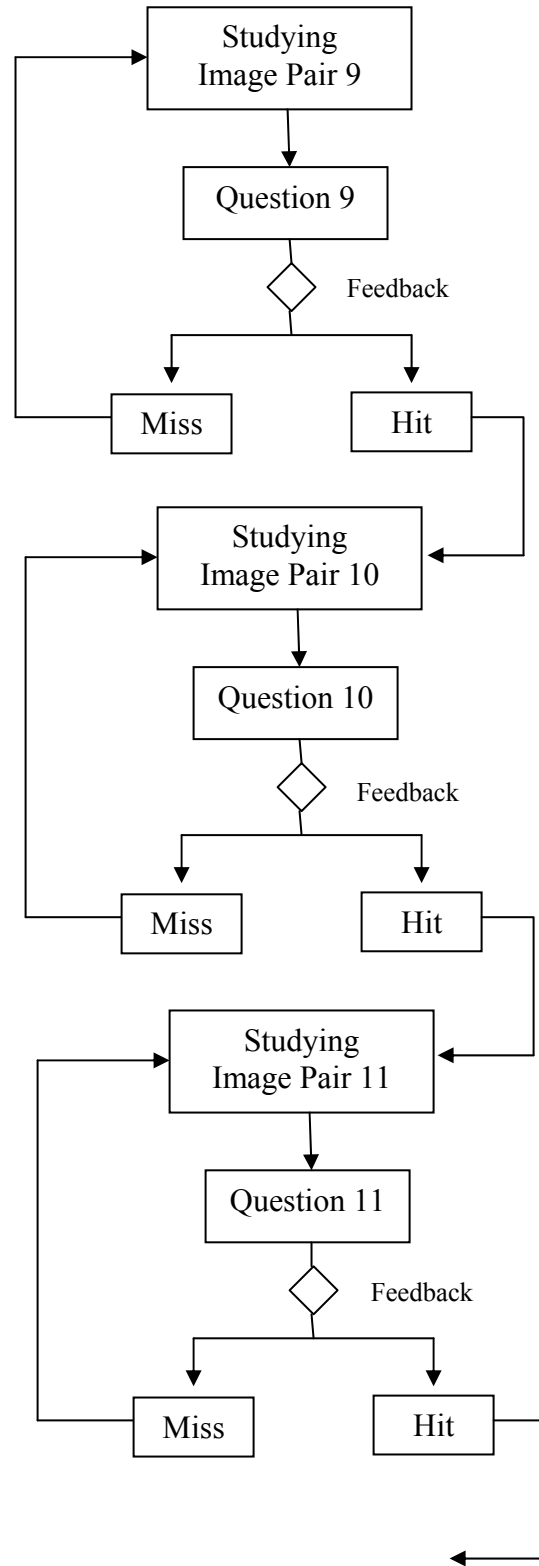
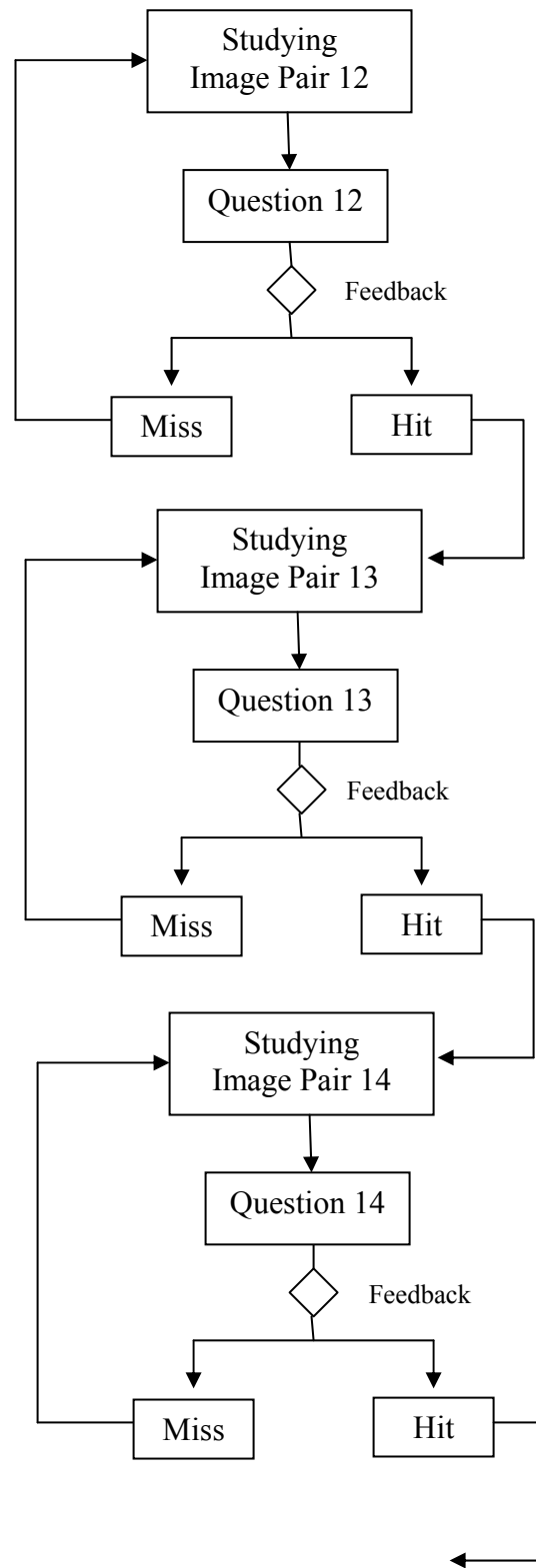


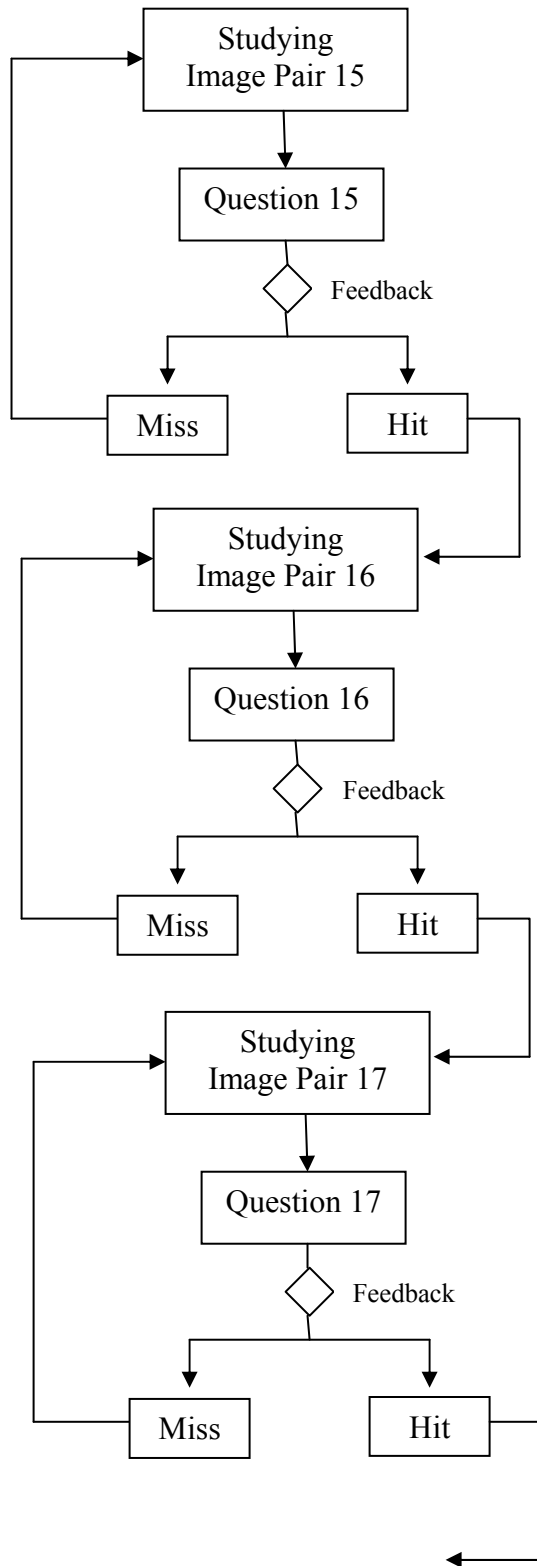
Figure 3.3. Flow chart illustrating the flow and structure of the three parallel versions of the program

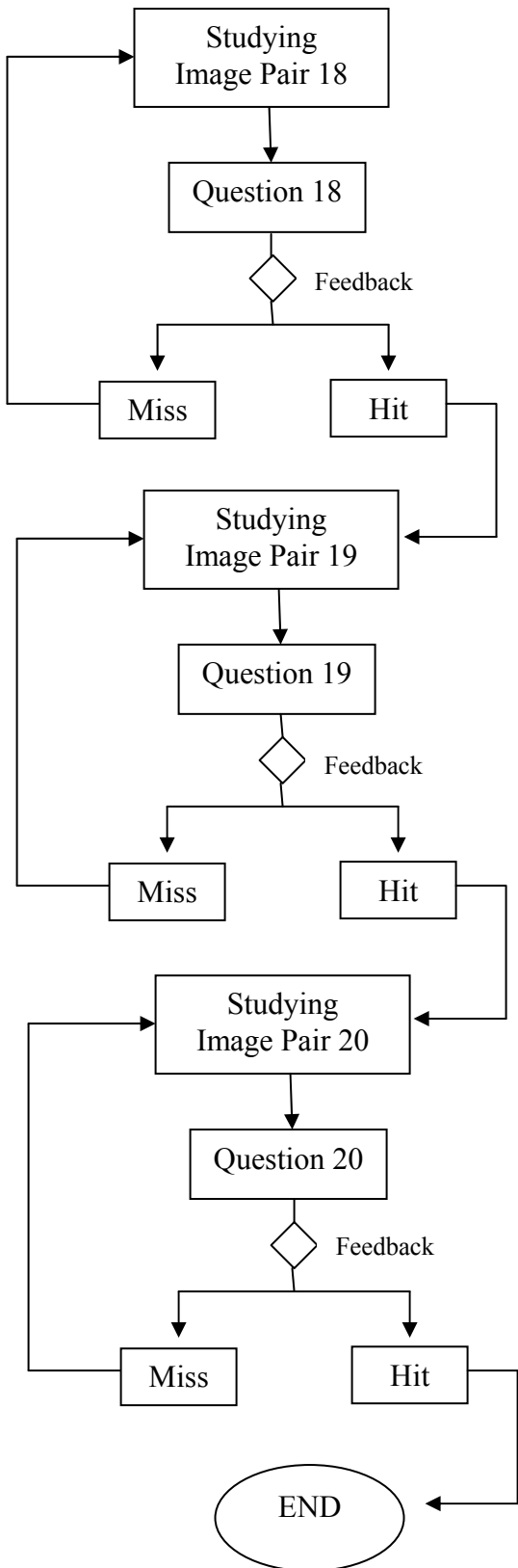












As demonstrated in Figure 3.3 on the general procedure of each study case in the study section, after the participants were instructed to complete the pretest of ten questions, they were instructed to study cases and identify abnormal patterns by looking for the change or difference in the two images of each case. Afterwards, the participants were given an inquiry of patterns upon each case, followed by feedback on the response(s) to each case. If they were wrong in identifying the change, they would be brought back to the previous images and activity and try again. If they responded correctly, they would move ahead to the next set of images. The same sequence was repeated until all the cases were completed. All of the images were in digital formats with high resolution, edited when necessary.

Three Independent Treatment Programs

There were three instructional strategies, so three versions of independent treatment programs were developed. Specifically, in the program with the comparison treatment (see Appendix B), the participants were instructed to compare the two images displayed on the same screen and identify the change(s) across the two images. The participants were instructed that if they thought they detected the change, they could then stop the study tasks, continue to go ahead with an assessment task and choose the malignancy on the image they had studied by clicking on the pattern they identified. If they correctly detected the pattern, they could move onto the next study case. Otherwise, they would be provided with the same case to study the same set of images with the same search task until they responded correctly and then they could move onto the next case. In the program with the flicker treatment (see Appendix C), the participants were instructed to search for the change in two alternatively flashed images with a blank screen in

between. The other parts of assessment and feedback models in the flicker treatment were the same as with the comparison method. In the program with the no-flicker treatment (see Appendix D), the participants were instructed to search for the change in two equivalent images alternatively flashed but without a blank screen in between. The other parts of assessment and feedback models in no-flicker treatment were the same as those in the comparison and flicker treatment. Instructions were provided about the tasks of studying images and searching for the changes that stand for patterns, the teaching points in the study.

The length of image display time 500 *milliseconds* proposed in the original proposal were modified and increased to about two to six *seconds* for the following reasons: First, the images in this study were not everyday scene pictures but complex radiographic images, so the duration of studying these images should consider the load of both the complexity of information and the number of items of objects contained in the images (Alvarez & Cavanagh, 2004; Phillip, 1974). Even radiologists tend to spend at least a few seconds to scan a radiograph, so longer duration of scanning needs to be given to novices in both the flicker and no-flicker treatment. Second, visual short-term memory differs from sensory memory, with the former lasting from 600 *milliseconds* to a few *seconds* and the later less than 300 *milliseconds* (Phillip, 1974). The lengthening of displays can put learning more in the area of visual short-term memory rather than sensory memory. Third, the 500 *milliseconds* proposed in the original study will probably cause stress and result in visual fatigue easily in learners because of the unstable nature of flickers in the flicker and no flicker treatment, so the duration was adjusted to what made it possible for the learner to view entire images. However, because of lack of evidence in

optimal display duration, the rule of thumb was used in making decisions to make it challenging for beginners as well as possible to observe, search, encode, and compare image patterns.

Furthermore, user control of rates of displays was provided to the participants and duration options were not limited to one time span of image display but three duration options were provided to learners in both flicker and no-flicker tasks, considering individual characteristics and the common rule of applying interactivity in educational software design and development.

The user control of display duration was considered as an integral part of different instructional strategies, so it would not influence the manipulation of the independent variable and control of extraneous variables. As parts of the flicker and no-flicker activities, the speed modes of displayed images were regarded as internally embedded interactions for students to choose from rather than one single speed. Otherwise, if argued from the other perspective and proposed just one speed option and provided that one to the participants, the speed modes would still vary from treatment to treatment, therefore it would still be a potential extraneous variable. That is, the displays in different instructional treatments vary in duration, which is one of the properties or nature of the treatments. In addition, the pilot studies indicated that duration options were individual-based and variations of selection of duration were found even for the participants who studied with the same instructional strategy. Therefore, the proposed perspective for duration options was to regard the speed modes as an internal part of treatments and provide the participants different duration options.

Instruments of Dependent Measures

To measure the participants' prior knowledge of recognizing abnormalities in mammograms and performance of what they learned about the visual category and instances, three criterion instruments were designed, developed, evaluated and implemented, including a pretest instrument (Appendix E), a posttest recognition instrument (Appendix F), and a posttest classification instrument (Appendix G). The three instruments contained three sets of questions, including another set of 10 proven images except the images investigated in the study session, and the pretest was composed of the same cases as those in posttests but in different orders. The two posttests consisted of 10 studied images for recognition questions and 10 unstudied images for classification questions.

The posttest instruments were developed according to how recognition and classification tests were defined in this study. As for each recognition question, participants were provided with images they examined in study sessions and were instructed to identify the malignancy they observed in study. For classification questions, participants reviewed images that they did not study in the study sessions and were asked to identify malignancy and classify instances as examples of the concept. The images used for classification questions did not appear in the study. Each of the questions in the two instruments counted for 2 points and there were 20 points for the ten questions in each of the two tests. Each item in the pretest was also counted for 2 points and thus 20 points were the total perfect score in the pretest.

In addition, to deal with the potential issue of the impact of short-term memory in study upon assessment and help clear the short-term memory, the first few images in the

recognition test were the images that the participants learned at the beginning of the study session. Because of the short span of about half a second to a few seconds' duration of visual short-term memory, the short-term memory was supposed to be cleared with this method. The other optional method to help clear short-term memory was three math problems that the participants were asked to solve.

Validating Instruments

To validate the independent and dependent measure materials, evaluation was conducted with evaluation instruments (Appendix I). The participants of the evaluation were subject area experts, IT experts, and the participants in the first pilot study. Four subject area experts (SME) with M.D.s were invited to assess the instruments and three of them provided their responses to the evaluation queries. One of them has specialty in radiology with more than twenty years' experience in instruction and research. The other two specialize in pathology and have more than ten years' experience in research and instruction. When one of them was provided with the instruments, the researcher was told that the instruments could not be opened, so the researcher went over the instruments together with experts. Beyond the assessment of the images and related issues, four IT experts were invited to provide their evaluation of the instruments. One of the IT experts has more than twenty years of instructional and research experience in programming, instructional design, and evaluation. Versions of prototypes were provided to this expert for formative and summative evaluation and the expert provided suggestions and comments that will be explained in the evaluation results section. One of the other IT experts has more than ten years of experience in instructional design and works as an instructional designer in a multimedia company. The other two IT experts have several

years of instructional design and technology study and working experience. The latter three IT experts were shown the programs and suggestions were solicited from them. In addition, seventy six college students participated in the evaluation of the instruments, going through the pretest, the study materials and the posttests and completing a usability survey. Open-ended questions were utilized for experts to evaluate the instruments for details and depth of information. Appreciations were expressed to the experts and professional participants and compensation gifts were given to the usability study participants.

To warrant that the instruments were good tests that could measure what were planned to measure and generate consistent scores, validity and reliability of the instruments were evaluated and validated. To search for evidence of validity, the following procedures were used to analyze both content validity and construct validity of these instruments.

First, to search for evidence of content and construct validity, a test blueprint was developed stating what were intended to learn and what each set of test questions should include. To guarantee content validity, the tests were supposed to test pattern recognition that learners learned in the study activities and should have contained the images, the image features and patterns that the study materials covered rather than irrelevant ones. Specifically, the learning objective of recognizing image features and patterns were provided to two subject area experts to examine whether the content of the visual category in tests matched that of the study cases. Regarding construct validity, the basic criterion is that the tests should be consistent with and reflect the construct of pattern recognition and assess the construct. The construct of pattern recognition was analyzed

and identified to consist of recognition and classification, which would be further explained in the following sections. Therefore, the definitions of the two terms of recognition test and classification test were offered to the two subject area experts to validate the construct validity of the dependent instruments. The researcher collected their opinions on the content and construct validity of the instruments.

Second, reliability of the criterion measures was examined through pilot studies. The test scores from the pilot studies were analyzed and found that the tests could discriminate learners and learning. The phenomenon of extreme low or high scores in the tests was uncommon, so the tests seemed to show variability among participants and discriminate performance among different learners. The internal consistency of the instruments was also considered and the Cronbach's alpha derived from the pilot study will be analyzed in more details in the follow-up section.

Results of Instrument Validation

The subject area experts (SMEs), IT experts, and the first pilot study participants identified that both the dependent and independent measures were valid according to the criteria and could be used to conduct the study later on. They simultaneously provided their revision suggestions and the researcher revised the instruments considering their comments and suggestions. When provided with the instruments to one of the SMEs, the expert told the researcher that the instruments could not be opened, so the researcher went through the instruments with the three experts individually by showing them the study cases, the tests, and the instruments. While they provided positive responses to the evaluation questions, one of them suggested that cases be reconsidered if the instruments were employed in real-life instructional and learning situations, which could be discussed

in the limitation section of this study. One reason was that edited images could be changed and edited cases did not sufficiently reflect clinical studies. The other reason was that it might take a group of experts to collect these cases in a long run, which was beyond the scope of this study. Furthermore, they provided suggestions on images. One of the experts suggested that one of the images in the study be changed because the case might be too difficult to naïve learners and the case was replaced with a more appropriate case in the level of difficulty. They also suggested that a few other images be changed, the patterns of which did not belong to the category that was instructed in the independent instrument. They thought that such demographic data as age to be collected because learners of different ages may vary in learning. The SMEs also recognized that the study materials had sufficient levels of breadth and depth, they were structured generally according to the difficulty levels of the cases from easy to difficult, the instructional methods used in the study may be useful for instruction and learning, and the material could be used independently or in a blended format.

The IT experts provided their evaluation suggestions and comments throughout prototype design and development. One of the IT experts, with more than twenty years of experience in instructional technology doctoral program mentoring, instruction, research, and design and development, has been mentoring and guiding the researcher to design and develop the programs from scratch. In the iterative design, development, and evaluation processes, the expert examined the program codes, identified problems in codes, and provided suggestions on programming work. Before the proposal defense, the focus of mentoring was to guide the design of IT affordance on the bases of human learning theories and instructional design principles. After the proposal defense, the focus

of mentoring was to guide the design and development to address the questions raised in defense outcomes through evaluation of codes, interface, interactivity, and the other portions of the programs. An example was that the expert noted that the duration measuring codes “compute a minute as equal to 100 seconds. You need to fix this for the actual study. For the pilot results I think that you will need to ignore the program-computed duration, make a valid computation using the start and finish times, and then re-run your analyses”. The researcher then revised the codes and the expert assessed the codes for computation accuracy. For the pilot study results, the Excel program was used to calculate the duration from the starting and finish time. Furthermore, the expert suggested to improve instruction messages for the study task in the programs and provided a revision example. The expert also provided comments and evaluation of the screen design and suggested to enhance background design that may make image features salient and help learners focus their attention on the images. Both the study instructions and the backgrounds of the programs were revised according to the expert’s comments and suggestions. With more than twelve rounds of formative and summative evaluation of many versions of prototypes of the programs and instruments as well as research, the expert also suggested that the researcher apply an item analysis to evaluate the images used in the test and study.

The other IT expert has more than ten years of experience in instructional design. This expert went through the programs and provided the suggestion that menus for navigating the programs be added. The researcher explained that menus were planned not to be developed for this experimental study in order to control variables. Two other experts have more than five years of study and working experience in instructional design

and technology. They were shown the programs and they offered their comments and suggestions on the programs. One of them provided a revision of the instructions in the introduction part of the programs. The other one gave suggestions on the placement of buttons to keep the consistency of screen design. Their suggestions were taken into consideration in revisions of prototypes of the programs.

Generally, the IT experts identified that the study materials and assessment materials have reliable theoretical support, with learning and instruction relevant and sufficient interactivity, including feedback. The screen design follows the basic principles of instructional design and the presentation of information can facilitate learners to become focused on study. They also recognized the appropriate chunking of information, understandable structure of the materials, and ease of use. About the load of the materials, they thought, for the groups that would be tested in the study, the material may be somewhat challenging but this challenge may be located in the participants' zones of proximity. They also commented that the screen design elements worked well, including the background, texts, colors, and other parts of presentation.

In addition, the 76 participants in the first pilot study rated the programs with the usability survey. Table 3.1 demonstrates the mean scores, standard deviations, minimum scores, and maximum scores that each group of the participants had. Generally, the mean scores of items ranged from 3.85 to 4.65, which indicated that about and more than 80% of the participants thought the programs easy to learn and efficient, with comprehensive structures, and simple and consistent in operation. The ratings of their overall impression of the programs indicated that the programs had reached a certain level in terms of usability and could be employed to conduct the experiments. More particularly, the

results of item 1 indicated that the participants in the comparison program demonstrated the highest rating of the ease of the program, mean score=4.50, those in the no-flicker treatment program had the lowest mean score among the three groups, mean score=3.96, and those in the flicker program had the mean score of 4.04. Interestingly, the mean scores of the other evaluation items also demonstrated the same pattern as that in the first item, with the highest score given to the comparison program, the lowest score to the no-flicker treatment program, and the flicker treatment program in between. However, it was unclear whether these rating differences had any statistical significance.

Table 3.1

Response Results by Treatment Group for the Usability Test Survey Items in the First Pilot Study

Question	Group	N	Mean	Std.		
				Deviation	Minimum	Maximum
Item 1	Comparison	26	4.50	.648	3	5
	Flicker	26	4.04	.958	2	5
	No-flicker	24	3.96	1.122	1	5
	Total	76	4.17	.944	1	5
Item 2	Comparison	26	4.27	.962	2	5
	Flicker	26	4.23	.815	3	5
	No-flicker	24	3.92	.974	2	5
	Total	76	4.14	.919	2	5
Item 3	Comparison	26	4.27	1.041	2	5
	Flicker	26	3.96	.916	2	5
	No-flicker	24	3.83	1.049	1	5
	Total	76	4.03	1.006	1	5
Item 4	Comparison	26	4.65	.745	3	5
	Flicker	26	4.42	.987	1	5
	No-flicker	24	4.17	.917	2	5
	Total	76	4.42	.898	1	5
Item 5	Comparison	26	4.35	.846	2	5
	Flicker	26	4.12	.864	2	5
	No-flicker	24	3.92	1.100	1	5
	Total	76	4.13	.943	1	5

To examine whether there were significant differences of group ratings towards different programs, an analysis of variance was conducted. Table 3.2 shows that there were no significant differences between groups for all of the items, $F=2.547$ and $p=.085$ for item 1, $F=1.093$ and $p=.341$ for item 2, $F=1.261$ and $p=.289$ for item 3, $F=1.878$ and $p=.160$ for item 4, and $F=1.311$ and $p=.276$.

Table 3.2

Results of ANOVAs for the Three Groups' Responses to Each Item of the Usability Test Survey in the First Pilot Study

		Sum of Squares	df	Mean Square	F	Sig.
Item 1	Between Groups	4.356	2	2.178	2.547	.085ns*
	Within Groups	62.420	73	.855		
	Total	66.776	75			
Item 2	Between Groups	1.844	2	.922	1.093	.341ns*
	Within Groups	61.564	73	.843		
	Total	63.408	75			
Item 3	Between Groups	2.537	2	1.269	1.261	.289ns*
	Within Groups	73.410	73	1.006		
	Total	75.947	75			
Item 4	Between Groups	2.962	2	1.481	1.878	.160ns*
	Within Groups	57.564	73	.789		
	Total	60.526	75			
Item 5	Between Groups	2.312	2	1.156	1.311	.276ns*
	Within Groups	64.372	73	.882		
	Total	66.684	75			

Note. *ns = not statistically significant ($p > .05$).

Furthermore, the study materials were validated through SME reviews. To determine whether the test items reflected the content that was planned to be covered and the construct that was intended to be examined through the tests, content and construct validity were investigated by examining the test items and the criteria of these two types of validity. The criteria of the content validity were the learning objectives, i.e., whether test items reflected the learning objectives to be learned. The criterion of the construct validity was that logically whether the test items reflected the construct interested to be examined through proposed measures.

Learning objectives were examined and identified through content analysis and task analysis. One of the learning objectives in the learning session was that when learners were provided with images that they studied in learning sessions, they could correctly identify the patterns in the images. The other learning objective was that when learners were provided with images that they did not view in study sessions, they could correctly identify the patterns in the images. Based on the learning objectives, two sets of test questions were created with one set testing with the image cases that appeared in study sessions and the other set testing with the image cases that did not appear in study sessions. To validate the content validity, the learning objectives were provided to the SMEs, who looked through the study cases and the test questions and gave their judgments on whether what were to be learned, were tested. The alignment of objectives with tests was identified to indicate that the tests were valid in terms of the content. That is, some or all of the image cases that appeared in the study were presented as test questions in the recognition test and all the cases in the recognition test were cases that learners would study in the learning session. The classification questions were questions that learners did not study before in case studies and all of the learned cases were not included in the classification test. Table 3.3 shows the relationship between learning objectives and the ascertained content validity:

Learning objectives	Study sessions	Assessment questions	Content valid or not
1. Provided with previously studied cases, participants can identify patterns with accuracy	Cases with the patterns planned to study and stated in instructional design of learning objectives	1. Questions are previously studied cases	Ok
2. Provided with cases not studied in study sessions, participants can identify the patterns they learn in study sessions		2. Questions are not previously studied cases	Ok

Construct validity means that the construct that was intended to be examined in the plan was measured with instruments. The construct that was planned to investigate in this study was pattern recognition. The two measurements used to measure pattern recognition in learners were respectively a recognition test and a classification test. Derived from previous research of radiology expertise, both recognition and classification were essential for diagnosis. Research of perceptual processes also indicated recognition as an important result and procedure in human perception (Goldstein, 2002). These theories helped establish the logical relevance between pattern

recognition and the proposed assessment methods. Logical evidence could also be provided by a further factor analysis of pattern recognition, indicating that pattern recognition included four componential results, illustrated in Figure 3.4.

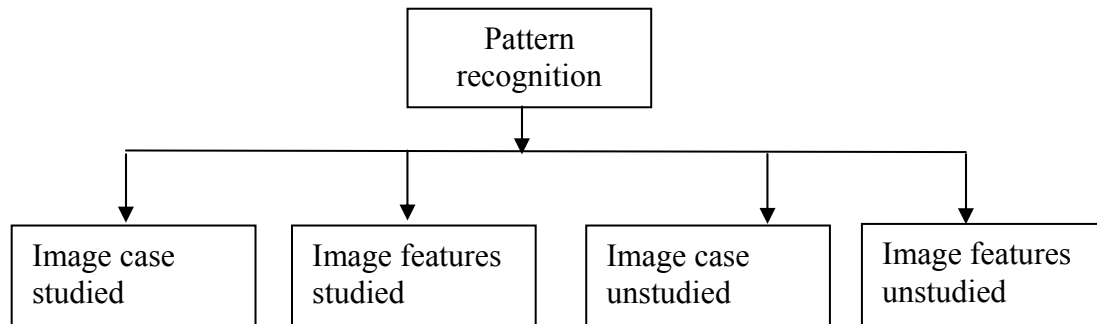


Figure 3.4 An analysis of the construct pattern recognition

In this study, the feature and the image background were regarded as one unit and hence one factor. Thus, two factors were considered, namely image cases studied before and those unstudied before. Figure 3.5 illustrates the factors and corresponding measures of the construct:

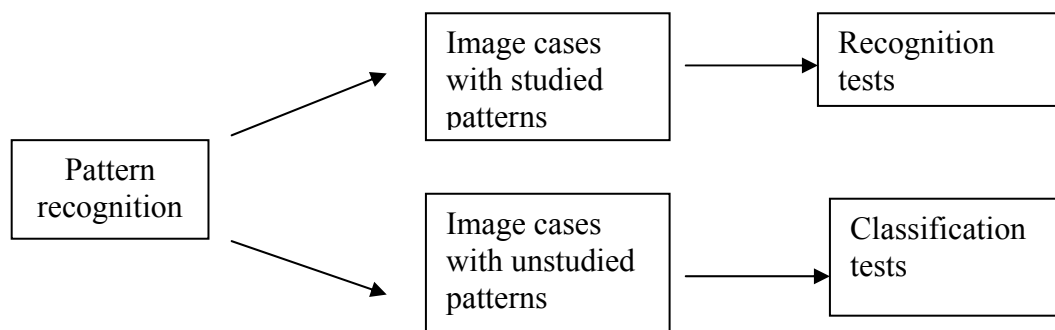


Figure 3.5 The construct pattern recognition and measures of the construct

The test items were examined based on the two factors identified in pattern recognition and the procedures identified. The recognition test items were examined that they contained the images that were reviewed in study sessions while the classification test items were checked to ensure they did not contain cases that were studied in the study

materials. With the theoretical and empirical evidence, the two sets of tests formed an investigation of the knowledge acquisition and higher-level performance of transfer of pattern recognition.

Methods and Results of the Instrument Reliability

Searching the literature through Medline database, there was no existing test instruments of recognition and classification of radiographic images and no evidence of instrument reliability. Therefore, the tests had to be initially designed, developed, and evaluated. In this situation, a more lenient criterion was used to examine the results of reliability tests.

The internal consistency of the instruments of criteria measures was investigated through the first pilot study, which was described in more details in the following section on pilot studies. The statistical software package SPSS was employed to process datasets of the participants' responses to each item respectively. Cronbach's alpha was applied to examine the test reliability, namely the internal consistency among the 10 items of the posttest one and the other 10 items of the posttest two. The reason was that Cronbach's alpha was appropriate to look at the correlation of performance test items. The results of Cronbach's alpha would range from 0 to 1, 0 meaning that a certain test does not work to measure anything and 1 represents that the scores obtained from the test are true scores without any errors.

The results from the first pilot study were that the Cronbach's alpha of the posttest one was .554 and that of the posttest two was .659. If interpreting the results with the accepted criterion of .80 in education, these two results would be considered as marginally satisfying. However, there were reasons to regard the results as reasonable

and acceptable for the following reasons: First, different from the tests of math, science, language, and other subject matters in education, tests in radiographs are far from the ease of measuring different levels of learning with few errors because these image patterns are either easy to identify or difficult to identify. That is, almost all students may get their answers right or wrong with similar tendency. Therefore, reliability score criteria may be set lower than those in the other subject areas in education. Second, existing literature and practice do not offer any instrument of recognition and classification and evidence of instrument reliability. Hence, it is almost impossible to have a high reliability test in a preliminary study of this sort.

Procedures

The general steps of the study included random assignment of the participants to the groups, administrating the pretest, instructing to learn with different interventions, administrating two posttests, and administrating a usability survey (in pilot studies only), and a demographic survey. For random assignment, the following steps were operated. First, the researcher asked the participants about their consent for participating in the study. Second, the participants were appreciated for their voluntary participation. Third, the participants were randomly assigned to one of the three groups by using a table of random numbers. The participants were provided with numbers in a bag to choose from. After they picked up a paper card with a number on it, they were asked to open the card and show the number to the researcher. The researcher then looked up the number table prepared before the study for the number and the corresponding group. Fourth, the researcher explained to the participants what would be included in the materials, including a pretest, a study session, and two posttests.

Then the instruments and study materials were administrated in the order as they were programmed. First, the participants were pretested on their knowledge and skills in recognizing radiographic image patterns. They responded to ten questions on these images by clicking on the spots that they thought abnormal patterns. After they completed pretest items, they were instructed to study cases by carefully comparing sets of images, responding to the questions on what they studied and provided with feedback on their responses. Immediately after the study session, three math questions were raised to decrease the influence of short-term memory. To clear short-term memory, another strategy used was that the first few images in the recognition test did not include the last few images in the study. Then the participants were post-tested on the recognition and classification of image patterns, with the recognition test preceding the classification test. After the posttests, the participants were instructed to complete the demographic survey (Appendix J). Finally, they were provided with the compensations. Appreciation was expressed and they were told that they could log off the program.

Ethical Considerations

In compliance with the regulations and guidelines of human subject protections, after the researcher passed her proposal defense, an institutional review board (IRB) package was written and compiled based on the proposal. The researcher sent it to her Major Professor for comments and suggestions, revised it accordingly, and sent to IRB for review and approval. The application was approved and a written approval was received from IRB (Appendix A).

Statistical Analysis Procedures

To obtain results and respond to the research questions raised in the study, it was planned to go through the following data coding, entry, observations, and initial computation procedures to prepare for statistical analysis after data collection of the formal study. First, raw data were coded, input and organized into datasets with the Excel program. Names of the fields were entered in the first row and each record could be identified by a unique identification number. The organized datasets could be conveniently imported into the statistical program package SPSS for Windows for further analysis. Second, the organized data were observed to identify the number of outliers. The records of extreme high scores, for example, scored 20 or 0 in all of the three tests, were disregarded in statistical analysis. Third, scores were also examined to check if there were any missing scores and unreasonable scores. If these scores were identified, the records containing these scores would be eliminated in the subsequent statistical tests.

Research questions were responded through hypothesis testing methods in statistics because statistical significance of manipulated variable effect and group differences in an experiment were supposed to be evaluated by using statistical methods. Hypothesis testing in this study was conducted with the statistical probability rate set at 95% and the alpha level set at .05. Considering the multiple dependent variables, one independent variable, and a covariance examined in the study, the test results collected from the experiment was analyzed with Multivariate Analysis of Covariance (MANCOVA) and Analysis of Covariance (ANCOVA) with SPSS, as shown in Figure 3.6, to respond to the first and second research questions and evaluate if and in which criterion measure the participants performed differently. The former procedure was

utilized to assess the overall differences and the latter one was used to determine the difference of each of the three dependent variables, both with pretest scores as the covariate to equate the initial state of the study across groups if the conditions of homogeneity hypothesis assumption could be satisfied. Furthermore, to ensure the appropriateness of using the covariate analysis method, two presumed aspects were checked. One respect was whether the pretest was related to the dependent variables. To check the relationship, a Pearson Correlation was calculated to seek for evidence of the relationship between pretest scores and performance scores. The other respect was to scrutinize one of the assumptions of MANCOVA: the homogeneity assumption. The homogeneity of the slopes of linear regression of the three groups was examined by conducting a homogeneity test. If significant differences among the slopes were not found, the assumption then would be regarded as satisfied. However, if significant differences were found, the conditions of the assumption could not be satisfied. In the former case, four types of tests then were used in hypothesis testing, including Pilla's test, Wilks' test, and Hotelling test. In the latter case, a special test was used for hypothesis testing. Furthermore, the other assumptions of MANCOVA and ANCOVA were watched, especially the independence assumption, meaning that the participants completed the study and tests independently. In addition, each group had the similar number of participants to improve the possibility of keeping the covariance assumption of MANCOVA.

In addition to these procedures to address the first three research questions, post-hoc adjusted mean tests were performed to evaluate and locate specific group differences in order to respond to the fourth research questions. The results from the post-hoc tests

were scrutinized to see where the participants' performance differences were located; the results would indicate in details between which groups significant differences of the performance scores were found and/or between which groups there were no significant differences statistically.

Beyond these tests of instructional strategy effect on recognition and classification of radiographic images, examinations of the dependent factors of duration, number of incorrect responses, number of trials in the study session were conducted to evaluate the group differences in these three factors, evaluate significant differences, and respond to the corresponding research questions. To implement the investigation of these three factors, the following statistical procedures were used: a comparison study of group differences of the factors was conducted. More particularly, one-factor analysis of variance (ANOVA) was individually used to test whether there was any statistically significant difference among the groups in terms of duration, the number of incorrect responses, and the number of trials. If significant differences were identified through comparing means across groups with ANOVA, then correlation studies would be conducted to evaluate how related these three factors were with the two criterion measures recognition scores and classification scores. If they were identified correlated with the scores, then these factors would be taken as covariates and further assess effect by ruling out the influence of these factors through another run of MANCOVA.

In summary, the computer-based software programs of Excel and SPSS were applied to enter, code, clean, process, and calculate key descriptive and referential statistics, the significant tests of MANCOVA and ANCOVA were performed to evaluate whether an overall significant difference exists or not and in which measures the

difference lie. To evaluate the pair-wise differences among groups, adjusted mean post hoc procedures were applied. To evaluate the group differences in duration and the number of incorrect responses and trials, ANOVAs were used and the results showed if there were any significant group differences in these three factors.

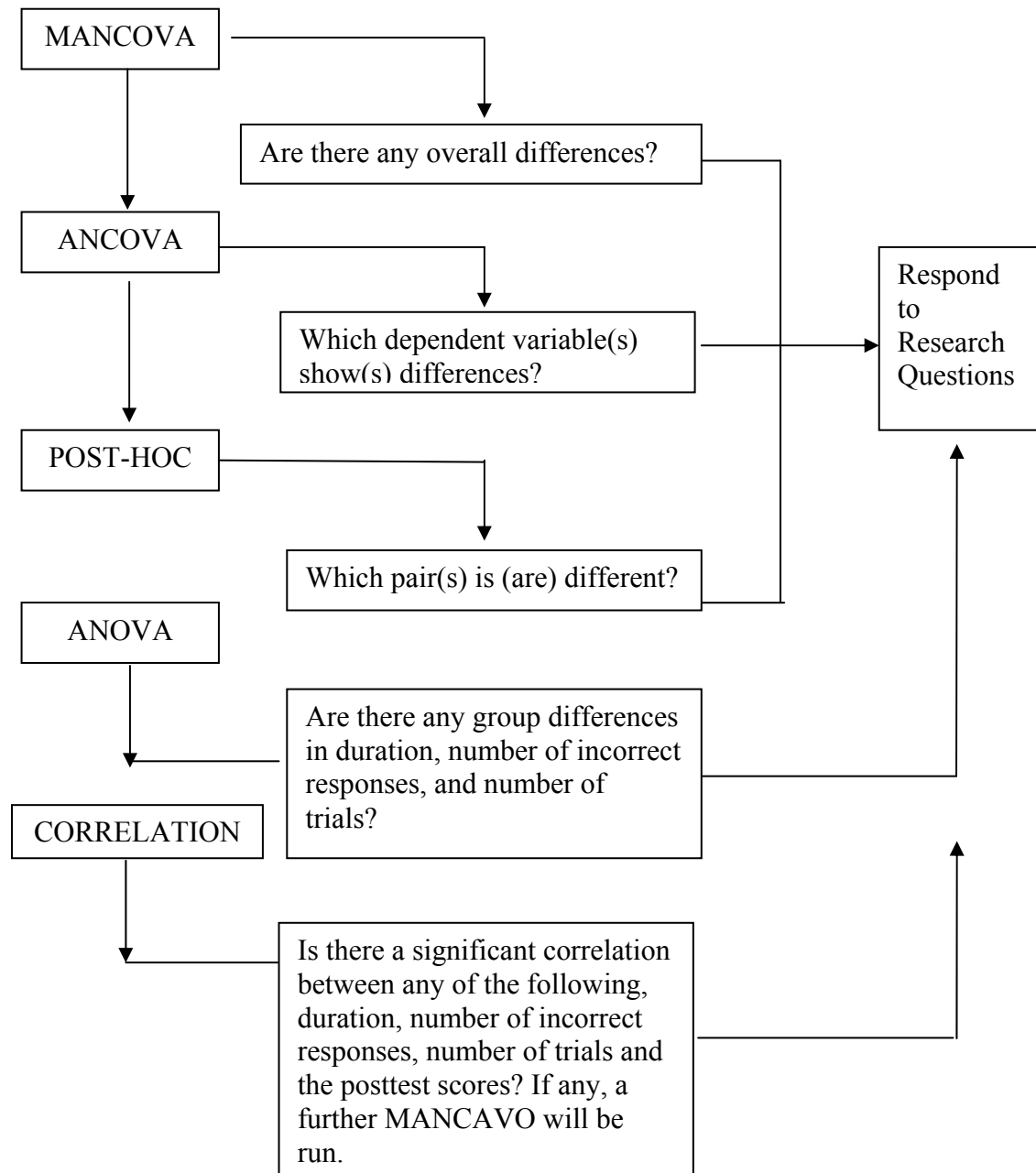


Figure3 6 Diagram of an overview of statistical analysis procedures

Pilot Studies

Two pilot studies were conducted to obtain data about the usability of the dependent and independent instruments and materials, the reliability evidence of the dependent instruments, the feasibility and implementation of the proposed study procedures, the comparison of effectiveness of the instructional strategies upon recognition and classification performance, the initial results of the group differences in on-task duration and the number of incorrect responses, and the duration options in the flicker and no-flicker groups.

The First Pilot Study

The primary purposes of the first pilot study were to observe the usability of the dependent and independent instruments, feasibility of data collection, and practice of proposed research procedures. In order to collect usability data, a usability survey (Appendix K) was implemented. This pilot study also functioned to scrutinize the internal consistency of the instruments used for criterion measures of recognition and classification and obtain data to practice and check the statistical analysis plan. Here was an itemized description of the first pilot study in terms of its participants, procedures, settings, and observations:

The sample of the first pilot study consisted of seventy six participants ($n = 76$), with the number of 26, 26, and 24 participants randomly assigned to the comparison group, flicker group, and no-flicker group. The participants were primarily undergraduate students, majoring in arts, science, and engineering

The study was conducted in a computer room with computers of similar configurations: the Window XP operating system and quality monitors. Students were

seated separately with spaces in between and there were boards between the seats so that participants worked independently on their studies.

Every participant was instructed to complete a demographic survey (Appendix J) after they completed the program. Figures 3.7, 3.8, 3.9, and 3.10 illustrate the demographic information specifications of the participants. Generally, the majority of the participants were female undergraduate students between the age of 15 and 25. More specifically, the frequency of the male and female participants was 32 and 44 and the percentages of males and females were 42.1 % and 57.9%. The components of age groups were that there were 67 (88.2%) of the participants between the age of 15 and 25, 8 (10.5%) participants in the age group of 26-35 and only one (1.3%) of the participants in the age group of 36-45. There were no participants in the other age groups. For ethnicity, 38 (50.0 %) of the participants were White, 13 (17.1%) were Black, 7 (9.2%) were Spanish, 9 (11.8%) were Asian, and 9 (11.8%) belonged to the other ethnical groups. For educational programs, 71 (93.4%) of the participants were undergraduate students while 5 (6.6%) of the participants were from graduate programs.

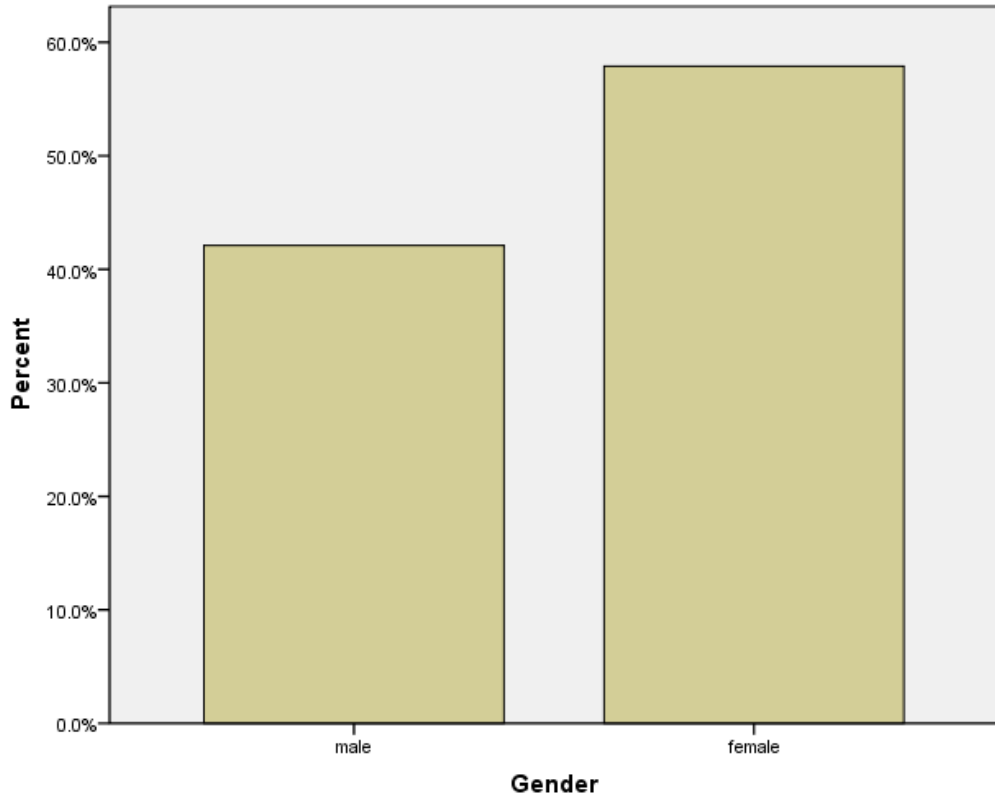


Figure 3.7 Participants gender distribution in the first pilot study

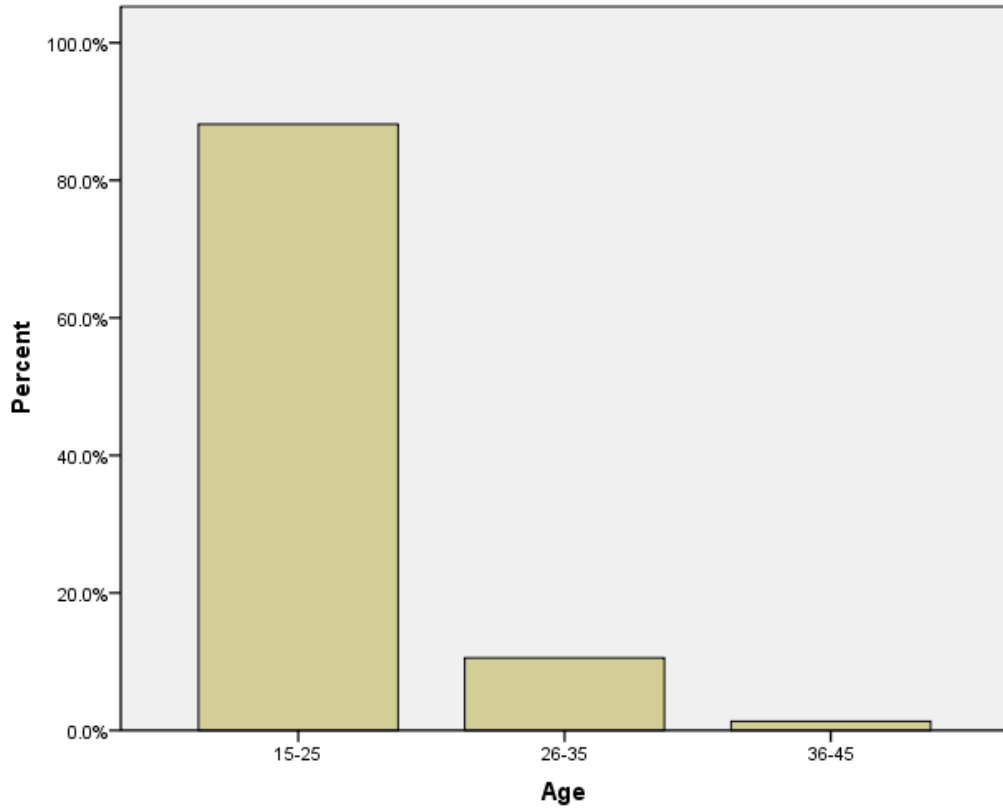


Figure 3.8 Participants age distribution in the first pilot study

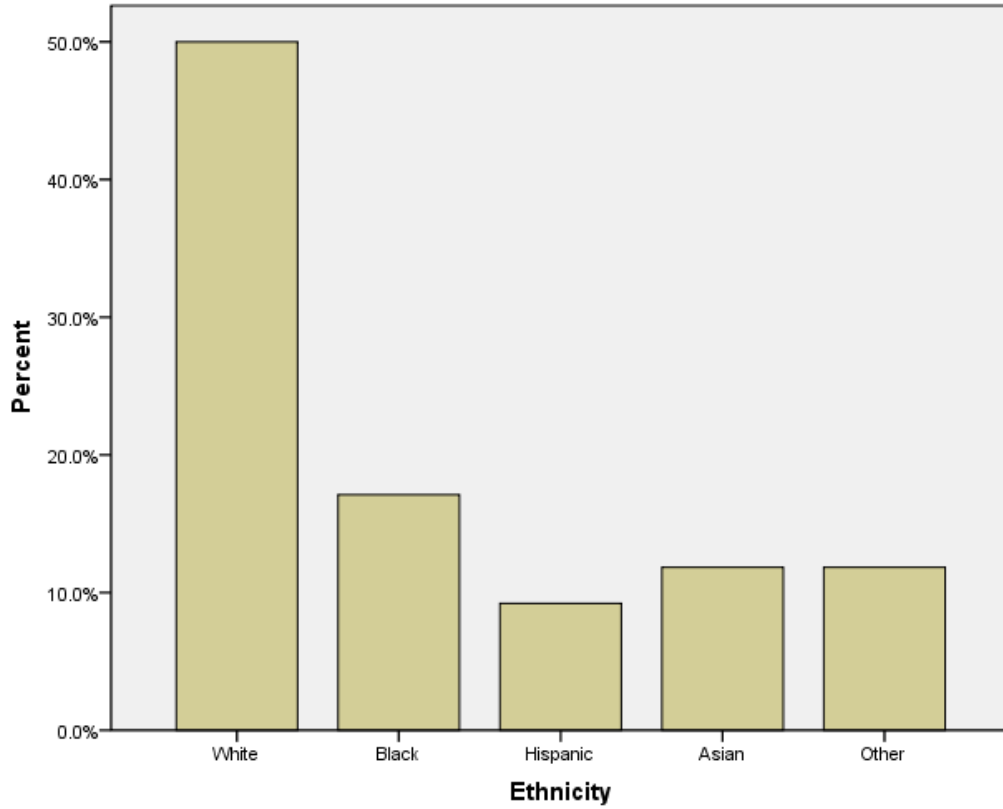


Figure 3.9 Participants ethnicity distribution in the first pilot study

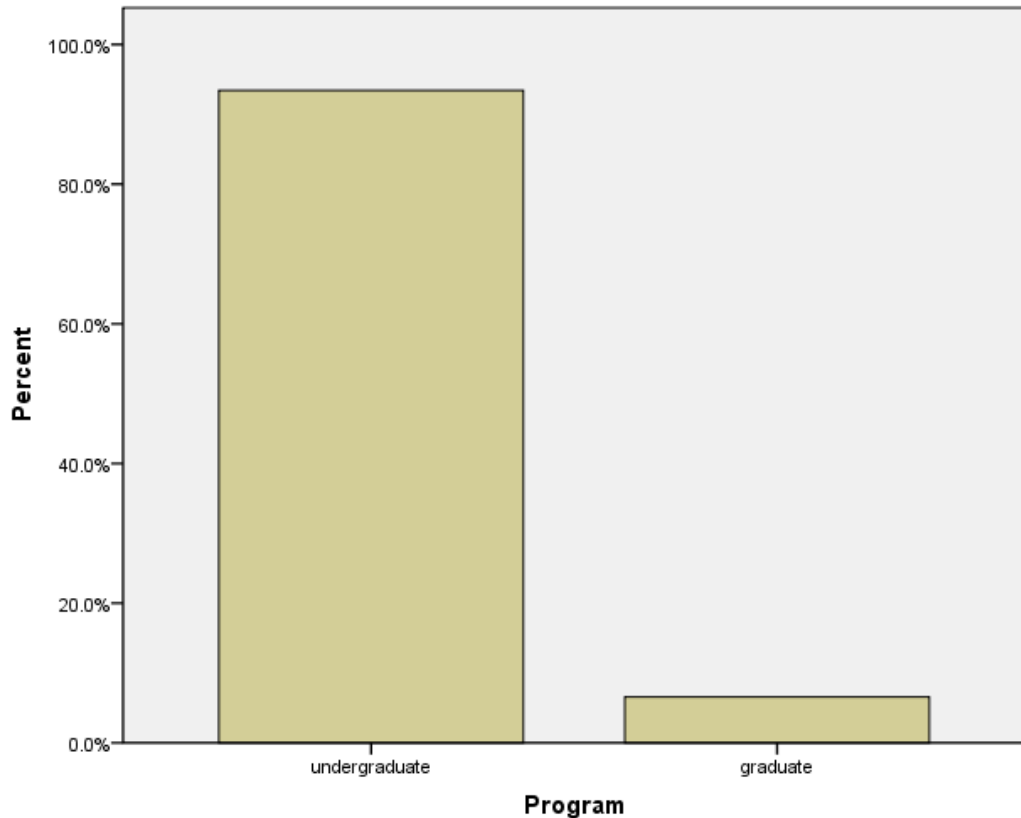


Figure 3.10. Participants program distribution in the first pilot study

Usability Tests

The usability of the program was tested with a usability survey (Appendix I). The results from the survey were presented in a previous section (See Table 3.4 and Table 3.5 and related explanations) about the mean scores, standard deviation, minimum, and maximum scores of each group for each item of the survey. What was not presented about the analysis of these results was that the mean scores of the overall impression of the programs' usability were respectively 4.35, 4.12, and 3.92.

Instrument Reliability

Reliability of the posttest instruments was scrutinized with the index of Cronbach's alpha, which indicated the internal consistency of test items within a test. Table 3.4 showed that the alpha value .554 for the items in the recognition test and .659 for the items in the classification test. The reliability coefficient for the first posttest indicated that the internal consistency level among the 10 items in posttest 1 was marginal and less than a usual satisfying level of consistency. The alpha statistics of .659 derived from the second posttest was higher than posttest 1 and closer to the satisfying level in education and may be regarded satisfying. Given the fact that the difficulty level of this type of images tend to be too difficult or too easy to identify, as explained in the previous section, the leniency in considering the internal consistency could be understandable. Furthermore, it took further research and time to build up more reliable test items if possible. There were no existing reliability test results in literature and practice.

Table 3.4

Test Reliability Coefficients of the Posttest Instruments

Criterion Tests	Number of Items	Cronbach's alpha
Recognition	10	.554
Classification	10	.659

Note. There are 10 items in each criterion test.

Table 3.5 shows that the Pearson Correlation of the pretest, recognition test, and classification test. The values indicated that there were significant relationships between the pretest and the recognition test, Pearson Correlation = .333 and $p=.003$ and the recognition test and the classification test, Pearson Correlation=.612 and $p=.000$.

Table 3.5
*Pearson Correlation Values Indicating the Relationship
 between the Pretest and Posttests in the First Pilot Study
 (n=76)*

		Pretest	Posttest1	Posttest2
Pretest	Pearson Correlation	1	.333**	.162
	Sig. (2-tailed)		.003	.163
	N	76	76	76
Posttest1	Pearson Correlation	.333**	1	.612**
	Sig. (2-tailed)	.003		.000
	N	76	76	76

Note. ** Correlation is significant at the 0.01 level (2-tailed).

Table 3.6 shows the number of participants, means, standard deviations, ranges, and the measures of kurtosis, and skewness of the participants' scores in pretest, posttest 1, and posttest 2 in each of the experimental and control groups.

Table 3.6

Mean, Standard Deviation, Sample Size and Other Descriptive Statistics Results of the Three Tests by Treatment Group in the First Pilot Study

Instrument	n	Mean	Std. Deviation	Range	Kurtosis	Skewness
Side-by-side comparison control group						
Pretest	26	6.62	4.826	18	-.501	.681
Posttest 1	26	15.85	2.588	10	-.071	-.087
Posttest 2	26	15.54	2.486	12	2.192	-1.015
Flicker experimental group						
Pretest	26	9.31	4.038	14	.292	-.970
Posttest 1	26	15.77	2.286	8	-.779	.591
Posttest 2	26	14.77	2.338	8	-.448	-.513
No-flicker experimental group						
Pretest	24	6.75	5.067	16	-1.322	.273
Posttest 1	24	15.42	4.624	20	5.881	-2.259
Posttest 2	24	13.00	4.755	20	1.994	-1.228

In the pretest, the distributions of the scores of the participants in the comparison control group and no-flicker group had relative positive skewness (.681 and .273), indicated with longer right tails while the distribution of the scores of the participants in the flicker group had relative negative skewness (-.970), indicated with a longer left tail. As for the peaks, the kurtosis values indicated that low peaks in the comparison group (-.501) and the flicker group (-.071) but slight high peak in the no-flicker group (2.192). In the recognition test, a positive skewness occurred in the comparison group (.591) and a negative skewness was identified in the comparison group (-.970) and no-flicker group (-.513). Furthermore, low peaks were indicated with the flicker group (-.779) and no-flicker group (-.448) while a slight high peak was identified with the comparison group (.292). For the classification test, negative skewness occurred with the flicker (-2.259) and no-flicker group (-1.228) and the peaks were low for the comparison group (-1.322) and high for the flicker (5.881) the no-flicker group (1.994).

Figures 3.11, 3.12, and 3.13 illustrate the group differences in terms of criterion measures of pretest, posttest 1, and posttest 2. It appeared that participants in the three groups performed similarly well in the recognition test but differently in posttest 2 across the three groups.

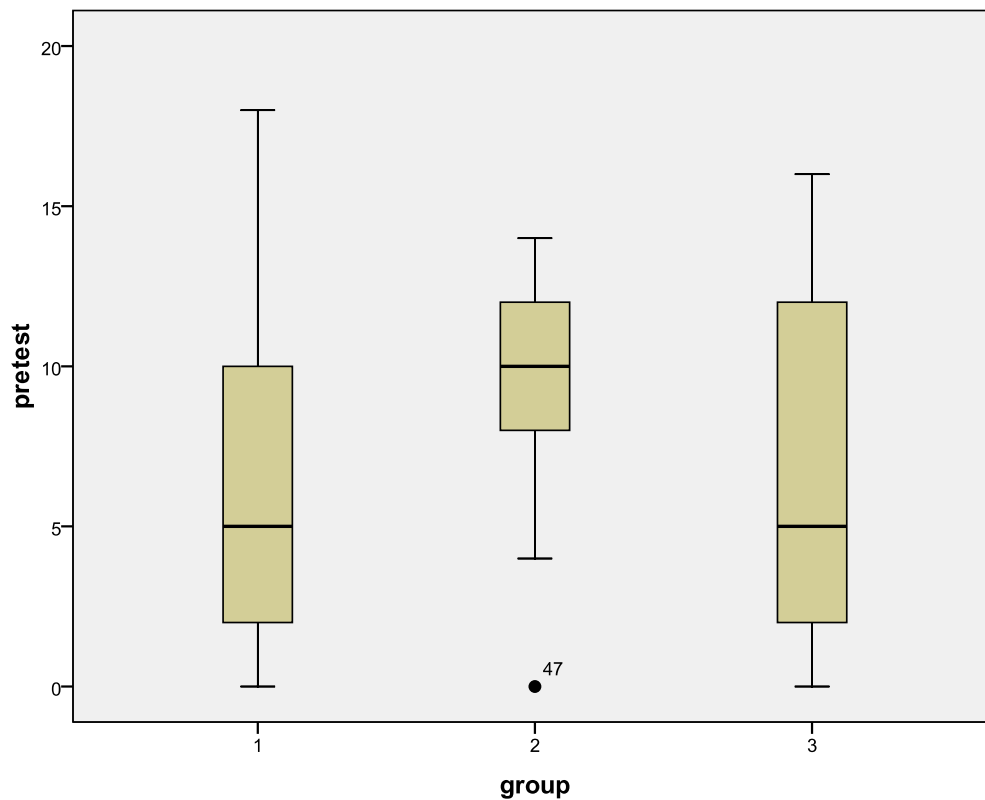


Figure 3.11. Pretest performance by group in the first pilot study shown with box plots

There were no outliers in the three groups in the pretest. The middle dark line in the boxes showed that the medium scores were not in the center of the boxes, indicating somewhat skewness across groups.

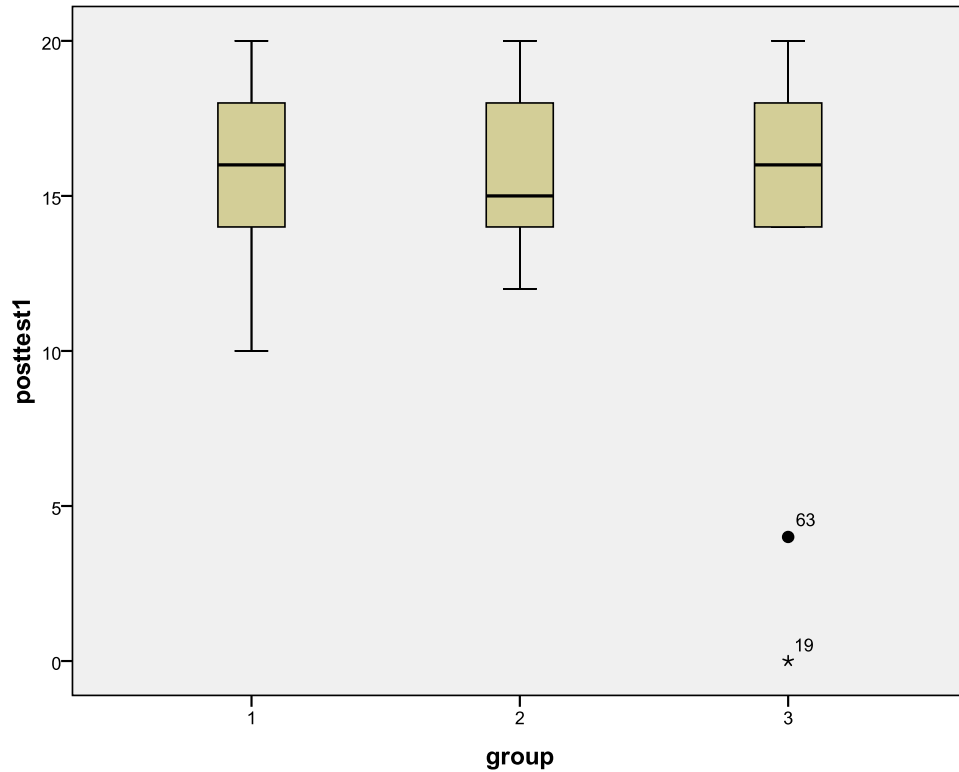


Figure 3.12. Recognition test performance by group in the first pilot study shown with box plots

Figure 3.12 indicates that there were two outliers in the no-flicker group. The medium lines indicate very slight skewness of score distribution in the comparison and no-flicker group but apparent skewness in the flicker group. The presence of only the upper whiskers for the no-flicker group indicates that 50% of the scores were above the boxes with the other 50% represented by the boxes.

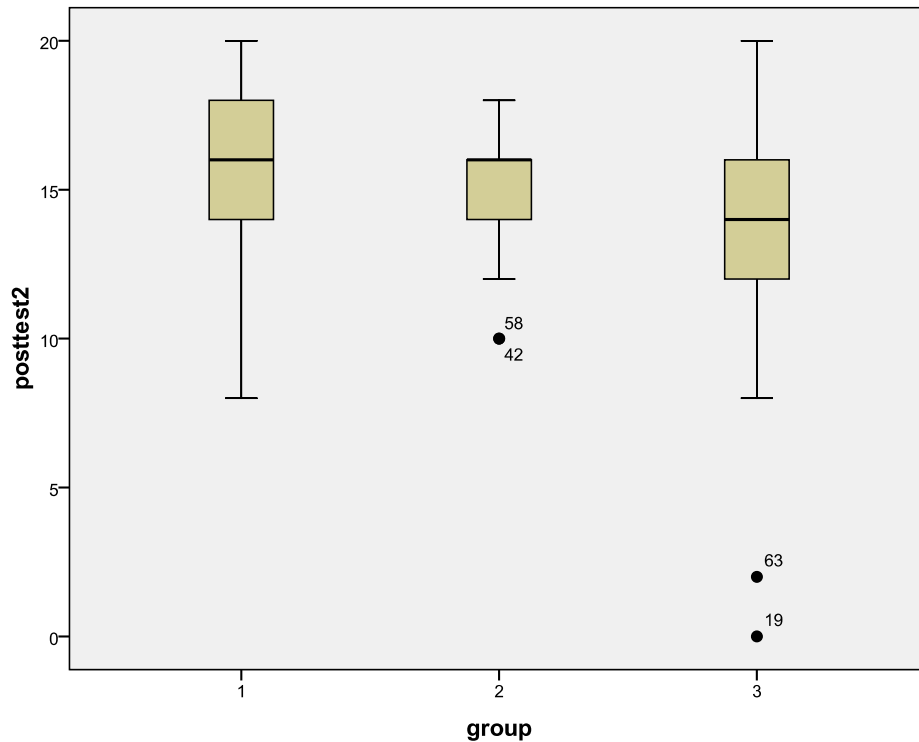


Figure 3.13. Classification test performance by group in the first pilot study shown with box plots

Figure 3.13 demonstrates that two outliers in the no-flicker group deviated from the group and two outlier somewhat deviated from the group distribution in the flicker group.

Before performing MANCOVA, two outliers in the no-flicker group identified with the box and whisker plots in the previous analysis were removed from the sample because MANCOVA test, especially the Box's test of homogeneity of covariance is highly sensitive to outliers. The implemented MANCOVA test showed that the result of Box's M was 6.409. There was no significant difference of covariance across the groups, $F(6, 105937) = 1.024, p = .407$ at the significance level of .05. The F ratio and p value indicated that there was no significant difference of covariance among the groups and the

assumption of homogeneity of variance was satisfied, so MANCOVA test could be performed to assess the overall group difference.

Table 3.7
Results of Multivariate Analysis of Covariance in the First Pilot Study

Effect		Value	F	Sig.	Observed Power ^b
Group	Pillai's Trace	.123	2.297	.062	.657ns*
	Wilks' Lambda	.879	2.294 ^a	.062	.656ns*
	Hotelling's Trace	.135	2.291	.063	.655ns*

Note. *ns = not statistically significant ($p > .05$).

Table 3.7 shows that there were no overall significant differences of the participants' performance in different groups, with the three tests of MANCOVA, including Pillai's Trace, $F(4, 140) = 2.297, p = .062$, Wilks' Lambda, $F(4, 138) = 2.294, p = .062$, and Hotelling's Trace $F(4, 136) = 2.291, p = .063$, among which Pillai's Trace is the most strict and robust test and Hotelling's Trace is the most frequently used test when there are two dependent variables. Therefore, the response to the first research question is: The participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method did not demonstrate any statistically significant differences in their overall performance as measured by recognition and classification posttest instruments.

Although no overall significant effect was identified in the above MANCOVA test, two follow-up univariate analysis of covariance was still conducted to test whether significant differences could be detected of the effect of instructional strategies upon the dependent measures because the mean scores and the box and whisker plots show some differences across the groups. To conduct these tests, a Levene's test of equality of error variance was carried out and no significant differences of variance were detected, $F(2,$

71)=.775 and $p=.465$, hence the assumption of homogeneous variance was satisfied.

Table 3.8 demonstrates the results of the univariate analysis of covariate for the recognition test, indicating no significant differences of their performance among groups in this test. Then the analysis result of no significant differences was derived, $F=1.834$ and $p=.167$, with the significance level alpha set at .05. The response to the second research question is: The participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method did not demonstrate any statistically significant differences in their recognition performance as measured by the recognition posttest instrument.

Table 3.8
ANCOVA Results of the Group Recognition Scores in the First Pilot Study

Source	Sum of Squares	Mean Square	F	Sig.	Observed Power ^b
Group	18.423	9.212	1.834	.167ns*	.370
Pretest	37.584	37.584	7.485	.008	.770
Error	351.507	5.022			
Total	19472.000				
Corrected Total	399.784				

Note. *ns = not statistically significant ($p>.05$).

To further assess the significant differences upon the classification test, another Levene's test of equality of error variance was carried out and no significant differences of variance were detected, $F(2, 71)=1.214$ and $p=.303$, hence the assumption of homogeneous variance was satisfied. Table 3.9 demonstrates the results of the univariate analysis of covariance for the classification test, indicating no significant differences of their performance among groups in this test. Then the analysis result of no significant difference was derived, $F=1.909$ and $p=.156$, with the significance level alpha set at .05.

The response to the third research question is: The participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method did not demonstrate any statistically significant differences in their recognition performance as measured by the classification posttest instrument.

Table 3.9
ANCOVA Results of the Group Classification Scores in the First Pilot Study

Source	Type III Sum of Squares	Mean Square	F	Sig.	Observed Power ^b
Group	26.782	13.391	1.909	.156ns*	.384
Pretest	3.944	3.944	.562	.456	.115
Error	490.951	7.014			
Total	16812.000				
Corrected Total	520.054				

Note. *ns = not statistically significant ($p > .05$).

Three duration options were embedded in both the flicker and no-flicker tasks. Specifically, the image display duration options of fast, medium, and slow modes in the flicker group were respectively 2 seconds, 4 seconds, and 6 seconds and the blank screen in between was displayed for 1, 2, and 3 seconds; the duration options of fast, medium, and slow modes in the no-flicker group were that images were displayed for .8 seconds, 2.4 seconds, and 4.1 seconds.

Table 3.10 demonstrates on average how many times the participants clicked a certain display speed option and how frequently they selected a certain rate of display when they studied with different instructional strategies. As for the participants in the flicker group, participants most frequently selected the medium duration option, mean

score=16.04, less frequently participants selected the slow duration option, mean score=13.77, and the least selected option in the flicker group was the fast option, mean score=8.58. Furthermore, the frequency of selecting the fast, medium, and slow options ranged from 0 to 42, 0 to 63, and 0 to 69. As for the participants in the no-flicker group, participants most frequently selected the fast duration option, mean score=11.91, less frequently participants selected the slow duration option, mean score=8.86, and the least selected option in the flicker group was the fast option, mean score=1.14. In addition, the frequency of selecting the fast, medium, and slow options in no-flicker group ranged from 0 to 28, 0 to 22, and 0 to 16. No duration options were embedded in the comparison method of instruction.

The proposal of the duration options for the formal study is as follows: The same speed options as those in the pilot study will be embedded in the flicker and no-flicker method of instruction. The decision of this in-package duration options can be justified with the reasons that were explained in the previous section, including the instructional design principles of user control and interactivity and the educational principle of individual differences. In addition, the results of the pilot study indicated that in practice the participants had their individual preferences in duration options.

Table 3.10

Mean, Standard Deviation, Sample Size, and Other Statistics of the Display Rates Selection Frequency in the First Pilot Study

		N	Mean	Std.		
				Deviation	Minimum	Maximum
Fast	Flicker	26	8.58	13.892	0	42
	No-flicker	22	11.91	8.949	0	28
Medium	Flicker	26	16.04	16.081	0	63
	No-flicker	22	8.86	8.747	0	22
Slow	Flicker	26	13.77	19.251	0	69
	No-flicker	22	1.14	3.655	0	16

Table 3.11 shows that the participants studying with the flicker method version of programs made the highest number of incorrect responses (mean=19.85) and those studying with the no-flicker method made the lowest number of incorrect responses (mean=1.88).

Table 3.11

Mean, Standard Deviation, Sample Size, and Other Statistics of the Number of Incorrect Responses by Treatment Group in the First Pilot Study

Group	N	Mean	Std.		
			Deviation	Minimum	Maximum
Comparison	26	12.31	19.903	0	92
Flicker	26	19.85	21.705	0	92
No-flicker	24	1.88	4.184	0	18
Total	76	11.59	18.667	0	92

Furthermore, an analysis of variance was used to examine whether there was significant differences in the number of incorrect responses across groups (see Table 3.12). An ANOVA test was used to assess whether significant differences could be

identified. It was identified that participants performed differently and made significantly different number of incorrect responses, $F(2, 73) = 6.695, p = .002$.

Table 12

Results of an Analysis of Variance of the Number of Incorrect Responses in the First Pilot Study

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4050.807	2	2025.404	6.695	.002s*
Within Groups	22083.548	73	302.514		

Note. *s=statistically significant ($p < .05$)

Table 3.13 shows that the participants studying with the flicker method version of programs made the highest number of trials (mean=39.85) and those studying with the no-flicker method made the lowest number of trials (mean=21.88).

Table 13

Mean, Standard Deviation, Sample Size, and Other Statistics of the Number of Trials by Treatment Group in the First Pilot Study

	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
Comparison	26	32.31	19.903	3.903	20	112
Flicker	26	39.85	21.705	4.257	20	112
No-flicker	24	21.88	4.184	.854	20	38
Total	76	31.59	18.667	2.141	20	112

Moreover, an analysis of variance was used to examine whether there was significant differences in the number of trials across groups (see Table 3.14). An ANOVA test was used to assess whether significant differences could be identified. It

was identified that participants performed differently and made significantly different number of trials, $F(2, 73) = 6.695, p = .002$.

Table 3.14

Results of an Analysis of Variance of the Number of Trials in the First Pilot Study

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4050.807	2	2025.404	6.695	.002s*
Within Groups	22083.548	73	302.514		

Note. *s=statistically significant ($p < .05$)

To assess the images in the tests, an item analysis was conducted through computing item difficulty and discrimination indices. The item difficulty index P represents the proportion making correct responses to a certain item. The discrimination index refers to how well the item distinguishes between knowledgeable and skillful learners from less knowledgeable and skillful learners. Here Item difficulty P was calculated with the following formula:

$$P = P_H + P_L$$

P_H stands for the proportion of correct responses in the highest third group while P_L stands for the proportion of correct responses in the lowest third group. In order to get these two indices, the total number of participants was divided by three to compute the number of participants in the highest and lowest groups. Then P_H and P_L values

were calculated through divided the number of correct responses to the item by the number of participants in the group.

The item discrimination index D stands for item discrimination, computed with the following formula:

$$D = P_H - P_L$$

Table 3.15 shows the item analysis results for each image in the posttest 1, including the difficulty level index and discrimination index.

Table 3.15
Results of Item Analysis of Recognition Test Images with Item Difficulty and Discrimination Indices

	Item	Item	Item	Item	Item	Item	Item	Item	Item	Item
	1	2	3	4	5	6	7	8	9	10
P	0.96	0.92	0.94	0.78	0.66	0.94	0.78	0.72	0.84	0.34
D	0.08	0.16	0.12	0.44	0.68	0.12	0.28	0.32	0.24	0.60

Generally speaking, the items in the recognition test well distinguished different individuals because the D values of the ten items were all beyond .10 except that of one item. Among them, item 4, 5, 8, and 10 made good discrimination and the others were fairly good. Furthermore, the range of P values indicated that images in the recognition test had different levels of difficulty. For an example, 78% of the participants responded to the fourth item correctly and the discrimination value .44 showed that the image had good quality in distinguishing individual learners.

Table 3.16
Results of Item Analysis of Classification Test Images with Item Difficulty and Discrimination Indices

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
<i>P</i>	0.88	0.94	0.56	0.74	0.88	0.96	0.96	0.38	0.66	0.08
<i>D</i>	0.24	0.12	0.64	0.52	0.24	0.08	0.08	0.60	0.60	0.08

Table 3.16 demonstrates that the items in the classification test well distinguished different individuals because the *D* values of the ten items were all beyond .10. The *D* values of three items below .10 were close to .10. Among them, item 3, 4, 8, and 9 made good discrimination and the others were fairly good. Furthermore, the range of *P* values indicated that images in the classification test had different levels of difficulty. For an example, 74% of the participants responded to the fourth item correctly and the discrimination value .52 showed that the image had good quality in distinguishing individual learners.

Table 3.17 shows the assessment of the images in the study, with the number of trials that the participants took to reach correct responses.

Table 3.17
Mean, Standard Deviation, Sample Size, and Other Descriptive Statistics of Number of Trials for Each Case in the Study Sessions of the Three Groups in the First Pilot Study

	N	Minimum	Maximum	Mean	Std. Deviation
case 1	76	1	9	1.46	1.527
case 2	76	1	10	1.53	1.527
case 3	76	1	11	1.28	1.292
case 4	76	1	13	1.72	2.017
case 5	76	1	25	1.42	2.763
case 6	76	1	4	1.11	.478
case 7	76	1	2	1.04	.196
case 8	76	1	1	1.00	.000
case 9	76	1	2	1.03	.161
case 10	76	1	2	1.03	.161
case 11	76	1	70	4.36	8.929
case 12	76	1	33	2.47	4.438
case 13	76	1	2	1.04	.196
case 14	76	1	6	1.13	.680
case 15	76	1	13	2.00	2.577
case 16	76	1	16	2.17	2.346
case 17	76	1	4	1.09	.437
case 18	76	1	37	2.34	5.005
case 19	76	1	2	1.03	.161
case 20	76	1	12	1.47	1.815
Valid N (listwise)	76				

Examining the means of the number of trials for the twenty study cases, participants used more trials for some of the cases, including case 11, 12, 15, and 16 but less for some of the other cases, such as case 3, 6, 7, and 8. Different levels of difficulty

of these cases seemed to be apparent, which may indicate the quality of the cases. In addition, the cases seemed to have a tendency to be arranged from easy to difficult although two of the cases in the middle seemed to have highest frequency of trials. Considering the number of cases for new learners, this sequence may somewhat encourage and motivate learners to learn continuously.

The Second Pilot Study

The second pilot study intended to scrutinize whether there were any group differences in terms of the factors of duration and the number of incorrect responses and trials in the study session. Duration was computed with the records of finish time of the task minus the starting time of the task. The number of incorrect responses was calculated by counting the number of the missed/incorrect responses that the participants made in study. The number of trials was computed by counting the correct and incorrect number of responses. The procedures and instruments were similar to the ones used in the previous pilot study.

The second pilot study's sample consisted of 14 participants from the same population as the previous pilot study. Generally, the majority of the participants were female undergraduate students between the age of 15 and 25. More specifically, the frequency of the male and female participants was 5 and 9. The components of age groups were that there were 10 of the participants between the age of 15 and 25, two participants in the age group of 26-35 and two of the participants in the age group of 36-45. There were no participants in the other age groups. For ethnicity, 7 of the participants were White, 3 were Black, 1 was Spanish, 1 was Asian, and 2 belonged to the other ethnic groups. For educational programs, 10 of the participants were undergraduate

students while 2 of the participants were from graduate programs and 2 were from the other programs.

Table 3.18 shows that the mean scores of the three tests of comparison group were 3.60, 17.60, and 16.40, those of flicker group were respectively 4.67, 12.67, and 14.67, and those of no-flicker group were respectively 5.00, 11.33, and 10.67.

Table 3.18
Mean, Standard Deviation, Sample Size, and the Other Descriptive Statistics by Test and Treatment Group in the Second Pilot Study

		N	Mean	Std. Deviation	Minimum	Maximum
Pretest	Comparison	5	3.60	1.673	2	6
	Flicker	3	4.67	3.055	2	8
	No-flicker	6	5.00	4.858	0	10
	Total	14	4.43	3.435	0	10
Posttest1	Comparison	5	17.60	1.673	16	20
	Flicker	3	12.67	1.155	12	14
	No-flicker	6	11.33	4.502	4	16
	Total	14	13.86	4.185	4	20
Posttest2	Comparison	5	16.40	.894	16	18
	Flicker	3	14.67	1.155	14	16
	No-flicker	6	10.67	5.007	2	16
	Total	14	13.57	4.164	2	18

Table 3.19 shows that the mean duration of the comparison group, flicker group, and no-flicker group was respectively about 7 minutes and 36 seconds, 8 minutes and 42 seconds, and 4 minutes and 35 seconds, so on average the participants studying with the flicker method spent the most time while those studying with the no-flicker method spent the least time to complete the study materials.

Table 3.19
Duration Results by Treatment Group in the Second Pilot Study

	N	Mean	Std. Deviation	Std. Error
Comparison	5	0:07:36.000	0:02:08.314	0:00:57.384
Flicker	3	0:08:42.333	0:04:21.251	0:02:30.834
No-Flicker	6	0:04:35.167	0:02:39.549	0:01:05.135
Total	14	0:06:32.714	0:03:12.748	0:00:51.514

To further assess the group differences in on-task duration, a one-way analysis of variance was performed. Table 3.20 shows the results of the analysis, indicating that there were no significant differences in on-task duration between groups, $F(2, 11)=2.558$ and $p=.122$.

Table 3.20
ANOVA Results of Group Differences in Duration in the Second Pilot Study

	Sum of Squares	Mean Square	F	Sig.
Between Groups	153333.357	76666.679	2.558	.122ns*
Within Groups	329641.500	29967.409		
Total	482974.857			

Note. *ns=not statistically significant ($p>.05$)

Table 3.21 shows that the participants studying with the flicker method version of programs made the highest number of incorrect responses (mean=36.33), those studying with the no-flicker method made the lowest number of incorrect responses (mean=2.00), and those studying with the comparison method made the number of incorrect responses in between (mean=7.50). As shown in Table 3.22, there were significant group differences in the number of incorrect responses across groups.

Table 3.21

Means, Standard Deviations and the Other Descriptive Statistics of the Number of Incorrect Responses by Treatment Group in the Second Pilot Study

	N	Mean	Std. Deviation	Minimum	Maximum
Comparison	5	2.00	1.000	1	3
Flicker	3	36.33	15.535	19	49
No-flicker	6	7.50	9.915	0	26
Total	14	11.71	16.112	0	49

Table 3.22

ANOVA Results of the Number of Incorrect Responses in the Second Pilot Study

	Sum of Squares	Mean Square	F	Sig.
Between Groups	2396.690	1198.345	13.476	.001s*
Within Groups	978.167	88.924		
Total	3374.857			

Note. *s=statistically significant ($p<.05$).

Table 3.23 shows significant differences between groups in the number of trials, $F=13.603$ and $p=.001$. Therefore, it took participants significantly different number of trials to reach the correct responses, studying with different instructional strategies.

Table 3.23

ANOVA Results of the Number of Trials in the Second Pilot Study

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2416.248	2	1208.124	13.603	.001s*
Within Groups	976.967	11	88.815		
Total	3393.214	13			

Note. *s=statistically significant ($p<.05$).

Conclusions

This chapter provides an overview and details of the research design, sample size, recruitment, instrumentation and instrument validation, procedures, ethical considerations, statistical analysis, and results of pilot studies. A pretest-posttest control group study is proposed to be conducted, with instructional strategy as the independent variable, recognition score, classification score, on-task duration, and number of incorrect responses as dependent variables, and pretest score as a covariate. The chapter also provides the results of the instrument evaluation that subject area experts, IT experts, peers, and participants carried out. The results of two pilot studies are reported. The first pilot study provides the results of usability test of the programs and reliability of the instruments. This pilot study also indicated no overall significant differences in MANCOVA and no effect was detected in the follow-up ANCOVA tests of the effect respectively upon the recognition and classification dependent variables. Duration options were examined in the study and are proposed to stay to be embedded as components of the methods of instruction. The chapter then reports pilot study 2, which identified significant group differences in the number of incorrect responses and trials. No significant differences in on-task duration were identified in this study, but the mean scores of duration across groups were different. Therefore, the formal study examined the effect of the treatments upon recognition and classification test scores with pretest score as the covariate. The factors of duration and the number of incorrect responses and trials were examined in the formal study and the duration options were embedded in the methods of instruction.

Chapter 4 Results of the Study

Introduction

Investigating the effectiveness of three instructional strategies in the three parallel CBI programs upon participants' performance in visual category learning measured with recognition and classification tests, as well as analyzing the group differences in the factors of duration, the number of incorrect responses, and the number of trials in study activities, an experimental study of pretest-posttest control group design was conducted as planned to collect data with the validated instruments and analyze data applying the proposed statistical models and methods with the statistical analysis software package SPSS Window version. This chapter provides the results of the statistical analyses from the formal study, including the information about the participants, the statistical responses to the research questions, and the rest of the analysis results of the exploratory study. First, the chapter presents the sample size and demographic information about the participants in the formal study. Second, descriptive statistics are provided about measures of mean scores, standard deviations, and the other facts of the study. Third, the chapter presents the analysis results of referential statistics, evaluating the null hypotheses with statistical hypothesis testing to address the research questions. Afterwards, the statistical analysis result of the main effect without the covariate is also presented.

Sample Size and Demographic Information of the Participants

Two hundred and forty seven college participants were recruited from the University in the Southeast of the United States with nineteen participants having extreme scores or incomplete sessions, so two hundred and twenty eight participants' records of performance were employed in statistical analysis because all of these records were complete and reasonable without missing data and extreme scores in every test.

Here is a presentation of the demographic information in percentages in each item. Generally speaking, almost all of the participants were undergraduate students, majoring in a great variety of subject areas from the programs of arts, science, and engineering. More specifically, there were 112 male participants and 116 female participants, with the percentages of males and females 49.1 % and 50.9%. The components of age groups were that 179 (87.8%) of the participants were between the age of 15 and 25, with 38 (9.2%), 9(3.1%), 1 (.45%), and 1 (.4%) in the age groups of 26-35, 36-45, 46-55, and 56-65. For ethnicity, 92 (40.4%) of the participants were White, 64(28.1%) were Black, 32(14.0%) were Hispanic, 22 (9.6%) were Asian, and 18 (7.9%) belonged to the other ethnic groups. For the educational programs, 194 (85.1%) participants were undergraduate students while 31 (13.6%) participants were from graduate programs, and 3 (1.3%) participants were from the other programs.

Analysis of the Relationship between the Pretest and the Posttests

In the research design, the pretest score was proposed to function as the covariate in assessing significant effect, suggesting that it was supposed to be related to the dependent measures of recognition and classification. To examine whether this

correlation assumption can be supported, Pearson's Correlation test was used to identify the relationship respectively between the pretest score and the recognition test score and the pretest score and the classification score.

Table 4.1 shows that there were significant relationships between pretest and posttest 1, with the Pearson Correlation value .221 and p value .001 while significance level set at .01 as well as pretest and posttest 2, with the Pearson Correlation value .236 and p value .001 while significance level set at .01.

Table 4.1
Pearson Correlation Values of the Relationships between the Pretest Scores and the Posttest Scores of Recognition and Classification Tests in the Experiment (n=228, All Items)

		Pretest	Recognition Test	Classification Test
Pretest	Pearson Correlation	1	.221**	.236**
	Sig. (2-tailed)		.01	.01

Note. ** Correlation is significant at the .01 level (2-tailed).

Descriptive Statistics

Among the two hundred and twenty eight participants, the similar number of participants was randomly assigned to each of the three groups, respectively 78 participants in the side-by-side comparison method group, 75 participants in the flicker method group, and 75 in the no-flicker method group.

Table 4.2

Mean, Standard Deviation, Sample Size and Other Descriptive Statistics Results by Treatment Group and Dependent Variable in the Study (n=228, All Items)

Instrument	n	Mean	Std. Deviation	Range	Kurtosis	Skewness
Side-by-side comparison group						
Pretest	78	7.28	5.217	18	-1.340	.149
Posttest 1	78	15.59	3.081	18	4.181	-1.418
Posttest 2	78	15.31	2.182	10	.282	-.376
Duration	78	435.83	200.649	1046	2.892	1.535
NIR	78	8.82	12.660	61	4.245	2.009
NT	78	28.82	12.660	61	4.245	2.009
Flicker group						
Pretest	75	6.27	5.223	20	-.778	.505
Posttest 1	75	15.95	2.546	10	-.595	-.110
Posttest 2	75	15.12	2.399	14	2.529	-.821
Duration	75	560.07	269.607	1318	.945	1.096
NIR	75	21.28	20.952	101	2.025	1.395
NT	75	41.28	20.952	101	2.025	1.395
No-flicker group						
Pretest	75	7.31	4.597	18	-.872	.089
Posttest 1	75	15.15	2.944	12	-.471	-.239
Posttest 2	75	14.19	2.654	12	1.420	-.924
Duration	75	235.91	110.030	476	2.339	1.609
NIR	75	2.17	5.134	28	14.198	3.602
NT	75	22.17	5.134	28	14.198	2.238

Note. NIR stands for number of incorrect responses that the participants made during their studies and NT stands for number of trials that include both the number of incorrect and correct responses the participants made in their image study assessment

Table 4.2 shows that the participants in the three groups performed similarly in the pretest, with mean scores of 7.28 and 7.31 in the comparison group and no-flicker group, although the mean score of the flicker group 6.27 had about one point difference from the mean scores of the other two groups. An ANOVA test was conducted to examine whether there was significant difference among the groups' pretest scores, no significant effect was identified, $F=1.113$, $p=.330$. It may reflect the validity of the study that was enabled by random assignment of the participants to groups before the study that

was used to equate groups. However, these scores were not exactly the same, so it was still necessary to use the pretest scores as the covariate to further equate groups and decrease measurement errors. Furthermore, in the recognition test, the participants in the three groups achieved similar mean scores, respectively 15.59 in the comparison group, 15.95 in the flicker group, and 15.15 in the no-flicker group. In the classification test, the participants raised about one point in their mean scores in the side-by-side comparison group and flicker task group over that in the no-flicker task group, respectively 15.31, 15.12, and 14.19. Scrutinizing the mean scores of the duration across the three groups by comparing these scores, the participants were found to use different lengths of time to study, with about 100 seconds difference in the mean duration of the comparison and flicker group and more than 300 seconds difference between the flicker and no-flicker group, with the participants in the flicker task group on average using the longest time to study cases, those in the comparison task group in between, and those in the no-flicker task group the least time to study. For the number of incorrect responses, the table shows that the participants made more than 10 points differences in their mean scores, with the participants in the flicker task group on average made the most number of incorrect responses and those in the no-flicker task group on average made the least number of incorrect responses. The trials that the participants made in study varied from group to group, with more than 10 points difference and the flicker group the highest number of trials and the no-flicker group the least number of trials.

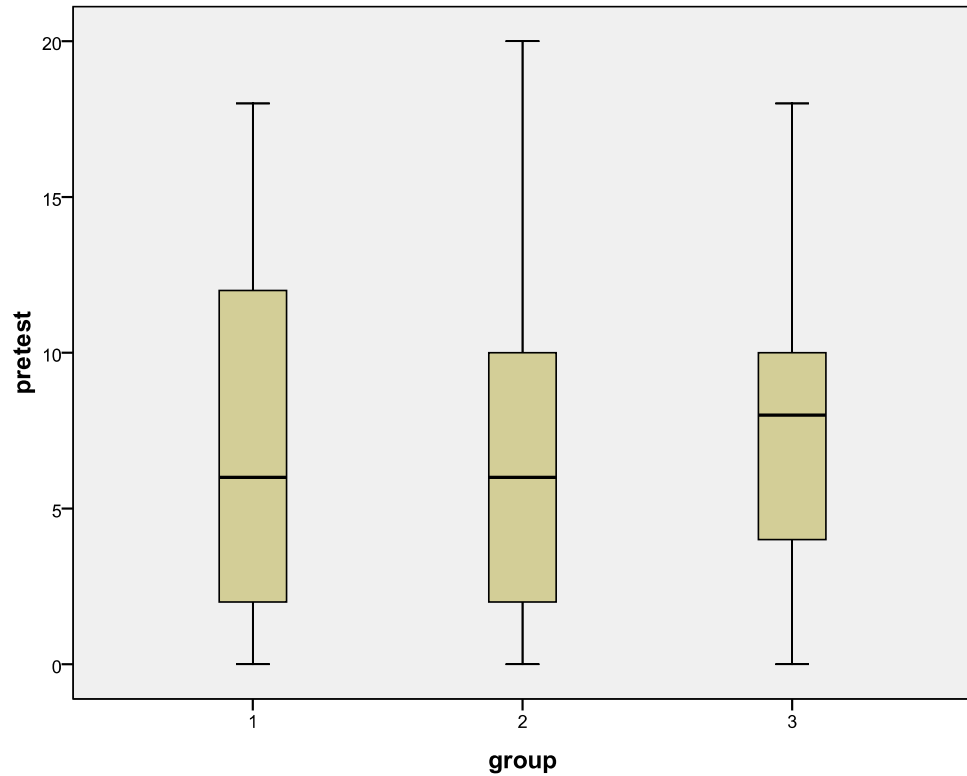


Figure 4.1. Pretest performance by group in the study shown with box plots

Figure 4.1 shows that there were no outliers in the three groups in the pretest. The middle dark line in the boxes shows that the medium scores were not in the center of the boxes, indicating somewhat skewness of distribution of the pretest scores among groups.

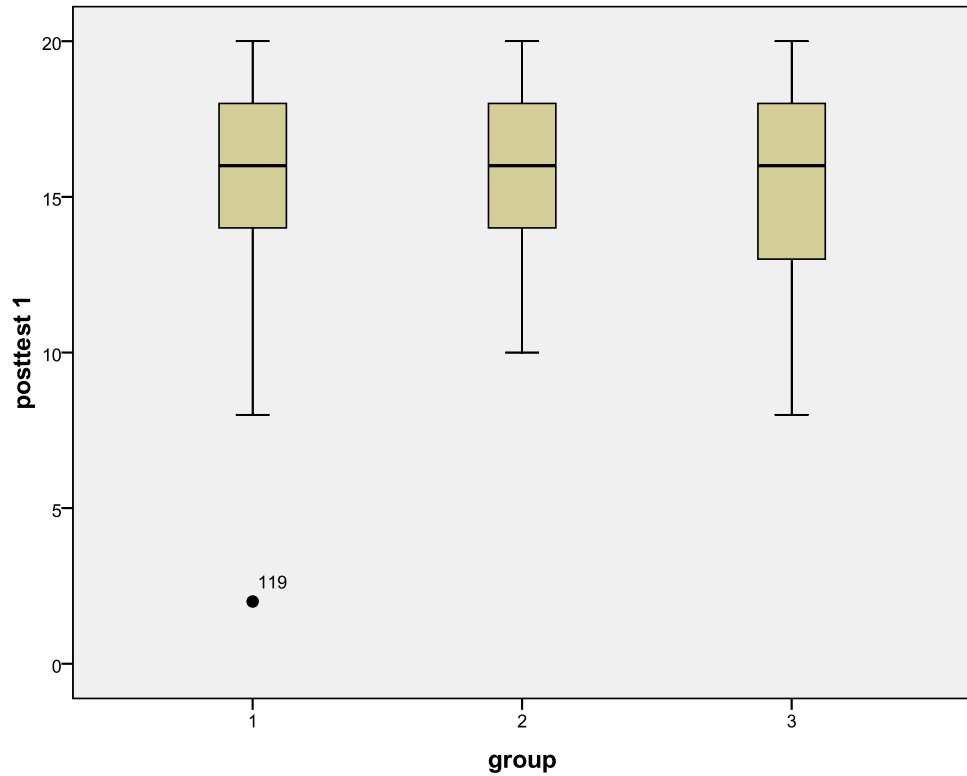


Figure 4.2. Recognition test performance by group in the study shown with box plots

Figure 4.2 demonstrates that there was one outlier in the comparison group in the recognition test. The median lines indicate slight skewness of score distribution in the no-flicker group but normal distribution in the other two groups.

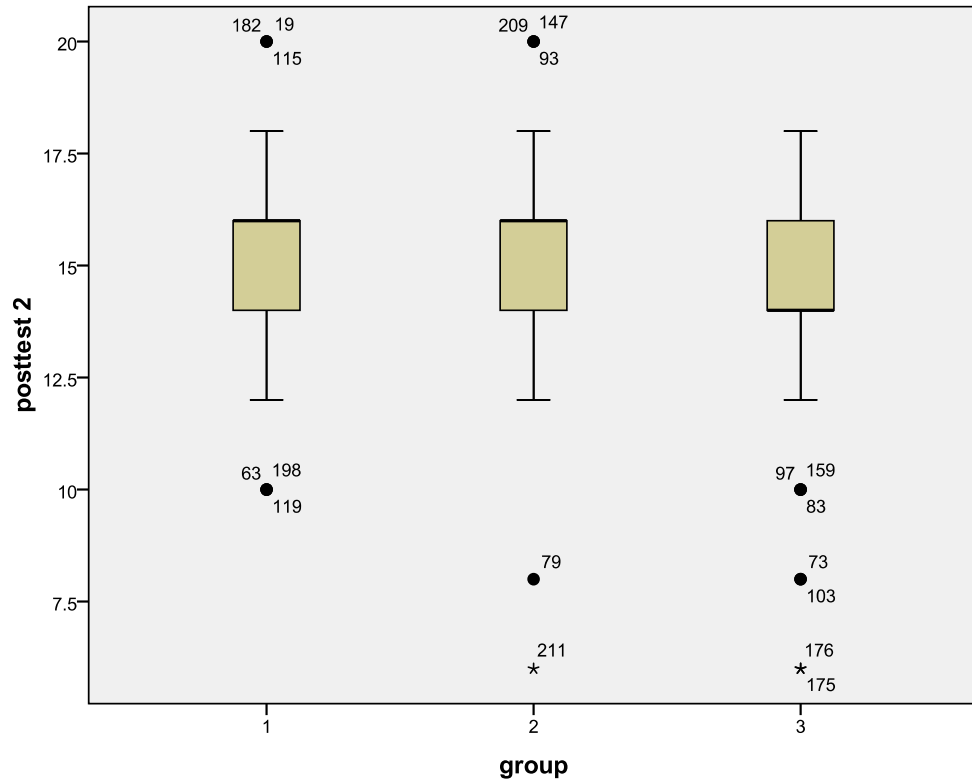


Figure 4.3. Classification test performance by group in the study shown with box plots

Figure 4.3 demonstrates that eighteen outliers in the three groups deviated from the groups. Three extreme outliers #211, 176, and 175, more than 10 points away from the mean scores, were eliminated from the data and the rest of the tests were conducted without these three records. Therefore, the number of participants in the comparison group, flicker group, and no-flicker group was respectively 78, 74, and 73.

Assessing Group Differences in the Outcome Measures

The independent variable in this study was instructional strategy, the dependent variables were recognition scores, classification scores, duration, the number of incorrect responses, and the number of trials in study, and pretest score was taken as a covariate. The research questions were posited to investigate whether there were significant effectiveness differences in a global sense, as well as individually in the recognition and classification tests. Group differences were also examined in a pair-wise fashion to identify the exact location of differences if significant differences were identified. In addition, the analysis of data would provide clear information to indicate whether significant group differences occurred in the factors of duration, the number of incorrect responses, and the number of trials. In the following sections, the proposed statistical hypotheses testing processes would be used to analyze the data with the General Linear Model and Analysis of Variance and research questions would function as the bases of the structure and content of this section.

Analysis of the Relationship between the Recognition Test and the Classification Test

Before the overall significance test with MANCOVA, a correlation test was employed to assess the relationship between the dependent variables recognition test score and classification test score that was measured with the two posttests. More particularly, the Pearson Correlation test was used to examine whether a certain level of correlation existed between these two criterion measures.

Table 4.3 demonstrates that the two posttests are correlated, Pearson Correlation = .672 and $p = .000$. Therefore, the condition of correlation between the two dependent variables in the proposed MANCOVA test was satisfied.

Table 4.3

Pearson Correlation Values Indicating the Relationship between the Posttests in the Experiment (n=228, All Items)

		posttest 1	posttest 2
posttest 1	Pearson Correlation	1	.364**
	Sig. (2-tailed)		.000

Note. ** Correlation is significant at the 0.01 level (2-tailed).

Effectiveness Testing

Question 1: Did the participants who studied visual patterns in CBI with the flicker method of instruction, no-flicker method, and comparison method demonstrate any statistically significant differences in their overall performance as measured by recognition and classification posttest instruments?

In order to address this research question, MANCOVA was employed and hence a null hypothesis was stated in the hypothesis testing in this statistical analysis procedure: There is no overall difference among the participants who studied radiographic images in CBI with the flicker method of instruction, no-flicker method, and comparison method in their performance as measured by recognition and classification posttest instruments.

MANCOVA was employed to assess if the three instructional strategies had an overall significant difference in their effects upon recognition and classification performance. Before further evaluating the hypothesis of this research question, assumptions of MANCOVA were assessed, primarily including the assumptions of normality (evidence provided by the previous boxplots and complementary analyses), independence of observation (evidence provided by the fact that the participants completed studies and tests independently) and homogeneity of variance.

Box's M Test was used to evaluate the assumption of homogeneity of covariance. That is, the test was employed to see whether the population regression slopes were the same across groups. According to the Box's M test, there were no significant differences of the covariance regression, with the Box's M statistics 5.499, $F(6, 1202929) = .905$, and $p = .490$. The assumption of homogeneity of the covariate pretest among the groups satisfied the requirement for the application of MANCOVA.

Table 4.4 shows that the participants studied the images and image features had an overall significant difference in their performance. Among the four MANCOVA tests, the most strict one Pillai's Trace F value was found significant at the .05 alpha level, with the prior knowledge controlled, $F(4, 442) = 2.762$, partial eta squared = .024, and $p = .027$. The Wilks' Lambda F value showed significance at the .05 alpha level, $F(4, 440) = 2.770$, partial eta squared = .025, and $p = .027$. The Hotelling's Trace demonstrated significant differences at the .05 alpha level, $F(4, 438) = 2.777$, partial eta squared = .025, and $p = .027$.

On the bases of the above findings, the null hypothesis of no significant difference among the participants who studied radiographic images in computer-based instruction with the flicker method of instruction, no-flicker method, and comparison method in their performance as measured by the recognition and classification posttest instruments was rejected with all the three tests. Therefore, the research question one about the overall significant difference in the participants' global performance was addressed with a positive response.

Table 4.4
Results of Multivariate Analysis of Covariance of the Overall Group Differences in the Study (n=228, All Items)

Effect	F	Sig.	Partial Eta Squared	Observed Power ^b
Pillai's Trace	2.762	.027	.024s*	.759
Wilks' Lambda	2.770 ^a	.027	.025s*	.760
Hotelling's Trace	2.777	.027	.025s*	.762

Note. *s=statistically significant ($p < .05$).

Question 2: Did the participants who studied visual patterns in CBI with the flicker method of instruction, no flicker method, and comparison method demonstrate any statistically significant differences in their recognition performance as measured by the recognition posttest instrument?

This question was examined because an overall significant difference was identified in the previous test with MANCOVA. In order to address this research question, ANCOVA was employed and hence a null hypothesis was stated in the hypothesis testing in this statistical analysis procedure: There was no significant difference among the participants who studied radiographic images in CBI with the

flicker method of instruction, no-flicker method, and comparison method in their recognition performance as measured by the recognition instrument.

To assess the significance of the three instructional strategies on the two dependent variables recognition and classification test scores, the Levene's Test was utilized to examine the cross group equivalence in the error variance of the dependent variables, recognition score and classification score.

The error variances of the two test scores were similar, respectively $F(2, 222) = .340$ and $p = .712$ in the recognition test and $F(2, 222) = .507$ and $p = .603$ for the classification test. No significance was identified in either group, so the equality of error variance was satisfied.

Table 4.5
ANCOVA Results of the Group Recognition Scores in the Experiment (n=228)

Source	Type III Sum of Squares	Mean Square	F	Sig.	Partial Eta Squared	Observed Power ^b
Pretest	98.131	98.131	12.751	.000	.055	.945
Group	29.078	14.539	1.889	.154ns*	.017	.390
Error	1700.796	7.696				
Total	56700.000					
Corrected Total	1819.129					

Note. *ns=not statistically significant ($p > .05$).

Table 4.5 demonstrated the results of the univariate analysis of covariance for the recognition test, indicating no significant differences of their performance among groups in this test, $F=1.889$, partial eta squared =.017, and $p=.154$.

On the bases of the above findings, the null hypothesis of no significant difference among the participants who studied radiographic images in CBI with the flicker method of instruction, no-flicker method, and comparison method in their recognition performance as measured by the recognition instrument failed to be rejected. Therefore, the research question two about the significantly different effects of the three instructional strategies upon the participants' recognition performance was addressed with a negative response.

Question 3: Did the participants who studied visual patterns in computer-based instruction with the flicker method of instruction, no flicker method, and comparison method demonstrate any statistically significant differences in their classification performance as measured by the classification instrument?

This question was examined because an overall significant difference was identified in the previous test with MANCOVA. In order to address this research question, ANCOVA was employed and hence a null hypothesis was stated in the hypothesis testing in this statistical analysis procedure: There was no significant difference among the participants who studied radiographic images in CBI with the flicker method of instruction, no-flicker method, and comparison method in their classification performance as measured by the classification instrument.

Table 4.6

ANCOVA Results of the Group Classification Scores in the Experiment (n=228)

Source	Type III Sum of Squares	Mean Square	F	Sig.	Partial Eta Squared	Observed Power ^b
Pretest	80.344	80.344	17.553	.000	.074	.986
Group	42.902	21.451	4.686	.010s*	.041	.782
Error	1011.564	4.577				
Total	51724.000					
Corrected Total	1128.996					

Note. *s=statistically significant ($p<.05$).

Table 4.6 demonstrated the results of the univariate analysis of covariance for the classification test, indicating significant differences of their performance among groups in this test, $F=4.686$, partial eta squared =.041, and $p=.010$.

On the bases of the above findings, the null hypothesis of no significance differences among the participants who studied images in CBI with the flicker method of instruction, no-flicker method, and comparison method in their classification performance as measured by the classification instrument was rejected. Therefore, the research question three about the significantly different effects of the three instructional strategies upon the participants' classification performance was addressed with assurance.

Question 4: Were there any statistically significant differences in their performance as measured by posttest instruments between students who studied visual patterns in computer-based instruction with the flicker method of instruction and the no-flicker method of instruction, those studying with the flicker method and the comparison method, and/or those studying with the no-flicker method and the comparison method?

This question was examined because a significant difference was identified among the participants in their performance in the classification test in the previous tests with ANCOVA. In order to address this research question, the post-hoc procedures of simple group comparison of adjusted means were employed and hence a null hypothesis was stated in the hypothesis testing in this statistical analysis procedure: There was no significant difference between students who studied radiographic images in CBI with the flicker method of instruction and the no-flicker method of instruction, those studying with the flicker method and the comparison method, and/or those studying with the no-flicker method and the comparison method.

Table 4.7 gave an idea of the exact location of the differences between the groups' performance and significant differences of treatments measured with posttest 1 and posttest 2. Significant differences were identified between those in the comparison group and no-flicker group in the posttest 2, mean difference=.904, $p=.010$ (<adjusted alpha .0167), as well as the flicker and no-flicker groups, mean difference=.963, $p=.007$ (<adjusted alpha .0167), in the classification test. However, there was no significant difference identified between the groups in the recognition test.

Table 4.7

Results of Group Contrast of Adjusted Means of Posttest Scores with the Pretest Scores as a Covariate (n=228)

Dependent Variable	Group	Group	Difference	Sig.
Recognition	Comparison	Flicker	-.547	.227ns*
	Comparison	No-flicker	.337	.456ns*
	Flicker	No-flicker	.884	.056ns*
Classification	Comparison	Flicker	.059	.865ns*
	Comparison	No-flicker	.904	.010s*
	Flicker	No-flicker	.963	.007s*

Note. *s=statistically significant ($p < .05$); *ns= not statistically significant ($p > .05$).

On the bases of the above findings, the null hypothesis of no significant differences among the participants who studied images in CBI with the flicker method of instruction and the no-flicker method of instruction, those studying with the flicker method and the comparison method, and/or those studying with the no-flicker method and the comparison method was rejected. Therefore, the research question four about between which groups' performance the significant performance difference could be found was responded: the participants studying visual patterns with the comparison method performed significantly better in the classification test than the participants studying visual features with the no-flicker method, mean difference=.904, $p=.010$; the participants studying visual patterns with the flicker method performed significantly better in the classification test than the participants studying visual features with the no-flicker method, mean difference=.963, $p=.007$

Question 5: Was there any statistically significant difference in their on-task duration among the participants who studied visual patterns in CBI with the flicker method of instruction, no-flicker method, and comparison method?

A post-hoc question of this question was added: If any significant effects were identified in duration, between which groups were the significant differences identified?

This question was examined although a non-significant difference was identified among participants in their study time in the second pilot study. In order to address this research question, the ANOVA procedures were employed and hence a null hypothesis was stated in the hypothesis testing in this statistical analysis procedure: There was no significant difference in their duration among the three groups of participants who studied images in CBI respectively with the side-by-side comparison method, the flicker method, and the no-flicker method.

Table 4.8
ANOVA Results of Group Differences in Duration (n=228)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3884505.703	2	1942252.851	46.080	.000s*
Within Groups	9357141.657	222	42149.287		
Total	1.324E7	224			

Note. *s=statistically significant

On the bases of the above findings, as shown in Table 4.8, the null hypothesis of no significant difference in on-task duration among the participants who studied images in CBI with the flicker method of instruction, no-flicker method, and comparison

method was rejected, $F=46.080$, $p=.000$. Therefore, the research question five about the significantly different group differences in duration was addressed with an affirmative response.

Table 4.9
Results of Multiple Comparisons of Group Duration with Tukey HSD
($n=228$)

(I) Group	(J) Group	Mean Difference (I-J)	Sig.
Comparison	Flicker	-124.57*	.001s*
Flicker	No-flicker	322.42*	.000s*
No-flicker	Comparison	-197.85*	.000s*

Note. *s=statistically significant

Table 4.9 shows significant differences of duration between groups. More specifically, each group spent a significantly different length of time from each other, with significant differences between the comparison group and the flicker group, mean differences = -124.57, $p=.001$; the comparison group and the no-flicker group, mean differences = 197.85, $p=.000$; and the flicker group and the no-flicker group mean differences = 322.42, $p=.000$.

Question 6: Was there any statistically significant difference in the number of incorrect responses and the number of trials they made in their study among the participants who studied visual patterns in CBI with the flicker method of instruction, no-flicker method, and comparison method?

The other post-hoc question of this question was added: If any significant group differences were identified in the number of incorrect responses and trials, between which groups were the significant differences?

This question was examined because a significant difference was identified among the participants in the number of incorrect responses and the number of trials in their study in the pilot studies. In order to address this research question, the ANOVA procedures were employed and hence a null hypothesis was stated in the hypothesis testing in this statistical analysis procedure: There were no significant differences in their number of incorrect responses and number of trials among the three groups of participants who studied images in CBI respectively with the side-by-side comparison method, the flicker method, and the no-flicker method.

Table 4.10

The ANOVA Results of Group Differences in the Number of Incorrect Responses (n=228)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	13195.297	2	6597.648	31.888	.000s*
Within Groups	45931.663	222	206.899		
Total	59126.960	224			

Note. *s=statistically significant ($p < .05$).

Table 4.11

The ANOVA Results of Group Differences in the Number of Trials (n=228)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	13195.297	2	6597.648	31.888	.000s*
Within Groups	45931.663	222	206.899		
Total	59126.960	224			

Note. *s=statistically significant ($p < .05$).

On the bases of the above findings (Shown in Table 4.10 and Table 4.11), the null hypothesis of no significant difference in the number of incorrect responses they made among the participants who studied images in CBI with the flicker method of instruction, no-flicker method, and comparison method was rejected, $F=31.888$, $p=.000$ (Table 4.10). The null hypothesis of no significant difference in the number of trials they made among the participants who studied images in CBI with the flicker method of instruction, no-flicker method, and comparison method was rejected, $F=31.888$, $p=.000$ (Table 4.11). Therefore, the research question six about the significantly different group differences in the number of incorrect responses and the number of trials was addressed with an affirmative response.

Table 4.12
Results of Multiple Comparisons of the Number of Incorrect Responses with Tukey HSD (n=228)

Group	Group	Mean Difference	Std. Error	Sig.
Comparison	Flicker	-12.07*	2.334	.000s*
Flicker	No-flicker	18.66*	2.373	.000s*
No-flicker	Comparison	-6.59*	2.342	.015s*

Note. *s=statistically significant ($p<.05$)

Table 4.12 presents the results of the group differences in the number of incorrect responses. Significant group differences were identified between the comparison group and the flicker group, mean difference=-12.07, $p=.000$; the comparison group and the no-flicker group, mean difference=6.59, $p=.015$; and the flicker group and the no-flicker group, mean difference=18.66, $p=.000$.

Table 4.13

Results of Multiple Comparisons of Group Differences in the Number of Trials with Tukey HSD

(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.
Comparison	Flicker	-12.07*	2.334	.000s*
Flicker	No-flicker	18.66*	2.373	.000s*
No-flicker	Comparison	-6.59*	2.342	.015s*

*s=statistically significant ($p < .05$)

Table 4.13 shows significant group differences in the number of trials in study between the comparison group and the flicker group, mean difference=-12.07, $p=.000$; between the comparison group and the no-flicker group, mean difference=6.59, $p=.015$; and the flicker group and the no-flicker group, mean difference=18.66, $p=.000$.

In addition to the above results, data of the selection frequency of display rates in the flicker and no-flicker group was also recorded and calculated to observe the differences in choosing each option. Table 4.14 provides the mean scores and standard deviation of the options of different pace of animation. The fast, medium, and slow columns respectively represent the number of selections/clicks of the fast rate display button, medium rate display button, and slow rate display button. In the flicker group, the fast pace is the least selected, the slow pace the most frequently selected, and the medium pace in between, with the mean times of selecting the fast, medium, and slow pace respectively 8.66, 13.04, and 19.15. In the no-flicker group, the mean times of selecting the fast, medium, and slow pace are respectively 12.08, 7.70, and 2.45.

Table 4.14

Results of Selection Frequency of Display Rates in the Flicker and No-Flicker Groups

Group		Fast	Medium	Slow
Flicker	Mean	8.66	13.04	19.15
	N	74	74	74
	Std. Deviation	16.290	12.319	22.575
No-flicker	Mean	12.08	7.70	2.45
	N	73	73	73
	Std. Deviation	9.049	9.635	7.307
Total	Mean	10.36	10.39	10.86
	N	147	147	147
	Std. Deviation	13.267	11.351	18.743

More Covariate Analyses: Necessary or Not

With the results of the significant differences in the duration, the number of incorrect responses, and the number of trials, it might be necessary to run another turn of the tests with these factors as covariates, joined with the pretest scores. Pearson correlation was examined, showing relationships among the factors, duration, number of incorrect responses, and number of trials, and the two posttests (as shown in Table 4.15). From the table, it is clear that there was no correlation between these variables and the posttest 2, so another round of statistical analyses with these variables as covariates was omitted.

Table 4.15

Correlation Coefficients of Duration, Number of Incorrect Responses, and Number of Trials with the Posttest Scores

		Recognition	Classification
Duration	Pearson Correlation	-.184**	-.042
	Sig. (2-tailed)	.006	.534
Number of Incorrect Responses	Pearson Correlation	-.147*	-.088
	Sig. (2-tailed)	.028	.190
Number of Trials	Pearson Correlation	-.147*	-.088
	Sig. (2-tailed)	.028	.190

Note. **Correlation is significant at the 0.01 level (2-tailed)

*Correlation is significant at the .05 level (2-tailed)

In order to indicate perspectives and give audience a more complete view of the study, another post-hoc question was raised in the analyses process: without the pretest score as the covariate, did the participants who studied visual patterns in CBI with the flicker method of instruction, no-flicker method, and comparison method demonstrate any statistically significant differences in their overall performance as measured by recognition and classification posttest instruments? Another statistical analysis method multiple analysis of variance (MANOVA) was used to respond to this post-hoc question. Thus, a null hypothesis was stated in the hypothesis testing in this statistical analysis procedure: Without the pretest score as the covariate, there were no overall significant differences in the participants' performance measured by the recognition and classification test.

Before conducting the MANOVA test, the Box's M test was used to examine the homogeneity hypothesis. No significance was identified in the Box's test indicating that

the covariance of the dependent variables were equal and the MANOVA study could be applied, $F(6, 1202929) = .905, p = .490$.

Table 4.16 indicated that there were no significant differences in the instructional strategies upon pattern recognition, Pillai's Trace F value was found insignificant at the .05 alpha level, $F(4, 444) = 2.152$, partial eta squared = .019, and $p = .074$. The Wilks' Lambda F value did not show significance at the .05 alpha level, $F(4, 442) = 2.154$, partial eta squared = .019, and $p = .073$. The Hotelling's Trace did not demonstrate significant differences at the .05 alpha level, $F(4, 440) = 2.150$, partial eta squared = .019, and $p = .073$. Therefore, the null hypothesis of no overall significant differences in the groups' performance without the pretest test score as the covariate failed to be rejected. The response to the seventh questions can be that, without the pretest score as the covariate, the participants who studied visual patterns in CBI with the flicker method of instruction, no-flicker method, and comparison method did not demonstrate any statistically significant differences in their overall performance as measured by the recognition and classification posttest instruments.

Table 4.16
Results of the Analysis of the Instructional Strategy Effects upon Learning without the Pretest Scores as the Covariate with MANOVA (n=228)

Effect Tests	F	Sig.	Partial Eta Squared	Observed Power ^b
Pillai's Trace	2.152	.074	.019	.636
Wilks' Lambda	2.154 ^a	.073	.019	.636
Hotelling's Trace	2.156	.073	.019	.637

Conclusions

About 247 participants were recruited and 228 participants' records were employed in the experimental study. They were undergraduate students who knew little about the visual category and patterns instructed in the study and assessed in the test materials. They were randomly assigned to the three experimental and control groups and were administered the materials with the proposed research procedures. Here in this chapter the results presented in order include the sample size and demographic information of the participants, from whom raw data were collected, the evaluation of statistically significant differences through processing collected raw data from the participants with the proposed statistical approaches, and the suggested responses to the research questions in this study, among which answering research questions served as the focus and structure guide of this chapter. Furthermore, this chapter provides the responses to the research questions through assessing the relevant null hypotheses at the significant alpha level of .05 with statistical procedures. In the first round of statistical analysis, with the first MANCOVA test, an overall significant group difference was assessed and identified in the participants' global performance because the null hypothesis of no difference was rejected. Furthermore, the chapter provides the assessment results of no significant group differences in the recognition test but significant group differences in the classification test with ANCOVAs. The other result with the test of simple contrast of adjusted means is the significant group differences between the comparison group and no-flicker group, as well as the flicker group and the no-flicker group in the classification test. Furthermore, evaluation of the statistically significant group differences in the on-task duration, the number of incorrect responses

and number of trials is provided. In addition, without the covariates, another statistical analysis MANOVA result is also provided.

Chapter 5 Discussion

Introduction

This chapter provides an integrative discussion of the experimental study findings, including presentation and interpretation of the findings on the bases of literature, implications for research and practice, limitations of the study, and recommendations for future research. First, an overview of the research findings is presented, including a summary of the findings and an explanation of the responses to research questions. Second, the chapter evaluates the flicker treatment and the other two treatments measured with the outcome variables. Third, the chapter provides the implications of the study for research and practice of instructional technology. Fourth, limitations of the study are discussed, including the cautions in generalization of the study results. Fifth, the chapter provides recommendations for future research.

Findings of the Experimental Study

This experimental study investigated the significantly different effects of three instructional strategies upon pattern recognition in CBI. The independent variable in this experimental study was instructional strategy, the dependent variables were recognition scores, classification scores, duration, the number of incorrect responses, and the number of trials, and the pretest score was considered as a covariate.

This research intended to respond to these six research questions: The first question is whether the participants demonstrated an overall significant difference in their

pattern recognition performance. The second and the third questions are to locate where the significant differences were, if any, in the recognition and/or classification test. As a follow-up of question number three, question four asks between which groups the significant difference(s) was/were. The fifth and sixth two questions are whether significant differences were identified in duration, number of incorrect responses, and number of trials in the study. Moreover, the following three post-hoc research questions were examined: First, between which groups the significant difference(s) was/were identified in duration? Second, between which groups the significant difference(s) was/were identified in the number of incorrect responses and number of trials? Third, did the participants demonstrate any significant differences in their overall performance measured by the recognition and classification posttests, without the pretest score as the covariate.

With the collected data from 228 participants, the effectiveness of three CBI methods of visual patterns, the flicker treatment, no-flicker treatment, and comparison treatment, was examined by analyzing the participants' performance in the recognition and classification tests with two rounds of statistical analyses, respectively with and without the pretest score as covariate. Furthermore, analyses were conducted with the outcome measures of duration, number of incorrect responses, and number of trials. The primary findings of the study are listed as follows:

1. With the pretest score as covariate, the participants who studied visual patterns in CBI with the flicker method of instruction, no-flicker method, and comparison method demonstrated statistically significant differences in their overall performance as measured by recognition and classification posttest instruments.

2. With the pretest score as covariate, the participants who studied visual patterns in CBI with the flicker method of instruction, no flicker method, and comparison method did not demonstrate any statistically significant differences in their recognition performance as measured by the recognition posttest instrument (knowledge acquisition).

3. With the pretest score as covariate, the participants who studied visual patterns in CBI with the flicker method of instruction, no flicker method, and comparison method demonstrated statistically significant differences in their classification performance as measured by the classification posttest instrument (transfer of learning).

4. With the pretest score as covariate, statistically significant differences were detected between the flicker group and the no-flicker group, as well as between the comparison group and the no-flicker group, in their performance in the classification test (transfer of learning). More specifically, the participants in the flicker group outperformed those in the no-flicker group while the participants in the comparison group outperformed those in the no-flicker group. No significant differences were identified between the flicker group and the comparison group in the classification test.

5. There were differences of statistical significance in their study duration among the three groups studying with the flicker, no-flicker, and comparison methods. The flicker group was found to spend significantly longer time in the study session than the comparison and no-flicker groups. Furthermore, the comparison group was found to spend significantly longer time than the no-flicker group.

6. There were differences of statistical significance among the three groups in their number of incorrect responses and number of trials in the study session. The flicker

group made significantly more errors and trials than the comparison and no-flicker groups while the comparison group made significantly more errors and trials than the no-flicker group. Therefore, the no-flicker group made significantly the least errors and trials among the three groups.

7. Without the pretest score as the covariate, no overall significant group differences were identified measured with the recognition and classification test.

The Participants in the Three Groups Learned

The data indicated that all the participants in all of the three groups learned significantly. The learning gains in the three groups were obvious, comparing their performance before the study sessions with that after study sessions. Furthermore, the three groups achieved higher accuracy scores and lower false alarm rates than the documented performance of trained residents and radiologists. They reached such accuracy with cost effectiveness.

Specifically, from the instructional design's perspective, the learning objectives of recognizing both studied and unstudied visual patterns were achieved with the three instructional strategies through all of the three programs although individual methods and programs differed in their effectiveness. The primary finding from the data analyses shows that the performance was increased in all of the three groups in both recognition and classification tests, indicating that the treatments and programs increased the performance of novice learners and cultivated their knowledge of visual patterns from none to a certain level of recognition and categorization. The three groups' mean scores in posttests show that the accuracy rate in both tests reached approximately 80% of the total accuracy rate. Moreover, compared the baseline scores in the pretest with the

posttest scores, the scores were almost or over doubled among all of the three groups.

The mean scores of the performance therefore indicate the effectiveness and usefulness of the programs. About this growth among the participants, the preliminary results from the pilot study were consistent with those of the formal study.

Moreover, through comparing the accuracy rates, or sensitivity, of pattern recognition in this study and those in literature, it can be concluded that all of three groups in this study learned effectively. Sensitivity means an observer's ability to discriminate the targeted stimulus from noise and recognize it. Computed accuracy rates show that the three groups performed better than the residents and radiologists documented in literature (Newstead, 2003) if merely comparing the absolute rates of accuracy of recognition without considering case varieties and familiarity. The recognition accuracy rates for the comparison, flicker and no-flicker groups were respectively $15.59/20=.7795$, $15.95/20=.7975$, and $15.15/20=.7575$. The classification accuracy rates for the comparison, flicker and no-flicker groups were respectively $15.31/20=.7655$, $15.12/20=.756$, and $14.19/20=.7095$. Compared with year-one to year-four residents' sensitivity, respectively 33%, 48%, 38%, and 54% (Newstead, 2003), and the residents', radiologists' and experts' average sensitivity, respectively 46%, 72%, and 82%, the participants in this study gained much higher sensitivity through about half an hour's image study with computer-based instruction. Particularly, the comparison group gained 77.95% and 76.55% accuracy respectively in recognition and classification, the flicker group gained 79.75% and 75.60% accuracy, and the no-flicker group gained 75.75% and 70.95% accuracy. Therefore, the training results were significant in comparison with the results of residents' training performance because all of the

performance outcomes of the three groups surpassed all of the residents' performance, ranging from 15% to 45% higher than the residents' sensitivity. Even when comparing the radiologists' sensitivity of 72% with the three groups' accuracy rates in both recognition and classification tests, almost all of the groups scored three to seven percent higher than this rate except that the no-flicker group gained approximately two percent lower.

In addition to the outcomes of recognition and classification performance, the other indices derived from the study can also support the argument of the significance of the three groups' learning outcomes. The times spent in study support that the groups were quick in learning with the methods compared with the years of time that the residents spent on image studies to reach the stated sensitivity performance. The comparison group on average spent 435.83 seconds, that is, 7.3 minutes on the study session, to reach about 80% accuracy rate. The flicker group spent significantly more time than the other two groups, but compared with the years of time the residents spent to reach much lower percentage of accuracy, they still spent very little time, with the average time in study as 560.07 seconds, that is, 9.33 minutes. The no-flicker group spent the least time of the three groups, averaging 235.91 seconds, that is, 3.93 minutes of time to complete the study session. Even though it is unclear exactly how much time the residents tend to spend reading mammograms, they surely study much longer time within their years of residency than the participants in the three groups. Hence, the efficiency of learning among the three groups may be significantly and practically noticeable.

The other indices, such as the number of incorrect responses and trials, also demonstrate the worth of the methods studied here. With the number of incorrect

responses divided by the number of trials, percentage of incorrect responses can be derived, resulting in $8.82/28.82=0.30$, $21.28/41.28=0.51$, and $2.17/22.17=0.09$. That is to say, among all of the trials in detection, the participants in the comparison, flicker, and no-flicker groups made false alarms respectively 30%, 51%, and 9% of the trials. In another word, they on average made accurate trials respectively 70%, 49%, and 91% of all the trials they made in study sessions. According to Newstead (2003), an average specificity was 72%, 68%, and 53% for residents, radiologists and experts. That is, their false alarm rates were 28%, 32%, and 47%. Therefore, the false alarm rate 9% the no-flicker group made was significantly lower than those made by the residents, radiologists, and experts. The false alarm rate that the comparison group made was somewhat higher than that made by the residents and radiologists but lower than that made by experts. The false alarm rate that the flicker group made was higher than those by the residents, radiologists, and experts.

The other data that may be supportive include the frequency of selections the participants made in choosing the learning paces of their image studies. The frequency data shows that the flicker group significantly more frequently selected the slow pace of learning while the no-flicker group significantly more frequently selected the fast pace of learning. This outcome implies that the participants had a tendency to study at a certain pace. That is, they adjusted their time according to what they thought the best for their learning. This can be regarded as a different situation indicated in literature (Bassett, 2003): more than 60% of the residents do not want to spend one fourth (1/4) of their residency time in studying mammograms, in which case motivation for learning is limited and learning becomes compromised.

The programs' effectiveness was intended and expected in the program design and validation before the experiment and could be further explained with the shared interactive system of the programs. First, as presented in chapter 3 of this document, the three programs were designed and validated with rounds of expert review and user tests. Although the major purpose of the program validation in this study was to ensure that the experimental study could have valid experimental materials, the program validation simultaneously ensured the criteria-based characteristics of the programs. Evaluation instruments used in the program validation was listed in Appendix I, indicating that a variety of instructional design principles and rules of thumb were followed in design and development and the programs were judged based on these criteria in validation. The usability test data reflected the soundness of the design and ease of use from the target audience's perspectives. This interpretation is congruent with the literature about how instructional design, including evaluation, is essential for effectively integrating technology affordance into education (Alessi & Trollip, 2001; Jonassen, 2008, 2004; Reigeluth, 1999). Second, the programs have similar interactive instructional systems, enhancing learning and instruction with the instructional strategies as the essence of the interactive systems. This interpretation is consistent with the literature about how different levels and types of interactivity can engage and enhance learning (Chou, 2003; Moore, 1989). The interactive instructional system shared among the three programs consisted of cases, puzzles of patterns, assessment, feedback, branched interactions, and user control. More details about the shared interactive system are elaborated as follows:

First, cases and puzzles of patterns in the study sections of the programs functioned as problems for the participants to immerse into authentic diagnostic and

detection environment. They could engage the participants in the real-world clinical problems differently from the other methods, such as providing them with a complex textbook to read through, an atlas to grasp the definitions of patterns, or simplified patterns with sketched images to memorize. All of these other methods have been in use in visual rich medical education, specifically in radiology education, but they seem to deviate learners from authentic situations and concrete experience of examining patterns in cases. The three programs in this study adopted a more problem-based approach with case study, asking the participants to figure out the patterns themselves by looking for and identifying the differences or changes between images and assigning meanings to the patterns they noticed. Although the problems were not posed by the participants, as suggested in traditional problem-based instruction, the cases and puzzles in the instructional framework of the three programs encouraged the participants to pursue the responses to the case problems, which more or less engaged them in closely examining the patterns of images. Moreover, teaching in problem-solving contexts will influence the transfer of knowledge to new situations (Bransford, Brown, & Cocking, 1999; Jonassen, 2004; Mayer, 2002). With problems, the participants not only learned a certain case or image but also the concept represented by the case, relating what had been learned to what was learned, which could facilitate learning of conditional knowledge (Bransford, Brown, & Cocking, 1999). The declarative knowledge of the concept and the conditional knowledge about when to apply the concept could enable the participants store and apply knowledge in new situations.

Furthermore, combined with the instructional methods, the puzzles of patterns and case study in the shared instructional system of the programs more or less enabled

generative learning. For example, the case study and puzzles continuously engaged the participants to connect the pattern in one image with the pattern in the other image, the patterns in their prior knowledge with those in the images viewed in the cases, especially in the conditions of the comparison and flicker methods. For the comparison group, the participants went through the patterns and compared the patterns one by one across images. By doing this, they compared across image patterns in the two images while ruling out the patterns that are not the searched patterns. The participants in the flicker group would go through similar processes of comparing across the images, but might not systematically compare across images as those in the comparison group. The images were flickered in animations and the movements and the unstable characteristics of the images might increase the difficulty of systematic comparisons. Anyhow, comparisons could enable the participants to connect patterns internally, make inferences, elaborate their generalization, and revise their generalization.

Compared with the case study and puzzles in the interactive system, the use of texts, atlas, sketched images, and other methods without images or with an annotated image or a few sketches to teach patterns will be less able to set a ground and goals for the participants and activate them to attend to features, make connections and hypotheses, and develop elaborations, hence generative learning is much less possible to occur with these methods than in the designed CBI system. With the atlas, text, and lecture methods, patterns tend to be told to learners directly. Pointing out the patterns directly to learners or even extracting patterns to more easily observed abstract forms of the patterns may hamper learners from constructing their own knowledge, which takes rounds of assimilation and accommodation. The processes of assimilation and accommodation are

crucial for learners to adjust their mental models and integrate patterns into their own knowledge structures.

Moreover, compared with these methods, the puzzles of patterns could promote the process of attending to details of the patterns, or extracting the detailed patterns into more abstract ones through retrieving knowledge from long-term memory and integrating knowledge. The process of input and output continues on with the generative activities and strategies (Grabowski, 2004; Stull & Mayer, 2007; Mayer, 2005; Wittrock, 1974) and different levels and types of mental processes. Studying images and patterns is not a one-shot project, but takes continuous efforts to enrich and extract, connect and communicate, and monitor and motivate. The problem-based contexts with the puzzles of patterns in the three treatments engaged the participants in generative learning, enabling them to continuously attend to, infer, integrate, organize, and evaluate visual information.

The cases and puzzles of patterns did not only activate the generation process but also facilitate the participants to overcome their limitations in visual perception, including limited attention, visual short-term memory capacity, and the lack of awareness, control, and monitoring in visual perception. Even if the puzzles of patterns were merely questions about what the patterns are, the participants still needed to invest much of their attention to the patterns because of the puzzles. Enabling the participants to search for changes, think of patterns, and connect what they found with what they wanted to define, the puzzles of patterns pulled the participants out of inertia so that attention, awareness, monitoring, and motivation were activated in figuring out the puzzles. The puzzles were a part of the interventions and instructional strategies although they became different when implemented in different instructional methods.

Second, next to the case study and the puzzles of patterns, the formative assessment and feedback in the shared instructional system promoted learning because the activities could facilitate the participants to distinguish salient patterns, connect their internal representations of patterns with what they viewed in quizzes, confirm their diagnostic decisions with provided responses, monitor their study progress, and motivate them to make continuing effort in study. For the participants in the flicker and comparison groups, assessment and feedback might have been regarded as “shortcuts” to correct responses, but still they would need to go through different patterns and make comparisons to reach correct responses. For the participants who emphasized more on image studies and got responses through studies, assessment and feedback would be useful for them to connect what they studied, represent internally, compare with assessment images, make their decisions and confirm their decisions. For the participants in the no-flicker group, they could have received direct answers to the puzzles, but they were also activated to retrieve from their memory the patterns they observed before assessment.

Formative assessment and feedback can benefit learning also because they may increase opportunities of accommodation and assimilation. Soon after errors were found and trials continued, the participants would adjust their thoughts, mental models of the newly detected patterns, until the patterns were evaluated as correctly identified in feedback. What matter are not the errors and trials but the internal constructive and discovery process and the meanings of patterns generated in this process. Furthermore, formative assessment and feedback could enhance the participants’ motivation and

engage them in reflecting what they learned and what they could revise and feedback could monitor and motivate this constructive learning process.

Third, in addition to the above-mentioned mental and instructional interactivity (Chou, 2003; Moore, 1989; Proske, Narciss, & Korndle, 2007), the branched interactions and user control provided the participants with options and interactions, the basis of effective CBI or WBT and individualized education, which could fit into the individual needs in visual perception of patterns and arouse curiosities and interest in experimenting with images and increase participation. Branching and user control take individual participants into consideration, providing different routes of progress and knowledge construction according to the learners' responses. This matches the essence of generative theories about learners rather than instructors as the center of learning and instruction. In this study, branched interactions were mainly located at the case study, assessment, and feedback of study sessions in the three programs. These interactions facilitated the participants to connect what they learned with what they were assessed and kept generating and revising patterns internally. User controls was mainly developed for the participants to select their preferred rates of animated images in both the flicker and no-flicker conditions. The options provided the participants possibilities to make comparisons between images to search for differences and/or changes and make inferences about patterns. Without user control of the speed, the participants might lack the mechanisms to observe the images, locate and identify patterns at their own paces, which otherwise might hamper generative learning. Furthermore, the user control interactivity here can also be classified as mental interactivity (Proske, Narciss, & Korndle, 2007). Usually linearity is regarded as irresponsible to different learners in

facilitating their learning processes (Alessi & Trollip, 2001; Jonassen, 2004). Lack of user control also weakens a CBI or WBT program because learners may become more engaged in learning when the instructional system responds to their thoughts and choices. Hence, higher interest, motivation, thinking, and individualized education, which promote generative processes, may become more possible with branched interaction and user control.

All of the three methods, combined with the shared interactive system, made learning occur. Although from appearance, no-flicker method was more a direct method, with responses to the puzzles directly demonstrated to the participants, the participants would still need to figure out the meanings of patterns for the following two reasons: First, the participants in the no-flicker group were only told that the change in a display indicated the pattern but they were not exactly told about what change indicated the pattern when animated patterns changed and popped out, resulting in a light color pattern swapping with a darker color pattern. Second, the participants in the no-flicker group also went through tests of the patterns after studying images. That is, they also needed to make sense of what they saw in the images and self-assessed the meanings they developed. By going through feedback, they were then confirmed of the meanings they created, in which they also somewhat tested their own representations. It may be argued that the participants in this group could mechanically view the images later on as they figured out the animated patterns, but the followed-up tests would still required an activated comparisons of their internal representations with what they viewed in tests.

Of course, short-term memory might be attributed as the testing results because the interval between the case study and study tests was immediate and brief in the no-

flicker group. The other factor that may compromise generative learning among this group could be that the participants might get only stuck in isolated patterns without connecting the patterns with one another to construct a more generic pattern of a category. This may explain why the no-flicker group performed significantly lower than the other two groups in the classification test.

In the flicker treatment, the puzzles of patterns were more challenging and the demands of internal processes and generation were higher than those in the no-flicker treatment. The urge and difficulty level of making sense and constructing connections of what they already know with new information were higher than those for the no-flicker group. The puzzles of patterns took more learning effort and connections of patterns to be figured out. Meaning making and mindful learning was crucial for these participants solving the puzzles of patterns. Focused attention, continuous comparisons between the images to identify patterns, the internal representation accompanied, and other generative learning processes must have been going on internally to make possible the solution to the puzzles.

In the comparison treatment, the puzzles of patterns motivated the participants to continuously search for differences between the side-by-side images. It was possible for the participants in this group to make systematic comparisons of the images in order to solve the puzzles. Consistent effort could be invested in this process. Meaning searching became persistent because the participants could solve puzzles in each set of cases by going through the tasks of searching for patterns, comparing patterns, generating meanings, and representing meanings for further study tests.

Therefore, the shared system of the puzzles of patterns, case study, assessment, feedback, branching and other interactions worked together with the three methods. All of the participants in the study learned because of the synergy of the shared interactive system and individual methods. The results of significant learning outcomes from the three groups supported that learning was fostered and promoted.

No Significant Effect in the Recognition Test

No significant effect was identified in the recognition test among the three groups of participants, even with the pretest score as covariate to decrease variance errors. Nevertheless, in the proposal of this study, the participants studying visual patterns with the flicker treatment was expected to outperform those studying with the comparison treatment as well as the no-flicker treatment. Of course, the mean scores of the tests showed that the flicker group achieved the highest recognition mean score, followed by the comparison group and then no-flicker group. However, the mean score differences of the recognition test between the flicker group and the comparison group, and the comparison group and the no-flicker group were less than .50 points and no statistical significance was detected. After adjusting means by deducting the pretest score's influence, the mean score difference between the flicker and no-flicker group stayed the highest among three pairs of groups and rose to .80, but still no significant difference was detected. This finding of insignificance in the recognition test is consistent with the finding of no significant difference in its recognition test in a recent change detection/flicker and no-flicker study with scenery pictures (Carlin, Soraci, & Strawbridge, 2005). This result is also somewhat coherent with the previous generation effect with pictures (Kinjo & Snodgrass, 2000; Peynircioglu, 1989) , in which generation

effect was identified stronger in free recall tasks while the effects identified in recognition tests were not as strong as those in free recall tests. However, it was somewhat inconsistent with the generation effect studies with texts (Slamecka & Graf, 1978; Jacoby, 1978), in which significant differences were identified in both recognition and free recall tests. The reason, as indicated by Kinjo and Snodgrass (2000), may be that pictures have more sensory cues for retrieval than texts (Paivio, 1971).

Significant Flicker Effect in the Classification Test

The results from the classification test were partly expected and partly unexpected in the proposal. Although the proposal did not hypothesize on the directionality of group differences in the classification test, it supported the flicker effect upon pattern recognition over the other two treatments. The proposal reasoned from different respects of learning to support this argument, mainly with the generative learning theory (e.g. Grabowski, 2004; Wittrock, 1974) and evidence in education and the generation theory and evidence (Slamecka & Graf, 1978; Jacoby, 1978) from psychology. With the results from the classification test, the significantly different effect was indeed detected between the flicker and no-flicker treatment, but no significant difference was identified between the flicker treatment and the comparison treatment. More specifically, with the pretest score as covariate, significant differences were detected in the classification test between the flicker group and the no-flicker group, adjusted mean difference=.963, $p=.007$.

Another unexpected result is that the comparison treatment was identified to have a significant effect over the no-flicker treatment in the classification test, with the pretest score as a covariate, adjusted mean difference=.904, $p=.010$. Of course, the proposal analyzed the merits of the comparison method and indicated that the comparison method

was commented as an engaging method in the subject area education (e.g., Roberts & Chew, 2003). Furthermore, the proposal also cited resources about the advantages of the comparison method (Schwartz and Bransford, 1998; Mayer, 2001). However, these studies were integrated into the framework to support the flicker effect rather than support the effectiveness of the comparison treatment.

Significant Differences in the Other Outcome Measures

Other than the effectiveness measurement of the three treatments with the recognition and classification test scores, three other outcome measures were examined to provide more evidence about the effectiveness and efficiency of the three treatments. The study results suggested that the participants studying with the three treatments spent significantly different time and made significantly different number of incorrect responses and trials.

The data indicated significant differences in study duration among groups: the participants in the flicker group spent significant more study time than those in the comparison and no-flicker groups while the participants in the comparison group spent significant more study time than those in the no-flicker group, hence those in the no-flicker group spent the least amount of time among the three groups.

The result of the longest duration among the flicker group participants is congruent with the change detection literature results (Carlin, Soraci, & Strawbridge, 2005; Philip, 1974; Rensink, O'Regan, & Clark, 1997; Rensink, 2002; Simon & Levin, 1997). In the change detection studies, the subjects in the flicker treatment tended to spend significantly longer time than those in the no-flicker treatment when change signals were "instantaneous and visible" (Simon & Ambinder, 2005). The task in previous

change detection studies/flicker studies asked the subjects to identify the changes between images and these changes could be anything and did not belong to a category or categories. Therefore, the changes were definitely unexpected and viewers would have no idea of what they would see in the images, let alone the changes among images. Nevertheless, in this study, the general content and tissues of the images were known to the participants, who had a general idea of what they would see in cases. Furthermore, the categorization task was used to make more expectations happen in learning. However, the newly learned complex images still took time for the participants. In addition, images were more complex in this situation than those of everyday images in former change detection and flicker studies. Of course, the significantly more time spent among the comparison group than the no-flicker group was not expected because the comparison method was anticipated as a method with which the participants could easily get the responses, almost as those in the no-flicker group.

Moreover, the study results show that, compared with the participants studying with the no-flicker treatment, the participants studying with the flicker treatment and the comparison treatment took significantly more trials and made more errors to solve the problems of recognizing, identifying, and locating changes between images. A part of the results were expected in the proposal because the flicker treatment was expected to challenge the participants in pattern recognition while the no-flicker treatment was anticipated as a direct method to provide correct responses. Meanwhile, the results of the comparison method was unanticipated because the treatment was originally regarded as a direct one, from which the participants could reach recognition accuracy with facile.

The results of a significantly higher number of incorrect responses and trials in the flicker and comparison groups than the no-flicker group can be explained with the following reasons: First, in the comparison and flicker treatments, learners needed to invest cognitive and metacognitive resources to identify patterns and changes between images, in which they probably could not correctly identify changes initially. Or even though they recognized the changes soon, they might have insufficient resources to attend to different patterns and it would take trials for them to compare across patterns and reach the correct responses through carefully viewing patterns. In the no-flicker treatment, the change of patterns were directly presented through animated patterns, which attracted learners' attention so that they identified the patterns and responded to the questions with significantly fewer trials and errors than those in the flicker and comparison treatments. This result was consistent with the results of change detection studies (Carlin, Soraci, & Strawbridge, 2005; Philip, 1974; Rensink, O'Regan, & Clark, 1997; Rensink, 2002; Simon & Levin, 1997). It tended to take longer time and alternations to detect changes, as in flicker conditions, although such changes as instantaneous ones could be identified with good detection, as in no-flicker conditions. The results imply that generative learning may cost more time, errors and trials than non-generative learning. Admittedly, the comparison treatment was not anticipated as a generative method in the proposal and the related hypothesis will be further discussed in the following section of theoretical implications.

It is worthy to note that significant differences were identified between the flicker group and the comparison group in duration, number of incorrect responses, and number of trials. The flicker group participants were found to spend longer time and took more

errors and trials to reach correct responses than the comparison group participants. However, these significant differences were not reflected in the recognition and classification posttest performance because the participants in these two groups were not found to outperform each other in the recognition test and the classification test. The reason may be that the flicker method is a novel strategy that may take more time to learn than the comparison strategy. The other reason can be that the comparison of sequential images in the flicker treatment cost more resources than that in the simultaneous comparison.

Theoretical Implications

The significant effect identified in this study can at least support and extend generative learning theory (Grabowski, 2004; Wittrock, 1974), generation effect theory (Carlin, Soraci, & Strawbridge, 2005; Kinjo & Snodgrass, 2000; Peynircioglu, 1989; Slamecka & Graf, 1978; Jacoby, 1978), and change blindness theory (Rensink, O'Regan, & Clark, 1997; Rensink, 2002; Simon & Levin, 1997). The former sections in this chapter have commented on how the significant differences of the flicker and no-flicker methods have been represented in the data of the outcome measures and how these results are consistent with the generative theory and change blindness theory and studies. This section will continue to interpret these results with potential explanations and will also propose theoretical grounds for the findings of the comparison treatment.

To begin with, the result that the flicker group significantly outperformed the no-flicker group in the transfer test of classification can be further explained with the multiple cue hypothesis, distinctiveness theory, coding specificity, and cognitive operation theory (e.g., Carlin, Soraci, & Strawbridge, 2005; Kinjo & Snodgrass, 2000;

Rensink, O'Regan, & Clark, 1997; Rensink, 2002; Simon & Levin, 1997), which tended to be used to interpret generation effect and generative learning phenomena and results.

First, in the flicker treatment, the participants needed to search and view across image patterns to identify the change and visual pattern, so they noticed more than one pattern. However, in the no-flicker treatment, the participants merely needed to notice the pop-up pattern to reach correct responses. The multiple cues used by the flicker group participants could benefit retrieval of information, which in turn could lead to better connection and sorting out of information to benefit categorization of patterns.

Second, the distinctiveness theory suggests that decision makings among multiple paths/responses may increase memory of the selected response. In the flicker treatment, the participants had to make a decision in each case study among potential stimuli and select one item as the pattern after comparing across the patterns. Of course, in the no-flicker treatment, the sensory distinctiveness of visuals that lied in the animated pattern and popped up could also increase memory. However, this latter distinctiveness did not have decision-making elements but merely meant sensory salience, which may probably not lead to higher order thinking of classification.

Third, when study tasks and test tasks were congruent, the study results would be effective. The flicker treatment included categorization tasks that demanded sorting out potential data into categories, so it resulted in better learning outcomes in the categorization test. However, in the no-flicker treatment, there were no learning processes of categorization although categorization was also suggested because the answer was provided.

Fourth, the cognitive operation theory emphasizes the importance of computations in learning processes and assumed that more operations with cognitive and metacognitive resources can lead to better learning than less. The flicker treatment took more effort than the no-flicker treatment, supported by the results of duration, number of incorrect responses, and number of trials. Hence, it could make better learning occur than the no-flicker treatment.

For all these reasons, the flicker treatment could cultivate the generation of mental models, analysis of visual patterns, elaboration and inference of patterns. Therefore, a significantly better transfer of learning occurred in the flicker treatment as a generative strategy than the no-flicker treatment as a direct strategy.

Moreover, as indicated in the recognition test, the no-flicker treatment may increase the retention of the patterns because of its merits in facilitating learning. The advantages of animated patterns and pattern changes in the no-flicker condition facilitated the participants to gain attention because the animated pattern change was helpful for selective attention and enabled the novice learners to learn efficiently, complete study sessions in shorter periods of time and meanwhile make less errors and achieve high accuracy rate in the study session. Moreover, according to psychological principles (Goldstein, 2002), motion can facilitate viewers to bring hidden images from clutters. When image features are not salient because of noise in the environment, learners' visual systems can be overloaded and the view of features may be obscured. Motion can be helpful in decreasing the load of the noise in signal detection (Goldstein, 2002). Hence, animation can be a good strategy when it is used with this type of complex images and image features for novice learners to decrease cognitive load. These

advantages of the no-flicker treatment are also reflected in the learning gain from the pretest to posttests.

Furthermore, evidence from this study supports that the comparison treatment can be hypothesized as a generative method, congruent with the existing generative learning and generation effect theory. More specifically, compared with the other two treatments, the comparison method has the optimal results taking consideration of all of the dependent variables, including the recognition score, classification score, duration, number of incorrect responses, and number of trials. This finding differs from what was expected about the flicker effect surpassing the comparison effect.

The following four facts from the study can partly reveal the nature of the comparison method. First, with the second least time and number of incorrect responses/trials, the comparison group performed the best in both the recognition and categorization tests except a slightly lower recognition test mean score than the flicker group. Compared with the flicker group and no-flicker group, the comparison group on average used significantly more time than the no-flicker group but less time than the flicker group to study the same set of cases with the same set of guidance, including assessment and feedback. Second, the comparison group was identified to perform significantly better than the no-flicker group in the classification test. Third, there were no significant differences between the comparison group and the flicker group in both the recognition and classification test. Fourth, it was found that the flicker group performed better in the transfer test than the no-flicker group, but the time spent and errors made in the learning processes were much higher than the no-flicker groups and also more than the comparison group.

Therefore, the comparison method can be proposed as a generative strategy for further investigation. This possibility will be further explored in the recommendation section of this chapter.

Implications for Practitioners

First, generative strategies of visual concept learning may improve learning effectiveness, considering the learning gains through the treatment of flicker and significant differences of the flicker treatment from the no-flicker treatment in the classification test, and potential return of investment (ROI) in training and education. As indicated in this study, transfer of learning can be better attained in studying with such a generative strategy as the flicker treatment and a potential generative method the comparison treatment. Of course, both generative and direct instruction may enhance learning, but the generative strategy and potential generative strategy can more significantly enhance classification performance than the no-flicker treatment.

Nevertheless, as well identified in this study, the practical differences between the flicker treatment and the no-flicker treatment and the comparison and no-flicker treatments did not appeal to attention, with merely about one point difference in the mean scores between the flicker and no-flicker group, and comparison and no-flicker group. One point in the study measured only half of the points of a question, indicating on average the flicker and comparison participants did not answer one more question correctly than the no-flicker group participants. The other point that can reflect the small practical significance among these treatments is the small effect size of the flicker and no-flicker difference, as well as the comparison and no-flicker difference (Cohen $d < .5$).

The following details about the effect sizes provide evidence of the practical significance of the three computer-based instructional methods.

First, Cohen's d values show that both the comparison method and the flicker method had small learning advantages over the no-flicker method in impacting classification performance because the effect sizes, respectively .46 and .37, were identified as small although significant differences existed. Second, Cohen's d values on the other hand demonstrated that the small learning advantages of the two methods were gained at high cost. Cohen's d values of time, number of incorrect responses, and number of trials showed that the significant differences of these three items were large between the comparison group and the no-flicker group (1.24, .69, and .69), the comparison group and the flicker group (-.52, -.72, and -.72), and the flicker group and the no-flicker group (1.57, 1.25, and 1.25). Therefore, the comparison and flicker groups did significantly better than the no-flicker group in the classification test. Nevertheless, the significant effect sizes were small and the cost for the small learning advantages over the no-flicker treatment was high. That is, relatively small learning advantages of the flicker group over the no-flicker group and the comparison group over the no-flicker group were achieved at relatively high cost of time, the number of incorrect responses (false alarms), and the number of trials (alternations).

The Cohen's d values also assist the practitioners in evaluating the magnitude of significance and making decisions about the practical values of the three computer-based instructional methods. Considering the relatively small learning advantages at relatively high cost, the following implications can be drawn for the practitioners. First, the flicker method was one of the methods significantly increasing classification performance but

with the method learners had to spend significantly longer time studying to gain negligible significant differences from the learners studying with the other two methods. Moreover, the misconceptions and the trials in study sessions are detrimental to clinical work because false alarms and alternations endanger patients and cause problems.

Therefore, one of the approaches to implementing the method can be computer-based simulations. Simulated cases and contexts can be designed and developed for practice in a virtual clinical environment to avoid potential problems. Second, the no-flicker method can be used as an aid to clinical study because the method can increase the efficiency of detection and decrease false alarm rates and patient recall rates. If this method were used to provide second opinions for residents and doctors in clinical study, these practitioners could extend their perception and see what they might not see efficiently and accurately. Hence, this method can be used as a method in aiding doctors in clinical study. Third, the comparison method has been in use in clinical environment and it may be complemented with the no-flicker method as a second-view method to improve the accuracy and efficiency of detection and diagnosis.

The comparison method may be preferred than the other two methods because it made significant differences in the transfer test of pattern recognition performance. Meanwhile, this significance did not result in too much time to spend and too many false alarms and alternations. However, this on-the-whole better method does not deny the usefulness of the other two methods. The flicker method did result in significant higher transfer scores than the group studying with no-flicker method. Of course, it led to significantly more time, false alarms, and alternations in the study process, but all of these may lead to better long-term memory, which was not tested in this study. With the

no-flicker method, on the other hand, the participants spent significantly less time than those in the other two groups as well as significantly less false alarms and alternations.

Therefore, the three methods have their own merits. One of the recommendations for the practitioners is that they can apply all of the three methods and try to take all of the measures to promote learning. There is no one optimal method but all of the methods may work in some respects. It is good to take all potential measures that may help improve learning but not just one method to solve an urgent problem of learning. It is even worse to wait until one “optimal” method is found finally then experiment with the method in instruction and learning.

It also depends on individual instructional designers to decide whether it is worthwhile to achieve the small effect size significance in classification by taking significantly longer time and made significantly more errors and trials with the flicker treatment and the comparison treatment; or ignoring the small effect sizes and pursue a less time and less error and trial study of images with the no-flicker treatment. Of course, in the future, more generative methods can be designed and evaluated in this area for CBI and/or WBT. In addition, such direct methods as the method of no-flicker can also be designed, complemented, and compared with generative methods of pattern recognition.

Second, the flicker treatment can challenge students in engaging their cognitive and metacognitive resources in studying complex images, resulting in constructing mental models, making sense of patterns, generating inferences, and evaluating models and inferences. The original change detection tasks were revised to decrease the cognitive load of learning materials with expectation of the changed objects' categories, user control of image display pace, and one type of images. Although the complexity of the

flicker treatment in this study may be still high, the top-down knowledge of the changed category provided learners expectations of the change. Therefore, the flicker treatment can be employed to increase effort in closely viewing and studying visual patterns to enhance pattern recognition performance. The caution in using the method is that naïve learners in complex images may take long time to figure out the strategy itself, as indicated in the study, so measures need to be taken if the treatment will be applied in real world instructional design.

Third, the comparison treatment can be a useful and effective method of teaching and learning to enhance fundamental pattern recognition knowledge and skills in CBI and/or WBT. The reason is that this method can engage learners in attending to possible patterns, discriminating these patterns, classifying objects into categories, and making connections among what they view with their prior knowledge. As indicated in literature (e.g., Schwartz & Bransford, 1998), comparison methods can facilitate learners to construct their concrete knowledge of patterns before they read more texts on these patterns for concept learning. The prior knowledge of patterns and schemas can support learners to make sense of what they read later. In this sense, the comparison method can prepare learners for their future meaningful learning (Schwartz, Martin, & Nasir, 2005). What is more important, the method can improve learners' transfer of learning in CBI which is a highly recommended learning outcome pursued in instructional design (Bransford, Brown, & Cocking, 1999; Jonassen, 2004; Mayer, 2002, 2005). Of course, as indicated previously, the practical learning gain difference was not outstanding because there was merely about one point difference between the comparison group and the no-

flicker group in both tests. This gain also accompanied high costs of time, errors, and trials.

Fourth, the no-flicker treatment can be integrated into real world instructional design because the no-flicker task can help students achieve a similar level of performance in recognition, but it may not be as effective as the other two methods in increasing students' performance in transfer tests. From appearance, the no-flicker task is a direct approach which may incur passiveness of learning. However, the no-flicker group participants in the study demonstrated learning gains with their doubled scores in both recognition and classification tests, compared with their pretest scores. Furthermore, the no-flicker treatment cost the least in study time, number of errors and trials, compared with the other two treatments.

Fifth, the use of different instructional methods in this situation may influence time cost, false alarm rates, and trials, the selection of methods may depend the contexts of learning and instruction. As elaborated at the beginning of this section, the no-flicker method has the advantage of cost-effectiveness and less false alarm rates and trials over the other two methods. Therefore, the method can be an option of learning and instruction, considering the contexts. As a method of practice or even as a clinical method as the second view method combined with the traditional methods, it probably has potential to improve learning with less time and false positive and trial rates than the other two methods.

In conclusion, the comparison method and the no-flicker method may be practical in real life instructional design. The differences of the three treatments generally indicate to the practitioners the balance of effectiveness and efficiency in instructional design.

Limitations of the Study

In generalization of the findings of the study, it is necessary to take cautions. First, the study cases may not reflect clinical instruction and learning requirements. In clinical situations, the difficulty level and the change of images may be more complex than the cases in this study. With the complexity increased, the effect findings in this study may become uncertain. Second, the edited images may not reflect the images in clinical contexts and cannot be generalized to clinical studies. The images in this study were edited only for the purpose of visual concept instruction but did not reflect clinical contexts. The images in clinical contexts are probably different from what the edited images in this study represented. The reason for the simplified edited images was that the researcher did not have expertise in the corresponding subject area.

The other limitations of this study are the population, background, motivation, knowledge, and the other factors that have not been investigated in this study. The study results are limited to the studied population, including their demographics, prior knowledge, expertise, and other characteristics. Therefore, the results cannot be generalized to the other populations without more examinations.

First, the study results only stand for the outcomes of the studied population in the metropolitan university in the Southeast of the United States. Demographics of the other populations in the other areas are probably different from those of this population, including the genders, ethnicity, ages, and the other aspects of demographics. These categorical variables may impact learning differently from those in this study. Therefore, the study results cannot be generalized to the other populations.

Second, the sample of the study was drawn from arts, science, and engineering programs, mostly undergraduate programs in the University. This population is different from the population from medical areas because their educational background contains very limited components of medicine. Medical school students and radiology residents have more solid educational foundations in medicine and trainings in medicine and probably have significantly different learning outcomes from the studied population. Thus, the results cannot be generalized to these populations with educational backgrounds in medicine and the other populations may have significantly different outcomes with the same sets of treatments.

Third, as an important factor of learning, motivation of populations influences study outcomes, so the learning outcomes of this study do not represent those of the other populations because the other populations may have higher or lower level of motivation in studying these visual patterns. As indicated at the beginning of this document, residents were identified as lack of motivation in studying mammograms. Then studying with the treatments may lead to results different from those in this study. For another example, medical school students, who may be interested in becoming residents in related areas, may be interested in learning the visual concepts and then invest in most affective, cognitive, and metacognitive resources in learning activities and perform significantly differently in the studied measures.

Fourth, the level of knowledge and expertise may largely influence learning outcomes. All of the participants in this study are novices in mammography interpretation. Therefore, the study results cannot be generalized to the populations with higher level of expert knowledge, such as radiology residents, radiologists, and experts.

The potential participants in this study were asked whether they had little knowledge of mammograms, thus novices were the study's target audience. It was unclear how these treatments would result in if participants had higher levels of knowledge in this area.

Recommendations for Future Research

Through this experimental study, an exploratory investigation was conducted to examine the impact of three treatments on pattern recognition in CBI and/or WBT. Three computer-based instructional programs were compared with each other, primarily on the different effects of the treatments upon recognition and classification of image patterns, as well as on the study duration, number of incorrect responses and number of trials. Based on this study, here are recommendations for future research:

In this study, the three treatments were compared in terms of their effects upon visual pattern recognition and learning efficiency. Two of the three treatments and their shared instructional systems, the flicker and no-flicker treatments, were rarely used in mammogram instruction. In real-life instructional design, concurrent images tend to be instructional designers' choice in displaying images. However, this is not to say that simultaneous image displaying is sufficient for learning. Furthermore, there is lack of empirical evidence to support the effectiveness of this display method and the simultaneous displaying method has rarely been studied with an interactive instructional system in mammogram instruction. Therefore, it is necessary to explore innovative and existing methods of instruction for improving learning outcomes of pattern recognition. With continuous studies of effective and efficient CBI and/or WBT methods, both students and instructors can benefit from instructional and learning strategies and make learning occur in technology-based environment. Hence, in the future, how to engage

learning, enhance performance, and improve efficiency in visual pattern recognition can be further studied.

First, innovative methods of instruction and learning activities can be created or discovered on a multitude of theoretical and empirical bases and these methods can be studied on their effectiveness of impacting visual pattern recognition and other learning factors. Reflecting upon the study, experiments with different methods of instruction can provide instructors with potential methods of instruction and students with potential learning strategies to increase learning in CBI and/WBT. Therefore, it is worthwhile to explore what other methods of instruction exist and/or can be created, at what levels these methods can promote pattern recognition, and whether there are methods of instruction that can significantly increase the effectiveness and efficiency of visual pattern recognition. For example, if providing learners the options of the flicker, no-flicker, and comparison methods, it may be interesting to examine what methods or method they will choose in learning visual concepts and how the method(s) will impact their learning outcomes. Research into multiple methods of visual pattern recognition may benefit students by providing them with different CBI/WBT approaches to learning, selecting what fit into their learning styles, and increasing the opportunities of engaging in studying images.

Second, the future research can be extended to testing the study results from this study with other populations, who have different demographics, prior knowledge, abilities, and other individual features. Specifically, the studies of these three methods can be retested with different populations who may represent different types of individuals. Through more studies on these factors researchers can provide more

knowledge to the practitioners about whether the methods can be useful for different populations and what methods more benefit types of populations. Although this study has already had evidence in a population of potential learners, it is far from a conclusive study and many more studies need to be conducted in the future, which will gradually validate and reconstruct computer-based instructional systems and methods of effective and efficient learning in pattern recognition. There could be totally different results of studies if populations were changed, which may lead to design and development of other options of instruction and learning of pattern recognition and thus establish other evidence.

The other individual factors include learners' motivation and expertise. Learners' perception of their motivation before and after studies may be collected to compare with the existing motivation data. Motivation is critical in learning and may significantly influence learning outcomes in computer-based instruction. However, this study did not collect data about how the programs impacted the participants' motivation in learning image patterns. Furthermore, future research can also include the other factors that have not been covered in this study, for example, expertise. It will be interesting to study the effects of these three methods among readers of different expertise. The participants in this study had little knowledge of what they learned and they ended up with largely increased performance in recognition and classification. It is unclear whether similar results of this type can be found among experts.

Third, to continuously integrate instructional affordances into radiology education and related areas, future research in instructional design and technology can be grounded in subject experts' experience and knowledge of instruction. Existing pedagogical

methods can be studied and researchers can examine what methods or components of methods possible to be integrated into computer-based instruction and/or Web-based training on the basis of human learning and instructional design theories and principles. Throughout years of teaching in a variety of modalities and with many students, experts in radiology must have had in-depth experience and practice in teaching pattern recognition and knowledge and skills beyond pattern recognition in this area. Their first-hand knowledge of what work and what do not work well can inform the future generations of novices in this area, helping them improve their learning strategies. This type of studies can inform practitioners of potential activities and tasks that can increase e-learning effectiveness and efficiency. Ranging from the fundamental thinking in this area to more complex and higher-levels of thinking and an integrative practice of expert knowledge and skills, researchers can further navigate the uncharted sea in this interdisciplinary area. Through the convergence of the perspectives of general knowledge of human learning and instructional design and specific first-hand knowledge of experience of instruction, researchers will have more solid foundations for the art and science of instruction and learning in this field.

Fourth, future researchers in instructional technology are supposed to continue with investigations of potential effective and efficient options for individual learners to engage in activities and improve their pattern recognition and related abilities. The three dimensions of thinking, including human learning-based individualized instruction, adaptive instructional methods, and rich technology affordances coping with learners, instructors, and instructional methods, are the grounds of future researchers. Many options of learning need to be designed and developed but the purpose of doing so is not

only examining which package or single method work significantly better than the others but comprehensive systems of instructional system will be designed and get empirical support, established, and refined. The design, development and research of instructional options and instructional systems for individualized computer-based instruction may lead to more integrative arts and science of instruction and learning. Research of pattern recognition and related areas in the context of technology is an unending process of achieving understanding through making sense of previous knowledge and new information and producing and examining new questions and hypotheses.

Fifth, reading images can be studied in other areas rather than mammograms, including images in the other areas of radiology, medicine, biology, math, chemistry, architecture, and languages. It is necessary to study the computer-based instructional methods that can increase learners' understanding, analysis, and evaluation of images in these areas. With these methods, learners can learn concepts, principles, and solve problems more effectively and efficiently by studying images in these areas.

Conclusions

This chapter provides a comprehensive evaluation of the research findings based on relevant literature. The chapter analyzes and interprets the effectiveness and efficiency of the three programs and the three CBI treatments for pattern recognition. Finally, the chapter provides implications for instructional designers, researchers, and related educators and researchers. Limitations and recommendations for future research directions accompany.

List of References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Alessi, S. M., & Trollip, S., R. (2001) *Multimedia for learning: Methods and development*. Boston, MA: Allyn and Bacon.
- Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, 32(8), 10-14.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Antonov, I., Antonova, I., & Kandel, E. R. (2001). The contribution of activity-dependent synaptic plasticity to classical conditioning in Aplysia. *Journal of Neuroscience*, 21(16), 6413-6422.
- Armato, S. G., Doshi, D. J., Engelmann, R., Croteau, C. L., & MacMahon, H. (2006). Temporal subtraction in chest radiography: automated assessment of registration accuracy. *Medical physics*, 33(5), 1239-1249.
- Auble, P. M., Franks, J. J., & Soraci, S. A., Jr. (1979). Effort toward comprehension: Elaboration or “aha”? *Memory & Cognition*, 7, 426–434.
- Azevedo, R. (1998) Expert problem solving in mammogram interpretation: A visual cognitive task. Ph.D. dissertation, McGill University (Canada), Canada. Retrieved June 22, 2007, from ProQuest Digital Dissertations database.
- Bangert-Drowns, R., Kulik, J., & Kulik, C. L. (1985). Effectiveness of computer-based education in secondary schools, *Journal of Computer-Based Instruction*, 3, 59–68.

- Barlow, W. E., Chi, C., Carney, P.A., Taplin, S. H., D'Orsi, C. J., Cutter, G., et al. (2004). Accuracy of screening mammography interpretation by characteristics of radiologists. *Journal of the National Cancer Institute*, 96, 1840–1850.
- Bassett, L.W., Monsees, B. S., Smith, R. A., Wang, L., Hooshi, P., Farria, D. M., et al. (2003). Survey of radiology residents: breast imaging training and attitudes. *Radiology*, 227(3), 862-9.
- Begg, I., Snider, A., Foley, F., & Goddard, R. (1989). The generation effect is no artifact: Generating makes words distinctive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 977–989.
- Berge, Z. L. (2002). Active, interactive, and reflective eLearning. *Quarterly Review of Distance Education*, 3(2), 181-90.
- Berk, L. E. & Winsler, A. (1995). Scaffolding children's learning: Vygotsky and early childhood education. Washington DC: National Association for the Education of Young Children.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*. 94. 115-147.
- Blok, H., Oostdam, R., Otter, M., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: a review. *Review of Educational Research*, 1, 101–130.
- Braden, R. (1996). Visual literacy. In D.H. Jonassen (Eds.), *Handbook of research for educational communications and technology*. New York: Macmillan.
- Brody, A. S. & Zerlin, J. M. (2000). Internet teaching files in radiology: A call to action. *Academic Radiology*, 7(9), 748-9.

- Bransford, J. D., Stein, B. S., Vye, N. J., Franks, J. J., Auble, P. M., Mezynski, K. J. & Perfetto, G. A. (1983). *Differences in approaches to learning: An overview*. Journal of Experimental Psychology: General, 3(4), 390-398.
- Bransford, J. D., Brown, A. L., and Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions. *Journal of the Learning Sciences*, 2(2), 141-178.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.
- Bruning, R. H., Schraw, G. J., Norby, M. M. & Ronning, R. R. (2004). *Cognitive psychology and instruction* (4th ed.). Upper Saddle River, NJ: Pearson.
- Buckley, D., Coleman, W., Cohen, M., & Stewart, R. (1999). Interactive multimedia learning environments: Tools to foster transition to the learning paradigm. Report: ED448702. Retrieved June 1, 2007, from the ERIC database.
- Bucy, E. P. & Tao, C. C. (2007). The mediated moderation model of interactivity. *Media Psychology*, 9 (3), 647-672.
- Carmody, D. P., Nodine, C. F., & Kundel, H. L. (1980). Global and segmented search for lung nodules of different edge gradients. *Investigative Radiology*, 15(3), 224-233.
- Carlin, M. T., Soraci, S. A., & Strawbridge, C. P. (2005). Generative learning during visual search for scene changes: Enhancing free recall of individuals with and without mental retardation. *American Journal on Mental Retardation*, 110(1): 13-22.

- Chan, S. & Gunderman, R. B. (2005). Emerging strategic themes for guiding change in academic radiology departments. *Radiology*, 236: 430-440.
- Chandler, P., & Sweller, J. (1990). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8, 293-332.
- Chew, F. S. (2001). The case-based radiology teaching conference for residents: Beneficial effect of previewing cases and using answer sheets. *Academic Radiology*, 8(10), 993-997.
- Cho, J.-R. & Mathews, R. C. (1996). Interactions between mental models used in categorization and experiential knowledge of specific cases. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 49(3), 572-595.
- Chou, C. (2003). Interactivity and interactive functions in Web-based learning systems: A technical framework for designers. *British Journal of Educational Technology*, 34(3), 265-79.
- Chung, H. & Zhao, X. (2004). Effects of perceived interactivity on Web site preference and memory: Role of personal motivation. *Journal of Computer-Mediated Communication*, 10 (1), 00-00.
- Clark, R. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53(4), 445-459.
- Clark, R. E. (1989). Current progress and future directions for research in instructional technology. *Educational Technology Research and Development*, 37(1): 57-66.
- Clark, R. E. (1994). Media will never influence learning. *Educational Technology, Research, and Development*, 42(2), 21-29.

- Clark, R. (2005). Re: Accuracy of screening mammography interpretation by characteristics of radiologists. *Journal of the National Cancer Institute*, 97(12): 936.
- Cobb, T. (1999). Applying constructivism: A test for the learner as scientist. *Educational Technology Research & Development*, 47 (3), 15-31.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.) *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Collins, J., Blankenbaker, D. G., Albanese, M. A., Stack, S. P., Heiserman, K. K., Primack, S. L., & Kazerooni, E. A. (1999). Chest radiology case exchange program: A paradigm for resident teaching and independent resident learning. *Academic Radiology*, 6, 34-39.
- Collins, J. (2000). Curriculum in radiology for residents: What, why, how, when, and where. *Academic Radiology*, 7(2), 108-13.
- Collins, J. (2006). Medical Education Research: Challenges and opportunities. *Radiology*, 240 (3), 639-647.
- Comer, P. G. & Geissler, C. (1998). A methodology for software evaluation. Report: ED421140. Retrieved May 10, 2007, from the ERIC database.
- Cook, D. A. (2005). The research we still are not doing: An agenda for the study of computer-based learning. *Academic Medicine*, 80(6), 541-548.

- Craik, F., & Lockhart, R. (1972). Levels of processing: a framework for memory research. *J Verbal Learn Verbal Behav*, 11, 671–684.
- Crutcher, R. J., & Healy, A. F. (1989). Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4), 669-675.
- Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. New York: Teachers College Press.
- Cuban, L. & Usdan, M. (2003). *Powerful reforms with shallow roots: Improving America's urban schools*. New York: Teachers College Press.
- Davidson-Shivers, G. V. (2002). Instructional technology in higher education. In R. A. Reiser & J. V. Dempsey (Eds.), *Trends and issues in instructional design and technology*. Upper Saddle River, NJ: Merrill Prentice Hall.
- Dee, K. E. (2002). MammoEd: digital interactive breast imaging education. *Medical Education*. 36(11), 1103-4.
- Deitte, L. (2006). Challenges to radiology resident education in the new era. *J Am Coll Radiol*, 3: 528-533.
- De Laurentiis, E. C. (1993). How to recognize excellent educational software. Report: ED355932. Retrieved June 2, 2007, from the ERIC database. Department of Defense (1996). High level architecture for modeling and simulation master plan, Version 1.7.
- Draves, W A (2000) *Teaching online* LERN Books, River Falls, Wisconsin.
- Dewey, J. (1902). *The child and the curriculum*. Chicago, University of Chicago Press.

- DiPietro, M., Ferdig, R., Boyer, J., & Black, E. (2007). Towards a Framework for Understanding Electronic Educational Gaming. *Journal of Educational Multimedia and Hypermedia*, 16(3), 225-48.
- Elissavet, G., & Economides, A. A. (2003). An evaluation instrument for hypermedia courseware. *Educational Technology & Society*, 6(2), 31-44. Retrieved July 15, 2005 from <http://ifets.ieee.org/periodical/6-2/4.html>.
- Elmore, J. G., Taplin, S. H., Barlow, W. E., Cutter, G. R., D'Orsi, C. J., Hendrick, R. E. et al. (2005). Does *litigation* influence medical practice? The influence of community radiologists' medical malpractice perceptions and experience on screening mammography. *Radiology*, 236(1): 37-46
- Ericsson, K. A., & Charness, N. (1997). Cognitive and developmental factors in expert performance. In Feltovich PJ, Ford KM, Hoffman RR, *Expertise in context* (pp. 3-41). Mahwah, NJ: Erlbaum.
- Erkonen, W. E., D'Alessandro M. P., Galvin, J. R., Albanese, M. A., & Michaelsen, V. E. (1994). Longitudinal comparison of multimedia textbook instruction with a lecture in radiology education. *Acad Radiol*, 1, 287-292.
- Friedman, A., Polson, M. C., & Dafoe, C. G. (1988). Dividing attention between the hands and the head: Performance trade-offs between rapid finger tapping and verbal memory. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 60-68.
- Friedman, R. B. (1996). Top ten reasons the World Wide Web may fail to change medical education. *Acad Med*, 7, 979-981.

- Gagne, R. M., Wager, W., Golas, K. C., & Keller, J. M. (2005). *Principles of instructional design (5th ed.)*. Belmont, CA: Wadsworth/Thomson Learning.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction (7th ed.)*. Boston: Pearson Education, Inc.
- Gardiner, J. M., & Rowley, J. M. C. (1984). A generation effect with numbers rather than words. *Memory & Cognition*, *12*, 443-445.
- Gardner, H. E. (1999). Multiple approaches to understanding. In C. M. Reigeluth, *Instructional-design theories and models, Volume II: A new paradigm of instructional theory* (pp. 69-89). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. Palgrave Macmillan.
- Gibson, E. J. (1969). Principles of perceptual learning and development. *New York: Appleton-Century-Crofts*.
- Goldstein, E. B. (2002). *Sensation and perception (6th ed.)*. Pacific Grove, CA: Wadsworth.
- Goldstone, R. L., Schyns, P. G., & Medin, D. L. (1997). Learning to bridge between perception and cognition. In R. L. Goldstone, P. G. Schyns, & D. L. Medin (Eds.), *Psychology of Learning and Motivation: Perceptual Learning, Vol. 36*. (pp. 1-14). San Diego, CA: Academic Press.
- Grabowski, B. L. (2004). Generative learning contributions to the design of instruction and learning. In D. H. Jonassen (Eds.), *Handbook of research on educational*

- communications and technology* (pp. 719-744). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Grace, P.E. (2006). Full-page versus partial-page screen designs in Web-based training: Their effects on learner satisfaction and performance. Ph.D. dissertation. The University of South Florida, United States – Florida. Retrieved August 30, 2008, from ProQuest Dissertations and Theses database.
- Gredler, M. E. (2004). Games and simulations and their relationships to learning. In D. H. Jonassen (Eds.), *Handbook of research on educational communications and technology* (pp. 719-744). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Gunderman, R. B., Kang, Y. P., Fraley, R. E., & Williamson, K. B. (2001). Instructional technology and radiologic education. *Radiology*, 221(1), 1-4.
- Gunderman, R. & Chan, S. (2003). Knowledge sharing in radiology. *Radiology*, 229,314-317.
- Gunderman, R. B., Heitkamp, D. E., Kipfer, H. D., Frank, M. S., Jackson, V. P., & Williamson, K. B. (2003). 2003 AUR Joseph E. And Nancy O. Whitley Award. Developing tomorrow's academic radiologists: a 3-month residency elective in education. *Academic Radiology*, 10(6), 650-656.
- Gunderman, R. B., & Wood, B. P. (2004). Trusting the student: Learner-centered education. *Journal of the American College of Radiology: JACR*, 1(12), 897-900.
- Hagg, B. B. (1995). *The effect of visual manipulation strategies within computer-based instruction on various types of learning objectives*. Ph.D. dissertation, The Pennsylvania University. Retrieved July, 2009, from ProQuest Digital Dissertations database.

- Hattie, J. (2004). Meta-analysis: The process and interpretation. Retrieved August 1, 2008, from <http://www.cemcentre.org/Documents>.
- Hannum, W. H. (2007). When computers teach: A review of the instructional effectiveness of computers. *Educational Technology*, March-April, 5-13.
- Heinich, R., Molenda, M., Russell, J. D., & Smaldino, S. E. (1999). *Instructional media and technologies for learning* (6th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Hilbelink, A. J. (2007). The effectiveness and user perception of 3-dimensional digital human anatomy in an online undergraduate anatomy laboratory. Ph.D. dissertation. The University of South Florida, United States – Florida. Retrieved August 30, 2008, from ProQuest Dissertations and Theses database.
- Hirumi, A. (2002). A framework for analyzing, designing, and sequencing planned Elearning interactions. *Quarterly Review of Distance Education*, 3(2), 141-60.
- Hyun, J., Woodman, G. F., Vogel, E. K., Hollingworth, A., & Luck, S. J. (2009). The comparison of visual working memory representations with perceptual inputs. *Journal of Experimental Psychology*, 35(4), 1140-1160.
- IEEE Learning Technology Standards Committee (2002). Standard for learning object metadata. Retrieved at <http://ieeeltsc.org>.
- Jacoby, L. L., & Craik, F. I. M. (1979). Effects of elaboration of processing at encoding and retrieval: Trace distinctiveness and recovery of initial context. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 1-22). Mahwan, NJ: Erlbaum.

- Jacoby, L. L., Craik, F. I. M., & Begg, I. (1979). Effects of decision difficulty on recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, 18, 585-600.
- Javidi, G. (2005). A comparison of traditional physical laboratory and computer-simulated laboratory experiences in relation to engineering undergraduate students' conceptual understandings of a communication systems topic. Ph.D. dissertation. The University of South Florida, United States – Florida. Retrieved August 30, 2008, from ProQuest Dissertations and Theses database.
- Jonassen, D. H. (1985). Interactive Lesson Designs: A Taxonomy. *Educational Technology* (June), 7-17.
- Jonassen, D. (1999). Designing constructivist learning environments. In C. M. Reigeluth (Eds.), *Instructional–design theories and models (volume II)* (pp. 215-240). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Jonassen, D. H. & Henning, P. (1999). Mental models: knowledge in the head and knowledge in the world. *Educational Technology*, 39(3), 37-42.
- Jonassen, D. H. (2004). *Learning to solve problems: An instructional design guide*. San Francisco, Calif.: Pfeiffer.
- Jonassen, D. H., Strobel, J., & Gottdenker, J. (2005). Model building for conceptual change. *Interactive Learning Environments*, 13(1-2), 15-37.
- Johns, E. E., & Swanson, L. G. (1988). The generation effect with nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 180-190.
- Keller, F. S., & Schoenfeld, W. N. (1950). *Principles of psychology: A systematic text in the science of behavior*. New York: Appleton-Century-Crofts.

- Kennedy, G. E. (2004). Promoting cognition in multimedia interactivity research. *Journal of Interactive Learning Research*, 15(1), 43-61.
- Kennewell, S., Tanner, H., Jones, S., & Beauchamp, G. (2008). Analyzing the use of interactive technology to implement interactive teaching. *Journal of Computer Assisted Learning*, 24(1), 61-73.
- Kim, S. & Astion, M. (2003). Patterns of image comparison and contrast feature in Urinalysis Tutor™. *Computerized Medical Imaging and Graphics*, 27, 157-164.
- Kinjo, H., & Snodgrass, J. G. (2000). Does the generation effect occur with pictures? *American Journal of Psychology*, 113, 95–121.
- Knowles, M. S. (1990). *The adult learner: A neglected species*. 4th ed. Houston, TX: Gulf Publishing.
- Knowles, M. S. (1998). *The adult learner: The definitive classic in adult education and human resource development*. Houston, TX: Gulf Publishing.
- Kopans, D. B. (1992). The positive predictive value of mammography. *AJR*, 158, 521.
- Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Journal of Educational Technology Research and Development*, 42(2), 7–19.
- Krawiec, K. (2007). Generative learning of visual concepts using multiobjective genetic programming. *Pattern Recognition Letters*, 28(16), 2385-2400.
- Kulik, C.-L., & Kulik, J. (1986). Effectiveness of computer-based education in colleges. *AEDS Journal*, 19(2–3), 81–108.
- Kulik, C.-L., & Kulik, J. (1991). Effectiveness of computer-based instruction: an updated analysis. *Computers in Human Behavior*, 7, 75–94.

- Kulik, J. (2003). Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say. Arlington, VA: SRI International.
- Kundel, H. L., Nodine, C. F., & Carmody, D. (1978). Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative Radiology*, 13, 175-181.
- Kundel, H. L. (1981). Visual sampling and estimates of the location of information on chest films. *Investigative Radiology*, 9(2), 87-93.
- Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: gaze-tracking study. *Radiology*, 242(2), 396-402.
- Krupinski, E. A. (1996). Visual scanning patterns of radiologists searching mammograms. *Academic Radiology*, 3(2), 137-44.
- Lesgold, A. Feltovich, P. J., Glaser, R., & Wang, Y. (1981). *The acquisition of perceptual diagnostic skill in radiology* (Tech. Rep. No. PDS-1). Pittsburgh: University of Pittsburgh, Learning Research and Development Center.
- Lesgold, A., Rubinson, H., Feltovitch, P., Glaser, R., Klopfer, D., & Wang, Y., (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. Chi, R. Glaser, & M. Farr, M. (Eds.), *The nature of expertise*, Erlbaum, Hillsdale, NJ. 311-342.
- Levin, D. & Simon, D. (1997). Failure to detect changes in attended objects in motion pictures. *Psychonomic Bulletin and Review*, 4, 501-506.
- Lieberman, G., Abramson, R., Volkan, K., & McArdle, P. J. (2002). Tutor versus computer: a prospective comparison of interactive tutorial and computer-assisted instruction in radiology education. *Academic Radiology*, 9(1), 40-9.

- Liu, M. & Bera, S. (2005). An analysis of cognitive tool use patterns in a hypermedia learning environment. *Educational Technology Research & Development*, 53, 5-22.
- Luo, P., Eikman, A. E., Kealy, W., & Qian, W. (2006). Analysis of a mammography teaching program based on an affordance design model. *Academic Radiology*, 13, 1542-1552.
- Luo, P., Szabunio, M. M., & White, A. J. (2008, March). *Interactivity in technology-enhanced mammogram reading instruction*. Paper presented at the meeting of the American Educational Research Association, New York, NY.
- Mack, A., & Rock, I. (1998). Inattention blindness. *Cambridge, MA: MIT Press*.
- Mandler, J. M. & Orlich, F. (1993). Analogical transfer: The roles of schema abstraction and awareness. *Bulletin of the Psychonomic Society*, 31, 485-487.
- Markett, C., Sanchez, I. A., Weber, S., & Tangney, B. (2006). Using short message service to encourage interactivity in the classroom. *Computers and Education*, 46(3), 280-293.
- Matteson, S. R. (2002). Radiographic diagnosis: How do our minds work? *Texas Dental Journal*, 119(5), 392-4.
- Matthews, D. E., VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., et al. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, 31(1), 3-62.
- Mathews, R. C., Lane, I. M., Roussel, L. G., Nagy, M. S., Haptonstahl, D. E., and Brock, D. B. (1996). Using conscious reflection, group processes, and AI to facilitate development of expertise. *Journal of Experimental and Theoretical Artificial Intelligence*, 8, 259-76.

- Mathews, R. C., Roussel, L. G., & Cochran, B. P. (2001). The role of implicit learning in the acquisition of generative knowledge. *Cognitive Systems Research, 1*(1-4), 161-174.
- Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.
- Mayer, R. E., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology game. *Journal of Educational Psychology, 94*(1), 171-185.
- Mayer, R. E. (2005). *The Cambridge Handbook of Multimedia Learning*. Cambridge, New York: Cambridge University Press.
- McCann, R. S., Besner, D., & Davelaar, E. (1988). Word recognition and identification: Do word-frequency effects reflect lexical access? *Journal of Experimental Psychology: Human Perception and Performance, 14*, 693-706.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insight from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419-457.
- Mello-Thoms, C., & Chapman, B. (2004). A preliminary report on the role of spatial frequency analysis in the perception of breast cancers missed at mammography screening. *Academic Radiology, 11*(8), 894-908.
- Moore, M. G. (1989). Three types of interaction. *The American Journal of Distance Education, 3*(2), 1-6.
- Morrison, G. R., & and others. (1994). Promoting generative learning with elaboration training in computer-based instruction. *Proceedings of the Association for Educational Communication and Technology, USA, 16*, 547-557.

- Myles-Worsley, M., Johnston, W. A., & Simons, M. A., (1988). The influence of expertise on X-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 553–557.
- Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
- Nieveen, N. (2007). Formative evaluation in educational design research. In J. Van den Akker, B. Bannan, A. E. Kelly, N. Nieveen & T. Plomp (Eds.), *Educational design research* (pp. 89-102). Routledge, London.
- Nodine, C. F., Kundel, H. L., Mello-Thomas, C., Weinstein, S. P., Orel, S. G., Sullivan, D. C., Conant, E. F. (1999). How experience and training influence mammography expertise. *Academic Radiology*, 6(10), 575-85.
- Norman, G. R. Coblenz, C. L., Brooks, L. R., & Babcock, C. J. (1992). Expertise in visual diagnosis: A review of the literature. *Academic Medicine*, 67(10), S78-S83.
- Paivio, A. (1990). *Mental representations: A dual coding approach* (2nd ed.). New York: Oxford University Press.
- Peynircioglu, Z. F. (1989). The generation effect with pictures and nonsense figures. *Acta Psychologica*, 70(2), 153-160.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics*, 16(2), 283-290.
- Piaget, J. (1970). *Science of education and the psychology of the child* (D. Coltman Trans.). New York, Orion Press. (Original work published 1969)

- Pinar, W. F., Reynolds, W. M., Slattery, P., Taubman, P. M. (2000). *Understanding curriculum: An introduction to the study of historical and contemporary curriculum discourses*. New York: Peter Lang.
- Proske, A., Narciss, S., & Korndle, H. (2007). Interactivity and learners' achievement in Web-Based learning. *Journal of Interactive Learning Research*, 18(4), 511-531.
- Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 88-94.
- Reed, S. K. (2006). Cognitive architectures for multimedia learning. *Educational Psychologist*, 41(2), 87-98.
- Reeves, T. C., Herrington, J., & Oliver, R. (2005). Design research: A socially responsible approach to instructional technology research in higher education. *Journal of Computing in Higher Education*, 15(2), 97-116.
- Reigeluth, C. M. (1999). The elaboration theory: Guidance for scope and sequence decision. In C. M. Reigeluth, *Instructional-design theories and models, Volume II: A new paradigm of instructional theory* (pp. 425-453). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Reiner, B., & Siegel, E. (2008). The potential for gaming techniques in radiology education and practice. *Journal of the American College of Radiology*, 5(2), 110-4.
- Resnick, L. B. (1987). Learning in school and out. *Educational Researcher*, 16 (9), 13-20.
- Ridley, R. T. (2007). Interactive teaching: a concept analysis. *Journal of Nursing Education*, 46 (5): 203-9.

- Rieber, L. P. (1994). *Computers, graphics, and learning*. Madison, Wisconsin: WBC Brown & Benchmark.
- Roberts, C. C. & Chew, F. S. (2003). Teaching radiology residents, and radiology residents as teachers. *Academic Radiology*, 10(suppl 1): S97-S101.
- Roelofs, A. A. J., Karssemeijer, N., et. al. (2007). Importance of comparison of current and prior mammograms in breast cancer screening. *Radiology*, 242(1), 70-77.
- Rogers, E. (1992). Visual interaction: A link between perception and problem solving. Ph.D. dissertation. Georgia Institute of Technology, United States. Retrieved June 1, 2007, from ProQuest Dissertations database.
- Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- Ross, S. M. & Morrison, G. R. (2004). Experimental research methods. In D. H. Jonassen (Eds.), *Handbook of research on educational communications and technology* (pp. 1021-1043). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Roubidoux, M. (2005). A Breast cancer detective: a computer *game* to teach breast cancer screening to Native American patients. *Journal of Cancer Education*, 20(1 Suppl), 87-91.
- Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. In D. E. Meyer & S. Kornblum (Eds.). *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 3-30). Cambridge, MA: MIT Press.
- Saddik, A. E. (2001). *Interactive Multimedia Learning*. Berlin: Springer.

- Sanocki, T. (1991). Effects of early common features on form perception. *Perception & Psychophysics*, 50(5), 490-497.
- Sanocki, T. (1993). Time course of object recognition: Evidence for a global-to-local contingency. *Journal of Experimental Psychology: Human Perception and Performance*. 19, 878-898.
- Sanocki, T., Bowyer, K. W., & Heath, M. D. (1998). Are edges sufficient for object recognition? *Journal of Experimental Psychology: Human Perception and Performance*, 24(1), 340-349.
- Sanocki, T. (1999). Constructing structural descriptions. *Visual Cognition, Special issue: The neuroscience of perceptual integration*, 6(3-4), 299-318.
- Scarsbrook, A. F., Graham, R. N. J., & Perriss, R. W. (2006). Radiology education: a glimpse into the future. *Clinical radiology*, 61(8):640-648.
- Shaffer, K. (2005). Radiology education in the digital era. *Radiology*, 235: 359-360.
- Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, 16(4), 475-522.
- Schwartz, D. L., Martin, T., Nasir, N. (2005). Design for knowledge evolution: Towards a prescriptive theory for integrating first- and second-hand knowledge. In P. Gardenfors and P. Johansson (Eds.), *Cognition, education, and communication technology* (pp. 21-54). Mahwah, NJ: Lawrence Erlbaum.
- Schuwirth, L. W., Verheggen, M. M., van der Vleuten, C. P., Boshuizen, H. P., & Dinant, G. J. (2001). Do short cases elicit different thinking processes than factual knowledge questions do? *Med Educ*, 35, 348-356.

- Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, 16(4), 475-522.
- Scott-Brown, K. C., Baker, M. R., & Orbach, H. S. (2000). Comparison blindness. *Visual Cognition*, 7(1/2/3), 253-267.
- Seitz, A. & Watanabe, T. (2005). A unified model for perceptual learning. *Trends in Cognitive Sciences*, 9(7), 329-334.
- Selcer, B. A. (1993). Computer-assisted interactive radiology courseware. *JVME*, 20 (3). Retrieved September 2, 2006, from <http://scholar.lib.vt.edu/ejournals/JVME/V20-3/selcer.html>.
- Shapiro, K. L., Arnell, K. M., & Raymond, J. E. (1997). The attentional blink: A view on attention and glimpse on consciousness. *Trends in Cognitive Science*, 1, 291-296.
- Sharples, M. (1991). Computer-based tutoring of visual concepts: from novice to expert. *Journal of Computer Assisted Learning*, 7, 123-132.
- Slamecka, N. J., & Graf, P. (1978). Generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), 592-604.
- Smith, P. L. & Ragan, T. J. (1993). *Instructional design*. New York: Merrill.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74, 498.
- Spear, N. E., & Riccio, D. C. (1994). *Memory: phenomena and principles*. Boston: Allyn & Bacon. 1994.

- Soraci, S. A., Carlin, M. T., Chechile, R. A., Franks, J. J., Wills, T., & Watanabe, T. (1999). Encoding variability and cuing in generative processing. *Journal of Memory and Language*, 41, 541–559.
- Srinivasan, M., Wilkes, M., Stevenson, F., Nguyen, T., & Slavin, S. (2007). Comparing problem-based learning with case-based learning: effects of a major curricular shift at two institutions. *Academic Medicine*, 82(1), 74-82.
- Stones, E. (1979). *Psychopedagogy*. Methuen, London.
- Stull, A. T. & Mayer, R. E. (2007). Learning by doing versus learning by viewing: Three experimental comparison of learner-generated versus author-provided graphic organizers. *Journal of Educational Psychology*, 99(4), 808-820.
- Su, T. J., & Shaffer, K. (2004). Reinventing the apprenticeship: The hot seat in the digital era. *Academic Radiology*, 11(11), 1300-1307.
- Sun, R., Zhang, X., & Slusarz, P. (2007). The interaction of implicit learning, explicit hypothesis testing learning and implicit-to-explicit knowledge extraction. *Neural Networks*, 20(1), 34-47.
- Sweller, J. (1999). Instructional design in technical areas. *Australian Education Review*, 43. 49-53.
- Swensson, R. G., Hessel, S. J., & Herman, P. G. (1978). Detection before attention in skilled visual search. Paper presented at the annual meeting of the Psychonomic Society, November.
- Tachakra, S. & Dutton, D. (2000). Long-distance education in radiology via a clinical telemedicine system. *Telemedicine Journal and E-Health*, 6(3), 361-5.

- Terrell, S. & Rendulic, P. (1996). Using Computer-Managed Instructional Software to Increase Motivation and Achievement in Elementary School Children. *Journal of Research on Computing in Education*, 26(3), 403-414.
- Tennyson, R. D. & Park, O. (1980). The teaching of concepts: a review of the instructional design literature. *Review of Educational Research*, 50, 55-70.
- Thomas, E. L. & Lansdown, E. L. (1963). Visual search patterns of radiologists in training. *Radiology*, 81, 288-292.
- Tourassi, G. D. (1999). Journey toward computer-aided diagnosis: Role of image texture analysis. *Radiology*, 213, 317-320.
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition, Special issue: Visual search and attention*, 14(4-8).
- Tuddenham, W. J. & Calvert, W. P. (1961). Visual search patterns in roentgen diagnosis. *Radiology*, 76, 255-256.
- Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology*, 77, 593-601.
- Twitchell, D. G. (2001). *A rapid prototyping model for the design and development of instructional systems in theory and practice: A case study*. Unpublished doctoral dissertation. Utah State University, Brigham City, Utah.
- VanDeventer, S. S. & White, J. A. (2002). Expert behavior in children's video game play. *Simulation & Gaming*, 33(1), 28-48.
- Vygotsky, L. (1986). *Thought and language* (A. Kozulin, Trans.). Cambridge, Mass.: MIT Press. (Original work published 1934)

- Wagner, E. D. (1994). In support of a functional definition of interaction. *The American Journal of Distance Education*, 8(2), 6-26.
- Wartman, S. A. (2004). Revising the idea of a national center for health professions education research. *Academic Medicine*, 79(10), 910-917.
- Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology*, 18, 158-194.
- Wattenmaker, W. D. (1999). The influence of prior knowledge in intentional versus incidental concept learning. *Memory & Cognition*, 27(4), 685-98.
- Whiteside, H. C. (1987). Can multi-image presentations be affective and effective in education? In Braden R. A., Beauchamp, D. G., & Miller, L. W. (eds) *Visible and viable: The role of images in instruction and communication*. International Visual Literacy Association, Commerce, TX.
- Wileman, R. E. (1993). *Visual communicating*. Englewood Cliffs, N.J.: Educational Technology Publications.
- Williamson, K. B., Gunderman, R. B., Cohen, M. D., & Frank, M. S. (2004). Learning theory in radiology education. *Radiology*, 233(1), 15-18.
- Wills, T. W., Soraci, S. A., Chechile, R. A., & Taylor, H. A. (2000). ‘‘Aha’’ effects in the generation of pictures. *Memory & Cognition*, 28, 939–948.
- Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist*, 19(2), 87-95.
- Wittrock, M. C. (1990). Generative processes of comprehension. *Educational Psychologist*, 24, 345-376.

- Wittrock, M. C. (1992). Generative learning processes of the brain. *Educational Psychologist, Fal 1992. Special issue: Brain and education*. 27(4), 531-541.
- Wood, B. P. (1999). Visual expertise. *Radiology*, 211, 1-3.
- Zaidel, M., Hopper, K., & Lyriboz, T. (1999). Interactive Web-based radiology teaching file. *Journal of Digital Imaging*, 12(2), Suppl 1, 203-4.
- Zhu, L. (2006). *Questions and feedback: Effects of direct manipulation and animation in facilitating student achievement on tests measuring different educational objectives*. Ph.D. dissertation, The Pennsylvania University. Retrieved June, 2009, from ProQuest Digital Dissertations database.

Appendices

Appendix A IRB Approval



DIVISION OF RESEARCH INTEGRITY AND COMPLIANCE

Institutional Review Boards, FWA No. 00001669
11901 Bruce B. Downs Blvd., MDC035 • Tampa, FL 33612-4799
(813) 974-5638 • FAX (813) 974-5618

July 1, 2009

Ping Luo
Secondary Education
14031 Halstead Ct., #316
Tampa, FL 33613

RE: **Exempt Certification** for IRB#: 108121 G
Title: *Evaluation of the Flicker Effect as a Generative Strategy in Learning to Evaluate Complex Radiographic Images*

Dear Ping Luo:

On June 30, 2009, the Institutional Review Board (IRB) determined that your research **meets USF requirements and Federal Exemption criteria one (1)**. It is your responsibility to ensure that this research is conducted in a manner reported in your application and consistent with the ethical principles outlined in the Belmont Report and with USF IRB policies and procedures.

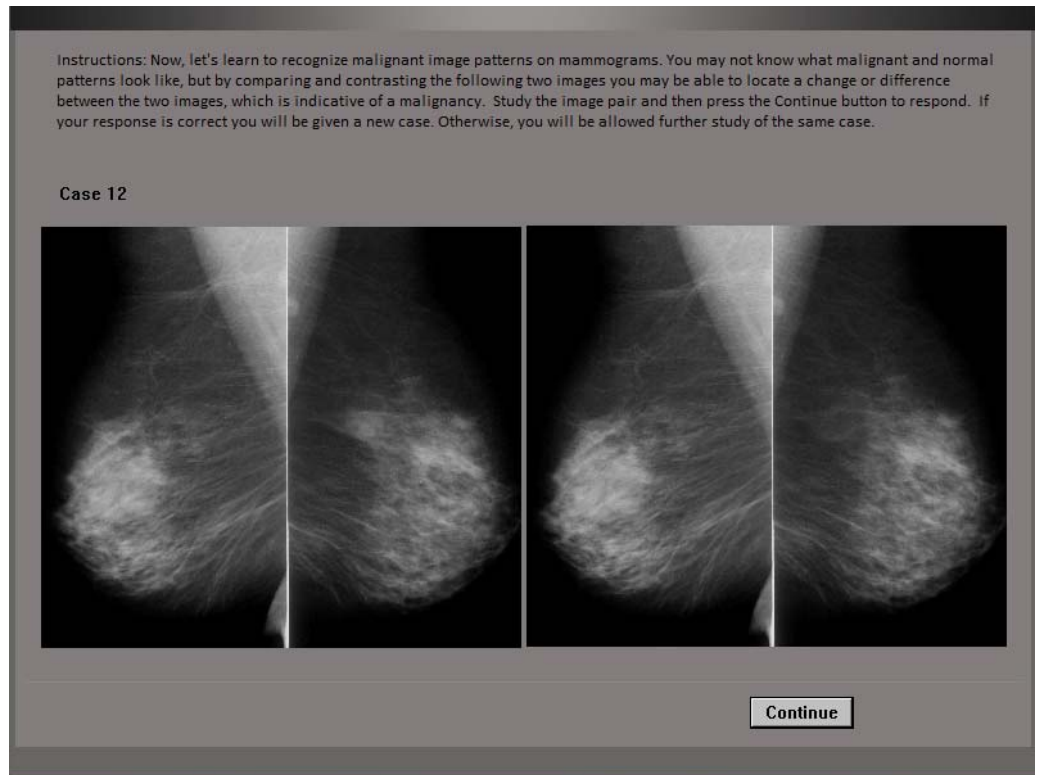
Please note that changes to this protocol may disqualify it from exempt status. It is your responsibility to notify the IRB prior to implementing any changes.

The Division of Research Integrity and Compliance will hold your exemption application for a period of five years from the date of this letter or for three years after a Final Progress Report is received. If you wish to continue this protocol beyond those periods, you will need to submit an Exemption Certification Request form at least 30 days before this exempt certification ends. If a Final Progress Report has not been received, the IRB will send you a reminder notice prior to end of the five year period; therefore, it is important that you keep your contact information current with the IRB Office. Should you complete this study prior to the end of the five-year period, you must submit a Final IRB Progress Report for review.

Please reference the above IRB protocol number in all correspondence regarding this protocol with the IRB or the Division of Research Integrity and Compliance. In addition, you can find the Institutional Review Board (IRB) Quick Reference Guide providing guidelines and resources to assist you in meeting your responsibilities in the conduction of human participant research on our website. Please read this guide carefully. It is your responsibility to conduct this study in accordance with IRB policies and procedures and as approved by the IRB.

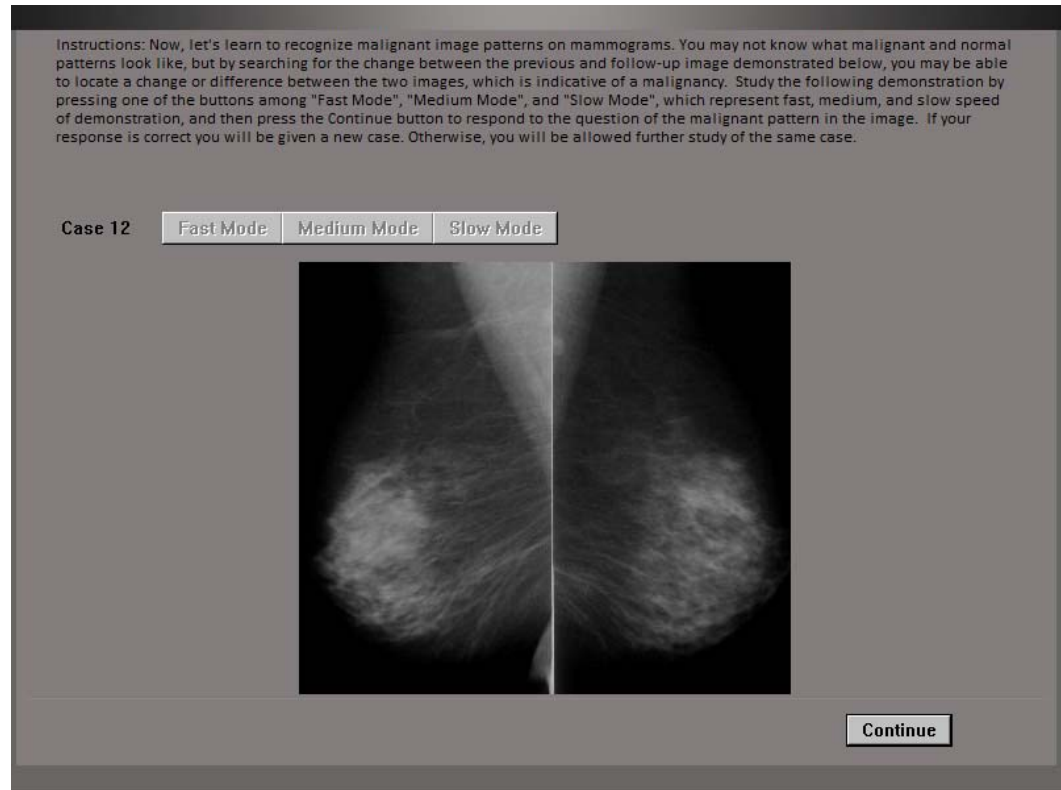
Appendix B A Screenshot of the Comparison Treatment

Figure B1 An Instructional Screen of the Comparison Treatment



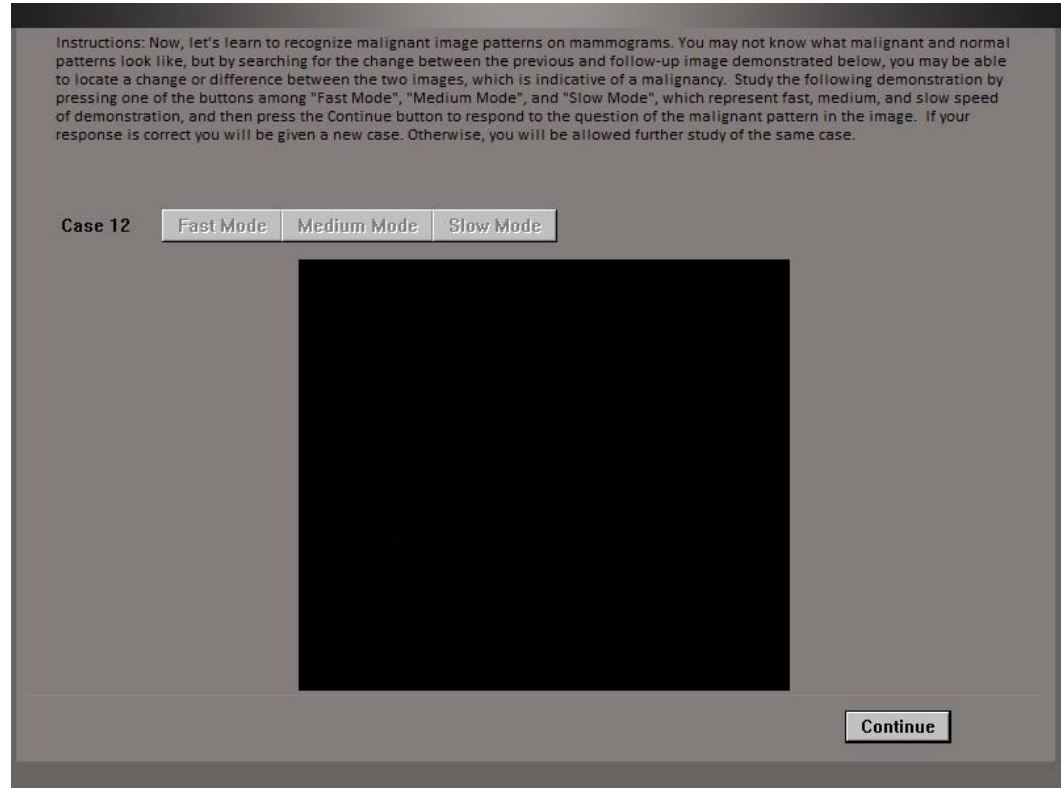
Appendix C Screenshots of the Flicker Treatment

Figure C1 The First Screen of a Case Study in the Flicker Treatment



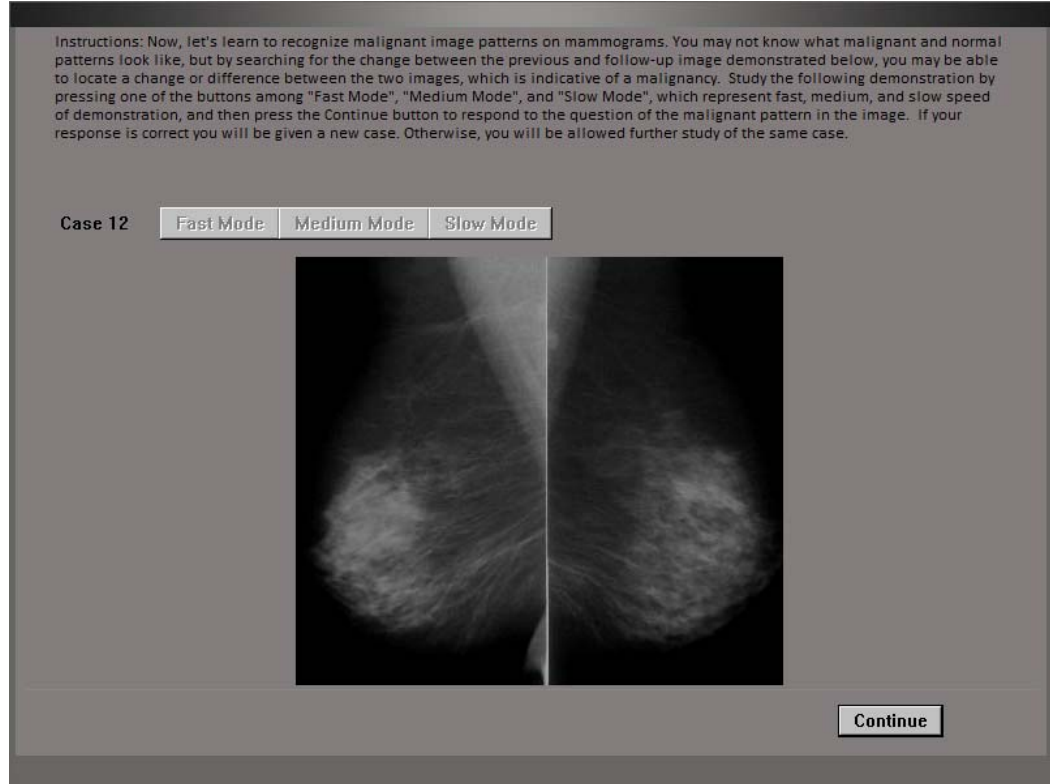
Appendix C (Continued)

Figure C2 The Second Screen of a Case Study in the Flicker Treatment



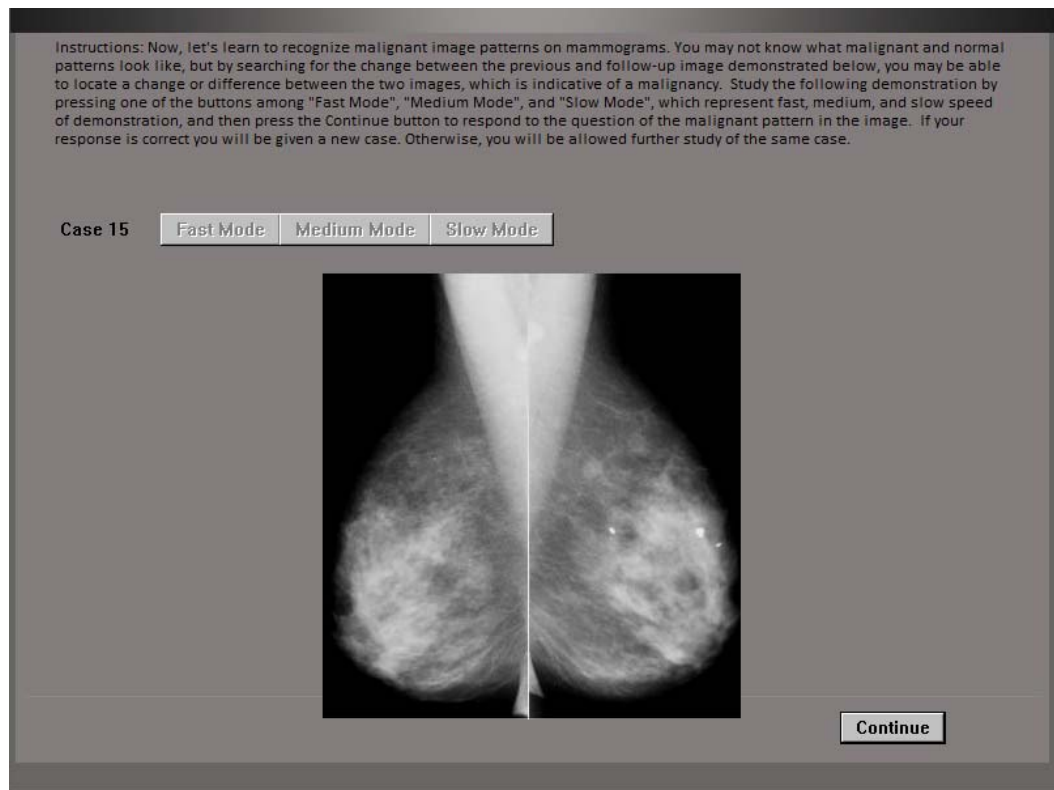
Appendix C (Continued)

Figure C3 The Third Screen of a Case Study in the Flicker Treatment



Appendix D Screenshots of the No-Flicker Treatment

Figure D1 The First Screen of a Case Study in the No-Flicker Treatment

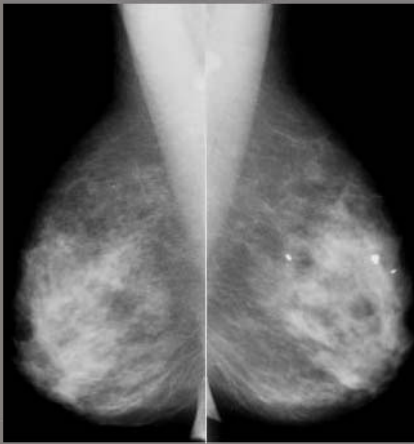


Appendix D (Continued)

Figure D2 The Second Screen of a Case Study in the No-Flicker Treatment

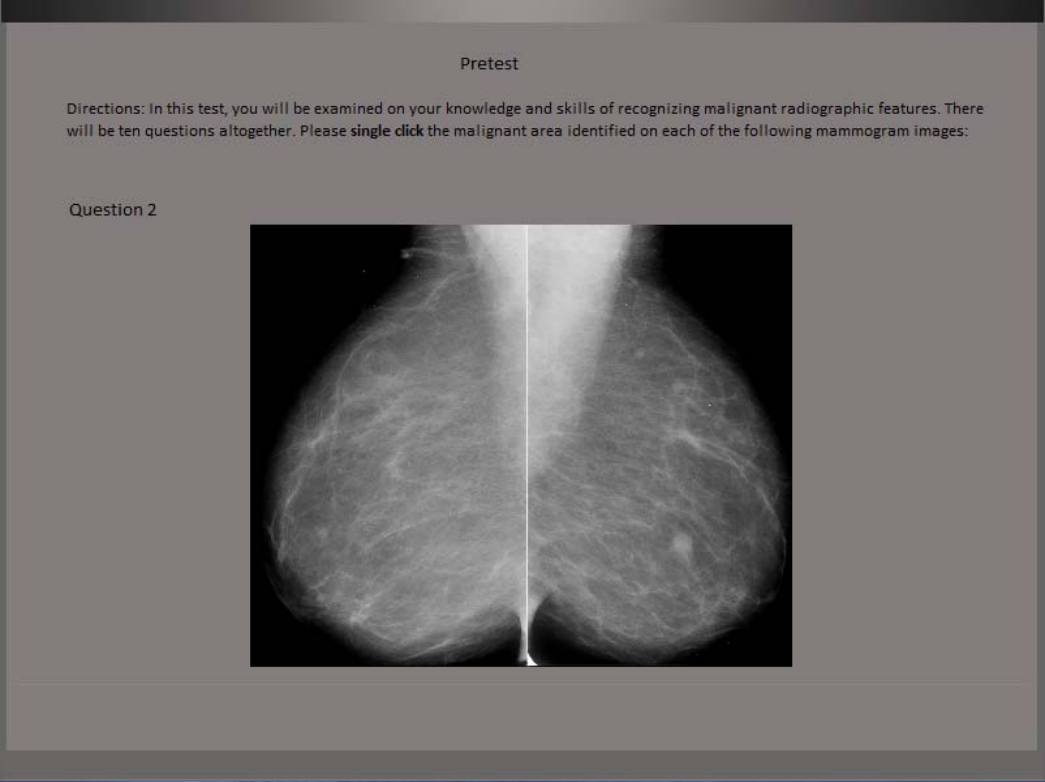
Instructions: Now, let's learn to recognize malignant image patterns on mammograms. You may not know what malignant and normal patterns look like, but by searching for the change between the previous and follow-up image demonstrated below, you may be able to locate a change or difference between the two images, which is indicative of a malignancy. Study the following demonstration by pressing one of the buttons among "Fast Mode", "Medium Mode", and "Slow Mode", which represent fast, medium, and slow speed of demonstration, and then press the Continue button to respond to the question of the malignant pattern in the image. If your response is correct you will be given a new case. Otherwise, you will be allowed further study of the same case.

Case 15



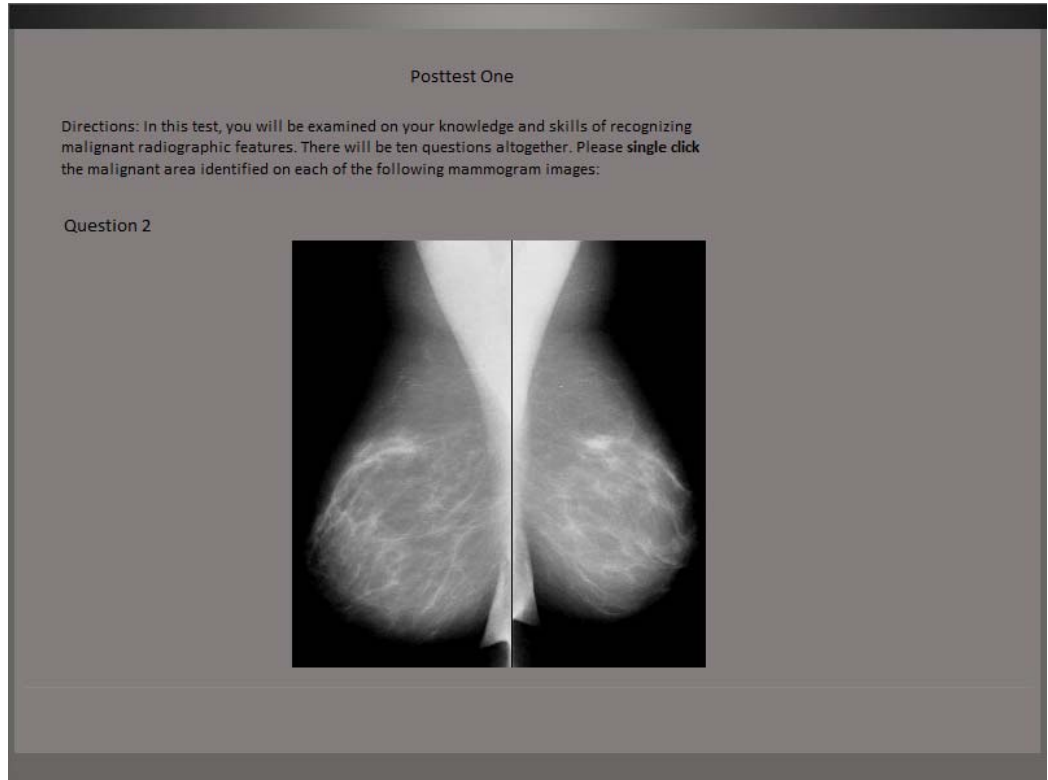
Appendix E A Screenshot of a Test Item in the Pretest

Figure E1 A Screenshot of a Test Item in the Pretest



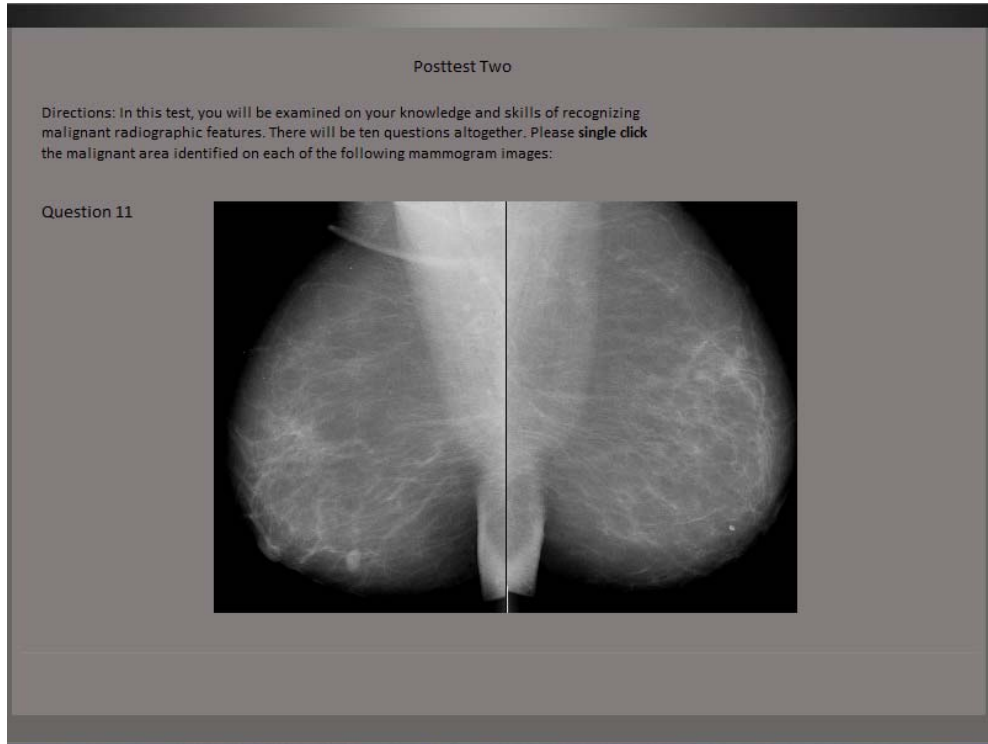
Appendix F A Screenshot of a Test Item in the Recognition Test

Figure F1 A Screenshot of a Test Item in the Recognition Test



Appendix G A Screenshot of a Test Item in the Classification Test

Figure G1 A Screenshot of a Test Item in the Classification Test



Appendix H A Recruitment Flyer

Secondary Education University of South Florida

Participants Needed for Research in Instructional Technology

If you have little knowledge and experience in interpreting radiographic images and have basic computer skills, you are invited to our computer-based research study that received an exemption certificate from the University Institutional Review Boards (IRB).

The purpose of the study is to examine how to integrate technology into higher education. In the study, you will experience a pretest, a study session, and two posttests. It will take you about less than half an hour to complete the entire session. Your participation in this study will be anonymous and voluntary. The study will be conducted at your convenience time. In appreciation of your participation, you will receive compensations.

For more information about the study, or to volunteer for the study, please contact your professor or Ping Luo at
813-343-0966
pluo@mail.usf.edu

Thank you for your voluntary participation!

Appendix I Evaluation Instruments

Table I1 The evaluation instrument for the subject matter expert (adapted from the evaluation instrument developed by Elissavet & Economides, 2003)

Items	Comments and Suggestions
1. Is the information in the instructional materials characteristic of sufficient scope and depth for one study session for naïve learners?	
2. In general, is it all right to say that the cases in these materials are arranged with increasing complexity?	
3. Do the instructional strategies used in the materials have potential to foster learning among naïve learners?	
4. Can the instructional materials be used by learners alone and/or blended with other types of learning materials	

Appendix I (Continued)

Table I2 The evaluation instrument for the instructional technology expert
(adapted from the evaluation instrument developed by Elissavet & Economides, 2003)

Items	Comments and Suggestions
1. Is the design of the materials based on reliable learning and instructional theories?	
2. Is the content structured in a clear and understandable manner?	
3. Does the interactivity of the materials foster learning?	
4. Does the program provide opportunities for interaction at least every three or four screens?	
5. Is the content chunked into small segments?	
6. Does the program provide feedback immediately after a response?	
7. Does the program provide feedback to verify the correctness of a response?	
8. Are screens designed in a clear and understandable manner?	
9. Can the presentation of information captivate the attention of students?	
10. Does the design overload students' memory?	
11. Are the principles of screen design followed?	
12. Are the texts in the materials grammatically correct?	
13. Are screens designed in a clear and understandable manner?	

-
14. Can the presentation of information captivate the attention of students?
 15. Does the design does overload students' memory?
 16. Is the use of space according to the principles of screen design?
 17. Does the design use proper fonts in terms of style and size?
 18. Does the use of text follow the principles of readability?
 19. Does the color of the text follow the principles of readability?
 20. Is the number of colors in each screen no more than six?
 21. Is there consistency in the functional use of colors?
 22. Is the quality of the images and graphics good?
 23. Are reasonable contrasts between graphics and background retained?
-

Appendix I (Continued)

Table I3 Usability Test Survey

This evaluation survey is adapted from Elissavet & Economides (2003). Please circle the number representing your opinions about the computer-based instruction that you experienced. In the scales below, “1” represents the lowest level and “5” stands for the highest level:

Items	Scales
1. The program is easy to learn	1 2 3 4 5
2. The program is efficient.	1 2 3 4 5
3. The structure of the program is comprehensive	1 2 3 4 5
4. The program is simple and consistent in its operation	1 2 3 4 5
5. Overall impression of the program	1 2 3 4 5

Appendix J A Demographic Survey

Instruction: This survey is anonymous and conducted for the purpose of research. Please answer the following questions about yourself to the best of your ability by circling the appropriate response:

1. Your gender is

- a. Male b. Female

2. Your age is

- a. 15-25 b. 26-35 c. 36-45 d. 46-55 e. 56-65 f. > 65

3. Your ethnicity is

- a. White b. Black c. Hispanic d. Asian e. Other: _____

4. Your current educational program is

- a. Undergraduate b. Graduate c. Other: _____

5. Your current major is _____

About the Author

Ping Luo initiated research projects to promote learning in visual recognition related areas, especially medical images. With her major professor's guidance, she designed and developed three computer-based software programs helping novices learn visual patterns.

With her enthusiasm in learning and research, she devoted her time studying organization of information, instructional design, psychology, and technology affordances at University of South Florida. She received her M.A. in library and information science and Ph.D. in curriculum and instruction with an emphasis in instructional technology. Throughout the years of her graduate studies, she gained awards to support her research in instructional technology, pedagogy, and learning.

Ping holds her B.A. in English literature at Wuhan University, China. She also has about ten years of teaching experience at Huazhong University of Science and Technology, China and University of South Florida, the United States.