RAPID DYNAMIC ASSESSMENT OF EXPERTISE: A COMPARISON OF PERFORMANCE AND MENTAL EFFICIENCY MEASURES IN ACCORDANCE WITH COGNITIVE LOAD THEORY

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE
UNIVERSITY OF MINNESOTA
BY

David Lloyd Johnson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Simon R. Hooper, Advisor

DECEMBER, 2007

ACKNOWLEDGEMENTS

The fear of the Lord is the beginning of knowledge ….[Proverbs 1:7]

But thanks be to God! He gives us the victory through our Lord Jesus Christ. [I Corinthians 15:57]

First, I would like to thank God for giving me the opportunity to research and write this dissertation and to thank Jesus Christ for being my Savior and Lord.

I would like to thank my advisor, Dr. Simon Hooper, for his patience and time and for sticking with me to the end.

I would like to thank my committee members, Dr. Judith Lambrecht, Dr. Aaron Doering, and Dr. Michael Harwell for their guidance and support.

I would like to thank Dr. Slava Kalyuga for his generosity providing me with information about his studies.

I would like to thank my brothers, Rick and Russ, and sister-in-law Sue for their support and my eight-year-old niece Kirsten, who is the joy of my life.

Finally, I would like to thank my parents, Lloyd and Yvonne Johnson, for their steadfast love and support and for providing me with excellent role models all of my life.

ABSTRACT

The purpose of this study was to determine whether the use of performance measures for instructional adaptation were more effective and efficient than the use of mental efficiency measures. Fifty-three undergraduate accounting students were randomly assigned to a performance group, a mental efficiency group, and a non-adapted control group. Participants were administered an initial diagnostic test, were placed in a training session about accounting cost-volume-profit analysis, and were administered a final diagnostic test, similar to the initial diagnostic test, and a mental effort rating of the training session.

Performance group participants were placed in the training session and allowed to skip certain training session stages based on the results of rapid verification tests administered during the initial diagnostic test. Mental efficiency group participants were placed in the training session and allowed to skip certain training session stages based on the results of rapid verification tests and mental effort ratings administered during the initial diagnostic test. The non-adapted control group participants were placed in the training session at the beginning and did not skip any stages.

The training session consisted of four difficulty levels, each with five stages. At each stage a faded worked example and a faded completion problem were provided and a rapid verification test and a mental effort rating were administered. Performance group participants advanced to the next stage or repeated the current stage based on the results of the rapid verification test. Mental efficiency group participants advanced to the next

stage or repeated the stage based on the results of the rapid verification test and mental effort ratings. The non-adapted control group did not repeat any training session stages.

The study produced no significant differences between any treatment groups for instructional time, final diagnostic test score, mental effort rating of the training session, or instructional efficiency (final diagnostic test score divided by mental effort rating of the training session). The author speculated that the non-significant results of the study were attributable to either an insufficient training session length or to the use of faded completion problems rather than conventional problems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

"The principal payoff from application of computer technology in instruction may come from their capabilities to tailor highly interactive environments to the needs of each individual learner."

Fletcher, *Learning From Technology-Assisted Instruction*, (2003)

Introduction and Statement of Purpose

According to Fletcher (2003) a third revolution in instruction may be underway, following the invention of writing and the printing press. This revolution is made possible by the rapid development of computer technology, such as computer-based instruction, interactive multimedia instruction, "intelligent" tutor systems, networked tutorial simulations, and web-based instruction. Fletcher contends that computer technology can offer individualized instruction, "designed to adapt and respond to the needs and intentions of individual learners" (p. 81), much like one-on-one tutoring. According to Bloom (1984), the capability to interact with learners and provide individualized instruction can raise achievement two standard deviations, the equivalent of moving from the 50^{th} to the 98^{th} percentile. The third revolution will substitute the capital of computer technology for the labor of human teachers.

Certainly hyperbole has accompanied all technological innovations of the past, and today's innovations are no different (Tyack & Cuban, 2000). However, recent research in cognitive load theory (CLT) has investigated methods of adapting and responding to the needs of individual learners (Clark, Nguyen, & Sweller, 2006;

Kirschner, Sweller, & Clark, 2006). CLT offers principles and methods to design and deliver efficient instructional environments that best utilize the limited capacity of an individual's working memory (Sweller, van Merriënboer, & Paas, 1998). CLT research suggests that instructional methods should change based upon the cognitive load imposed on working memory.

Instructional methods effective for novices are not necessarily effective for advanced learners (Kalyuga, Ayres, Chandler, & Sweller, 2003). CLT assumes a limited capacity working memory that includes partially independent subcomponents for auditory/verbal information and visual information and assumes an unlimited capacity long-term memory holding schemas that vary in their degree of automation. Because novices lack organized and automated schematic knowledge, novel information can overwhelm working memory. Therefore, worked examples provide scaffolding for novices to overcome working memory capacity constraints. Worked examples provide a step-by-step solution to a problem (Sweller et al., 1998). In contrast, advanced learners possess some organized and partially automated schematic knowledge, which frees up working memory capacity. Therefore, conventional problems provide opportunities for advanced learners to construct and automate schemas. Conventional problems require the learner to provide all of the solution steps to a problem (van Merriënboer, 1997). Both worked examples and conventional problems are discussed in more detail next.

According to CLT, worked examples provide scaffolding for novices, foster a forward-directed problem solution, and prevent learners from employing a means-ends strategy, to which novices typically resort when attempting to solve conventional problems (Cooper & Sweller, 1987; Sweller & Cooper, 1985). Means-ends problem solving involves comparing a problem's goal state with its current state and through a series of iterations, progressively eliminating the differences. Such a backwards-oriented problem solution strategy is not an effective way to construct schemas (Sweller et al., 1998). Worked examples alleviate the need for novices to hold solution steps in working memory and allow novices to begin constructing a forward-directed problem solution.

However, according to the expertise-reversal effect, worked examples are not as effective as conventional problems for advanced learners. The scaffolding provided to novices by worked examples is redundant for advanced learners, requiring advanced learners to integrate the redundant information with existing knowledge in long-term memory (Kalyuga et al., 2003). Because advanced learners possess some organized and partially automated schematic knowledge, conventional problems provide schema strengthening activities for advanced learners, foster linking of new with existing knowledge, and promote principle elaboration for application to other problems. Conventional problems offer opportunities for advanced learners to retrieve knowledge and solve novel problems.

If worked examples are effective for novices, and conventional problems are effective for advanced learners, then some method of assessing expertise is required. According to CLT, evidence of expertise is the learner's ability to apply domain knowledge to solve a problem with little mental effort (or with mental effort comfortably within working memory capacity). Mental effort represents the cognitive effort the learner directs to a task. To measure the ability to apply domain knowledge to solve problems, researchers have developed performance measures, including constructed response tests (open-ended problems), selected response tests (e.g. multiple choice or true-false questions), interviews, think aloud procedures, and rapid assessments (quick ways to measure performance in real-time) (Kalyuga, 2006b). To measure mental effort, researchers have developed seven or nine-point Likert scales (Paas, 1992).

Performance, mental effort, and expertise are interrelated. A learner with low expertise must invest high mental effort to achieve a given level of performance. A learner with high expertise can invest low mental effort to achieve the same level of performance. Using performance measures alone may not provided reliable inferences about expertise (Sweller et al., 1998). Two learners may achieve the same performance, but one may have invested much less mental effort and therefore possess higher expertise. Similarly, using mental effort measures alone may not provide reliable inferences about expertise either, because learners may merely indicate their level of self-confidence or comfort level rather than their mental effort (Kalyuga, 2006b). Therefore, CLT researchers have developed a dual measurement approach, called mental efficiency, which combines the performance and mental effort measures (Paas, 1992; Paas & van Merriënboer, 1993).

Although researchers believe that measuring expertise with mental efficiency is a better adapting mechanism than measuring expertise with performance alone, the research has not been able to consistently prove that this is the case. Some researchers (Camp, Paas, Rikers, & van Merriënboer, 2001; Salden, Paas, Broers, & van Merriënboer, 2004) have observed that measuring performance alone is a better adapting mechanism. Experimental results indicated that adapting instructional methods with performance measures achieved higher instructional efficiency than adapting instructional methods with mental efficiency measures. Other researchers (Kalyuga, 2006a; Kalyuga & Sweller, 2005) have observed the reverse, that measuring mental efficiency is a better adapting mechanism than measuring performance alone. Experimental results indicated that adapting instructional methods with mental efficiency measures achieved higher instructional efficiency than adapting instructional methods with performance measures. Therefore additional research is required.

The purpose of the experimental study described herein is to test the theory that instructional methods that change dynamically based upon a measure of mental efficiency, a combined measure of task performance and mental effort, will reduce cognitive load during instruction and produce higher learner knowledge gains and more efficient instruction than a measure of task performance alone.

Rationale for the Study

The purpose of this section is to describe in greater detail the prior research that has led the author to propose the study's research question. This section contains discussions of: 1) the elements of cognitive load and how each promotes or impedes learning, 2) the ways that certain instructional methods increase or decrease cognitive load, 3) the methods used to measure cognitive load and adapt instructional methods accordingly, 4) the mixed results that researchers have obtained with the mental efficiency measure, which seems to contradict some elements of CLT, and 5) the need for more research to investigate the use of the mental efficiency measure.

Cognitive Load Theory and Learning

According to CLT, the aim of all instruction is to alter long-term memory (Kirschner et al., 2006), and learning may be inhibited if the learner devotes working memory resources to activities that do not directly relate to schema construction and automation (Tuovinen & Sweller, 1999). CLT assumes a limited working memory and an unlimited long-term memory containing schemas that vary in degree of automation. A schema, also called a mental model (Clark et al., 2006), is a memory structure located in long-term memory that is the basis for expertise. A schema categorizes elements of information according to the manner in which it will be used. Schema construction and automation help alleviate working memory capacity constraints (Sweller et al., 1998). Schema construction involves chunking schemas in long-term memory as single elements in working memory. Automation involves processing long-term memory schemas with few working memory resources. Therefore, instructional methods should promote schema construction and automation and accommodate working memory capacity constraints.

Because of working memory capacity constraints, CLT focuses on the work instructional methods impose on working memory. The work imposed on working memory is called cognitive load and consists of three elements: intrinsic load, extraneous load, and germane load:

Intrinsic cognitive load. Intrinsic cognitive load is work imposed on working memory because of content complexity, or in CLT terms, the level of element interactivity of the content. Element interactivity consists of the number of elements that must be processed simultaneously. An example of high element interactivity is speaking or reading a sentence, which requires understanding the meaning of individual words, grammatical rules, and sentence structure simultaneously. An example of low element interactivity is learning vocabulary word definitions, which can be learned in isolation. For learners with low prior knowledge, content with high element interactivity will cause high intrinsic cognitive load, which will impede learning (Sweller et al., 1998).

Extraneous cognitive load. Extraneous cognitive load is work imposed on working memory that does not contribute to learning. Proper instructional design can reduce extraneous cognitive load. For example, if the instructions to a diagram appear on one page, and the diagram appears on another page, a split attention effect occurs, which imposes high extraneous cognitive load. The learner must store information in working memory when switching from the page containing the instructions to the page containing the diagram. Therefore, the instructions should be integrated with the diagram on a single page (Sweller et al., 1998).

Germane cognitive load. Germane cognitive load is work imposed on working memory that contributes to learning by constructing or automating schemas. Ways to increase germane cognitive load include providing example variability and practice with varied examples (Sweller et al., 1998). The varied examples allow learners to apply a solution schema to problems that contain different surface features.

Cognitive Load Theory and Instructional Methods

Instructional methods impose different cognitive loads on learners. Novices benefit from the scaffolding provided by worked examples. Since novices lack solution schemas in long-term memory, novel information can overwhelm working memory capacity. Therefore, novice learners require a worked example to display the solution steps, so novices can offload the information from memory. Advanced learners benefit from the schema strengthening opportunities provided by conventional problems, which require the learner to provide all of the solution steps to the problem. Since advanced learners can retrieve solution schemas from long-term memory, excess working memory is available, and advanced learners may link novel information with existing schematic knowledge.

In domains with high element interactivity, such as algebra, geometry, and computer programming, Sweller and Cooper (1985) and van Merriënboer (1990) demonstrated that learners achieve higher test scores with less training time by studying worked examples or performing completion problems than by solving conventional problems. Completion problems are problems that provide some of the steps of a worked example and require the student to fill in the missing steps. Completion problems help alleviate the disadvantage of worked examples, which do not force careful study by learners (Sweller et al., 1998). Worked examples and completion problems reduce extraneous cognitive load because all solution steps are visible to the learner and not required to be held in memory (Sweller et al., 1998).

Kalyuga, Chandler, and Sweller (1998) subsequently observed that although worked examples are more effective than conventional problems for novices, the reverse is true for advanced learners. In fact, worked examples can constrain the learning of advanced learners. The expertise reversal effect occurs because the added detail provided to novices is redundant to experts, who already possess schematic knowledge in long-term memory and must correlate the redundant information with existing schematic knowledge (Kalyuga et al., 2003). Due to this expertise reversal effect, Kalyuga et al. (2003) recommend that worked examples be provided to novices and conventional problems be provided to advanced learners.

Renkl et al.(2002) recommend guidance fading as learners gain expertise to provide a transition from worked examples to conventional problems. Guidance fading progressively omits steps from worked examples until all steps are omitted, and the worked example becomes a conventional problem. Renkl et al.(2002) compared forward fading (eliminating solution steps from the beginning to the end), backward fading (eliminating solution steps from the end to the beginning), and conventional problem strategies. Results indicated that both fading methods produced better test results than conventional problems, and that backward fading produced lower cognitive load than forward fading. Therefore, as learners gain expertise, they should progressively solve more worked example steps through backward fading until they gain sufficient expertise to solve conventional problems.

Adapting Instructional Methods to Learner Expertise

Because worked examples are more effective for novices, and conventional problems are more effective for advanced learners, researchers have recently recommended adapting instruction to learner expertise (Clark et al., 2006; Kalyuga, 2006b; Kalyuga & Sweller, 2005; van Merriënboer & Ayres, 2005). However, adaptive learning models and aptitude by treatment interactions (ATI) were studied extensively in the past. Aptitude treatment interactions occur when certain instructional treatments (methods) produce different results in individuals because of their different aptitudes (prior knowledge, mental abilities, personalities, etc.). Park and Lee (2004) categorize past adaptive instruction research into: 1) macro adaptive instructional models that change pedagogy over an entire course (e.g. measuring elementary student reading proficiency over the semester and changing reading programs when necessary), 2) aptitude by treatment interaction (ATI) models that change instructional methods to student aptitudes (e.g. measuring student visualizer/verbalizer cognitive style and providing diagrams for visualizers and text for verbalizers), and 3) micro adaptive instructional models that change content presentation sequence to student knowledge (e.g. measuring student performance and repeating instruction when necessary).

CLT extends the past research on adaptive instruction by adding measures of mental effort to measures of prior knowledge, one of the most significant individual difference predictors (Jonassen & Grabowski, 1993). CLT researchers use a micro adaptive model to adapt instruction, in which the instructional method (worked example, completion problem, and conventional problem) is adapted to learner expertise using an assessment that combines a measure of task performance and mental effort. Thus, although CLT researchers adapt instruction based on prior knowledge (measured by task performance), which was used to adapt instruction in the past, CLT researchers have added the following to the research base: 1) use of mental effort measures to adapt instruction and 2) prescription of instructional methods, such as worked examples, completion problems, and conventional problems, to match learner prior knowledge and mental effort levels. (Kalyuga, 2006b).

Assessing Expertise

To adapt instruction, CLT researchers assess expertise by measuring performance and mental effort. Paas (1992) and Paas and van Merriënboer (1993) coined the term "mental efficiency" as the combined measure of task performance and mental effort. To measure performance, researchers use open-ended problems, think aloud procedures, interviews, and rapid assessments discussed below. To measure mental effort, researchers use a nine-point Lickert scale (Paas, 1992). A learner who performs well with low mental effort on an assessment should possess more expertise and more highly developed/automated schemas than a learner who performs well with high mental effort. The latter would indicate that the learner possesses less highly developed and automated schemas.

Rapid Assessment Techniques

To measure performance in real-time, Kalyuga and Sweller (2005) propose using rapid performance assessments, quick ways to measure performance and adapt instruction. Kalyuga and Sweller have investigated two rapid performance assessments:

the first-step method and the verification method (Kalyuga, 2006a, 2006b, 2006c, 2006d; Kalyuga & Sweller, 2005).

First-step method. The first-step method is a rapid assessment test that attempts to assess long-term schematic knowledge in working memory. Learners are instructed to provide their first-step solution to a problem. Answers that indicate step-skipping are indicative of high level schemas and high expertise (Kalyuga, 2006b).

Verification method. The verification method quickly measures performance in a computer environment. The verification method has been used for problems not suitable for the first-step method because of technological challenges. Examples include problems that require drawing diagrams, employ multiple solution paths, etc. Using the verification method, learners are presented with a task, which is subsequently removed from view, and learners are presented with a solution step. The learner is asked to verify from memory whether the solution step is correct. Each solution step presented is cumulative, asking students to progressively verify more steps until the final answer is reached. The solution steps are randomly determined in advance to be half correct and half incorrect, (Kalyuga, 2006b).

Adapting Instruction with Mental Efficiency Measures

CLT researchers have measured the effectiveness of various adaptive methods using instructional efficiency. Instructional efficiency is a technique to compare instructional methods by combining mental effort with performance measures. Raw data is converted into z-scores and placed in a graph, where the y-axis is performance, and the x-axis is mental effort. Instructional efficiency is the distance from the point on the graph to a diagonal line at which performance is equal to mental efficiency. High instructional efficiency results when performance exceeds mental effort (Sweller et al., 1998) and the point is above the diagonal line.

Recent CLT research has compared adapting instruction using mental efficiency measures (combining performance and mental effort) and performance measures alone. (Mental efficiency is used for real-time adaptation and should not be confused with instructional efficiency, which is used after the conclusion of the experiment to compare adaptive method results.) Research on the effectiveness and efficiency of the mental efficiency measure has been mixed. Camp, Paas, Rikers, and van Merriënboer (2001) and Salden, Paas, Broers, and van Merriënboer (2004) found that using mental efficiency to adapt instruction was more efficient than not adapting instruction, but that mental efficiency was no more effective or efficient than adapting with a measure of performance alone. In the Camp et al. study using a measure of performance alone achieved higher instructional efficiency than using a measure of mental efficiency, but in the Salden et al. study there were no statistical differences between using a measure of performance alone or using a measure of mental efficiency. Kalyuga (2006a) also found that using mental efficiency to adapt instruction was more efficient than not adapting instruction, but there were no statistical differences between using a measure of performance alone and not adapting instruction. In addition, Kalyuga also found that there were no statistical differences between using a measure of performance alone and using a measure of mental efficiency. To determine whether adding mental effort ratings to performance measures improves the efficiency of adapted instruction, additional research is necessary (Rikers, 2006).

Research Question

The following research question is proposed: Will the use of a mental efficiency measure, which combines a rapid verification test of performance with a subjective rating of mental effort, to adapt instruction lead to higher post-test scores and more efficient instruction than the use of a performance measure alone?

Significance of Study

This study will determine whether the use of a mental efficiency measure to adapt instruction leads to higher knowledge gains and more efficient instruction than the use of a performance measure alone. The results of this study could advance research in adaptive training. As Clark, Nguyen, & Sweller (2006) explain, e-learning presents the potential to tailor training to individual learner needs, and in the $50 billion training market (Clark et al., 2006), making instruction more efficient could result in vast economic savings. Adaptive online training could be used to tailor instructional methods to learner expertise. According to Clark et al. (2006), a dynamic rapid testing method is needed, and, "Since the rapid testing method is very new, we will need more research, including experiments and field trials, before recommending it for all types of instructional goals" (p. 285).

The study described herein could add to the relatively recent research performed on schema-based rapid diagnostic testing (Kalyuga, 2006b). Kalyuga's most recent study (Kalyuga, 2006a) indicates that adaptive learning through use of rapid verification tests and subjective cognitive load ratings provides more efficient learning than non-adapted learning. However, the combined use of performance and cognitive load measures was no more efficient than using the performance measure alone, leading Rikers (2006) to state, "Therefore, it remains an empirical question whether the approach advocated by Kalyuga (2006a) proves to be of additional value as compared to previous adaptive methods (Camp et al., 2001; Salden et al., 2004)" (p. 362). The study described herein could help clarify the issue and contribute to research on rapid testing techniques. Therefore, this study is timely and fits into research currently being performed around the world.

Delimitations of Study

The materials used in this study will present instruction about performing cost-volume-profit analysis. The focus of the study is to determine whether adaptation of instructional methods using measures of mental efficiency (a combination of measures of performance and cognitive load) is more effective than using measures of performance alone. However, the results may not be generalizable to other instructional materials. The results may only apply to content with the same number of interactive elements, or to computer-based training, or to the domain of accounting.

CHAPTER 2

REVIEW OF THE LITERATURE

Recently, cognitive load theory (CLT) researchers have investigated adapting instruction to learner expertise (Clark et al., 2006; Kirschner et al., 2006; van Merriënboer & Sweller, 2005). Typically, expertise is assessed with a measure of task performance or mental efficiency (a combined measure of task performance and mental effort exerted to perform the task). The measure of expertise is generally used to place learners in an appropriate stage of instruction, advance learners through instruction, and require learners to repeat instruction. The purpose of this study is to determine whether assessing expertise and adapting instruction with a mental efficiency measure will reduce cognitive load during instruction and produce higher posttest scores and more efficient instruction than assessing expertise and adapting instruction with a task performance measure.

This chapter provides a framework and rationale for the research proposed in this study and includes the following sections: cognitive load theory and learning, cognitive load theory and instructional methods, adapting instructional methods to learner expertise, assessing learner expertise, and adapting instruction with mental efficiency measures. The cognitive load theory and learning section describes the underlying CLT principles that explain how people learn and why working memory constraints impede learning. The cognitive load theory and instructional methods section explains why certain instructional methods either accommodate or exceed working memory limitations depending on learner expertise and consequently, why instructional methods should be adapted to learner expertise. The adapting instructional methods to learner expertise section describes how adaptive methods in the past attempted to accommodate learner characteristics, why the adaptive methods in the past produced limited success, and how recent CLT research can improve adaptive instruction. The assessing learner expertise section describes the methods that CLT adaptive learning research uses to measure expertise. Finally, the adapting instruction with mental efficiency measures section describes the results of CLT adaptive learning research studies that employ mental efficiency measures and explains why further research is needed.

Cognitive Load Theory and Learning

Cognitive load theory (Sweller, 1988; Sweller et al., 1998) is a set of principles and guidelines to design and deliver instructional environments that promote learning by utilizing the limited capacity of working memory and minimizing working memory overload. The theory assumes a limited working memory of seven elements for storing information, on which only two to four elements may be processed (Miller, 1956) and a virtually unlimited long-term memory (Simon & Gilmartin, 1973) holding large numbers of schemas (Chi, Feltovich, & Glaser, 1981), which can vary in degree of automaton (Kotovsky, Hayes, & Simon, 1985). Working memory is the information processor responsible for both high-level cognitive processing (such as constructing mental models) and short-term maintenance of information involved in cognitive processing (Baddeley, 1992). Working memory is very limited in capacity and duration. For example, people can remember no more than approximately seven serially presented random numbers

(Miller, 1956) for no longer than several seconds unless the numbers are intentionally rehearsed.

Long-term memory, on the other hand, is virtually unlimited in both capacity and duration (Simon & Gilmartin, 1973). Knowledge is stored in long-term memory in hierarchically organized, domain-specific structures called schemas. According to Marshall (1995), a schema is a vehicle of memory that permits people to categorize information in the manner in which it will be used. For example, when presented with a problem involving balancing a cart on an inclined plane, an expert physicist recognizes the problem as a balance-of-forces problem, while a novice will probably consider the problem to be a cart or inclined plane problem. The expert physicist can identify the deep structure of a problem and is not influenced by the problem surface features, as is a novice. The expert physicist has solution schemas organized by problem type, can recognize the problem type and applicability of the organized schemas, and can retrieve an appropriate solution procedure from memory (National Research Council, 2001).

According to CLT, the goal of all instruction is to alter knowledge in long-term memory (Kirschner et al., 2006). However, since working memory constraints limit information processing to two to four elements, becoming an expert physicist or an expert in any domain would be virtually impossible because of the need to simultaneously process hundreds of interrelated facts, concepts, and procedures when solving a problem. In fact, learning is made possible through schema construction and automation, which alleviate working memory constraints (Sweller, 1988; Sweller et al., 1998). Schema construction involves building more complex schemas by combining lower-level schemas into higher-level schemas (Sweller et al., 1998). Schema construction occurs mainly through elaboration and induction. Elaboration is the process of enriching existing knowledge with new information, whereby learners search memory for existing knowledge that will provide a general structure for understanding new information (van Merriënboer, 1997). For example, when a child encounters a cat, the child may search memory and find a stuffed animal schema that classifies small furry things as animals. Therefore, since the cat is furry, the cat is an animal.

Induction is the process of generalization and discrimination. Generalization abstracts general information from concrete experiences, and discrimination classifies general information into categories from concrete experiences. For example, a child may notice a cat and a dog and generalize that all animals have four legs. A child may also notice that a cat is different than a dog and discriminate that some four legged animals are cats and some are dogs. Constructing higher level schemas through elaboration and induction allows people to overcome the limitations of working memory by "chunking" multiple elements of information into a single, higher-level element (Chi et al., 1981; Larkin, McDermott, Simon, & Simon, 1980). Therefore, despite the two to four element processing limitations of working memory, each element can consists of very large and interconnected schemas.

Automation is the act of converting declarative knowledge to procedural knowledge following extensive practice (Anderson, Fincham, & Douglass, 1997). Declarative knowledge is formed from representations of the outside world and consists of simple schemas (facts, concepts, procedures, principle) or complex schemas (conceptual models, goal-plan models, and causal and functional models). Procedural knowledge is formed from a goal-based process that applies declarative knowledge to solve a problem. After extensive practice, procedural knowledge requires very little mental attention and allows people to overcome the limitations of working memory by effectively bypassing working memory (van Merriënboer, 1997).

Schema construction and automation help overcome working memory constraints. Although working memory can only process a few pieces of information at a time, those pieces of information can be quite large through the process of schema construction. In addition, working memory can be effectively bypassed through the process of schema automation. Building expertise is the process of constructing and automating ever more complex schemas. Therefore, instructional methods should help learners build expertise and overcome working memory constraints.

Cognitive Load

Because working memory capacity constraints exist and can impede learning, CLT focuses on the work instructional methods impose on working memory. The work imposed on working memory is called cognitive load and includes three types: intrinsic load, germane or effective load, and extraneous or ineffective load (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). The three types of cognitive load are additive, so total cognitive load equals the sum of intrinsic load, germane load, and extraneous load.

Intrinsic Load

Intrinsic load relates to the complexity of the learning material. Complex material is difficult for novices to learn because novices do not possess very well organized or developed schemas for the subject. Consequently, the two to four elements of free working memory are inadequate because the elements that novices retrieve do not contain "chunks" of information but merely small bits of information (Sweller, 1994). Intrinsic load depends on the complexity of the to-be-learned content, or more specifically, the element interactivity, which is the number of elements that must be processed simultaneously in working memory. Content contains low element interactivity if the content can be learned in isolation, not simultaneously with other content. For example, learning foreign-language vocabulary words provides an example of low element interactivity because students can learn words independent of other words. Therefore, learning low element interactivity content, such as foreign language vocabulary words, is not as effortful as learning high element interactivity content. Content contains high element interactivity if the content can only be learned simultaneously with other content. For example, speaking a foreign language provides an example of high element interactivity because foreign language speech requires simultaneous consideration of word meaning, pronunciation, and sequence.

Designers cannot reduce intrinsic load without reducing learning (van Merriënboer & Sweller, 2005). Designers can artificially reduce the complexity of the content, but ultimately the only way for learners to fully understand the content is to be presented the content with the full element interactivity present. One way designers can reduce the intrinsic load is to pursue an isolated-followed-by-interacting-elements approach (van Merriënboer & Sweller, 2005), in which isolated elements capable of being processed serially are first presented, then all information is presented at once, including the interactions among the elements. For example, learning to speak a foreign language typically first requires learning vocabulary words, then combining the words into spoken sentences. The element interactivity of speaking sentences is initially reduced by learning the vocabulary words in isolation, but ultimately, the high element interactivity of speaking sentences must be embraced to learn how to speak a foreign language. For learners with low prior knowledge, content with high element interactivity will cause high intrinsic load and will impede learning because of working memory constraints.

Germane Load

Germane load relates to learning activities that promote schema construction and automation (Paas & van Merriënboer, 1994), such as providing a highly varied sequence of worked examples and completion problems (Paas, 1992) and requiring self explanations (Renkl & Atkinson, 2003). A sequence of worked examples or completion problems that contains highly varied surface features helps learners determine the range of applicability of constructed schemas. However, determining the applicability of constructed schemas requires the mindful engagement of learners and increases cognitive load (germane load). Prompting learners to self explain requires learners to conduct a mental dialog with themselves asking what, why, and how questions. Self explanations also promote the mindful engagement of learners through elaboration of current knowledge with new information. Therefore, increasing germane load makes learning more effortful. However, the goal of instruction should be to impose germane load.

Consequently, when increasing germane load, designers must be careful to not overload working memory, especially for novice learners (van Merriënboer & Sweller, 2005).

Extraneous Load

Extraneous load relates to poorly designed learning activities that provide unnecessary, distracting, disorganized, or redundant information that does not promote schema construction or automation (Sweller et al., 1998). Instructional designers should strive to reduce extraneous load (van Merriënboer & Sweller, 2005). An example of disorganized material causing extraneous load is the so called split attention effect (Sweller, Chandler, Tierner, & Cooper, 1990), discussed next, followed by an example of redundant material causing extraneous load, called the redundancy effect (Chandler & Sweller, 1991).

Split attention effect. The split attention effect occurs when a diagram is separated from the text that explains the diagram. If a diagram is placed on one page, and the text describing the diagram is placed on another page, high cognitive load (extraneous load) is imposed, suppressing learning (Sweller et al., 1990). Learners are required to keep both the text and diagram in working memory when relating one to the other, using up working memory resources and preventing schema construction. Integrating the diagram and explanatory text, by placing the text next to the diagram on the same page, eliminates the extraneous load and improves learning.

Redundancy effect. The redundancy effect occurs when integrating a diagram with unnecessary explanatory text. Chandler and Sweller (1991) observed that when text is provided that merely re-describes a diagram, the text provides redundant information and imposes high cognitive load (extraneous load). However, when both the text and diagram together are required to make the information intelligible, the text does not provide redundant information. For example, text such as "Angle ABC = Angle XYZ" cannot be interpreted without a diagram. Therefore, integration of the text with the diagram does not provide redundant information. In contrast, text such as "Blood flows from the heart to the lungs and back to the heart," can be interpreted without a diagram. Therefore, integration of the text with a diagram of a heart, lung, and arrows depicting blood flow provides redundant information (Sweller, 2005). By eliminating redundant information, and the consequent extraneous cognitive load, working memory is freed to process germane load, and learning is improved.

In summary, CLT is a set of principles and guidelines for instruction to promote learning and avoid working memory overload. CLT assumes a limited working memory of seven elements (Miller, 1956) and an unlimited long-term memory (Simon & Gilmartin, 1973), containing schemas, vehicles of memory containing information organized in the manner in which the information will be used. According to CLT, the goal of all instruction is to alter knowledge in long-term memory (Kirschner et al., 2006). However, working memory constraints would make learning complex material virtually impossible were it not for schema construction and automation. Schema construction combines lower-level schemas into higher-level schemas, chunking multiple elements of information into a single, higher-level element. Schema automation converts declarative knowledge into procedural knowledge following extensive practice. Procedural knowledge requires very little mental attention and allows learners to effectively bypass working memory (Anderson et al., 1997).

Building expertise is the process of constructing and automating ever more complex schemas. Instructional methods should help learners build expertise, while maintaining cognitive load, the work imposed on working memory, within working memory resource constraints. There are three types of cognitive load: intrinsic, germane, and extraneous (Paas et al., 2003). Intrinsic load relates to the complexity of the material and cannot be lowered without reducing learning. Germane load relates to learning activities that promote schema construction and automation, and imposition of germane load, within working memory constraints, should be the object of instruction. Extraneous load relates to poorly designed activities that do not promote schema construction and automation, and elimination of extraneous lead should be a goal of instruction, thereby freeing working memory resources to process germane and intrinsic load. CLT focuses on the effect instructional methods have on cognitive load, seeking to maintain total cognitive load (sum of intrinsic, germane, and extraneous load) within learner working memory resource constraints.

Cognitive Load Theory and Instructional Methods

Introduction

The cognitive load a learner will experience is affected by the instructional method used and the learner's prior knowledge (Sweller & Cooper, 1985). The use of conventional problems (open-ended problems that require the learner to solve all of the steps) imposes high extraneous load on novices, but the use of worked examples (problems with all solution steps provided and explained) imposes low extraneous load on novices (Sweller & Cooper, 1985). Worked examples provide scaffolding for novices and to not overtax working memory. In contrast, conventional problems can overwhelm the working memory of novices because novices lack solution schemas in long-term memory and cannot supply all of the solution steps of a conventional problem from memory. Therefore, novice learners are aided by a worked example that displays all of the solution steps, reducing the need for novices to maintain all of the information in working memory.

However, researchers have also observed an expertise reversal effect (Kalyuga et al., 2003), where worked examples impose high extraneous load on advanced learners due to the need to process redundant information, and conventional problems impose low extraneous load, and high germane load, on advanced learners. Therefore, researchers have suggested that as learners gain expertise, the scaffolding provided by worked examples should be gradually reduced through a process called fading guidance, in which a series of completion problems (worked examples with blanks that learners fill in) are used that progressively omit, and require the learner to provide, more solution steps (Renkl & Atkinson, 2003). As novices gain expertise, faded completion problems and finally conventional problems are presented.

To more fully discuss how instructional methods affect cognitive load, the following four sections are provided. 1) The worked examples effect section discusses how providing worked examples, not conventional problems, can benefit novice learners.

2) The completion problem effect section discusses how substituting completion problems for some worked examples can also benefit novice learners. 3) The expertise-reversal effect section discusses how providing conventional problems, not worked examples or completion problems, can benefit advanced learners. 4) The fading guidance effect section discusses how first providing worked examples, then a series of faded completion problems, and finally a conventional problem offers appropriate scaffolding as learners gain expertise. Therefore, these four sections discuss the instructional methods that are appropriate for novice to advanced learner expertise levels.

Worked Examples Effect

Worked examples reduce extraneous cognitive load for novices by focusing attention on useful solution steps and preventing learners from employing a means-end strategy, which places a high demand on learner cognitive resources (van Merriënboer & Sweller, 2005). A means-end strategy imposes high extraneous load because the learner uses a trial and error procedure to successively reduce differences between the problem's given and goal state. The learner may solve the problem but not learn the proper solution procedure. For example, if a student does not possess an algebra solution schema, the student may solve the problem, $3X + 1 = 7$ by inserting different numbers for X in the equation until the problem is solved. Perhaps, on the second try, the student will insert 2 in the formula and find the answer. However, the student does not learn how to solve the problem in a forward directed manner, such as subtracting 1 from both sides and dividing both sides by 3. Worked examples present all of the solution steps for a problem and facilitate learner construction of solution schemas. Because the solution steps are presented, learners do not need to resort to a means-ends strategy but can focus their attention on how to solve the problem and in the process acquire a solution schema.

Sweller and Cooper (1985) observed that using worked examples was a faster method for middle school students to learn math than using conventional problems and resulted in fewer student errors on a posttest. The researchers randomly assigned 22 middle school students to either a worked example group or a conventional problem group. Each group was presented a lesson consisting of eight algebra problems of the type C(A+D)/F = G; solve the variable A. The worked examples group received a worked example followed by a conventional problem, and the sequence was repeated four times. The students had five minutes to solve the conventional problems and were provided the correct answers at the end of the time period. The conventional problem group received eight conventional problems. At the end of the lesson, students from both groups completed a posttest, requiring the students to solve six conventional problems similar to those used in the lesson. Results showed that the conventional problem group took six times longer to complete the lesson and fifty percent longer to complete the posttest. The worked example group made half as many errors as the conventional problem group on the posttest and made an average of 0 errors during instruction, compared to an average of 2.7 errors for the conventional problem group. The researchers concluded that worked examples provide a more efficient and effective method of learning math than conventional problems.

Cooper and Sweller (1987) conducted a similar study to the Sweller and Cooper (1985) study but with two differences: 1) the Cooper and Sweller (1987) study used simpler problems that reduced intrinsic load from lower element interactivity, freeing working memory for other use; and 2) the Cooper and Sweller (1987) study used a transfer problem for the posttest. Results showed that the worked example group spent significantly less time and committed significantly fewer errors than the conventional problem group on the transfer problem. Nine (out of 12) worked example participants were able to complete the transfer problem within the five-minute limit. In contrast, zero (out of 12) conventional problem participants were able to complete the transfer problem. The researchers concluded that worked examples can facilitate transfer performance by fostering schema automation. Both groups resorted to a means ends strategy when presented with the transfer problem. The researchers hypothesized that the nine participants of the worked examples group were able to complete the transfer problem because the worked examples group possessed sufficient working memory capacity to conduct the means ends strategy despite its heavy cognitive load. The unused working memory capacity was created because schemas were automated from studying the worked examples. In contrast, no participants of the conventional problem group were able to complete the transfer problem because of insufficient working memory capacity to conduct the means-ends strategy. Insufficient working memory capacity resulted because schemas were not automated from solving the conventional problems. Therefore, worked examples can foster schema automation, freeing working memory capacity to undertake novel problems.

The two studies just discussed (Cooper & Sweller, 1987; Sweller & Cooper, 1985) demonstrate that learners can achieve higher test scores with less instruction time by studying worked examples compared to solving conventional problems. In fact, Zhu and Simon (1987) report that a Chinese algebra and geometry class using worked examples was able to complete a three-year course in two years. Worked examples reduce extraneous load compared to conventional problems because worked examples display solution steps, which prevent learners from employing working memory resources to conduct a means-ends approach to solve the problem. Rather, learners can devote working memory resources to schema construction and automation when studying worked examples.

Completion Problem Effect

One disadvantage with worked examples is that students may not fully process the worked example, merely giving the worked example a cursory review, or skipping the worked example altogether. To direct attention to worked example steps, completion problems may be used (van Merriënboer, 1990). Completion problems are worked examples with one or more solution steps omitted, requiring the learner to complete the omitted steps. Completion problems ensure that learners do not ignore the solution steps to the example, while providing scaffolding support to limit working memory overload.

Van Merriënboer (1990) observed the benefits of using a completion problem strategy in a 10-week study of high school students learning to program. Forty-two students were placed in one of two groups: a completion strategy group and a generation strategy group. The completion strategy group learned to program by modifying and extending well defined existing programs. The generation strategy group learned to program by designing and coding new programs. All students completed five lessons over a ten week period. Each completion strategy group lesson required the group to study factual information about new commands and syntactical details and solve four completion assignments. Each completion assignment contained three components: 1) a problem specification, 2) a partial solution consisting of a well-structured program, and 3) several questions concerning the structure and method of the program and instructions to complete the program. Each generation strategy group lesson required the group to study the same factual information as the completion strategy group and solve one generation assignment. The generation assignment contained two components: 1) a model program, which employed the newly learned commands and syntax, and 2) a program specification and guidelines for designing and testing a new program.

After the 10 week period, all participants completed three posttests: 1) a program construction test, similar to the generation group lesson assignment, in which students were asked to design and code complete, new programs, 2) a factual knowledge test to measure passive knowledge of commands and syntax, and 3) a program comprehension test to measure proficiency interpreting programs. The results favored the completion strategy group, who scored significantly higher on the program construction test, including a higher percentage of correctly coded programs (the completion strategy group scored 81.3% correct, and the generation strategy group scored 68.2% correct), higher number of features used correctly, and a higher rating of semantic and syntactic program correctness. There were no significant differences between the groups for the factual knowledge test and program comprehension test. Van Merriënboer concluded that a completion problem strategy was more effective than a generation (conventional problem) strategy to learn programming because the completion problem strategy provided scaffolding support, which freed working memory for students to construct solution schemas. In contrast, the generation strategy required more working memory to remember the newly learned commands and syntax and limited the students' ability to construct solution schemas.

Previous studies have compared worked example to conventional problem use (Cooper & Sweller, 1987; Sweller & Cooper, 1985) and completion problem to conventional problem use (van Merriënboer, 1990). Paas (1992) compared all three methods: worked examples, completion problems, and conventional problems. Paas randomly assigned a class of 46 technical school students to a worked example group (14), a completion problem group (15), and a conventional problem group (13). The domain was statistics (compute mean, median, mode). During the instruction, each group was presented 12 problems and then completed a 24-item posttest. The problems provided during instruction varied by group. For the worked example group a three-problem sequence format was presented four times. The three-problem sequence format contained two worked examples followed by a conventional problem. For the completion problem group a three-problem sequence format was also presented four times. However, the three-problem sequence format contained two completion problems followed by a conventional problem. For the conventional problem group all twelve problems were conventional problems.

Table 1

Comparison of Worked Example, Completion Problem, and Conventional Problem Strategies in Statistics

	Worked example group	Completion problem group	Conventional problem group
Average posttest score (0 – 24)	18.8 *	16.1 *	12.4 * *
Average time in training (minutes)	32.3 * *	39.8 *	42.5 *

* Statistically significant difference with paired figure, $p < 0.05$
Source: Pass, F. (1992) Training Strategies for Attaining Transfer of Problem Solving Skill in Statistics: A Cognitive Load Approach.

The results of the Paas experiment, displayed in Table 1**,** show that both the worked example and completion problem groups scored significantly higher on the posttest than did the conventional problem group. Also, the worked example group spent significantly less time in training than did the completion problem and conventional problem groups. Paas concluded that both worked examples and completion problems are more effective than conventional problems for novices, but completion problems were not necessarily more effective than worked examples. Both worked examples and completion problems provide scaffolding support to free working memory to devote to schema construction and automation. In addition, worked examples seem to require less time to study than solving completion problems or conventional problems. Therefore, with the time savings, worked examples present an opportunity to increase example variability, which in turn should increase germane load and further promote schema construction and automation.

Expertise Reversal Effect

Although researchers have discovered cases where the use of worked examples are more effective for novices than the use of conventional problems, Kalyuga, Chandler, Tuovinen, and Sweller (2001) discovered cases where the opposite occurred: the use of conventional problems was more effective for advanced learners than the use of worked examples. This expertise reversal effect (Kalyuga et al., 2003) takes place when learners gain expertise and no longer need the schema support that worked examples provide. Worked examples serve as schema substitutes for novice learners who lack solution schemas in long-term memory. However, worked examples become redundant and counterproductive for advanced learners who do possess the solution schemas. If advanced learners are unable to avoid the instructional guidance, the advanced learners will have to cross reference and integrate the redundant information with existing schematic knowledge in memory, requiring cognitive resources and possibly resulting in cognitive overload. As a result, instructional guidance provided to novices may have negative consequences when provided to experienced learners.

Kalyuga et al. (2001) conducted a series of experiments over a five week period, in which 24 trade apprentices were randomly assigned to a worked example group (12) and a problem-solving group (12). The domain was programming relay circuits with programmable logic controllers (PLC). Twenty-four relay circuit problems were presented to the groups. The worked example group was presented 24 circuit problems as 12 worked examples and 12 problem-solving exercises, in which the participants were granted three attempts to write a program for each circuit. If unsuccessful after the third attempt, the participants were provided the correct steps to program the circuit. Therefore, the worked example group studied 12 worked examples and attempted 12 problems. The problem-solving group was presented all 24 circuits as problem solving exercises. At the end of the training, all participants received 12 posttest questions similar to the problem solving exercises.

Five stages of training were provided, each separated by one week. At each stage, the 24 relay circuits became progressively more complex. Results of the five-week training are provided in Figure 1. When the participants possessed little experience at the beginning of the training, the participants who studied worked examples, and received scaffolding support, benefited the most. Consequently, during Stages 1 through 3, the worked example group scored higher on the posttest questions than the problem solving group. However, with more experience in the domain over Stages 4 and 5, the worked example effect disappeared, and the problem solving group scored higher than the worked example group. The scaffolding support provided by the worked examples became redundant for the now advanced learners and provided no benefit. The researchers concluded that instructional methods should adapt with learner knowledge. Novice learners should receive worked examples, and as learners gain knowledge, advanced learners should receive conventional problems (or exploratory learning activities) to enhance and extend skills.

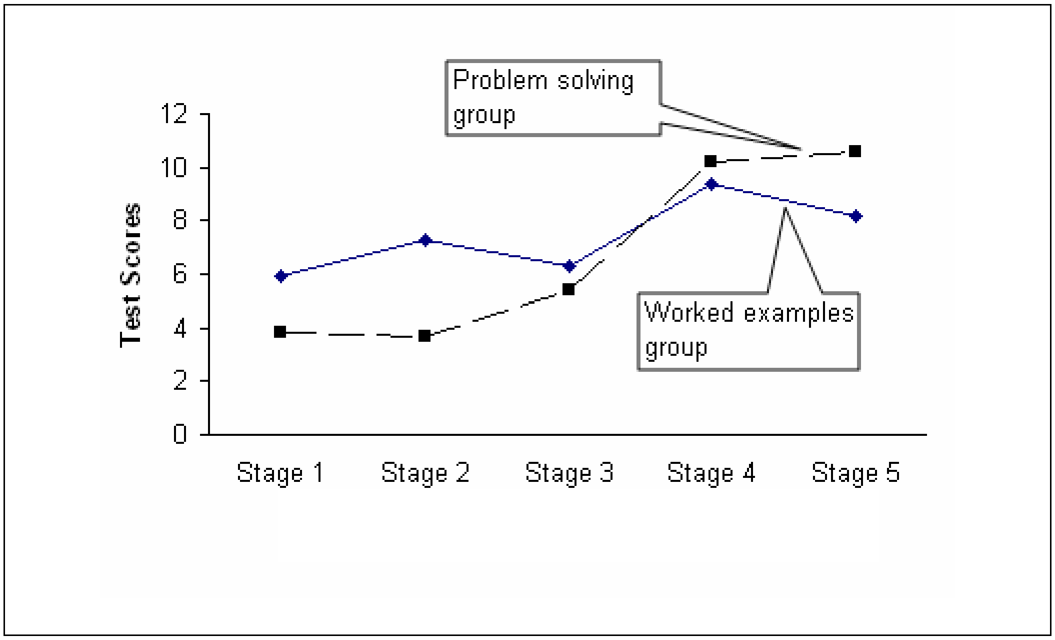


Figure 1. Test score comparison showing the expertise reversal effect in the domain of PLC programs

Source: Kalyuga, S., Chandler, P., Tuovinen, J., and Sweller, J. (2001) When Problem Solving is Superior to Studying Worked Examples.

Fading Guidance Effect

Because worked examples seem to benefit novices, and conventional problems benefit advanced learners, Atkinson, Renkl, and Merrill (2003) and Renkl, Atkinson, Maier, and Staley (2002) proposed fading guidance to provide a transition between the use of worked examples and conventional problems. With fading guidance, progressively more solution steps are omitted from worked examples, requiring the learner to provide the omitted steps from memory. Fading can be performed in a forward or backward manner. To illustrate forward fading, assume a worked example contains five solution steps. The first worked example presented does not omit any steps. The second worked example presented omits Step 1. The third worked example omits Steps 1 and 2, and the process continues until all the steps are omitted, and a conventional problem is presented. To illustrate backward fading, the first worked example presented does not omit any steps. The second worked example presented omits Step 5. The third worked example omits Steps 4 and 5, and the process continues until all the steps are omitted, and a conventional problem is presented.

Fading guidance emulates the model of skill acquisition developed by Van Lehn (1996) and Anderson, Fincham, and Douglass (1997). In the first stage, learners refer to known examples to solve problems by analogy. In the second stage, learners develop abstract declarative rules to guide problem-solving. In the third stage, learners automate declarative rules through practice and form procedural rules that can bypass working memory. In the fourth stage, learners practice many different problems and are able to retrieve entire solution schemas from memory. Renkl et al. (2002) suggest that worked examples are appropriate for the first and beginning of the second stages, and conventional problems are appropriate for the third and fourth stages, acknowledging that stages overlap.

Renkl et al. (2002) tested ways to transition from studying worked examples to solving conventional problems as novices gain expertise. Forty-five university students were randomly assigned to a foreword fading group (15), a backward fading group (15), and an example/problem pair group (15). The researchers presented a computer-based lesson in probability calculation. In the foreword fading group, students were provided a total of four problems as follows: 1) a worked example, 2) a completion problem with the first step required to be supplied by the student and the final two steps provided, 3) a completion problem with the first two steps required to be supplied by the student and the final step provided, and 4) a completion problem requiring the student to supply all three solution steps. In the backward fading group, the steps were omitted in reverse sequence. In the example/problem pair group the students were provided a worked example and conventional problem to solve and repeated this process once. After training, a posttest was administered that contained 12 conventional problems, consisting of six near- and six far-transfer problems.

The results of the experiments are shown in Table 2. Both fading groups performed significantly better than the example-problem pair group on the near-transfer posttest. Only the backward fading group performed significantly better than the example-problem pair group on the far-transfer posttest. The researchers concluded that fading provides scaffolding support for learners, freeing working memory for schema construction. Pairing a worked example with a conventional problem requires the learner to make an abrupt transition that is too severe. The learner progresses from a worked example, with all solution steps displayed, to a conventional problem, with no solutions steps displayed, requiring the learner to supply all of the steps from memory. The conventional problem causes working memory overload and suppresses learning. In contrast, a fading procedure requires the learner to make a smooth transition. The learner receives scaffolding support, which frees working memory for schema construction. The researchers also concluded that cognitive load is reduced even more with backward fading compared to forward fading because backward fading provides additional scaffolding by displaying all the steps that precede the step the learner must supply. In contrast, forward fading displays the steps that follow the step the learner must supply and does not provide as much support.

Table 2

Comparison of Fading Procedure Effectiveness in Statistics

	Average test scores		
	Forward fading group	Backward fading group	Example/problem pair group
Near transfer posttest (18 total)	12.4 *	11.6 *	6.5 * *
Far transfer posttest (18 total)	6.0	7.9 *	4.1 *

* Statistically significant difference with paired figure, $p < 0.05$
Source: Renkl, A., Atkinson, R., Maier, U., Staley, R. (2002) From Example Study to Problem Solving : Smooth Transitions Help Learning (Experiment 3).

In the Renkl et al. (2002) study previously discussed, both backward and forward fading produced significantly higher near-transfer posttest scores than example/problem pairing, but only backward fading produced significantly higher far-transfer posttest scores than example/problem pairing. Backward fading seemed to provide the most scaffolding and to lower cognitive load more than forward fading or example/problem pairing. Therefore, to employ the free working memory resources provided by backward fading, Atkinson, Renkl, and Merrill (2003) tested the effect of adding self-explanatory prompts to a backward fading strategy to increase germane load and improve far-transfer. The researchers conducted two experiments in the domain of statistical probability. In Experiment 1, seventy university students were randomly assigned to a backward fading group (19), an example/problem pair group (19), a backward fading with self-explanatory prompts group (20), and an example/problem pair with a self-explanatory prompts group

(20). The self explanatory prompts required the participants to answer multiple-choice questions about the probability rule or principle a solution step employed. The correct answer to the self-explanatory prompt was displayed along with the correct solution step to the problem.

Like the Renkl et al. (2002) study, a posttest was provided with six near- and six far-transfer questions. The results of the posttest showed that the backward fading group scored significantly higher than the example-problem pair group. In addition, the participants who received self explanatory prompts scored significantly higher on the posttest than the participants who did not receive the prompts.

In Experiment 2, only a backward fading strategy was tested with and without adding self-explanatory prompts. Forty high school students were randomly assigned to a backward fading group or a backward fading with self-explanatory prompts group. The self-explanatory prompts group outperformed the no-prompts group for near- and far-transfer with statistically significant results. The researchers concluded that fading worked examples with self-explanatory prompts induces reflection in accordance with the cognitive apprenticeship (Collins, Brown, & Newman, 1989) and zone of proximal development (Vygotsky as cited in Atkinson et al., 2003) models, respectively. Backward fading provides scaffolding, which lowers cognitive load and frees working memory resources, and self explanatory prompting employs the free working memory resources to increase germane load and induce schema construction.

In summary, instructional methods impose different cognitive loads on learners. Novices benefit from the scaffolding provided by worked examples. Since novices lack solution schemas in long-term memory, novel information can overwhelm working memory capacity. Therefore, novice learners require a worked example to display the solution steps, so novices can offload the information from memory. Advanced learners benefit from the schema strengthening opportunities provided by conventional problems, which require the learner to provide all of the solution steps to the problem. Since advanced learners can retrieve solution schemas from long-term memory, excess working memory is available, and advanced learners may link novel information with existing schematic knowledge.

Novice learners achieve higher test scores with less training time by studying worked examples or performing completion problems than by solving conventional problems (Sweller & Cooper, 1985; van Merriënboer, 1990). Completion problems are problems that provide some of the steps of a worked example and require the student to fill in the missing steps. Completion problems help alleviate the disadvantage of worked examples, which do not force careful study by learners. Worked examples and completion problems reduce extraneous cognitive load because all solution steps are visible to the learner and not required to be held in memory (Sweller et al., 1998).

Although worked examples are more effective than conventional problems for novices, the reverse is true for advanced learners. In fact, worked examples can constrain the learning of advanced learners. The expertise reversal effect occurs when the added detail provided to novices becomes redundant to experts, who already possess schematic knowledge in long-term memory and must correlate the redundant information with existing schematic knowledge (Kalyuga et al., 2003). Due to this expertise reversal effect, worked examples should be provided to novices and conventional problems be provided to advanced learners.

Renkl et al. (2002) recommend fading guidance as learners gain expertise to provide a transition from worked examples to conventional problems. Fading guidance progressively omits steps from worked examples until all steps are omitted, and the worked example becomes a conventional problem. There are two forms of fading guidance: 1) forward fading (eliminating solution steps from the beginning to the end) and 2) backward fading (eliminating solution steps from the end to the beginning). Both fading methods seem to produce better test results than conventional problems, and backward fading seems to produce lower cognitive load than forward fading. Therefore, as learners gain expertise, learners should progressively solve more worked example steps through backward fading until sufficient expertise is gained to solve conventional problems.

Adapting Instructional Methods to Learner Expertise

The preceding section described how worked examples and completion problems provide scaffolding support for novice learners and decrease cognitive load but provide redundant information for advanced learners and increase cognitive load. Because of the expertise reversal effect (Kalyuga et al., 2003), instructional methods that reduce cognitive load for novices are redundant and counterproductive for advanced learners. As a result, CLT researchers have recently investigated ways to adapt instructional methods to learner expertise (Clark et al., 2006; Kalyuga, 2006b; Kalyuga & Sweller, 2005; van Merriënboer & Sweller, 2005). This section will review the previous research conducted about adaptive instruction and discuss the advances that CLT has made that remedies some of the deficiencies inherent in the previous adaptive instruction research.

Adaptive learning has an extensive research base according to Park and Lee (2004). Numerous studies have investigated ways to adapt instructional methods and procedures to a variety of student characteristics. From the research base, the authors have identified three broad models of instructional adaptation: 1) macro-adaptive instructional models, 2) aptitude-by-treatment interaction models, and 3) micro-adaptive instructional models. Each model will be discussed next.

Macro-Adaptive instructional Models

Macro-adaptive instructional models adapt curricular variables to student needs over the span of an entire course, providing students with different instructional goals, depth of topic coverage, and method of instructional delivery. For example, the Keller Plan, developed by Columbia University and used during the 1960s and 1970s, required students to master a unit before moving to the next unit, allowed students to work at their own pace, used textbooks and workbooks as the primary instructional means, and used student proctors to assess performance and provide feedback. Since most school systems advance students uniformly through the course material, macro-adaptation models were difficult to integrate into existing systems, requiring a great deal of work by the teachers (Park & Lee, 2004).

Another example of a macro-adaptive instructional model is Curriculum-Based Measurement (CBM). Developed in 1987, CBM was created to improve special education instruction by administering repeated rapid tests across time and adapting the curriculum accordingly (Fuchs & Deno, 1991). Marston (1989) provides the following examples of curriculum based measures: a) for reading, counting the number of words a student reads correctly during a one minute interval; b) for writing, counting the number of words or letters written correctly during a three-minute interval; and c) for math, counting the number of digits written correctly during a two-minute interval.

According to Burns, MacQuarrie, and Campbell (1999), the motivation behind CBM was not to provide alternative instructional strategies or teaching techniques but to monitor instructional/teaching effectiveness. The CBM tests are short enough to be administered frequently, one to five times per week, and the test data can be graphed and charted systematically to monitor student progress. Shinn & Bamonto (1998) report that CBM was designed only to assess student performance in the basic skills areas of reading, spelling, mathematics computation, and written expression.

Aptitude by Treatment Interaction Models

Aptitude by treatment interaction (ATI) models attempt to adapt instruction to specific student characteristics, such as fluid/crystallized intelligence (i.e. general intelligence versus domain knowledge) and impulsive/reflective styles (i.e. high willingness versus low willingness to commit an error) (Jonassen & Grabowski, 1993). Jonassen and Grabowski note that, “The concept of aptitude-treatment interactions (ATI) is one of the best known in the educational research field” (p. 23). Although past ATI research is extensive, current ATI research activity has declined because: a) previous ATI research was often conceived without an understanding of the psychological processes of the interaction between traits and treatments; b) very few replications of ATI research yielded significant interactions; c) most ATI research lacked external validity because the studies were not classroom based; and d) interactions between aptitudes and treatments were too complex to apply in the classroom, requiring thousands of different lessons to match instructional methods with all possible individual differences (Jonassen, 1988; Park & Lee, 2004).

Micro-Adaptive Instructional Models

Micro-adaptive instructional models diagnose specific student learning needs during instruction and adapt primarily by varying the amount and presentation sequence of the content (Park & Lee, 2004). Park and Lee recommend a two-stage model, which 1) initially adapts to performance on a pretest, and 2) continuously adapts to performance during training. Pretest scores initially place learners in a stage of instruction, and test scores obtained during instruction advance learners to the next stage or require learners to repeat the current stage.

The micro-adaptive model draws from the ATI research on prior knowledge. As Jonassen and Grabowski (1993) report in their book, "One of the strongest and most consistent individual difference predictors is prior knowledge. Research on the effect of this difference is vast, consistent and significant" (p. 420). The authors observe, "As level of prior knowledge rises, the need for instructional support decreases, conversely, as level of prior knowledge decreases, the need for instructional support rises" (p. 419). Therefore, the micro-adaptive model initially measures knowledge with a pretest, placing learners in a state of instruction appropriate for the learner knowledge level and continually measures knowledge during instruction, varying the amount and presentation of the instruction accordingly.

Park and Lee report that micro-adaptation models in the past have been limited by: a) the ability to accommodate only simple learning tasks, such as a concept and rule learning; b) the need to instantly assess student knowledge and task difficulty; and c) the need to provide appropriate alternative instructional methods. However, recent CLT research has remedied some of the micro-adaptation model limitations. Similar to what Park and Lee (2004) prescribed, CLT researchers (Kalyuga, 2006a; Kalyuga & Sweller, 2004, 2005) have employed a micro adaptive instructional model that initially adapts to performance on a pretest and continuously adapts to performance during training. In addition, the model used in CLT research employs a rapid assessment to measure performance, and the results are used to prescribe instructional methods, such as use of worked examples/completion problems for novices and conventional problems for advanced learners. This in part overcomes some of the limitations of the micro adaptation model cited by Park and Lee (2004), which included the need for instant assessment and alternative instructional methods.

The model used in CLT research adapts not only to the scores of performance tests but also to the ratings of mental effort expended on the tests. Paas (1992) and Paas and van Merriënboer (1993) coined the term "mental efficiency" as the combined measure of task performance and mental effort. To measure performance, researchers (Kalyuga, 2006a; Kalyuga & Sweller, 2005)use rapid assessments discussed in the next section. To measure mental effort expended on the assessment, researchers use a nine-point Lickert scale (Paas, 1992). A learner who performs well on an assessment with low mental effort should possess more expertise and more highly developed/automated schemas than a learner who performs well with high mental effort. The latter would indicate that the learner possesses less well developed and automated schemas and should receive more instruction (Kalyuga, 2006a; Kalyuga & Sweller, 2005).

Thus, although CLT researchers adapt instruction based on prior knowledge (measured by task performance), which was used to adapt instruction in the past, CLT researchers have added the following enhancements to adaptive instruction: 1) use of rapid assessments to measure task performance and adapt instruction, 2) use of mental effort ratings to measure mental effort expended on a task and adapt instruction, and 3) prescription of instructional methods, such as worked examples, completion problems, and conventional problems, to match learner expertise levels. Therefore, CLT researchers believe that assessing expertise with mental efficiency measures and providing instructional methods that match expertise levels adds capability to the micro-adaptive instructional model that was not available in the past. The next section will discuss the reliability and validity of the mental efficiency measures (rapid assessments and mental effort ratings) used in CLT adaptive instruction research. The section that follows will discuss the results of CLT research using mental efficiency measures to adapt instruction to learner expertise levels.

Assessing Learner Expertise

The adaptive model used in CLT research adapts instruction to the expertise of the learner. CLT researchers and others contend that expertise is manifested by the ability to swiftly retrieve solution schemas from long-term memory. Therefore, assessments should evaluate the schematic knowledge of a learner in long-term memory. To adapt instruction, CLT researchers assess expertise by measuring mental efficiency, a combined measure of task performance and mental effort expended to achieve that performance. To measure task performance, researchers can use traditional tests (multiple choice, true/false, open ended problems) or rapid assessments, quick methods to measure schematic knowledge in memory (Kalyuga, 2006b). To measure mental effort, researchers can use psychophysical, secondary task, and rating scale techniques (Paas, 1992). The reasons why researchers (Kalyuga, 2006b) have chosen to use rapid assessments to measure task performance is explained next, followed by a section explaining the reasons why researchers (Paas, 1992) have chosen to use rating scales to measure mental effort.

Measuring Performance with Rapid Assessments

To measure performance in real-time for adaptive instruction, Kalyuga and Sweller (2005) propose using rapid assessments of schematic knowledge in long-term memory. Since experts can rapidly retrieve previously acquired domain-specific knowledge structures from long-term memory and apply those structures in a forward directed, cognitively efficient way (Chi et al., 1981; Larkin et al., 1980), measurement of knowledge structures (schemas) in long-term memory should be one of the major purposes of assessment (Kalyuga, 2006d). Rapid assessments attempt to measure the sophistication of long-term memory schematic knowledge brought into working memory.

For example, the learner may be asked to supply the first step the learner would make to solve a problem or asked to verify whether certain solution steps are correct after the problem information has been removed from view. The level of sophistication of schematic knowledge is measured by the number of steps skipped and number of steps correctly verified, respectively (Kalyuga, 2006a; Kalyuga & Sweller, 2005).

In contrast, traditional tests, such as multiplication, true/false, and open ended problems, may not always reveal the sophistication of the learner's schematic knowledge (Kalyuga, 2006b; Kalyuga & Sweller, 2005). A student may answer the question correctly, but the answer may not reveal the process the student used to answer the question: through a means-ends approach (i.e. trial and error) or a forward-seeking schematic solution procedure. Therefore, Kalyuga and Sweller (2005) propose using rapid assessments to measure schematic knowledge in long-term memory.

Measuring schematic knowledge in long-term memory is recommended by the National Research Council (2001), as well as members of the assessment community (Mislevy, 1993). The idea of a schema-based assessment that measures long-term memory structures has attracted the attention of researchers in recent years (Marshall, 1995; Singley & Bennett, 2002) but has not yet resulted in widely usable testing methods. One impediment is that assessments cannot directly access long-term memory. However, assessments can access long-term working memory, a new memory structure that is formed when schemas held in long-term memory are brought into working memory (Ericsson & Kintsch, 1995). Long-term working memory is durable, interference proof, and virtually unlimited in capacity. For example, reading text requires a continuously updated mental representation of the text in working memory while retrieving knowledge components from long-term memory. This long-term working memory is durable enough to survive temporary interruptions in reading.

In the past, researchers attempted to assess long-term working memory by using interviews, think-aloud procedures, and tests that required the solution of a series of problems (Kalyuga, 2005; Kalyuga & Sweller, 2005). However, these assessment techniques were too time consuming to be used in an adaptive environment (e.g. e-learning environments), where real-time assessment of learner expertise was required. Therefore, a faster assessment method was necessary and rapid assessments were developed (Kalyuga, 2006a, 2006c, 2006d; Kalyuga & Sweller, 2004, 2005).

CLT researchers (Kalyuga, 2006a, 2006b, 2006c, 2006d; Kalyuga & Sweller, 2004, 2005) have devised several new, rapid assessments methods that measure learners' knowledge structures in long-term working memory for numerical problem solving domains, such as algebra, geometry, and physics. Other CLT researchers (Camp et al., 2001; Salden et al., 2004) have created air traffic control simulation software that measures learner performance using a series of keyboard commands to guide aircraft to designated gates. The software does not explicitly attempt to measure knowledge structures in long-term working memory but does measure performance after the task is completed on the basis of number of commands issued, time aircraft spent outside of airspace, separation of aircraft, and successful gate hits. Therefore, the air traffic control simulation software follows a more traditional testing approach. The rapid assessment methods discussed next do not include the Camp et al and Salden et al. studies but rather, include the studies that attempt to measure knowledge structures in long-term working memory. The rapid assessment methods discussed include: the first-step method and rapid verification method. Studies that have employed the two rapid assessment methods are presented in Table 3 and discussed individually in the sections that follow.

Table 3

Studies Measuring Reliability and Validity of Rapid Assessments

Study	Domain	First-step method	Verification method
Kalyuga and Sweller (2004)	Grade 9 algebra/geometry	X	
Kalyuga (2006d)	Grade 8 arithmetic word problems	X	
Kalyuga (2006c)	Grade 7 reading proficiency		X

First-Step Method

The first-step method requires learners to indicate what their first-step would be to complete a task. Kalyuga and Sweller (2004) believe the first-step method reveals the highest level of schematic knowledge that learners possess. For example, if learners are presented an algebra task and asked to indicate their first-step to solve for "X" in the following equation: $(2X - 3) \div 2 = 3$, some learners may multiply both sides of the equation by 2 to indicate their first-step is $(2X - 3) \div 2 \times 2 = 3 \times 2$; others may perform the preceding step in their heads and indicate the result, $2X - 3 = 6$, as their first-step; still others may perform not only the preceding step in their heads but the next step in their heads as well, adding 3 to both sides, and indicate the result, $2X = 9$, as their first-step;

and still others may possess high expertise and process three steps in their heads, indicating the result, $X = 9 \div 2$, as their first-step or the final answer, $X = 4.5$. Step skipping is a characteristic of higher levels of expertise (Blessing & Anderson, 1996).

Two studies evaluated the external validity of the first-step method: 1) Kalyuga and Sweller (2004) conducted three experiments in the domain of elementary algebra and geometry, and 2) Kalyuga (2006d) conducted an experiment in the domain of arithmetic word problems. These two studies will be discussed next.

Kalyuga and Sweller (2004) conducted three experiments to evaluate the external validity of the first-step method. The first two experiments compared the results of a traditional test to a rapid assessment using the first-step method. Both experiments administered a traditional test to 9th grade students, requiring the students to completely solve a series of algebra equations, and administered a rapid assessment using the first-step method, requiring the students to provide only their first steps to solve a series of algebra equations. Results for the two experiments showed that the rapid assessment test times were significantly lower than the traditional test times, with an effect size of 2.23 for Experiment 1 and 1.33 for Experiment 2. In addition, the rapid assessment scores were highly correlated with the traditional test scores, with a Pearson product-moment correlation of 0.92 for Experiment 1 and 0.85 for Experiment 2.

The third experiment employed a rapid assessment using the first-step method in geometry to determine if the assessment could detect the expertise reversal effect. Using a rapid assessment, students were placed in a low or high knowledge group. Participants in the two groups were then randomly assigned to a worked example or conventional problem group, yielding four experimental groups: low knowledge worked example, low knowledge conventional problem, high knowledge worked example, and high knowledge conventional problem. Each group received instruction by the designated method and took a rapid assessment using the first-step method at the end of the instruction. Results showed a strong expertise reversal effect with the low knowledge worked example group scoring significantly higher than the low knowledge conventional problem group, and the high knowledge conventional problem group scoring higher than the high knowledge worked example group. Kalyuga and Sweller (2004) concluded that rapid assessments using the first step method possessed high validity in the domain of algebra because 1) the scores of the rapid assessments using the first step method were highly correlated to the scores of traditional tests and 2) rapid assessments were able to differentiate between high and low knowledge students.

Kalyuga (2006d) assessed the external validity of the first-step method in the domain of 8th grade arithmetic word problems. Two classes participated. One class received a rapid assessment using the first-step method, followed by a traditional test. The other class received a traditional test, followed by a rapid assessment. The traditional test required students to write complete solutions to 20 arithmetic word problems. The rapid assessment included 20 similar problems and required students to write their first solution step. Results showed that students took significantly less time to complete the rapid assessment, and the rapid assessment scores were highly correlated to the traditional test scores with a Pearson product-moment correlation of 0.72. Kalyuga

(2006d) concluded that the rapid assessment using the first-step method possessed high external validity in the domain of 8th grade arithmetic word problems.

Verification Method

The rapid verification method presents tasks to students for a limited time, removes the task information from view, and requires students to rely on memory to verify whether solution steps presented are correct. Students are presented with a sequence of correct and incorrect solution steps. Each series of steps is presented for a limited time, and students are required to immediately respond whether the steps are correct or incorrect.

Kalyuga (2006b) devised the verification method because the first-step method was not technically feasible for certain tasks, such as attempting to assess verbal responses or graphical representations (drawings) by computer. Rather, Kalyuga (2006b) claims the verification method is suitable for all domains and all tasks, “The rapid verification approach allows designing a computerized rapid schema-based assessment for practically any task domain and any type of knowledge” (p. 91).

Only one study has evaluated the external validity of the rapid verification method. Kalyuga (2006c) compared the results of a traditional test to a rapid assessment using the verification method in the domain of 7th grade reading. Thirty-four Grade 7 students received a traditional test and a rapid assessment using the verification method. The traditional test contained eight reading passages, each of which was assessed with four to six multiple choice questions, yielding a total of 42 questions. The rapid assessment using the verification method contained 18 sentences increasing in difficulty, each of which was assessed with four brief statements requiring verification whether the statement was correct or incorrect, yielding a total of 72 statements to verify. Results showed that the rapid verification assessment took significantly less time to complete than the traditional test, and the rapid assessment scores were correlated to the traditional tests scores, with a Pearson product-moment correlation of 0.66. Kalyuga (2006c) concluded that the rapid assessment using the verification method possessed a "sufficiently high degree of predictive utility" (p. 623).

The two methods of rapidly assessing schematic knowledge in long-term working memory using the first-step and verification methods are used by CLT researchers to measure task performance in numerical problem solving domains. In addition, CLT researchers have added a measure of mental effort expended to achieve the indicated task performance, reasoning that mental effort measures add insight about a learner's expertise (Paas & van Merriënboer, 1993). High mental effort to achieve a given task performance may indicate that a learner does not possess highly automated schemas, and consequently possesses relatively low expertise, compared to a learner who requires low mental effort to achieve the same task performance. The next section discusses the methods that CLT researchers use to measure mental effort.

Measuring Mental Effort (Cognitive Load)

To determine expertise levels, CLT researchers measure task performance and mental effort expended to achieve the task performance (Paas & van Merriënboer, 1993). Mental effort is a proxy for cognitive load. A learner that can perform a task with little mental effort, or low cognitive load, is considered to possess higher expertise than a learner that can only perform the same task with a great deal of mental effort, or high cognitive load. According to Paas, Tuovinen, Tabbers, and Van Gerven (2003), the most common methods to measure mental effort include the psychophysiological, secondary task, and rating scale techniques. These three techniques will be discussed next.

Psychophysiological Techniques

Psychophysiological techniques assume that changes in mental effort will affect physiological processes, such as heart, brain, and eye activity. Measures of heart activity include the cardiovascular spectral analysis (CARSPAN) of heart-rate variability (Paas & van Merriënboer, 1994). According to Paas & van Merriënboer, measures of heart activity are intrusive, invalid, and insensitive to subtle fluctuations in cognitive load. Measures of brain activity include neural imaging techniques such as positron emission tomography and functional magnetic resonance imaging (fMRI). These techniques are technically too complex, inconclusive, and impractical for use in authentic learning environments (Brunken, Plass, & Leutner, 2003). Measures of eye activity include the SensoMotoric Instruments (SMI) Remote Eyetracking Device (RED) to monitor pupillary dilation (Van Gerven, Paas, van Merriënboer, & Schmidt, 2004). Although monitoring pupillary dilation seems sensitive to cognitive load fluctuations, especially for young adults, the technique does not perform equally well for older adults. Therefore, measures of eye activity have not been used very often in cognitive load research (Brunken et al., 2003).

Secondary Task Techniques

Secondary task techniques require that subjects perform two tasks concurrently, a primary and a secondary task. The primary task requires the learner to use the instructional method under consideration. The secondary task requires the learner to perform a simple activity, such as detecting a visual or auditory signal. For example, when the color of a letter changes in a small frame above the main task pane, the learner reacts by clicking the mouse or pressing a key on the keyboard (Kalyuga, 2006b). Performance on the secondary task, measured by reaction time and accuracy, is supposed to reflect the level of cognitive load imposed by the primary task. However, Paas et al. (2003) claim secondary task techniques can interfere with the primary task, and therefore, the techniques are not used very often in cognitive load research.

Rating Scale Techniques

Rating scale techniques assume that people can subjectively report the mental effort expended on a task. The technique usually employs a questionnaire with a seven or nine-point rating scale, ranging from very, very low mental effort (1), to very, very high mental effort (9). According to Paas & van Merriënboer and Adam (1994), two studies have evaluated the usefulness of the rating scale: Paas (1992) and Paas and van Merriënboer (1994). Descriptions of the two studies follow.

Paas (1992) evaluated the internal and external validity of the mental effort rating scale in the domain of statistics. Participants in the study were randomly assigned to 1) a conventional group, required to solve 12 conventional problems; 2) a worked group, required to study 8 worked examples and to solve 4 conventional problems; and 3) a completion group, required to solve 8 completion problems and 4 conventional problems. After each example or problem, the participant rated the perceived amount of mental effort invested in studying the example or solving the problem using a 9-point rating scale. After the training session, all participants completed a 12-problem near-transfer test and 12-problem far-transfer test and rated the mental effort expended on each problem. Therefore, there were a total of 28 common problems that the participants solved. Each group solved at least 4 conventional problems during training plus 12 near-transfer and 12 far-transfer problems during the posttest. Using these 28 problems, a coefficient of reliability (Cronbach alpha) of 0.90 was obtained for the mental effort rating scale. To measure the sensitivity of the rating scale, the researchers found significant differences in mental effort by group. Perceived mental effort was significantly higher for the conventional problem group than for the worked example and completion problem groups. Therefore, the rating scale was sensitive to instructional method differences predicted by CLT. The researchers also found significant differences in mental effort by problem type. Perceived mental effort was highest for the far-transfer test problems and lowest for the training problems. For the test problems, the perceived mental effort was higher for the far-transfer problems than the near-transfer problems. Therefore, the mental effort ratings increased with task complexity as predicted by CLT. Paas (1992) concluded that the mental effort rating scale possessed high internal and external validity.

Paas and van Merriënboer (1994) evaluated the internal and external validity of the rating scale in the domain of geometrical problem solving with computer numerical controlled (CNC) programming. Participants in the study were randomly assigned to four groups: 1) a high variability conventional (HVC) group, required to solve 6 high variability conventional problems; 2) a low variability conventional (LVC) group, required to solve 6 low variability conventional problems; 3) a high variability worked (HVW) group, required to study 3 high variability worked examples and to solve 3 high variability conventional problems; and 4) a low variability worked (LVW) group, required to study 3 low variability worked examples and to solve 3 low variability conventional problems. After each example or problem, the participant rated the perceived amount of mental effort invested studying the example or solving the problem using a 9-point rating scale. After the training session, participants completed a 6-problem transfer test. Therefore, there were a total of 6 common problems (the 6 transfer problems) that participants solved. Using the 6 problems, a coefficient of reliability (Cronbach alpha) of 0.82 was obtained for the mental effort rating scale. To measure the sensitivity of the rating scale, the researchers found significant differences in mental effort invested by group. Perceived mental effort was significantly higher for the combined conventional problem groups (HVC and LVC) compared to the combined worked example groups (HVW and LVW). Therefore, the rating was sensitive to instructional method differences predicted by CLT. The researchers also found significant differences in mental effort invested by problem type. Perceived mental effort was higher for the transfer problems than the training problems. Therefore, the mental ratings increased with task complexity. Paas and van Merriënboer (1994) agreed with

Paas (1992) that the mental effort rating scale possessed high internal and external validity.

In summary, CLT researchers (Kalyuga, 2006a; Kalyuga & Sweller, 2005) have recently assessed learner expertise with measures of task performance using rapid assessments and measures of mental effort using rating scales. Researchers measure task performance with rapid assessments to determine the sophistication of schematic knowledge in long-term working memory, a memory structure formed when schemas held in long-term memory are brought into working memory (Ericsson & Kintsch, 1995). Rapid assessments also provide timely data to adapt instruction in real-time. Two rapid assessment methods used in research include the first-step and verification methods. The first-step method requires the learner to indicate the first-step to solve a problem and evaluates the sophistication of the step by counting the number of steps skipped. The first step method was evaluated and found to possess high external validity, because the first-step method resulted in scores that were highly correlated with traditional test scores, and the first-step method was able to differentiate between high and low knowledge students. In addition, the first-step method was significantly faster to administer than a traditional test (Kalyuga, 2006a; Kalyuga & Sweller, 2005).

The verification method requires the learner to verify a series of solution steps to a problem, for which the information has been removed from view, and evaluates the sophistication of learner expertise by counting the number of steps verified correctly. The verification method was evaluated and found to possess a high degree of predictive utility because the verification method resulted in scores that were correlated with a traditional test. The verification method was also significantly faster to administer than a traditional test.

Researchers measure mental effort with a seven or nine-point rating scale. Other methods to measure mental effort, such as psychophysiological techniques and secondary task techniques, are generally not used in CLT research because the techniques are too intrusive, complex, impractical, or prone to interfere with the research's primary task. Mental effort rating scales have been evaluated and found to be sensitive, valid, reliable, and unintrusive measures.

CLT researchers measure expertise by combining performance scores and mental effort ratings into a measure called mental efficiency. It is theorized that using a mental efficiency measure should provide a more effective means to adapt instruction than using either a performance measure or mental effort rating alone. The next section describes the research results of adapting instruction with mental efficiency measures.

Adapting Instruction with Mental Efficiency Measures

CLT researchers have recently investigated assessing expertise with mental efficiency measures and adapting instructional methods to expertise levels (Camp et al., 2001; Kalyuga, 2006a, 2006b; Kalyuga & Sweller, 2005; Salden et al., 2004). This section provides an overview of the studies, followed by a description of the individual studies. The section concludes by discussing the conflicting results from the studies that have provided the impetus to conduct the present study.

Overview of Research Studies Employing Mental Efficiency Measures

Recent CLT research has investigated the use of mental efficiency measures of expertise to adapt instructional methods. Mental efficiency measures combine performance measures and mental effort ratings. Performance measures (usually traditional tests or rapid assessments) determine how well a learner accomplishes a task. Mental effort ratings (usually 9-point rating forms) determine how much controlled cognitive processing a learner expends to perform a task (Paas & van Merriënboer, 1993). CLT suggests that using a mental efficiency measure of expertise should provide a better basis to adapt instruction than using either a performance measure or mental effort rating alone.

Rationale for Use of Mental Efficiency Measure

According to CLT, mental efficiency should provide a more useful measure of expertise to adapt instruction than either a performance measure or a mental effort rating alone (Camp et al., 2001; Kalyuga, 2006a, 2006b; Kalyuga & Sweller, 2005; Salden et al., 2004). In other words, the two measures combined should be better than either measure separately. To illustrate, assume that two learners are able to solve a problem correctly, thereby achieving the same task performance score. Learner A has experience in the domain and solves the problem effortlessly, and Learner B has little experience in the domain and must labor mightily to solve the problem. Learner A probably possesses more extensive and automated schemas than Learner B and is able to solve the problem with lower mental effort. If a performance measure alone were used to assess expertise and adapt instruction, both Learner A and Learner B would advance in the instruction because both were able to solve the problem. However, if a mental efficiency measure were used, combining the performance score and mental effort rating, only Learner A would advance in the instruction (because Learner A both solved the problem and rated the mental effort low), and Learner B would repeat the instruction to further develop and automate schemas until Learner B was able to solve the problem with low mental effort.

Alternatively, assume that two learners expend the same mental effort on a problem, but Learner A solves the problem, while Learner B does not. Learner A probably possesses more extensive schemas than Learner B because Learner A was able to solve the problem. If a mental effort rating alone were used to assess expertise and adapt instruction, both Learner A and Learner B would advance in the instruction because both rated the mental effort low. However, if a mental efficiency measure were used, Learner A would advance in the instruction (because Learner A both rated the mental effort low and solved the problem), and Learner B would repeat the instruction until Learner B was able to solve the problem. Therefore, CLT researchers have proposed using measures of mental efficiency to adapt instruction dynamically.

Description of Research Study Comparison Groups and Adaptation Methods

To evaluate the effectiveness of using mental efficiency measures to adapt instruction, CLT researchers have typically used three or four treatment groups in the adaptive instruction studies: a performance group, mental efficiency group, and a non-adapted control group, and sometimes a mental effort group. Instructional adaptation for the performance group is made in response to test scores. Instructional adaptation for the mental efficiency group is made in response to a combination of test scores and mental effort ratings of the tests. No instructional adaptation is made for the control group. Finally, instructional adaptation for the mental effort group (when used in the study) is made in response to mental effort ratings of the tests. Instructional adaptation is typically made by initially placing learners in a stage of instruction based upon results and/or mental effort ratings of a pretest and requiring learners to repeat a stage or advance to the next stage based upon results and/or mental effort ratings of tests administered during instruction.

The groups are compared with a dependent variable called instructional efficiency. Instructional efficiency is computed in a variety of ways. Using the simplest formula, instructional efficiency is calculated as performance divided by mental effort. For example, if the average performance score on a posttest for Participant A is 15 (out of 20), and the mental effort rating by Participant A of the posttest is 5 (out of 9, which is the highest mental effort), then the instructional efficiency score for Participant A is 15/5 = 3. High instructional efficiency is achieved with a high posttest score and a low mental effort rating. The origin of the instructional efficiency score and various methods of computing instructional efficiency are discussed next.

Description of Instructional Efficiency Dependent Variable

In the past, CLT researchers have measured the effectiveness of instructional methods using “instructional condition efficiency” (Tuovinen & Paas, 2004), referred to herein as instructional efficiency. Recently, CLT researchers have measured the effectiveness of adaptive instruction methods using instructional efficiency, as well. Instructional efficiency has typically been calculated at the conclusion of an experiment with two variables; P and E. P is a performance measure, typically obtained from posttest scores. E is a mental effort rating, typically obtained from 9-point surveys of mental effort expended on the training session, or alternatively, on a posttest. Tuovinen and Paas (2004) report that most research studies in the past have used mental effort expended on the training session, not on a posttest, because the studies have evaluated instructional methods that attempt to reduce extraneous cognitive load during the training session. Therefore, the researchers were attempting to identify the instructional method that generated the highest performance score on the posttest combined with the lowest mental effort expending during the training session. In contrast, a few studies (e. g., Paas & van Merriënboer, 1993) have used mental effort expended on the posttest, not on the training, because the studies were principally interested in evaluating instructional methods that increased germane cognitive load during the training session to foster high performance on a far-transfer posttest. Therefore, the researchers were attempting to identify the instructional method that generated the highest performance score on the posttest combined with the lowest mental effort expended during the posttest, reasoning that high mental effort expended during the training session should help lower mental effort expended during the posttest.

According to Paas and van Merriënboer (1993), instructional efficiency is typically calculated with a so-called “two-dimensional” (2D) measure, combining performance scores and mental effort ratings (of the training session or a posttest). Because the performance scores and mental effort ratings use nonequivalent scales, raw data is converted into z-scores (standardized with a mean of 0 and a standard deviation of

1). The z-score coordinates (performance score, mental effort rating) are placed in a graph, where the y-axis represents performance, and the x-axis represents mental effort. Instructional efficiency (equation provided below) is calculated as the distance from the z-score coordinates point on the graph to a 45° diagonal line extending from the lower left quadrant of the graph through point (0,0) to the upper right quadrant. The diagonal line represents Performance = Mental Effort (P=E). High instructional efficiency results when performance exceeds mental effort, and the resulting point is above the diagonal line. Instructional efficiency is calculated as $(P - E)|/\sqrt{2}$ where P is the average performance z score, and E is the average mental effort z score. The equation is a derivation of the general formula to calculate the distance from a point with coordinate (x, y) to the line, $ax + by + c = 0$ in an X-Y coordinate system.

As mentioned earlier, the 2D instructional efficiency measure uses either a mental effort rating of the posttest or mental effort rating of the training session. To incorporate both measures into the computation of instructional efficiency, Tuovinen and Paas (2004) suggest using the following equation: $(P - E_L - E_T)/\sqrt{3}$, where P represents the average z-score for test performance; E_L represents the average z-score for mental effort rating of the training; and E_T represents the average z-score for mental effort rating of the posttest. Tuovinen and Paas contend that using both mental effort ratings (of the training and the posttest) provides a more representative indicator of the overall instructional process, including both the training session and posttest.

Three methods of computing instructional efficiency have been used in CLT-based adaptive instruction research: 1) Camp et al. (2001), using an air traffic control simulation, converted raw data into z-scores and computed instructional efficiency as (P-E)/√2, where P represents performance scores on training session tasks, and E represents mental effort ratings of training session tasks; 2) Salden et al. (2004), in a partial replication of the Camp et al. study, converted raw data into z-scores and computed instructional efficiency as (P–E–T)/ √3, where P represents performance scores on a posttest, E represents mental effort ratings of training session tasks, and T represents total training time; and 3) Kalyuga (2006a), in a kinematics study, used raw data to compute instructional efficiency as P/E, where P represents knowledge gains (posttest – pretest scores), and E represents mental effort ratings of the training session tasks.

Camp et al. used measures of performance and mental effort of training session tasks to determine which adaptive method resulted in the ability of participants to best perform the training session tasks with low mental effort. Salden et al. used posttest scores, rather than training session task scores, to determine which adaptive method resulted in the ability to apply skills to new situations provided by a far-transfer posttest. Salden et al. also included total training session time as a variable, reasoning that two adaptation methods could result in the same performance scores and mental effort ratings, but if one adaptation method required less total training time, the method should be considered more efficient. Kalyuga elected not to convert raw data into z-scores and to use a simpler formula than the formulas used in the other studies, which computed the distance from the instructional efficiency coordinate on a graph to the P = E line.

Summary Results of Research Studies Employing Mental Efficiency Measures

Instructional efficiency scores are used in CLT adaptive instruction research to compare the different treatment groups. As discussed earlier, CLT would suggest that adapting instruction with mental efficiency measures should result in higher instructional efficiency than adapting instruction with performance measures alone. However, research results have not conclusively verified the utility of using mental efficiency measures relative to using performance measures alone (Camp et al., 2001; Kalyuga, 2006a; Salden et al., 2004). Camp et al. (2001), Kalyuga (2006a), and Salden et al. (2004) found that using a mental efficiency measure to adapt instruction was more efficient than not adapting instruction but that using a mental efficiency measure was no more effective or efficient than using a performance measure alone. In fact, the Camp et al. study found that using a measure of performance alone achieved higher instructional efficiency than using a measure of mental efficiency. In the Salden et al. and Kalyuga (2006a) studies, there were no statistical differences between using a measure of performance alone and using a measure of mental efficiency. Because conflicting results have been found between the studies and the results predicted by CLT, each individual study will be discussed in more detail next.

Description of Research Studies Employing Mental Efficiency Measures

Two groups of CLT-based adaptive instruction studies have been conducted. The first group (Camp et al., 2001; Salden et al., 2004) employed air traffic control simulation software to measure learner expertise and adapt instruction, and the second group (Kalyuga, 2006a; Kalyuga & Sweller, 2005) employed rapid assessments to measure learner expertise and adapt instruction. The adaptive model used in the air traffic control simulation studies initially placed all learners in the first difficulty level of the training session and measured expertise with a composite score of task performance during the training session to advance learners forward or backward in difficulty level. The adaptive model used in the rapid assessment studies measured expertise with a pretest and initially placed learners in a training session difficulty level and stage corresponding to learner expertise and measured expertise with rapid assessments administered during the training session to advance learners to the next stage or require learners to repeat the existing stage. The studies employing the air traffic control simulation software will be discussed next, followed by the studies employing rapid assessments.

Camp, Paas, Rikers, and van Merriënboer Study

Camp et al. (2001) conducted a study in the domain of air traffic control training. The researchers randomly assigned participants to one of four groups: performance, mental effort, mental efficiency, and a non-adapted control group. The study employed a computer simulator in which the participants entered keyboard commands to guide simulated aircraft safely to an assigned gate. The researchers created a database of 70 tasks at 10 difficulty levels. The difficulty levels varied by the number, direction, and altitude of the aircraft controlled by the learner.

All groups started at Difficulty Level 1. Tasks were chosen at random for the performance group from a difficulty level that was determined by the performance score of the previous task. The performance score (1 to 5) was based on the number of commands issued, time aircraft spent outside airspace, separation of aircraft, and successful gate hits. Depending on the performance score, the next difficulty level chosen was easier or harder by up to four difficulty levels. Tasks were chosen at random for the mental effort group from a difficulty level that was determined by the mental effort rating of the previous task. Depending on the mental effort rating, the next difficulty level chosen was easier or harder by up to four levels. Tasks were chosen at random for the mental efficiency group from a difficulty level that was determined by the mental efficiency score (Performance – Mental Effort) of the previous task. Depending on the mental efficiency score, the next difficulty level chosen was easier or harder by up to four levels. The non-adapted control group sequentially completed 20 tasks, two at each difficulty level. The three adapted groups finished the training when they had reached Difficulty Level 10 or completed twenty tasks.

Results from a series of ANOVA procedures showed that the mental efficiency group achieved a significantly higher average instructional efficiency score than the non-adapted control group, but contrary to expectations, the mental efficiency group achieved a significantly lower average instructional efficiency score than the performance group. The performance group achieved a significantly higher average instructional efficiency score than all of the other groups (mental effort, mental efficiency, and non-adapted control group), and the mental effort group achieved a significantly higher average instructional efficiency score than the mental efficiency and non-adapted control groups. Camp et al. (2001) concluded that dynamic problem selection leads to more efficient training than a fixed problem sequence, but dynamic problem selection based on mental efficiency does not necessary lead to more efficient training and better transfer than selection based on performance or mental effort alone. The researchers explained that the mental efficiency group achieved lower performance scores and expended higher mental effort than the performance and mental effort groups (resulting in lower instructional efficiency) because the mental efficiency group made larger jumps in difficulty levels. Therefore, the performance and mental effort groups received easier tasks and did not reach as high a level of difficulty as the mental efficiency group. Camp et al. suggested that more research was needed to better match task difficulty to expertise level.

Salden, Paas, Broers, and van Merriënboer Study

Salden et al. (2004) partially replicated the Camp et al. study, making two changes: 1) Salden et al. reduced the jump size in difficulty levels for the adapted groups from four to two, and 2) Salden et al. calculated instructional efficiency differently, using a 3D measure consisting of posttest performance, mental effort ratings of training, and training time (the Camp study used a 2D measure of training performance and mental effort ratings of training). Results from an ANOVA procedure showed that the combined average instructional efficiency score for the performance, mental efficiency, and mental effort groups was significantly higher than the non-adapted control group. However, there were no significant differences in average instructional efficiency scores for the mental efficiency group and a combination of the performance and mental effort groups. Salden et al. concluded, that like the Camp study, dynamic task selection leads to more efficient training than a fixed problem sequence, and that selection based on mental efficiency does not necessarily lead to more efficient training and better transfer than selection based on performance or mental effort alone. Salden et al. also attributed the non-significant differences in instructional efficiency between the mental efficiency group and combined performance/mental effort groups to the larger jumps in difficulty made by the mental efficiency group, who reached the highest difficulty levels faster and practiced more difficult training tasks than the mental effort and performance groups. Salden et al. concluded, "The potential of mental effort and mental efficiency as variables to be used in dynamic task selection needs further study" (p. 171). Therefore, both the Camp et al. and Salden et al. studies found that adapted instruction using either a performance or mental efficiency measure of expertise resulted in higher instructional efficiency than not adapting instruction but could not demonstrate that using a mental efficiency measure of expertise was more effective or efficient than using a performance measure alone.

Kalyuga & Sweller Study

The preceding two studies employed air traffic control simulation software to measure learner performance. In contrast, two recent studies have employed rapid assessments to measure learner performance by assessing schematic knowledge in long-term working memory as learners solve numerical problems in algebra (Kalyuga & Sweller, 2005) and kinematics (Kalyuga, 2006a), using the first-step method and verification methods, respectively. Kalyuga & Sweller (2005) conducted a study in the domain of 10th grade algebra (solving algebra equations of the type: $-3x = 7$, $4x + 3 = 2$, and $(2x + 1)/3 = 2$). The researchers randomly assigned participants to one of two groups: mental efficiency and control. The study consisted of an initial diagnostic test, training session with four difficulty levels, and a final diagnostic test. At each training session difficulty level, students were presented four exercises as follows: Difficulty

Level 1, two worked examples and two conventional problems; Difficulty Level 2, two completion problems (1 step to complete) and two conventional problems; Difficulty Level 3, two completion problems (2 steps to complete) and two conventional problems; and Difficulty Level 4, four conventional problems.

All students took an initial diagnostic test and rated the mental effort of the test using a 9-point rating scale. Students in the mental efficiency group were initially assigned to a difficulty level in the training session based upon a mental efficiency score (performance score ÷ mental effort rating). The performance score was obtained from a rapid assessment using the first step method. Participants were asked to type their first step to solve an algebra equation. More points were awarded for a more advanced first step. Participants were also required to rate the mental effort expended on the rapid assessment using a 1-9 rating scale. Students in the control group were randomly paired with an adapted group student and were initially placed at the same difficulty level as their yoked partner. This yoked design required the control group participants to take the same training sequence as their adapted group counterparts.

Students in the mental efficiency group advanced to the next difficulty level based upon a mental efficiency score from a rapid assessment and mental effort rating administered at the end of the preceding difficulty level. If the mental efficiency score fell below a pre-established level, the participant was required to study two full worked examples and retake the rapid assessment and re-rate the mental effort. If the mental efficiency score fell below a slightly higher pre-established level, the participant was required to study four shortened worked examples and retake the rapid assessment and re-rate the mental effort. If the mental efficiency score fell above the two pre-established levels just mentioned, participants were advanced to the next difficulty level. Students in the control group received the same exercises as their yoked partner in the adapted group but didn't take the rapid assessments or rate the metal effort of the assessments. At the end of the training session all students took a final diagnostic test.

Results from a series of ANOVA procedures showed that the mental efficiency group scored higher efficiency gains than the control group with a significant difference, $p < .05$. Efficiency gains were computed by subtracting initial diagnostic test scores divided by mental effort ratings from final diagnostic test scores divided by mental effort ratings. The mental efficiency group also scored higher knowledge gains (final minus initial diagnostic test scores) than the control group with a significant difference, $p < .10$. Mental effort ratings were not compared between groups. Kalyuga & Sweller (2005) concluded that learner adapted formats using rapid assessments (first-step method) proved to be more effective than non-adapted formats.

Kalyuga Kinematics Study

Kalyuga (2006a) conducted a study in the domain of 11th grade kinematics. He randomly assigned participants to one of three groups: performance, mental efficiency, and non-adapted control. The study consisted of an initial diagnostic test; training session with five difficulty levels, each with five stages; and a final diagnostic test. The five difficulty levels contained kinematics problems with two motion vectors of the following type: 1) one vector was parallel (0°) to the other; 2) one vector was in the opposite direction (180°) to the other; 3) one vector was perpendicular (90°) to the other;

4) one vector was 60° relative to the other; and 5) one vector was 120° relative to the other. Each stage of the difficulty level provided a faded worked example to study and a conventional problem to solve, followed by a rapid assessment and a mental effort rating of the assessment. With each subsequent stage of the difficulty level, the worked example was progressively faded in a foreword fading sequence, gradually eliminating solution steps, until finally at the fifth stage, all solution steps were eliminated, and the worked example only provided the numerical answer.

All participants took an initial diagnostic test and rated the mental effort of the test using a 9-point rating scale. The initial diagnostic test contained five problems corresponding to the five difficulty levels of the training session. The initial diagnostic test employed a rapid assessment using the verification method, in which a problem was presented to the participant, the problem information was removed from view, and a series of solution steps were presented to the participant, requiring the participant to verify whether the series of steps were correct or incorrect. The score of the rapid assessment was used to place performance group participants in the training session. For example, if the participant correctly verified Steps 1 through 3 for Problem 1 on the initial diagnostic test, the participant was placed in Difficulty Level 1, Stage 4 of the training session, bypassing Stages 1 through 3, which correspond to Steps 1 through 3 for Problem 1. A similar procedure was followed for the mental efficiency group participants, but the mental effort ratings of the rapid assessments were used as well. If the mental effort ratings of the rapid assessment were rated a 5 or above (on a 9-point scale), the effect was the same as an incorrect verification on the assessment. The non-adapted control group participants started the training at Stage 1 of the first difficulty level and did not bypass any stages.

At the end of each difficulty level stage, performance group participants took a rapid assessment. If the performance group participants correctly verified the steps in the rapid assessment, the performance group participants advanced to the next stage. Incorrect verifications required the performance group participants to repeat the worked example/ conventional problem pair and retake the rapid assessment. At the end of each difficulty level stage, the mental efficiency group participants also took a rapid assessment and in addition, rated the mental effort expended on the assessment. If the mental efficiency group participants correctly verified the steps in the rapid assessment or rated the mental effort low (a 4 or less on a 9-point scale), the mental efficiency group participants advanced to the next stage. Incorrect verifications or high mental effort ratings (5 or above) required the mental efficiency group participants to repeat the worked example/conventional problem pair, retake the rapid assessment, and re-rate the mental effort of the assessment. The control group participants progressed through each training session stage in sequence, not repeating any stage. Once the five stages for a difficulty level were completed, participants advanced to the next difficulty level, and once the five difficulty levels were completed, participants took the final diagnostic test.

A series of ANOVA procedures for the Kalyuga (2006a) study showed that the mental efficiency group achieved a significantly higher average instructional efficiency score than the non-adapted control group. Like the Salden et al. study and unlike the Camp et al. study, the Kalyuga (2006a) study found no significant differences for instructional efficiency between the mental efficiency group and the performance group. However, unlike the Camp et al. and Salden et al. studies, the Kalyuga (2006a) study found no significant differences for instructional efficiency between the performance group and non-adapted control group. Kalyuga (2006a) concluded that adapting instruction was more efficient than not adapting instruction. However further research was needed to determine the merits of using a mental efficiency measure compared to using a performance measure.

In summary, CLT research suggests that adapting instruction based upon a mental efficiency measure of expertise (a combined measure of performance scores and mental effort ratings) should lead to higher instructional efficiency (higher performance with lower mental effort) than adapting instruction based upon either a performance or mental effort measure of expertise. Adapting with mental efficiency measures will provide additional instruction to learners who achieve high performance with high mental effort because high mental effort could indicate poorly constructed or automated schemas. In contrast, adapting with performance measures will advance learners who achieve high performance with high mental effort to the next stage of instruction, even though the learners may have poorly constructed or automated schemas. Similarly, adapting with mental efficiency measures will provide additional instruction to learners who achieve low performance with low mental effort because low performance could also indicate poorly constructed or automated schemas. In contrast, adapting with mental effort measures will advance learners who achieved low performance with low mental effort to the next stage of instruction, even though the learners may have poorly constructed or automated schemas. However, research has not conclusively demonstrated that using a mental efficiency measure is more effective than using a performance or a mental effort measure. Some research (Camp et al., 2001) has shown that adapting instruction using a performance measure produces higher instructional efficiency than adapting instruction using a mental efficiency measure. Other research (Camp et al., 2001; Kalyuga, 2006a; Kalyuga & Sweller, 2005; Salden et al., 2004) has shown that adapting instruction using a mental efficiency measure produces higher instructional efficiency than not adapting instruction. However, in none of the studies did adapting instruction using a mental efficiency measure achieve higher instructional efficiency than adapting instruction using a performance measure.

Results of the Camp et al. (2001), Salden et al. (2004), and Kalyuga (2006a) studies seem to conflict. Camp et al. showed significantly higher instructional efficiency for the performance group compared to the mental efficiency group. Salden et al. and Kalyuga showed no significant differences between the performance group compared to the mental efficiency group. Rikers (2006), commenting on the results of the Kalyuga (2006a) study, concludes, "it remains an empirical question whether the approaches advocated by Kalyuga proves to be of additional value as compared to previous adaptive methods (e.g. Camp et al.(2001); Salden et al. (2004))" (p. 362). Therefore, additional research is necessary to determine whether adaptive instruction based on a combination of performance test scores and mental effort ratings leads to higher posttest scores and more efficient instruction than adaptive instruction based on performance test scores alone.

Summary of Literature Review

Recently, CLT researchers (Clark et al., 2006; Kirschner et al., 2006; van Merriënboer & Sweller, 2005) have investigated adapting instruction to learner expertise. CLT is a set of principles and guidelines to design and deliver instructional environments that promote learning, utilize the limited capacity of working memory, and minimize working memory overload (Sweller, 1988). CLT assumes a limited working memory of seven elements (Miller, 1956) and an unlimited long-term memory comprised of schemas that vary in complexity and automation. Learning occurs when schematic knowledge is altered in long-term memory. Learning is inhibited when working memory becomes overloaded. Schema construction and automation help alleviate working memory constraints (Sweller, 1988). Schema construction involves building more complex schemas by combining lower level schemas into higher level schemas, thereby "chunking" multiple elements of information into a single element. Automation is the act of converting information into procedural knowledge through problem solving practice. Procedural knowledge requires very little mental attention and effectively bypasses working memory.

Because of working memory constraints, CLT focuses on the work instructional methods impose on working memory, which is called cognitive load. There are three types of cognitive load: intrinsic, germane, and extraneous (Paas et al., 2003). Intrinsic cognitive load relates to the complexity of the material, characterized by element interactivity. Instructional design techniques can not reduce intrinsic cognitive load without also reducing learning. Germane cognitive load, or effective load, contributes to learning and can be increased by instructional design, such as providing example variability or replacing worked examples with conventional problems. Extraneous cognitive load, or ineffective load, does not contribute to learning, is due to suboptimal instructional procedures, and can be reduced by instructional design.

In an attempt to reduce extraneous cognitive load, Sweller and Cooper (1985) discovered that worked examples provide scaffolding support for novice learners. Worked examples provide and explain all solutions steps to a problem. To mitigate the tendency of learners to study worked examples superficially, van Merriënboer (1990) suggested using completion problems. Completion problems are worked examples with blanks that learners fill in. However, the expertise reversal effect (Kalyuga et al., 2001) demonstrated that worked examples or completion problems do not reduce extraneous cognitive load for advanced learners but rather provide redundant information that can actually increase cognitive load. Therefore, worked examples are suggested for novices, and conventional problems are suggested for advanced learners. Conventional problems are open-ended problems, which require learners to solve all of the steps. To transition from novice to advanced learner states, Renkl et al. (2002) suggest fading guidance in a backwards manner, in which worked example solution steps are progressively omitted backwards from the last to the first solution step until all steps are omitted, and the learner is required to solve a conventional problem.

Rather than classifying all students as novices and requiring all students to study worked examples and uniformly follow a fading guidance process, recent CLT research

(Camp et al., 2001; Kalyuga, 2006a; Kalyuga & Sweller, 2005; Salden et al., 2004) has investigated the effectiveness of adaptive learning. Student expertise is assessed with a pretest and the results used to initially place students at different points in the instruction along a worked example-conventional problem continuum. Student expertise is also assessed during instruction and the results used to require repetition of an instructional stage or advancement to a more difficult stage.

Adaptive learning has been extensively researched in the past (Park & Lee, 2004), and from the aptitude-by-treatment interaction research, Jonassen and Grabowski (1993) have cited the importance of prior knowledge and a gradual reduction of instructional support as knowledge increases. However, cognitive load theory extends this research in three ways; 1) CLT provides a unifying theory that helps explain why prior knowledge interacts with instructional methods, 2) CLT provides a measure of cognitive load that should augment measures of prior knowledge and adapt students to different instructional methods more effectively and efficiently, and 3) CLT provides instructional methods, such as worked examples, completion problems, and conventional problems, that improve learning at different prior knowledge levels.

To adapt learning, CLT researchers (Kalyuga, 2006a; Kalyuga & Sweller, 2005) propose measuring expertise with rapid assessments of task performance and mental effort ratings of the effort expended on the task. Rapid assessments measure schematic knowledge structures in long-term memory by assessing long-term working memory through the first-step and verification methods. The first-step method (Kalyuga & Sweller, 2005) asks learners to provide their first-step to solve a problem. Learners ostensibly possess more extensive and automated schemas if the learners can mentally perform many steps and provide an advanced first step. The verification method (Kalyuga, 2006a) asks learners to verify solution steps to a problem. The problem text disappears from learners' view, and the information must be retained in working memory while learners verify the solution steps. Mental effort ratings measure cognitive load imposed while performing a task by assessing the level of mental effort exerted with a 9-point subjective rating scale.

Measuring expertise with rapid assessments and mental effort ratings is called measuring mental efficiency. Using mental efficiency measures (combining performance and mental effort measures) should be a more effective method of adapting instruction than using only a performance measure (Camp et al., 2001; Kalyuga, 2006a; Kalyuga & Sweller, 2005; Salden et al., 2004). If two learners perform a task at the same level and Learner A performs the task with low mental effort and Learner B performs the task with high mental effort, Learner B probably possesses less well developed and automated schemas than Learner A. Therefore, Learner B should repeat instruction, and Learner A should advance in the instruction. If only a performance measure were used to assess expertise, both learners would advance in the instruction. Therefore, CLT suggests that using mental efficiency measures should be a more effective adaptation device than using performance measures alone.

Regarding the efficiency and effectiveness of using mental efficiency measures to adapt instruction, the research is inconclusive (Camp et al., 2001; Kalyuga, 2006a; Kalyuga & Sweller, 2005; Salden et al., 2004). Research has demonstrated that adapting instruction with mental efficiency measures is more efficient and more effective than not adapting instruction, but adapting instruction with mental efficiency measures is not more efficient and effective than adapting instruction with performance measures alone. In fact, some research (Camp et al., 2001) has shown the opposite to be true, that adapting instruction with performance measures alone is more effective than adapting instruction with mental efficiency measures. The results are contrary to that expected by CLT and therefore, additional research is necessary (Rikers, 2006).

CHAPTER 3

METHOD

Introduction

This study tested the theory that instructional methods that adapt dynamically based upon a combined measure of task performance and mental effort will reduce cognitive load during instruction and produce higher learner knowledge gains and more efficient instruction than a measure of task performance alone. The study compared two methods of adaptation (performance and mental efficiency) to a control group and measured final diagnostic test scores, mental effort ratings of the training session, instructional efficiency scores, and instructional times for financial accounting students at a Midwestern university.

Participants were randomly assigned to one of three groups: two adapted groups and one non-adapted control group. A computer-based initial diagnostic test assessed participant knowledge of cost-volume-profit (CVP) analysis, an introductory accounting topic. The results of the initial diagnostic test were used to place the two adapted group participants in a computer-based training session. Members of those two groups skipped sections of the training session for which they demonstrated expertise by their performance on the initial diagnostic test. The two adapted groups were also required to repeat sections of the training session based upon the results of assessments taken during the training session. The control group was placed in the training session at the beginning and worked through the training session sections in sequence, neither skipping nor repeating any sections.

After the training session all groups took a final diagnostic test, identical in structure to the initial diagnostic test, and then rated the overall mental effort expended on the computer based training session (mental effort rating of the training session). Final diagnostic test scores, mental effort ratings of the training session, and instructional times were compared for the three groups to determine the relative effectiveness and efficiency of the groups.

This chapter describes the methods used to collect the data and discusses the following topics: participants, materials (paper-based pretest, computer program, instruments, and method of adaptation), data scoring, design, setting, procedures, and statistical analysis.

Participants

This study was conducted during the Fall semester 2006 at a small mid-western university. Sixty students enrolled in two sections of ACCT 205, Introduction to Accounting, participated in the study. The students were primarily sophomores or juniors majoring in a variety of business subjects. The course met twice a week for 75 minutes in a computer lab over a 14 week semester. The author, who taught the course, had taught the course for 18 years, and had used the computer lab as the classroom for the last 3 years.

A power computation using SPSS revealed that with a power of 0.80, a sample size of 51 was needed to detect an effect size of 0.17, which was the Cohen's f-index for knowledge gains, the lowest effect size found in the Kalyuga (2006a) study, one very close in nature to the study proposed here. Therefore, the sample size of 60 participants formed by the two classes seemed to be sufficient for the proposed study.

Materials

The study phases included a paper-based pretest and a computer program that consisted of an initial diagnostic test, a training session, a final diagnostic test, and a mental effort rating of the training session. Please refer to Figure 2 for a schematic diagram of the study phases. The paper-based pretest was used to assess participant prior knowledge and to evaluate the validity of the initial diagnostic test. The initial and final diagnostic tests were identical in structure. Each test contained four problems varying in difficulty, and each problem contained five steps. The training session consisted of stages, each with a worked example and completion problem pair to instruct the participant step-by-step how to solve problems similar to those provided during the initial and final diagnostic tests. The mental effort ratings of the training session employed a nine-point Lickert scale. The following sections will describe the materials in detail.

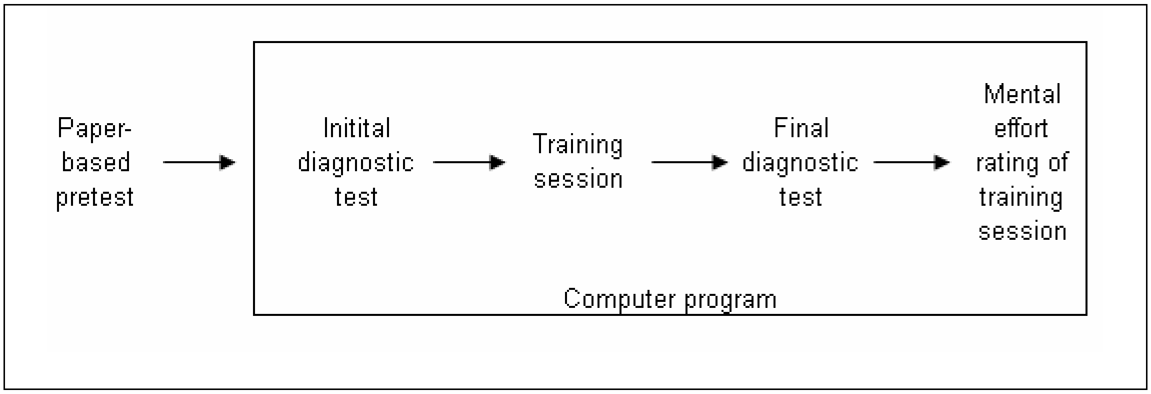


Figure 2. Schematic diagram of study phases

Paper-Based Pretest

A paper-based pretest was administered to assess participants' knowledge of CVP terms and concepts. The test consisted of five multiple choice questions about CVP terminology, five multiple choice questions about the CVP equation, five multiple choice questions about the use of the CVP equation, and one short open-ended problem. Student responses on the paper-based pretest were correlated with student responses on the initial diagnostic test to determine whether the initial diagnostic test was a valid instrument to measure CVP knowledge.

Computer Program

The author used Flash MX to write the web-based computer program to deliver CVP analysis instruction to accounting students. The computer program included an initial diagnostic test, a training session, a final diagnostic test, and a mental effort rating of the training session.

Initial Diagnostic Test

The initial diagnostic test consisted of four problems, each with five steps. The four problems increased in difficulty. For each problem, the procedure was as follows: The participant read the problem text on screen and clicked a "Next" button. Then, the computer removed the problem text, and the participant answered five rapid verification tests, each with a mental effort rating. A rapid verification test presents the results of one or more steps to solve a problem and requires the student to click a "Correct," "Incorrect," or "Don't Know" button. After each test, the participant was presented with a 9-point mental effort rating form to rate the mental effort expended on the test. Each of the five rapid verification tests was cumulative. The participant was asked to verify the results of the current plus any previous steps. For example, Rapid Verification Test 1 presented the results of Step 1 to solve the problem. Rapid Verification Test 2 presented the results of Steps 1 and 2. Rapid Verification Test 3 presented the results of Steps 1, 2, and 3. Rapid Verification Test 4 presented the results of Steps 1, 2, 3, and 4. Rapid Verification Test 5 presented the final solution to the problem, essentially requiring the participant to verify Steps 1 though 5.

To illustrate, a screen shot of Problem 1 for the initial diagnostic test is provided in Figure 3. The participant was required to read the problem text and to click "Next" when finished. Once the participant clicked "Next," the text of the problem shown earlier was removed, and the first of five rapid verification tests was presented, as shown in Figure 4. The first rapid verification test required the participant to verify the results of the first step to solve the problem (determine the target profit). Although the traditional income statement remained on the screen to provide some reference information, the information needed to solve the problem was contained in the problem text removed from view, requiring the participant to maintain

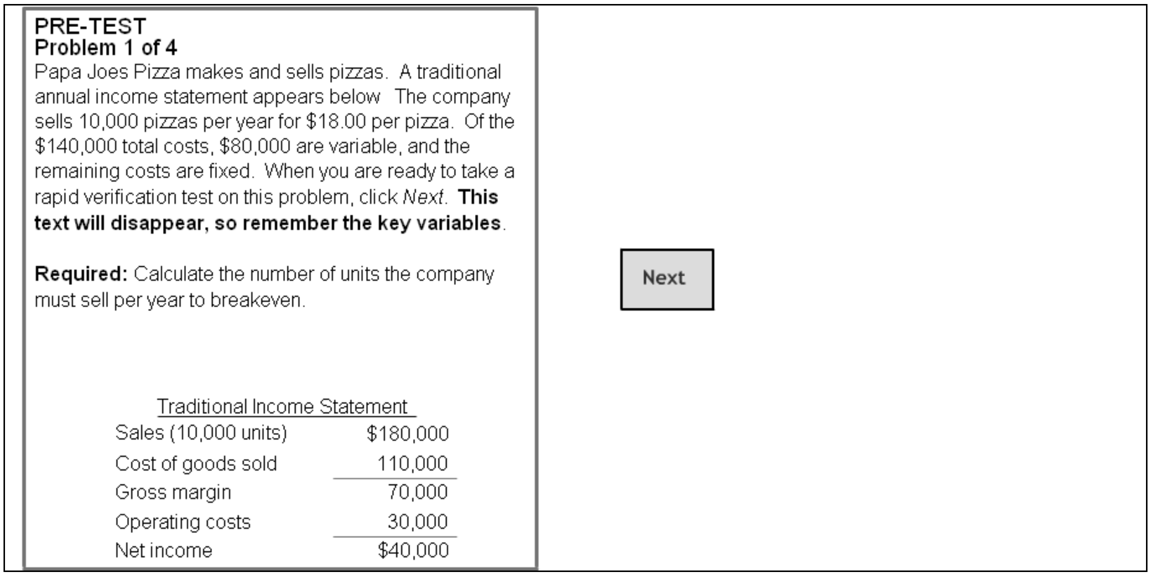


Figure 3. Screen shot of problem text for Problem 1 of the initial diagnostic test (The initial diagnostic test is entitled “Pre-Test” in the software.)

the information in working memory. The five verification tests for each problem were randomly mixed with roughly half correct and incorrect answers. The “Don’t Know” button was neither a correct nor incorrect verification and was automatically scored as an incorrect response. For Problem 1 there were two correct and three incorrect answers. For Problem 2 there were three correct and two incorrect answers. This sequence was repeated for Problems 3 and 4. The answer for the example displayed in Figure 4 is correct.

Once the participant clicked “Correct,” “Incorrect, “or “Don’t know,” he or she was asked to rate the mental effort expended on the verification test from extremely easy to extremely difficult, as shown in Figure 5.

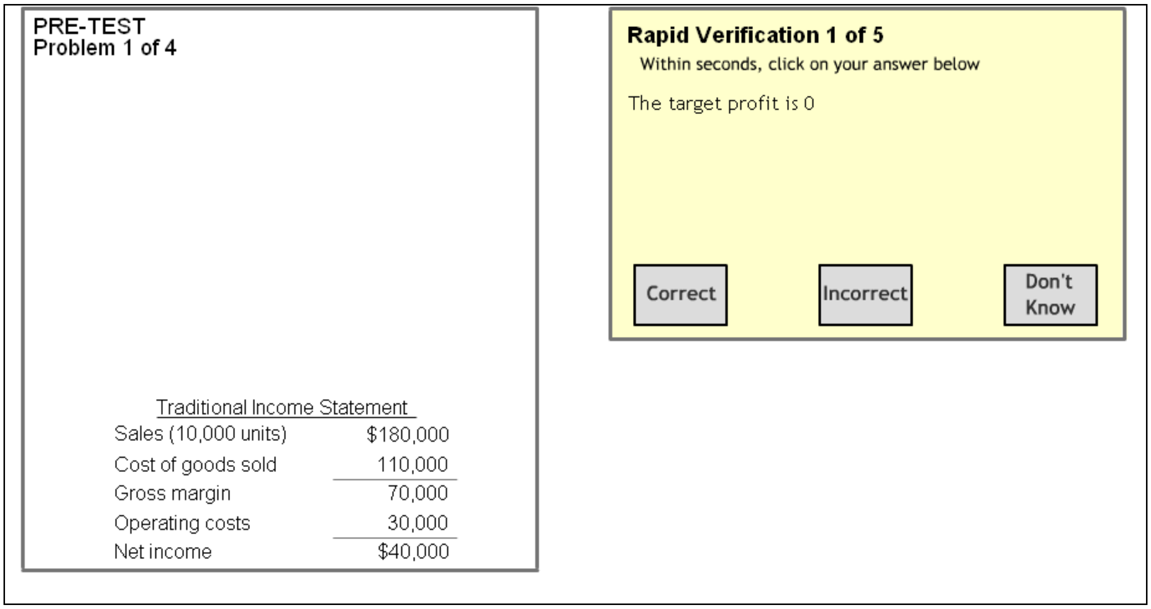


Figure 4. Problem 1 Rapid Verification Test 1 for the initial diagnostic test

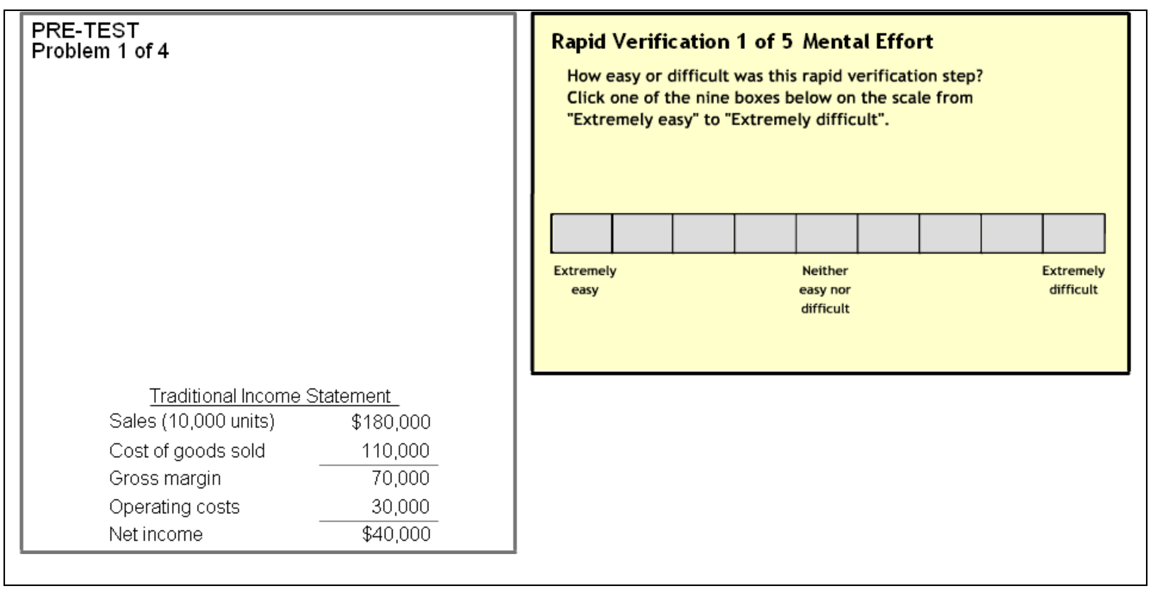


Figure 5. Mental effort rating for Problem 1 Rapid Verification Test 1

Once the participant rated the mental effort expended on the verification test, the second rapid verification test appeared as shown in Figure 6. The second

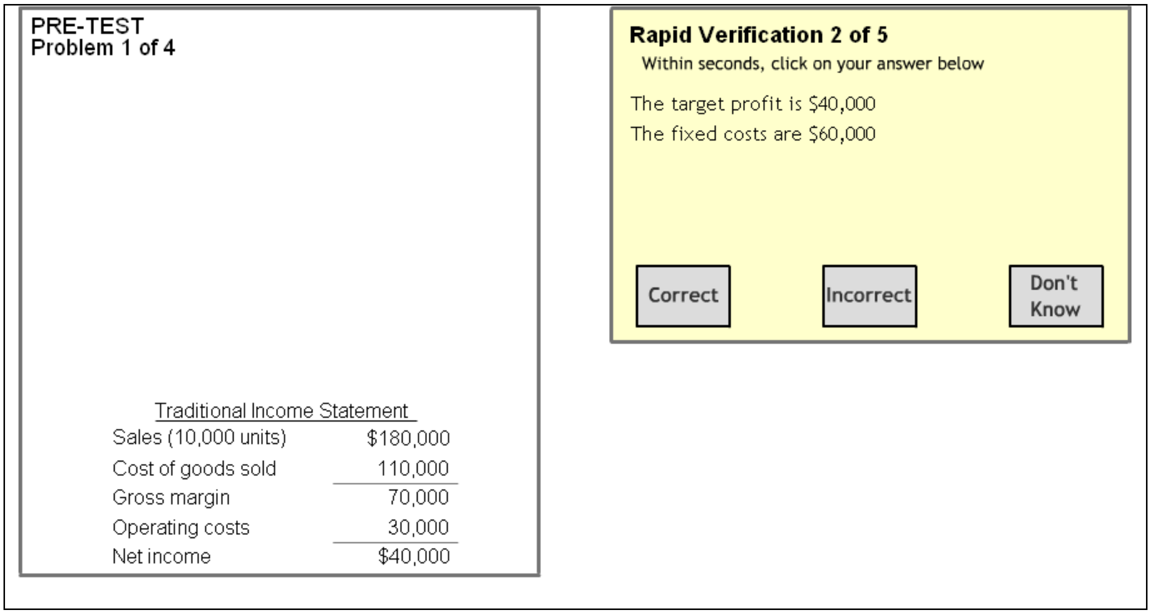


Figure 6. Problem 1 Rapid Verification Test 2

verification test required the participant to cumulatively verify the results of two steps for Problem 1: the target profit (previously verified) and the fixed costs. This required the participant to maintain the result of the previous step (target profit) in working memory as the participant mentally performed the next step, to determine the fixed cost. The answer to the second verification test is incorrect. The target profit is $0, not $40,000 as shown. After the verification test, the participant rated the mental effort expended on the verification test using the same mental effort rating scale used for the previous verification test.

The third rapid verification test, as displayed in Figure 7, required the participant to cumulatively verify the results of three steps for Problem 1: the target profit, the fixed costs, and the unit contribution margin. This required the participant to maintain the results of the previous two steps (target profit and fixed costs) in

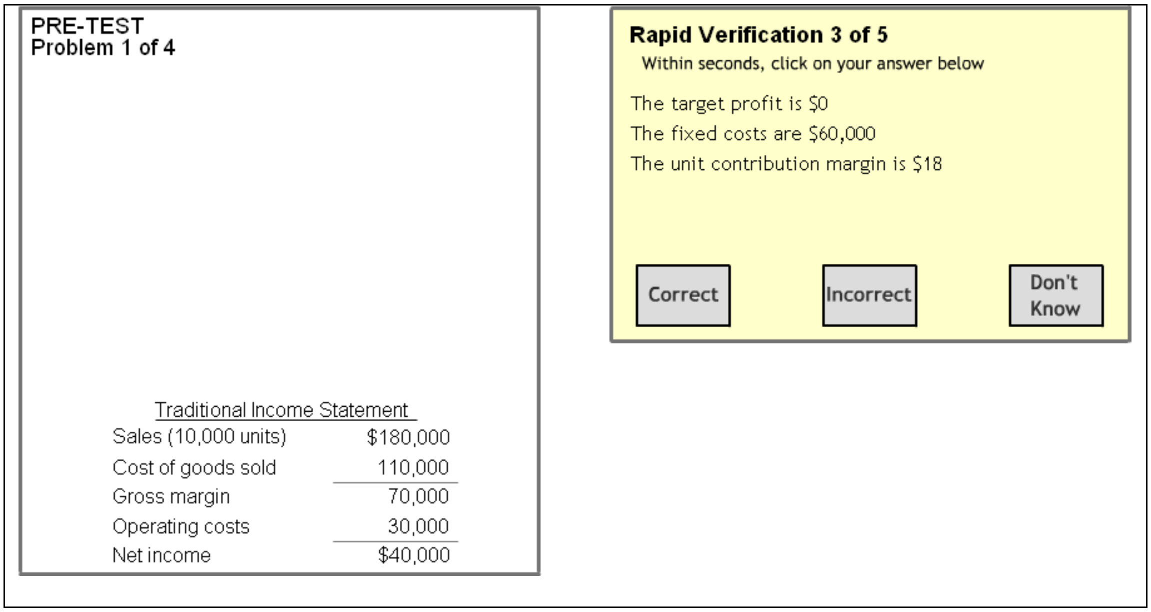


Figure 7. Problem 1 Rapid Verification Test 3

working memory as the participant mentally performed the next step, to determine the unit contribution margin. The answer to the third verification test is incorrect. The unit contribution margin is $10.00, not $18.00 as shown. (The target profit and fixed costs are correct.) After the verification test the participant rated the mental effort expended on the test.

The fourth verification test, as displayed in Figure 8, required the participant to cumulatively verify the results of four steps for Problem1: the target profit, fixed costs, unit contribution margin, and correct placement of the variables in the CVP equation. This required the participant to maintain the results of the previous three steps (target profit, fixed costs, unit contribution margin) in working memory as the participant mentally performed the next step, to place the three variables in the CVP equation. The answer to the fourth verification test is incorrect. The fixed costs are

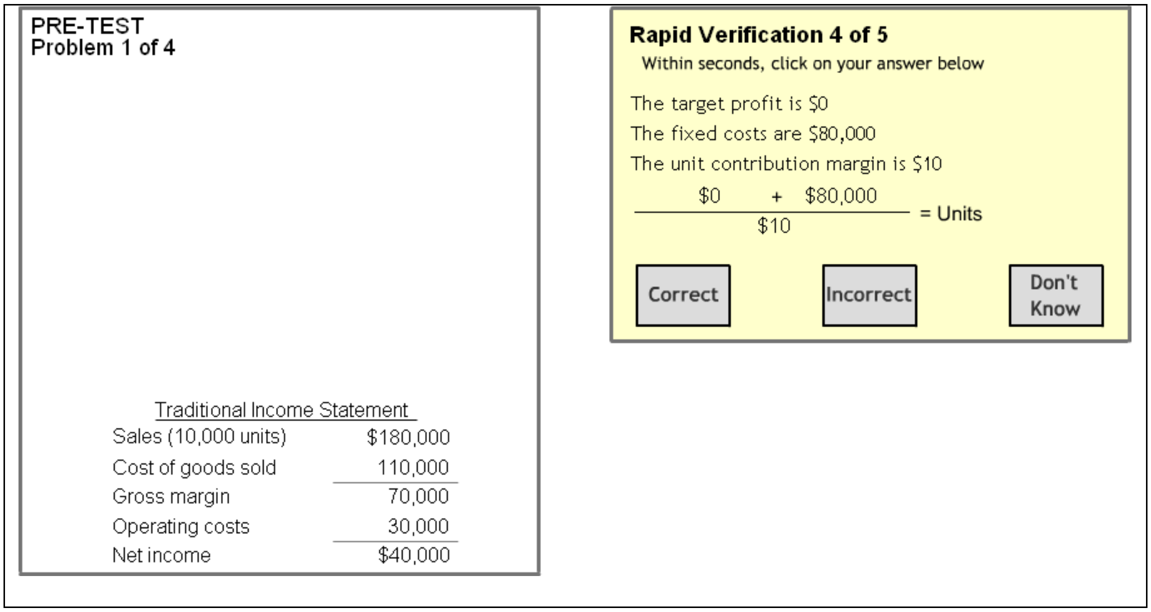


Figure 8. Problem 1 Rapid Verification Test 4

$60,000, not $80,000 as shown. (The target profit, unit contribution margin, and placement of the variables in the equation are correct.) After the verification test, the participant rated the mental effort expended on the test.

Finally, as displayed in Figure 9, the fifth rapid verification required the participant to verify the final answer. This required the participant to mentally place the variables in the CVP equation and compute the final answer. The answer to the fifth verification test is correct. After the verification test, the participant rated the mental effort expended on the test.

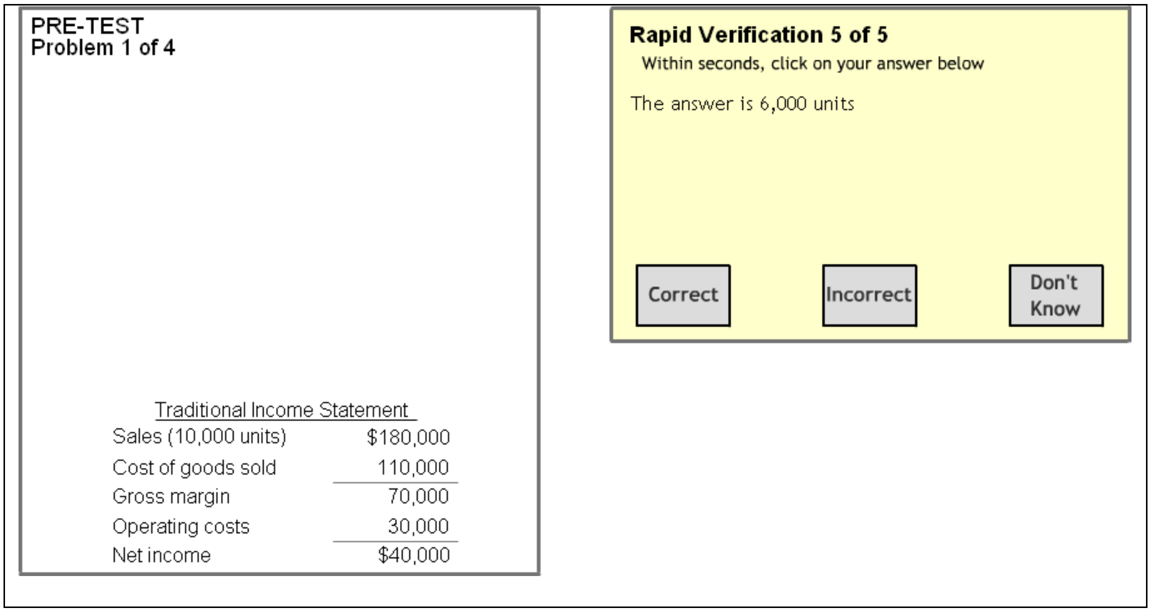


Figure 9. Problem 1 Rapid Verification Test 5

Problem difficulty. Each of the four problems provided in the initial diagnostic test contained five rapid verification tests and five mental effort ratings, as was discussed earlier for Problem 1. The four problems increased in difficulty as shown in Table 4. Problem 1 required the participant to compute the number of units a company must sell to breakeven. Since the target profit is zero, the participant did not expend much effort maintaining the variable in working memory but had to maintain the other two variables (fixed costs and unit contribution margin) in working memory.

Table 4

Initial Diagnostic Test Problems

Problem number	Task
1	Calculate units required to breakeven (i.e., target profit = 0)
2	Calculate units required to earn a target profit (i.e., target profit > 0)
3	Change sales price, unit variable cost, or fixed costs and calculate the units required to breakeven
4	Change sales price, unit variable cost, or fixed costs and calculate the units required to earn a target profit

Problem 2 required the participant to compute the number of units a company must sell to earn a target profit greater than zero. Therefore, Problem 2 was more difficult than Problem 1 because the participant had to remember a number other than zero for target profit in working memory, along with two other variables (fixed costs and unit contribution margin).

Problem 3 required the participant to change one variable (sales price, unit variable cost, or fixed cost) and calculate the number of units the company must sell to breakeven. Problem 3 was more difficult than Problem 2 because the participant had to perform an additional processing operation for Problem 3, not required by Problem 2. For Problem 3 not only was the participant required to remember the fixed costs, unit contribution margin, and target profit equal to zero, the participant had to also mentally perform an additional processing step. For example, if the problem text stated that the fixed costs increased by $10,000, the participant had to first determine the original fixed cost (e.g. $140,000 - $80,000 = $60,000), add $10,000 to the amount ($60,000 + $10,000 = $70,000), and maintain the resulting amount in working memory, along with the unit contribution margin and target profit equal to zero.

Problem 4 required the participant to change one variable (sale price, unit variable cost, or fixed costs) and calculate the number of units the company must sell to earn a target profit greater than zero. Problem 4 was more difficult than Problem 3 because Problem 4 required the participant to maintain in working memory a number greater than zero for the target profit. (For Problem 3 the target profit was zero, requiring much less effort to remember.)

Training Session

The training session consisted of four problem levels, corresponding in difficulty to the four problems provided during the initial diagnostic test. Each of the four training session problem levels contained five stages, corresponding to the five verification tests provided for each problem during the initial diagnostic test. Each stage contained a faded worked example, a completion problem, a rapid verification test, and a mental effort rating, each of which will be described below. Please refer to Figure 10 for a schematic diagram of a typical stage for a problem level. At each subsequent stage the worked example skipped an additional step and each completion problem required the participant to solve an additional step. This process will be more fully explained later. At each problem level the participant worked through the five stages. After completing Stage 5, the participant advanced to the next problem level. After the fourth problem level, the participant advanced to the final diagnostic test. The faded worked examples, completion problems, rapid diagnostic tests, and mental effort ratings are discussed next.

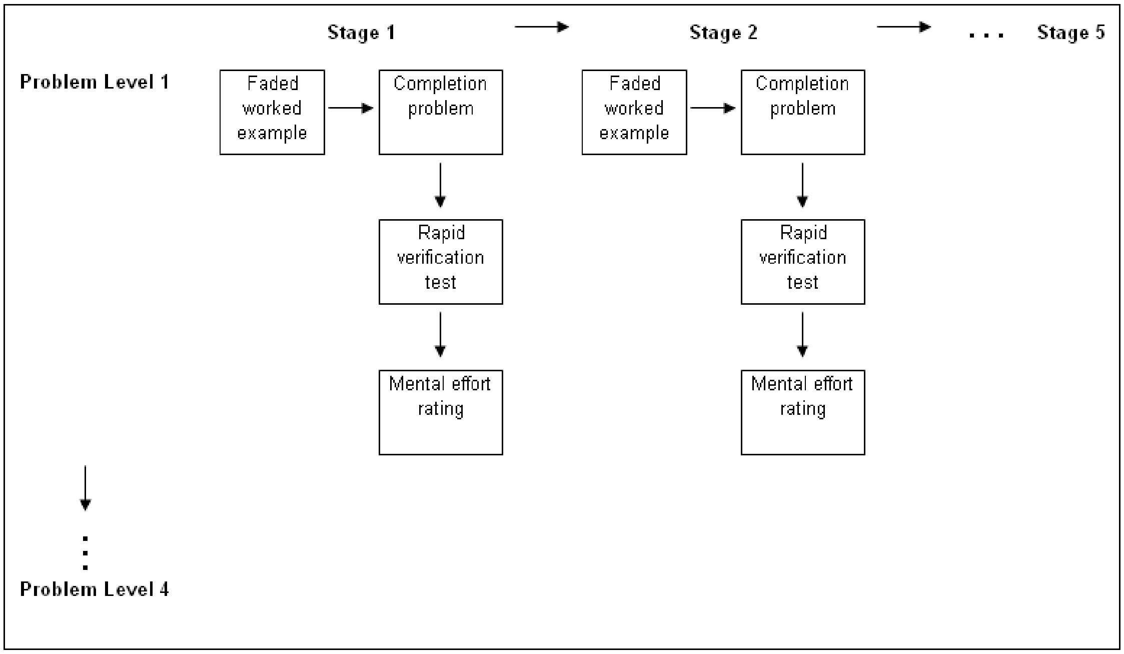


Figure 10. Schematic diagram of training session

Faded worked examples. Faded worked examples initially display all solutions steps, then progressively omit additional steps (Renkl et al., 2002). For the proposed study at Stage 1 the faded worked example provided all solution steps. At Stage 2 the faded worked example omitted the first solution step and provided the remaining four steps. At Stage 3 the faded worked example omitted the first two solution steps, providing the remaining three steps. An example is shown in Figure 11. At Stage 4 the faded worked example omitted the first three solution steps, providing the remaining two steps. At Stage 5, the faded worked example omitted the first four solution steps, merely providing the final answer.

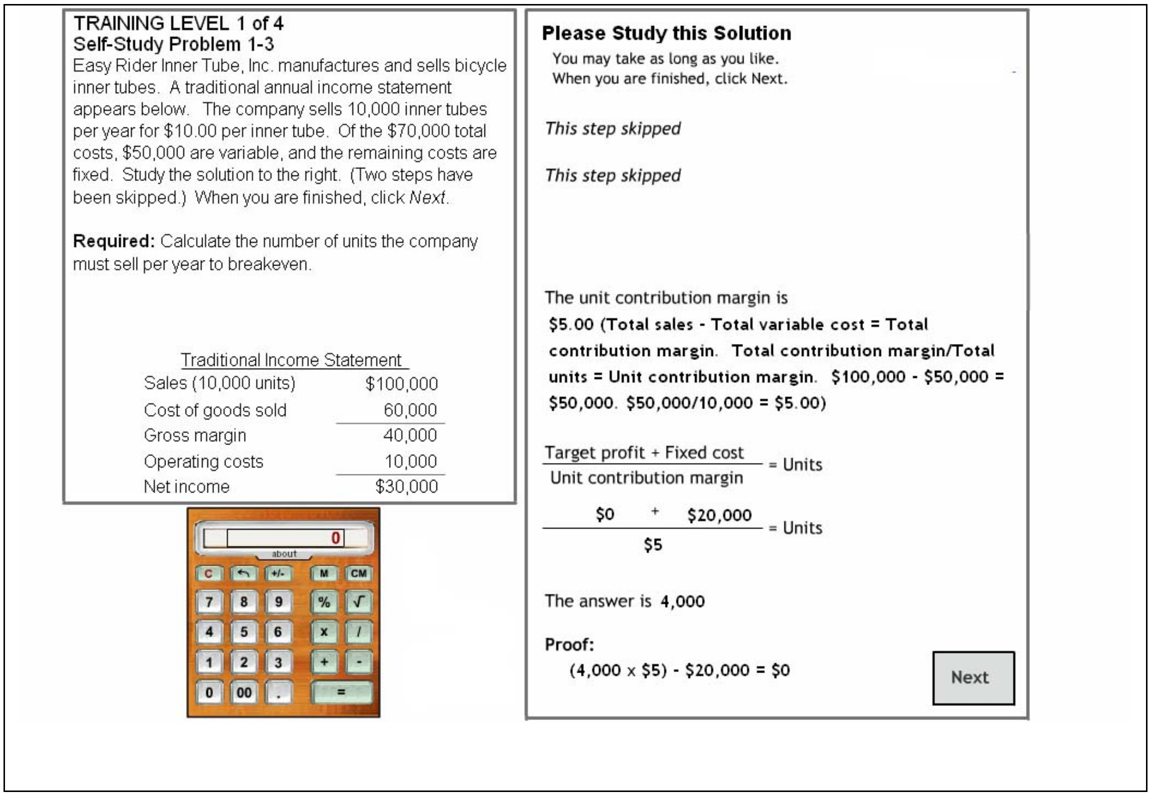


Figure 11. Faded worked example for Stage 3 of Problem 1
(called Self-Study Problem in the software.)

Completion problems. Completion problems show solution steps with one or more steps omitted in a backward fading manner, requiring the participant to provide the answer to the missing steps (van Merriënboer & de Croock, 1992). At Stage 1 the completion problem omitted the last solution step, requiring the participant to provide the answer to the missing step. Participants were permitted three attempts to provide the answer to the missing step and if unsuccessful after those attempts, were provided the correct answer. At Stage 2 the completion problem omitted the last two solution steps, requiring the participant to provide the answer to the missing two steps. At Stage 3 the completion problem omitted the last three solution steps, requiring the participant to provide the answer to the missing three steps, and at Stage 4 the completion problem omitted the last four solution steps, requiring the participant to provide the answer to the missing four steps. At Stage 5, the completion problem omitted all the solution steps, requiring the participant to provide the answer to all five steps, essentially solving a conventional problem.

Rapid verification tests. Following the completion problem at each stage, the participant had to take a rapid verification test, similar to the rapid verification tests provided during the initial diagnostic test with the following difference. The rapid verification tests provided during the initial diagnostic test displayed the problem text, removed the problem text, and displayed five verification tests in sequence. In contrast, the rapid verification tests provided during the training session displayed only one verification test, not five, after the problem text was removed from view.

To illustrate, assume the participant begins at Problem Level 1, Stage 1. The participant studies a worked example, finishes a completion problem, and is presented with the problem text for a rapid verification test. After the participant clicks “Next,” the problem text is removed from view, and the student must verify the results of one step. After the student clicks “Correct,” “Incorrect,” or “Don’t Know,” a mental effort rating appears. Then, the student begins Problem Level 1 Stage 2. The participant studies a worked example with one step skipped, finishes a completion problem with two steps skipped, takes a rapid verification test with two steps to verify, and rates the mental effort expended on the rapid verification test, etc. At Stage 3 the rapid verification test requires the participant to verify the results of three steps. At Stage 4 the rapid verification test requires the participant to verify the results of four steps (target profit, fixed costs, unit contribution margin, and correct placement of variables in CVP equation). At Stage 5 the rapid verification test requires the participant to verify the correct answer.

Mental effort rating. At each stage following the rapid verifications test, the participant was required to rate the mental effort expended on the test using a rating scale identical to the rating scale used during the initial diagnostic test.

Final Diagnostic Test

The final diagnostic test was identical in structure to the initial diagnostic test but with reworded questions that contain different values. Similar to the initial diagnostic test, the final diagnostic test contained four problems, each with five verification tests and mental effort ratings.

Mental Effort Rating of the Training Session

After participants finished the final diagnostic test, they rated the overall effort expended during the training session. The mental effort rating of the training session was a rating form almost identical to the mental effort rating form used during the initial and final diagnostic tests and training session. As shown in Figure 12, the mental effort rating of the training session asked the participant to rate the mental effort expended during the *entire* training session. The validity and reliability of the instruments used in the proposed study are discussed next.

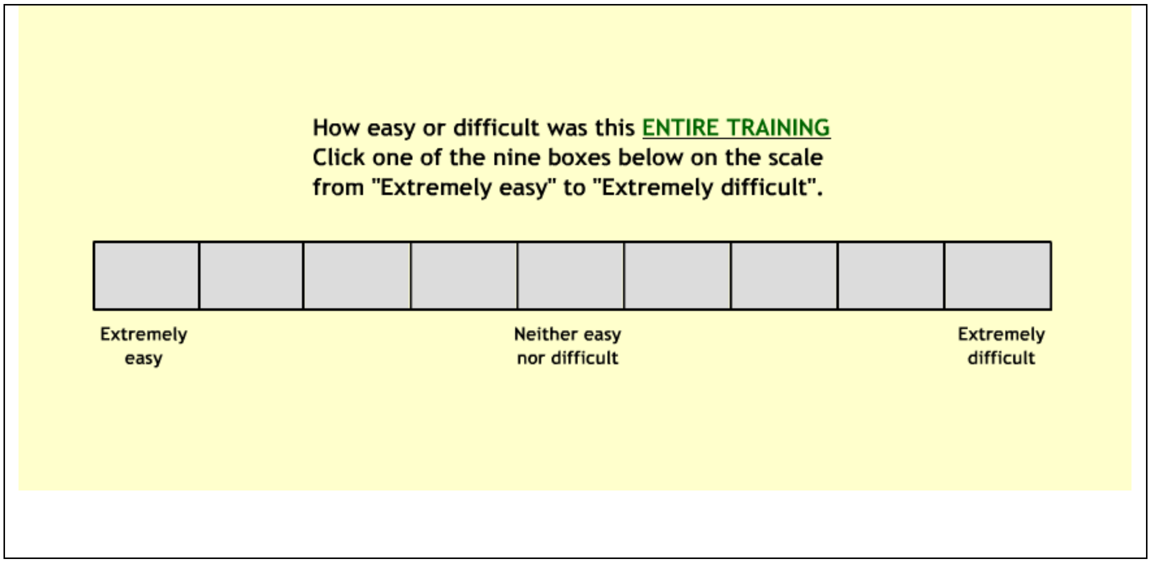


Figure 12. Mental effort rating of the training session form

Instruments

Three types of data collection instruments were used during this study: a paper-based pretest, computer-based rapid verification tests, and computer-based mental effort ratings. The paper-based pretest included traditional multiple choice questions and a short open-ended problem to test participants' knowledge of CVP terms and concepts. The computer-based rapid verification tests, as discussed earlier, presented learners with a problem, removed the problem text from view, and presented learners with the results of one or more solution steps, requiring the learners to verify whether the step or steps were correct (Kalyuga, 2006b). The computer-based mental effort ratings, as discussed earlier, employed a 9-point Likert scale, with which the learners rated the difficulty of the verification tests from extremely easy to extremely difficult (Paas, 1992). Each instrument is described below.

Paper-Based Pretest

A paper-based pretest assessed participants' knowledge of CVP terms and concepts. The test consisted of five multiple choice questions about CVP terminology, five multiple choice questions about the CVP equation, five multiple choice questions about the use of the CVP equation, and one short open-ended problem. Scoring of the tests is discussed in a later section. Student responses on the paper-based pretest (20 points total) were correlated with student responses on the initial diagnostic test (60 points total) to determine whether the initial diagnostic test was a valid instrument to measure CVP knowledge.

Validity and reliability of paper-based pretest. Confirming validity of the paper-based pretest was difficult because no established test to measure the knowledge of CVP was available. Face validity refers to the extent that interested stakeholders consider a test to be a reasonable measure. The paper-based pretest had face validity because it was developed by the author, who is a Certified Public Accountant with eighteen years of experience teaching accounting, and the questions were compiled from end-of-chapter questions and test banks from leading accounting textbooks.

The test-retest reliability of the paper-based pretest was measured by administering the test to participants two separate times: first during the study and second during a class fourteen days later. Reliability refers to the ability of a measure to produce consistent results. Five students were missing from class due to illness during the second administration of the test resulting in a sample size of 48. The test-retest correlation was $r = .60$, $p < .0001$, one tail. Estimates of Cronbach's coefficient alpha were .55.

Cronbach's alpha measures the internal consistency of a test by indicating the homogeneity among the items of the instrument. The internal consistency of the paper-based pretest was lower than the traditionally acceptable .7 to .8. The low score of the paper-based pretest could indicate that some of the items of the test did not differentiate between high and low achievement in the same manner as the entire test, possibly due to a low number of items (16).

Rapid Verification Tests

As was described earlier, rapid verification tests were used during the initial diagnostic test, the training session, and the final diagnostic test. The rapid verification tests were very similar to those used by Kalyuga (2006a) in kinematics (i.e., vector addition motion problems). For example, Kalyuga provided five problems (which he called "tasks") in his initial diagnostic test, and each problem contained five steps (which he called "subtasks"). The problem text was displayed, removed from view, and five verification tests appeared in sequence. To illustrate, Problem 4 of the initial diagnostic test states, "A boat is traveling 6 m/s. A passenger runs across the deck at 6 m/s in a direction of 60° relative to the direction of the boat. What is the velocity of the passenger relative to the water?"

Validity and reliability of rapid verification tests. Several studies (Kalyuga, 2006a, 2006c, 2006d; Kalyuga & Sweller, 2004) have developed and tested the validity of rapid diagnostic tests, using both the first-step and verification methods. The first-step method requires students to indicate their first-step to solve a problem. The verification method, as described earlier, requires students to verify whether one or more solution steps are correct. According to Kalyuga (2006c), "The rapid verification approach allows for the design of online rapid tests for practically any domain and type of knowledge" (p. 617).

Kalyuga and Sweller (2004) conducted two experiments to measure the validity and reliability of a rapid assessment using the first step method. In the first experiment the researchers provided 9th and 10th grade algebra students with a traditional assessment that required students to solve 12 algebra equations of the type: $-4x = 5$. The students were also presented with 12 similar equations and asked to provide their first step only to solve the equations. A correlation of .92 between the traditional and rapid assessment was found. Estimates of Cronbach's alpha for the traditional assessment was .78 and for the rapid assessment was .63. In the second experiment the researchers required 9th grade geometry students to solve 12 coordinate geometry tasks of the type: if lines AC and BC are parallel to the *x*- and *y*-axes, respectively, find the lengths of AC and BC. The students were also presented with 12 similar tasks and asked to provide their first step to find the lengths of AC and BC. A correlation of .85 between the traditional and rapid assessment was found. Estimates of Cronbach's alpha for the traditional assessment was .68 and for the rapid assessment was .57.

Kalyuga (2006d) also conducted an experiment to measure the validity and reliability of rapid assessments using the first step method. Two classes of 8th grade students were presented with two assessments, each containing 20 arithmetic word problems. One assessment represented a traditional assessment in which the students were asked to provide complete solutions to the problems. The other assessment represented a rapid assessment in which the students were asked to present only their first step to solve the problems. Half the students received the traditional assessment followed by the rapid assessment, and half received the rapid assessment followed by the traditional assessment. A correlation of 0.72 was obtained between scores for the traditional and rapid assessments. Estimates of reliability using Cronbach's coefficient alpha were .59 for the traditional assessment and .65 for the rapid assessment.

Kalyuga (2006c) conducted an experiment to measure the validity and reliability of rapid assessments using the verification method. 7th grade students were presented with two assessments, a traditional reading comprehension assessment and a rapid assessment using the verification method. The traditional reading comprehension assessment presented students with eight text passages and asked the students to answer 42 multiple choice questions that measured comprehension of the passages. The rapid assessment using the verification method was computer based and presented students with 18 sentences for a limited time, followed by four brief statements (correct and incorrect) about the sentence, one statement at a time. Students indicated their agreement with the statement by pressing a key on the keyboard and indicated their disagreement with the statement by pressing a different key. A correlation of .66 was obtained between scores for the traditional and rapid assessments. Estimates of reliability using Kudor-Richardson formula (KR20) were .27 for the traditional assessment and .73 for the rapid assessment.

To measure the validity and reliability of rapid assessments using the verification method contained in this proposal, the author presented the participants with a traditional paper-based pretest, discussed earlier and a series of computer-based rapid assessments using the verification method, called the initial diagnostic test. This initial diagnostic test presented participants with four accounting problems, removed most of the information from view, and asked the participants to verify whether a series of steps presented to solve the problems were correct. A correlation of .29 ($p < .035$) was obtained between scores for the traditional and rapid assessments. Estimates of reliability using Cronbach's coefficient alpha were .55 for the traditional assessment and .47 for the rapid assessment.

The relatively low correlation score of .29 did not indicate a high validity for the rapid assessments used in the initial diagnostic test. The initial diagnostic test was provided before the training session, and the final diagnostic test was provided after the training session. To measure whether the initial and final diagnostic tests were able to detect an effect from the training session, a paired sample *t* test was performed for all the participants. Descriptive statistics for this test are found in Table 5. There was a significant differences between the mean scores for the final and initial diagnostic tests ($t[52] = 6.607$, $p < .00001$). An estimate of the effect size using Cohen's d was 1.05, a large effect. Therefore, the initial and final diagnostic tests appeared to detect some effect, but there is question whether the tests provided a valid measure of accounting knowledge compared to a traditional accounting test.

Table 5

Final and Initial Diagnostic Test Paired Sample *t* Test Descriptive Statistics

Assessment	N	Mean	SD	Std. Error Mean
Final diagnostic test	53	36.68	11.03	1.52
Initial diagnostic test	53	25.89	9.47	1.30

Mental Effort Rating

The mental effort rating form used in the proposed study, as discussed earlier, consisted of a 9-point Lickert scale ranging from extremely easy to extremely difficult. The intent of the scale was to measure the mental effort expended when taking a rapid verification test. This scale was a replica of the scale used by Kalyuga (2006a), which in turn was evaluated by Paas and van Merriënboer (1993) and Paas, van Merriënboer, and Adam (1994).

Validity and reliability of mental effort rating. According to Paas and van Merriënboer (1993), mental effort ratings when taking a test or performing a task is a reliable estimate of cognitive load. Paas et al. (1994) evaluated physiological and subjective techniques to measure cognitive load. The physiological techniques assume that changes in mental effort will affect physiological variables, such as heart, brain, and eye activity. According to Paas et al. (1994) measures of heart activity are intrusive, invalid, and insensitive to subtle fluctuations in cognitive load. Measures of brain activity, which include neural imaging techniques such as positron emission tomography and functional magnetic resonance imaging (fMRI), are technically too complex, inconclusive, and impractical for use in authentic learning environments. Measures of eye activity, which include pupillary dilation, are sensitive to fluctuations in cognitive load, but could be affected by other factors such as attention or motivation.

Paas et al. (1994) also evaluated a subjective technique to measure cognitive load, drawing from Paas (1992) and Paas and van Merriënboer (1994). The subjective technique uses a rating scale, which participants use to indicate their mental effort performing a task or solving a problem. Paas et al. (1994) claimed that rating scales are sensitive to relatively small differences in cognitive load, are valid, reliable and non-intrusive, and are the most frequently used measures of cognitive load in cognitive load research.

Method of Adaptation

The method of adaptation chosen for the proposed study was identical to the method of adaptation employed by Kalyuga (2006a). The performance and mental efficiency groups were adapted, and the control group was not adapted. The adaptation of the performance and mental efficiency groups determined at which training session stage the two groups began, which training session stages they skipped, and which training session stages they repeated.

The performance group participants began the training session at a stage determined by rapid verification test results obtained during the initial diagnostic test, skipped training session stages determined by rapid verification test results obtained during the initial diagnostic test, and repeated training session stages determined by rapid verification test results obtained during the training session. The mental efficiency group participants began the training session at a stage determined by rapid verification test results and mental effort ratings of the rapid verification tests obtained during the initial diagnostic test; skipped training session stages determined by rapid verification test results and mental effort ratings of the rapid verification tests obtained during the initial diagnostic test, and repeated training session stages determined by rapid verification test results and mental effort ratings of the rapid verification tests obtained during the training session. In contrast, the control group began the training session at the beginning and progressed through in sequence, neither skipping nor repeating any stage. The following section describes the adaptation for the performance and mental efficiency groups and the instructional sequence provided for the control group.

Performance Group

Rapid verification test results obtained during the initial diagnostic test determined 1) the stage of training at which the performance group was placed initially and 2) stages the performance group subsequently skipped. Rapid verification test results obtained during each stage of the training session determined whether the participant repeated that stage. To make the conditions similar for all three groups, the performance group also rated the mental effort expended on the rapid verification tests taken during the initial diagnostic test and training session, but the mental effort ratings were not used for adaptation. The initial diagnostic test and training session adaptations are discussed next.

Performance group initial training stage placement. The results of the rapid verification tests taken during the initial diagnostic test were used to initially place the performance group in the training session. To reiterate, the initial diagnostic test provided the participants four problems, each with five verification tests. The training session provided the participants with four problems levels, corresponding to the initial diagnostic test problems, and each training session problem level contained five stages, corresponding to the five rapid verification tests provided for each problem during the initial diagnostic test.

Performance group participants were initially placed in the training session problem level and stage corresponding to the first verification test not answered correctly during the initial diagnostic test. For example, if the participant correctly answered the first two verification tests for Problem 1 of the initial diagnostic test and incorrectly answered the third verification test, the participant began the training session at Problem Level 1, Stage 3, skipping the first two stages of Problem Level 1. Similarly, if the participant correctly answered the first four verification tests for Problem 1 and the first verification test for Problem 2 but incorrectly answered the second verification test for Problem 2, the participant began the training session at Problem Level 2, Stage 2, skipping all the stages of Problem Level 1 and one stage of Problem Level 2.

Performance group training stages skipped. The results of the rapid verification tests taken during the initial diagnostic test were also used to skip subsequent training session stages for the performance group after their initial placement in the training session. For each training session problem level the performance group skipped stages that corresponded to consecutive verification tests answered correctly for the corresponding problem on the initial diagnostic test. Referring back to the first example in the previous paragraph, assume that the participant correctly answered the first two verification tests of Problem 1 on the initial diagnostic test but incorrectly answered the third verification test. As was stated before, the participant began the training session at Problem Level 1, Stage 3. Also assume that the participant answered the verification tests for subsequent problems on the initial diagnostic test as shown in Figure 13. If the participant correctly answered the first three verification tests for Problem 2 on the initial diagnostic test, the participant continued the training session at Problem Level 2, Stage 4, skipping training session Stages 1 to 3 of Problem Level 2. If the participant correctly answered the first two verification tests for Problem 3 on the initial diagnostic test, the participant continued the training session at Problem Level 3, Stage 3, skipping training session Stages 1 to 2 of Problem Level 3. If the participant did not correctly answer the first verification test of Problem 4 on the initial diagnostic test, the participant continued the training session at Problem Level 4, Stage 1.

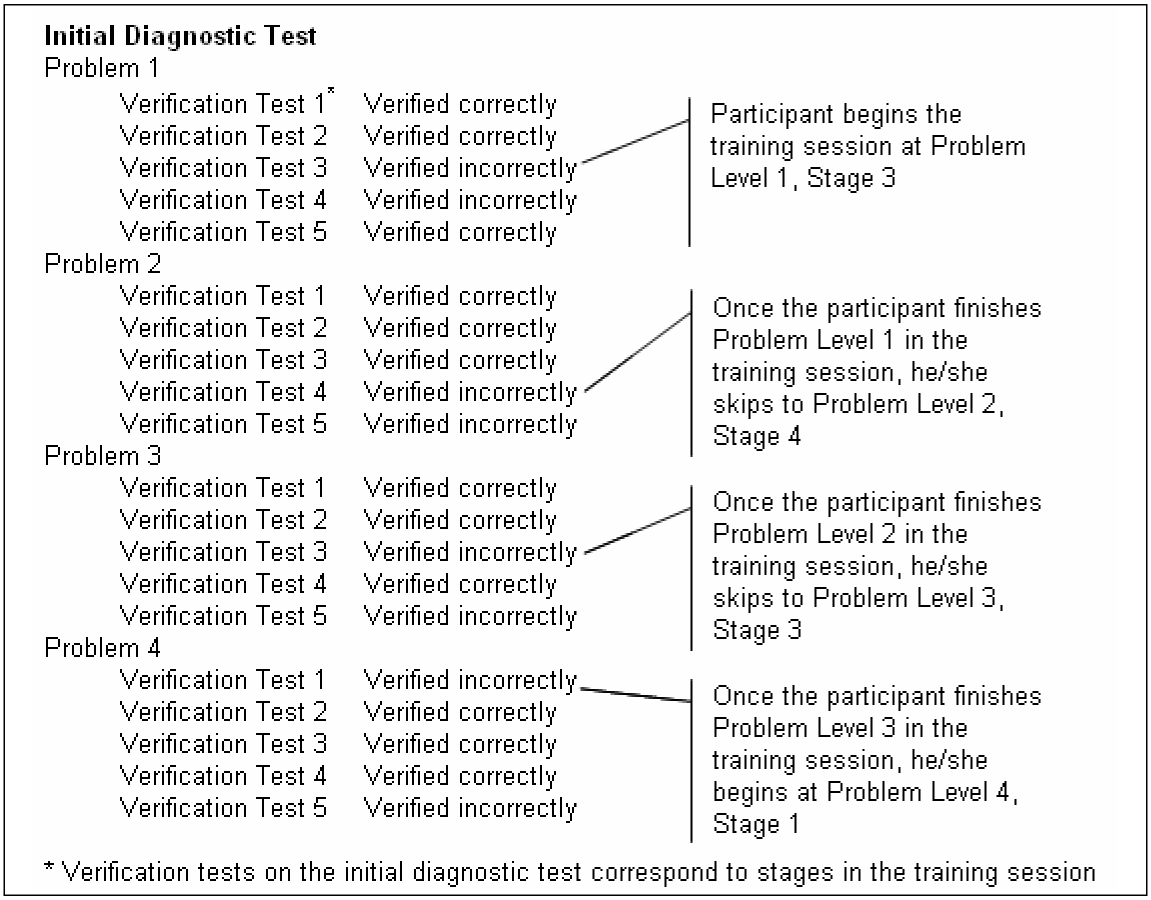


Figure 13. Example to illustrate adaptation for performance group during the training session based upon results of the initial diagnostic test

Performance group training stages repeated. To reiterate, at each training session stage the participant was required to study a faded worked example, finish a completion problem, answer a rapid verification test, and rate the mental effort expended on the rapid verification test. If the performance group participant correctly answered the verification test, he/she received the feedback, “Right, now study the next example” and advanced to the next stage with a new worked example, completion problem, etc. The performance group participant rated the mental effort of the verification test, but the results were not used for adaptation. If the participant did not correctly answer the verification test, he/she received the following feedback, “Sorry, that’s incorrect. Try studying the examples again” and had to study the same worked example, finish the same completion problem, and take the same rapid verification test (and rate the mental effort) again. After the second try of the rapid verification test, the participant was automatically advanced to the next stage. If the participant correctly answered the verification test on the second try, he/she received the same right answer feedback as above. If the participant incorrectly answered the verification test on the second try, he/she received the following feedback, “Sorry. Although that’s incorrect, let’s try the next example.”

Mental Efficiency Group

The mental efficiency group adaptation to verification test answers was exactly the same as the performance group. However, the mental efficiency group also adapted to ratings of the mental effort expended on the verification tests. If the participant rated the effort expended as a 5 or above, the effect was the same as answering the verification test incorrectly. A mental effort rating of 5 corresponds to “Neither easy nor difficult’ on the rating scale reproduced in Figure 14. “Extremely easy” is a mental effort rating of 1, and “Extremely difficult” is a mental effort rating of 9 in Lickert scale measurement units. Therefore, even if the mental efficiency group participant correctly answered the verification test, if he/she rated the mental effort a 5 or above, the effect was the same as an incorrect answer.

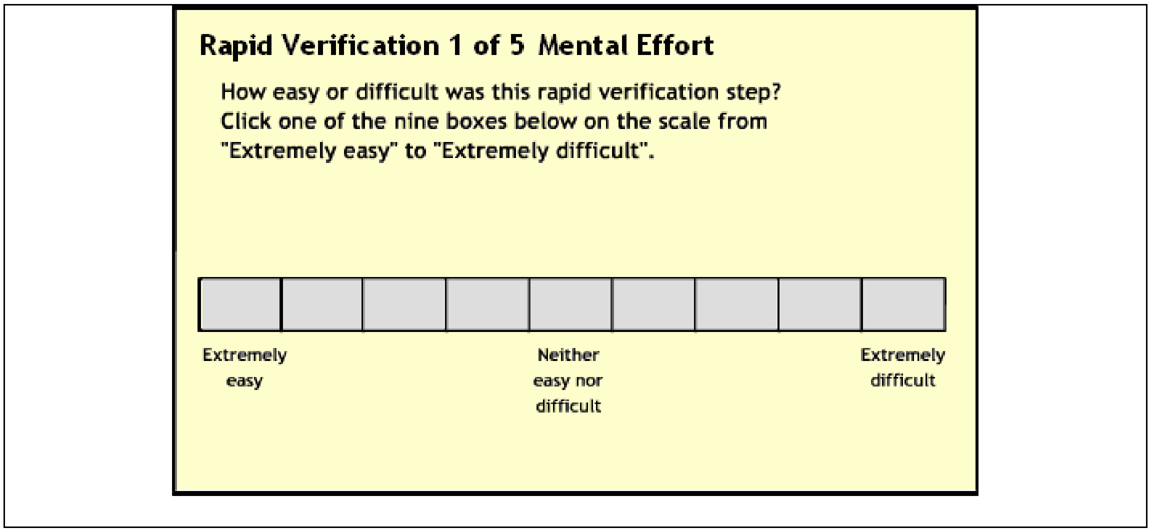


Figure 14. Mental effort rating for rapid verification tests

Mental efficiency group initial training stage placement. The mental efficiency group was initially placed in the training session based upon the results of the rapid verification tests taken during the initial diagnostic test and also the ratings of the mental effort expended on the rapid verification tests. Mental efficiency group participants were initially placed in the training session problem level and stage corresponding to either the first verification test not answered correctly during the initial diagnostic test *or* the first mental effort rating of 5 or above.

To use a similar example as that provided for the performance group earlier, refer to Figure 15. Assume the participant correctly answered the first two verification tests for Problem 1 of the initial diagnostic test and incorrectly answered the third verification test. However, also assume the participant assigned a mental effort of 4 for the first two verification tests but assigned a mental effort rating of 5 for the third verification test. Therefore, the participant began the training session at Problem Level 1, Stage 2, skipping the first stage of Problem Level 1.

Initial Diagnostic Test

Problem 1			
Verification Test 1	Verified correctly	Mental effort rated 4	Participant begins the training session at Problem Level 1, Stage 2
Verification Test 2	Verified correctly	Mental effort rated 5	
Verification Test 3	Verified incorrectly	Mental effort rated 5	
Verification Test 4	Verified incorrectly	Mental effort rated 5	
Verification Test 5	Verified correctly	Mental effort rated 5	
Problem 2			
Verification Test 1	Verified correctly	Mental effort rated 4	Once the participant finishes Problem Level 1 in the training session, he/she skips to Problem Level 2, Stage 3
Verification Test 2	Verified correctly	Mental effort rated 4	
Verification Test 3	Verified correctly	Mental effort rated 5	
Verification Test 4	Verified incorrectly	Mental effort rated 5	
Verification Test 5	Verified incorrectly	Mental effort rated 5	
Problem 3			
Verification Test 1	Verified correctly	Mental effort rated 4	Once the participant finishes Problem Level 2 in the training session, he/she skips to Problem Level 3, Stage 2
Verification Test 2	Verified correctly	Mental effort rated 5	
Verification Test 3	Verified incorrectly	Mental effort rated 5	
Verification Test 4	Verified correctly	Mental effort rated 5	
Verification Test 5	Verified incorrectly	Mental effort rated 5	
Problem 4			
Verification Test 1	Verified incorrectly	Mental effort rated 4	Once the participant finishes Problem Level 3 in the training session, he/she begins at Problem Level 4, Stage 1
Verification Test 2	Verified correctly	Mental effort rated 5	
Verification Test 3	Verified correctly	Mental effort rated 5	
Verification Test 4	Verified correctly	Mental effort rated 5	
Verification Test 5	Verified incorrectly	Mental effort rated 5	

Figure 15. Example to illustrate adaptation for mental efficiency group during initial diagnostic test

Mental efficiency group training stages skipped. The results of the rapid verification tests taken during the initial diagnostic test and ratings of the mental effort expended on those rapid verification tests were also used to skip subsequent training session stages for the mental efficiency group. For each training session problem level the mental efficiency group skipped stages that corresponded to consecutive verification tests answered correctly for the corresponding problem on the initial diagnostic test *and* rated a mental effort of 4 or less. To illustrate, again refer to Figure 15. Assume that the participant correctly answered the first three verification tests for Problem 2 during the initial diagnostic test. However, the participant assigned a mental effort of 4 for the first two verification tests of Problem 2 but assigned a mental effort of 5 for the third verification test of Problem 2. Therefore, once the participant finished Problem Level 1of the training session, the participant began Problem Level 2 at Stage 3, skipping Stages 1 to 2.

Mental efficiency group training stages repeated. Similar to the performance group, the mental efficiency group took a rapid verification test at each training session stage following the completion problem. If the mental efficiency group participant correctly answered the verification test and rated the mental effort a 4 or below for the test, he/she received the feedback, “Right, now study the next example” and advanced to the next stage with a new worked example, completion problem, etc. If the mental efficiency group did not correctly verify the step, no matter how he/she rated the mental effort, he/she received the following feedback, “Sorry, that’s incorrect. Try studying the examples again” and had to study the same worked example, finish the same completion problem, and take the same rapid verification test and rate the mental effort again. If the mental efficiency group correctly verified the step, but rated the mental effort a 5 or above, he/she received the following feedback, “Right, but that seemed difficult for you. Try studying the examples again” and had to study the same worked example, finish the same completion problem, take the same rapid verification test and rate the mental effort again. After the second try of the rapid verification test and mental effort rating, the participant was automatically advanced to the next stage. If the participant correctly answered the verification test on the second try and rated the mental effort a 4 or below, he/she received the same right answer feedback as above. If the participant correctly answered the verification test on the second try and rated the mental effort a 5 or above, he/she received the following feedback, "Right. Although that seemed difficult for you, try studying the next example." If the participant incorrectly answered the verification test on the second try, no matter how the participant rated the mental effort of the verification test, he/she received the following feedback, "Sorry, that's incorrect. Let's try the next example."

Control Group

The control group took all of the verification tests and rated the mental effort expended on the tests, but the results of the verification tests and mental effort ratings were not used to adapt the control group. The control group started the training session at Problem Level 1, Stage 1 and progressed through the training stages, studying all the worked examples, finishing all the completion problems, taking all the verification tests, and rating the mental effort for all the verification tests. The control group neither skipped any stages nor repeated any stages.

During the training session at each stage, the control group took the rapid verification test and rated the mental effort of the test but advanced to the next stage regardless of the verification test answer or mental effort rating. If the control group participant correctly answered the verification test (no matter how he/she rated the mental effort), the participant received the following message, "Right, now study the next example" and advanced to the next stage with a new worked example, completion problem, etc. If the control group participant incorrectly answered the verification test (no matter how he/she rated the mental effort), the participant received the following feedback, “Sorry, that’s incorrect. Try studying the next example” and advanced to the next stage.

Data Scoring

Paper-Based Pretest

For the paper-based pretest the multiple choice questions were scored one point each, and the short-answer response was scored a maximum of 5 points, with partial credit for incomplete or partially correct answers. The maximum possible score for the paper-based test was 20.

Initial and Final Diagnostic Tests

To reiterate, there were four problems provided during the initial and final diagnostic tests, each with five verification tests. Each verification test required the participant to verify the results of one or more steps to solve a problem. Verification Test 1 required the participant to verify the target profit. Verification Test 2 required the participant to verify the target profit and the fixed costs. Verification Test 3 required the participant to verify the target profit, fixed costs, and unit contribution margin. Verification Test 4 required the participant to verify the target profit, fixed costs, unit contribution margin, and the correct placement of the variables in the CVP equation. Verification Test 5 required the participant to verify the final answer with no variables or equation displayed.

As participants progressed through each of the five verification tests, the participants had to hold the results of an additional problem solution step in working memory. Therefore, points were awarded to each verification test based upon the number of solution step results held in working memory. The first rapid verification test was worth 1 point, the second 2 points, the third 3 points, the fourth 4 points, and the fifth 5 points. There were 60 total points possible for each of the initial and final diagnostic tests. Each problem had five rapid verification tests scored as follows: 1 + 2 + 3 + 4 + 5 =15. Since there were four problems, 15 x 4 = 60 total points.

Design

The study employed an experimental non-equivalent control group design (Cohen, Manion, & Morrison, 2003). The participants were not randomly selected from the population of college students but rather were selected from two accounting courses. However, the participants from the two courses were randomly assigned to method of adaptation groups.

The independent variable was the method of adaptation: performance, mental efficiency, and control. All participants were randomly assigned to one of three groups: a performance group, which was adapted with rapid verification tests; a mental efficiency group, which was adapted with a combination of the rapid verification tests and mental effort ratings; and a control group, which was not adapted. This method of adaptation was similar to the Kalyuga and Sweller (2005) and Kalyuga (2006a) studies. The Kalyuga and Sweller (2005) study used two groups: mental efficiency and control, and the Kalyuga (2006a) study used three groups: performance, mental efficiency, and control.

The dependent variables were: final diagnostic test scores; mental effort ratings of the training session; instructional times, which were measured from the beginning to the end of the computer-based training session; and instructional efficiency scores, which were computed by dividing final diagnostic test scores by mental effort ratings of the training session. To reduce error variability and to increase power, the paper-based test scores were used as a covariate in an effort to remove the effects of prior knowledge.

Setting

The study was conducted during a normally scheduled class in a computer lab with Dell X86 PCs, each equipped with Windows XP Professional and a 1024 x 768 monitor. All participants had a computer to use.

Procedures

Before the participants began using the computer program, they took the paper-based pretest. The paper-based pretests were handed out face-down so that all participants started at the same time. The participants were given 10 minutes to complete the pretest.

Next, participants launched Microsoft Internet Explorer and were provided a URL to download the computer program for the study. After the computer program was launched, the web browser navigation bar disappeared, and the participants used the navigation provided by the computer program. Participants selected their name from a dropdown box to log in. The computer program employed the method of adaptation assigned to each participant in advance.

After the participant completed the computer program, the participant's session variables (initial diagnostic rapid verification test scores and mental effort ratings, final diagnostic rapid verification test scores, training session rapid verification test scores and mental effort ratings, mental effort rating of the training session, and total instructional time) were automatically saved to a MySQL database on the Internet through a PHP script and were also printed out as a precaution against the data not saving to the database properly online. Students submitted the printout sheet to the author when exiting the computer lab.

Statistical Analysis

The statistical analysis of the study included the computation of descriptive statistics and a series of balanced, completely between-subjects, one-way, fixed-effects analyses of covariance (ANCOVA) with method of adaptation (performance, mental efficiency, and control) as the independent variable and final diagnostic test scores, mental effort ratings of the training session, instructional efficiency scores, and instructional times as the dependent variables, and paper-based test scores as the covariate to reduce error variability and increase power. A separate ANCOVA was conducted for each dependent variable. An alpha level of .05 was used to determine if there was a significance difference between the three group means.

CHAPTER 4

RESULTS

Introduction

This study was designed to test the theory that instructional methods that change dynamically based upon a combined measure of task performance and mental effort will reduce cognitive load during instruction and produce higher learner test scores and more efficient instruction than an assessment of performance alone. The study compared the independent variable, method of adaptation (control, mental efficiency, and performance), to the dependent variables: instructional times, mental effort ratings of the training session, final diagnostic test scores, and instructional efficiency scores for fifty-three financial accounting students at a Midwestern university.

Participants were randomly assigned to one of three method of adaptation groups (control, mental efficiency, and performance). All participants were given a paper-based pretest immediately before the study's computer-based phase. An Analysis of Covariance (ANCOVA) was performed individually for each of the dependent variables using the paper-based pretest scores as a covariate to reduce error variability and increase power.

Study Results

Since the participants were randomly assigned to the three method of adaptation groups, the groups are assumed to be equivalent in size and demographic characteristics. The groups are also assumed to be equivalent in prior knowledge. A paper-based pretest was administered to participants to determine the level of prior knowledge. The group equivalencies, results of the paper-based pretest, and a series of ANCOVA procedures that use the paper-based pretest scores as a covariate to reduce error variability and increase power are discussed next.

Sample Sizes and Equivalencies

Sample Sizes

Participants in the study were drawn from a pool of fifty-eight students enrolled in two financial accounting classes at a Midwestern University. The original intent was to randomly assign the participants into three groups with as equal sample sizes as possible: 19 in the control group, 19 in the mental efficiency group, and 20 in the performance group. The actual sample size for the control group was the same as planned, 19. However, the actual samples sizes for the other groups differed slightly as follows: 18 in the mental efficiency group and 16 in the performance group, resulting in 5 fewer participants than expected. Four participants were not able to attend class on the day of the study, and one experienced software problems and was omitted from the results.

Sample Equivalencies

Because the subjects were randomly assigned to groups, one can assume that the groups were equivalent in terms of demographic variables, such as age, years in school, number of accounting courses taken and gender. These demographic variables were obtained from the participants by survey, and an Analysis of Variance (ANOVA) performed with method of adaptation group as the independent variable and each of the following separately as the dependent variable: age, years in school, number of accounting courses taken and gender. There were no significant differences ($F < 1$, $p >$

.05) among the three treatment groups for age, years in school, number of accounting courses taken, or gender. This confirmed that the three groups were equivalent in terms of all of the demographic variables.

Paper-Based Pretest Scores

All participants were provided a paper-based pretest before the study's computer-based phase. The pretests were scored and a mean calculated for each group. Descriptive statistics for the pretest scores are provided in Table 6. An ANOVA was conducted to determine if any differences existed among the three groups for mean paper-based pretest score. The results are provided in Table 7. There were no significant differences between the control, mental efficiency, and performance groups ($F = 0.804$, $p > 0.05$) for mean paper-based pretest score.

Table 6

Descriptive statistics for paper-based pretest scores

Method of Adaptation	Mean	SD	N
Control	10.63	3.55	19
Mental efficiency	9.17	4.03	18
Performance	10.00	2.78	16

Table 7

ANOVA table for paper-based pretest scores

Source	Sum of Squares	df	Mean Squares	F	p
Method of adaptation	19.91	2	9.95	0.804	0.453
Error	618.92	50	12.38		
Total	638.83	52			

Instructional Times

Instructional time was captured for each participant from the beginning to the end of the computer-based training session. Descriptive statistics are provided in Table 8. The mean instructional times were analyzed using an ANCOVA with method of adaptation (control, mental efficiency, and performance) as the independent variable, instructional times as the dependent variable, and paper-based pretest scores as the covariate. Results of the ANCOVA are provided in Table 9. There were no significant differences between the control, mental efficiency, and performance groups ($F = 0.405$, $p > 0.05$) for mean instructional time.

Table 8

Descriptive statistics for instructional times
(in seconds)

Method of Adaptation	N	Mean	S D	Adjusted Mean
Control	19	2636.79	405.51	2674.86
Mental efficiency	18	2673.86	816.46	2630.99
Performance	16	2816.24	693.58	2819.38

Table 9

ANCOVA table for instructional times, adjusted for paper-based pretest scores

Source	Sum of Squares	df	Mean Squares	F	p
Method of adaptation	324,211.90	2	162,105.95	0.405	0.669
Pretest	1,888,343.22	1	1,888,343.22	4.716	0.035
Error	19,620,658.08	49	400,421.59		
Total	21,833,213.20	52			

Mental Effort Ratings of the Training Session

After completing the final diagnostic test, all participants rated the overall mental effort expended on the training. Descriptive statistics are provided in Table 10. The mean mental effort ratings were analyzed using an ANCOVA with method of adaptation (control, mental efficiency, and performance) as the independent variable, mental effort ratings of the training session as the dependent variable, and paper-based pretest scores as the covariate. Results of the ANCOVA are provided in Table 11. There were no significant differences between the control, mental efficiency, and performance groups ($F = 0.162$, $p > 0.05$) for mean mental effort rating of the training session.

Table 10

Descriptive statistics for mental effort ratings of the training session
(on a 9-point scale)

Method of Adaptation	N	Mean	S D	Adjusted Mean
Control	19	5.05	1.65	5.07
Mental efficiency	18	4.78	1.63	4.76
Performance	16	5.00	1.83	5.00

Table 11

ANCOVA table for mental effort ratings of the training session, adjusted for paper-based pretest scores

Source	Sum of Squares	df	Mean Squares	F	p
Method of adaptation	0.95	2	0.48	0.162	0.851
Pretest	0.41	1	0.41	0.140	0.709
Error	143.65	49	2.93		
Total	145.01	52			

Final Diagnostic Test Scores

A final diagnostic test was administered to participants after the training session. Descriptive statistics for the final diagnostic test are provided in Table 12. The mean final diagnostic test scores for the control, mental efficiency, and performance groups were analyzed using an ANCOVA with method of adaptation (control, mental efficiency, and performance) as the independent variable, final diagnostic test scores as the dependent variable, and paper-based pretest scores as the covariate. Results of the ANCOVA are provided in Table 13. There were no significant differences between the control, mental efficiency, and performance groups ($F = 1.488$, $p > 0.05$) for mean final diagnostic test score.

Table 12

Descriptive statistics for final diagnostic test scores
(out of total of 60 points)

Method of Adaptation	N	Mean	S D	Adjusted Mean
Control	19	34.37	10.11	33.60
Mental efficiency	18	36.44	11.40	37.31
Performance	16	39.69	11.63	39.63

Table 13

ANCOVA table for final diagnostic test scores, adjusted for paper-based pretest scores

Source	Sum of Squares	df	Mean Squares	F	p
Method of adaptation	322.34	2	161.17	1.488	0.236
Pretest	765.34	1	765.34	7.064	0.011
Error	5,308.97	49	108.35		
Total	6,396.65	52			

Instructional Efficiency Scores

Instructional efficiency scores were computed by dividing the final diagnostic test scores by the mental effort ratings of the training session. Descriptive statistics are provided in Table 14. The mean instructional efficiency scores were analyzed using an ANCOVA with method of adaptation (control, mental efficiency, and performance) as the independent variable, instructional efficiency scores as the dependent variable, and paper-based pretest scores as the covariate. Results of the ANCOVA are provided in Table 15. There were no significant differences between the control, mental efficiency, and performance groups ($F = 0.606$, $p > 0.05$) for mean instructional efficiency score.

Table 14

Descriptive statistics for instructional efficiency scores

Method of Adaptation	N	Mean	S D	Adjusted Mean
Control	19	8.07	5.44	7.72
Mental efficiency	18	9.25	6.56	9.65
Performance	16	9.58	6.42	9.56

Table 15

ANCOVA table for instructional efficiency scores, adjusted for paper-based pretest scores

Source	Sum of Squares	df	Mean Squares	F	p
Method of adaptation	42.43	2	21.22	0.606	0.550
Pretest	164.70	1	164.70	4.70	0.035
Error	1,716.60	49	35.03		
Total	1,923.73	52			

Summary

This study was designed to test the theory that instructional methods that change dynamically based upon a combined measure of task performance and mental effort will reduce cognitive load during instruction and produce higher learner test scores and more efficient instruction than an assessment of performance alone. The results of the study found no significant differences between method of adaptation groups (control, mental efficiency, and performance) for mental effort ratings of the training session, final diagnostic test scores, and instructional efficiency scores. Therefore, the results of this study indicate that instructional methods that change dynamically based upon a combined measure of task performance and mental effort do *not* reduce cognitive load during instruction, produce higher learner test scores and more efficient instruction than an assessment of performance alone

CHAPTER 5

DISCUSSION AND RECOMMENDATIONS

This chapter reports and interprets the findings of this study, outlines potential implications for adaptive instruction, presents recommendations for future research, and discusses the limitations of the study when interpreting the results. The research question that framed this study is as follows:

> Will the use of a mental efficiency measure, which combines a rapid verification test of performance with a subjective rating of mental effort to adapt instruction, lead to higher post-test scores and more efficient instruction than the use of a performance measure alone?

Discussion

The purpose of this study was to determine whether assessing expertise and adapting instruction with a mental efficiency measure (a combined measure of task performance and mental effort expended to perform the task) would reduce cognitive load during instruction and produce higher posttest scores and more efficient instruction than assessing expertise and adapting instruction with a task performance measure alone. The study included a pretest, training session, and a posttest. Fifty-four participants were randomly assigned to one of three groups: two adapted groups (performance and mental efficiency) and one non-adapted group.

Expertise of the performance group was assessed with a rapid verification test, which required participants to verify whether a series of solution steps to a problem were correct. Expertise of the mental efficiency group was assessed with a rapid verification test and a mental effort rating form, which required participants to rate the mental effort expended on the rapid verification test. To adapt instruction, expertise was assessed during the pretest and during the training session. The results of the expertise assessments made during the pretest were used to place adapted group participants in an appropriate stage of the training session. The results of the expertise assessments made during the training session were used to advance adapted group participants to the next training session stage or require adapted group participants to repeat the current stage. Expertise assessments for the non-adapted control group participants were not used to place participants in the training or to require the non-adapted control group participants to advance to the next stage or repeat the current stage. Rather, the non-adapted control group participants were placed in the training session at the beginning and worked through the training session stages in sequence, neither bypassing nor repeating any stages.

The training session included four difficulty levels, each containing five stages. Each stage contained a faded worked example, a faded completion problem, and an expertise assessment. As the participant advanced through each difficulty level, the worked example omitted an additional step and the completion problem required the participant to solve an additional step. For example, at Stage 1, the participants studied a worked example with all solution steps provided and solved the last step of a completion problem, which provided the preceding four steps. At Stage 2, the participants studied a worked example with the first step omitted and solved the last two steps of a completion problem, which provided the preceding three steps, etc. After the training session, all groups took a posttest, identical in structure to the pretest, and then rated the overall mental effort expended on the training session. Posttest scores, mental effort ratings of the training session, and instructional times were compared for the three groups to determine the relative effectiveness and efficiency of the groups.

The present study results indicated that the use of a mental efficiency measure to adapt instruction was neither more effective nor more efficient than the use of a performance measure alone. The average posttest score for the mental efficiency group was not significantly different than the score for the performance group. In addition, the average mental effort rating of the training session for the mental efficiency group was not significantly different than the rating for the performance group. The posttest scores and mental effort ratings of the training session were used to compute instructional efficiency. The average instructional efficiency score for the mental efficiency group was not significantly different than the score for the performance group. In addition, the average instructional time for the mental efficiency group was not significantly different than the time for the performance group.   The present study could not provide any evidence that using a mental efficiency measure of expertise to adapt instruction is more effective or efficient than using a performance measure of expertise. Therefore, the present study can only agree with Sweller (2006), "that using performance as the sole measure during adaptive instruction is just as good as using the more complex efficiency measure—an instructive finding" (p. 357).

There are some studies with findings that are similar to the present study and other studies with findings that conflict. Similar to the findings of Kalyuga (2006a) and Salden et al. (2004), the present study did not find any significant differences for instructional efficiency scores between the two adapted groups (mental efficiency and performance). However, in contrast to the findings of Camp et al. (2001), Kalyuga and Sweller (2005), Kalyuga (2006a), and Salden et al. (2004), the present study did not find any significant differences for instructional efficiency scores between the adapted groups (mental efficiency and performance groups) and the non-adapted control group. The Camp et al. (2001), Kalyuga and Sweller (2005), Kalyuga (2006a), and Salden et al. (2004) studies found that at least one adapted group achieved a significantly higher instructional efficiency score than the non-adapted control group. The Camp et al. (2001) study found that the mental efficiency, performance, and mental effort groups individually achieved higher instructional efficiency scores than the non-adapted control group. The Salden et al. (2004) study found that mental efficiency, performance, and mental effort groups combined achieved a higher instructional efficiency score than a non-adapted control group. The Kalyuga and Sweller (2005) and Kalyuga (2006a) studies found that mental efficiency groups achieved higher instructional efficiency scores than non-adapted control groups.

The following question must then be asked: Why did previous adapted instruction research studies find differences in instructional efficiency between the adapted groups and non-adapted control group, and the present study found no such differences? To discuss this question, the present study will be compared to the Kalyuga (2006a) study, whose methods and procedures the present study emulated.

The present study employed the same research design and procedures as the Kalyuga (2006a) study. Both studies used similar treatment groups: a performance, mental efficiency, and a non-adapted control group. Both studies used similar expertise assessments: rapid assessments using the verification method and mental effort ratings of the assessments. Both studies used a similar pretest and posttest and used expertise assessments to place adapted group participants in the training session and to advance or require participants to repeat training session stages.

However, there were three notable differences between the two studies. First, the domain employed in the Kalyuga (2006a) study was kinematics, whereas the present study was accounting. Yet, both domains require numerical problem solving, and domains requiring numerical problem solving have historically been the subjects for CLT research (Kalyuga, 2006c). Second, the Kalyuga (2006a) study training session provided five difficulty levels, each with five stages, and the present study training session provided four difficulty levels with five stages. Therefore, the present study provided fewer problems for participants to study/solve. Kalyuga (2006a) thought that a longer training session with additional problems would "amplify" the differences among the three treatment groups. Since the present study provided fewer problems than the Kalyuga (2006a) study, a lack of sufficient problems may have contributed to the non-significant differences for instructional efficiency found among the three treatment groups. Third, the Kalyuga study paired a faded worked example with a conventional problem at each training session stage, whereas the present study paired a faded worked example with a faded completion problem. Therefore, the present study substituted a faded completion problem for the conventional problem. It is possible that using faded completion problems, rather than conventional problems, resulted in the present study's lack of significant differences for instructional efficiency between the adapted groups and non-adapted control group. This speculation will be discussed in more depth next, along with research findings that lend support to this speculation.

Instructional Procedure Difference between Present and Kalyuga (2006a) Study

The present study closely followed the instructional procedures followed by Kalyuga (2006a) with two important differences. First, the present study used a faded worked example and faded completion problem pair at each training session stage, whereas the Kalyuga study used a faded worked example and conventional problem pair at each training session stage. Therefore, the present study substituted faded completion problems for conventional problems. Second, the present study provided correct answer feedback after each step supplied by the participant for the completion problems. Using a series of faded completion problems with step-by-step correct answer feedback is called a "fading guidance procedure" (Renkl & Atkinson, 2003; Renkl et al., 2002). A fading guidance procedure is an application of the completion strategy (van Merriënboer, 1990; van Merriënboer & de Croock, 1992), in which training begins with a worked example that includes all of the solution steps to a problem, proceeds with a series of completion problems that provide fewer and fewer steps, and ends with a conventional problem that provides no solution steps. A fading guidance procedure provides feedback after each step the learner supplies, and the completion strategy provides feedback after all of the steps have been supplied by the learner.

Renkl et al. (2002) found that the use of a fading guidance procedure results in higher near-transfer posttest scores than the use of worked example-conventional problem (example-problem) pairs. Renkl's results may help explain why the present study did not achieve significant differences for instructional efficiency between the adapted groups and non-adapted control group, while the Kalyuga (2006a) study did achieve significant differences. The premise is as follows: First, Renkl et al. (2002) have shown that a fading guidance procedure is more effective building near-transfer problem solving skills than an example-problem pair procedure because the fading guidance procedure provides better scaffolding and learner support. Second, the adaptation methods of the Kalyuga (2006a) study may work with a procedure that provides relatively little scaffolding (use of example-problem pairs) because the adapted groups can repeat the example-problem pairs and obtain scaffolding from the repetition of the problem, whereas the non-adapted control group cannot. Third, the adaptation methods of the Kalyuga (2006a) study may not work with a procedure that provides relatively more scaffolding (use of a fading guidance procedure) because the adapted groups gain no further scaffolding from the repetition; complete scaffolding was provided with the first exposure to the problem. Fourth, the adaptation methods used in both the present and Kalyuga (2006a) studies may not be more effective than a fading guidance procedure, but additional research is necessary to test this conjecture. Each of these assertions will be explained next.

Reasons for Effectiveness of Fading Guidance Procedure

Renkl et al. (2002) explained why a fading guidance procedure produced higher test scores and was favored over an example-problem pair procedure. The reason is twofold: First, a fading guidance procedure facilitates successive approximations to a target skill, as recommended by Anderson, Corbett, Koedinger, and Pelletier (1995) and employs methods consistent with the cognitive skill acquisition model, developed by Anderson et al. (1997) and VanLehn (1996). Full scaffolding support is initially provided to the learner and gradually faded out as the learner builds schematic knowledge. According to Renkl et al. (2002), this is compatible with the situated learning model and cognitive apprenticeship approach (Collins et al., 1989), which proposes a smooth transition from modeling, to scaffolded problem solving, to independent problem solving, while instructional support is faded during the transition. In contrast, use of example-problem pairs results in “quite abrupt changes with respect to the demands placed on the learners. After a first example, the learners have to solve a whole problem totally on their own. Under a fading condition, the first problem-solving demand is to generate just a single step, and the demands are only gradually increased” (Renkl et al., 2002, p. 298). Therefore, solving a whole problem early in the learning process places high cognitive load on the learner and impedes schema construction, and solving a faded completion problem places low cognitive load on the learner and facilitates schema construction (Renkl et al., 2002).

Second, a fading guidance procedure provides step-by-step feedback as learners supply the omitted steps for the completion problem. Anderson el al. (1995) recommend that errors should be avoided (or immediately corrected if they occur) when learning near-transfer problem solving rules. Therefore, the fading guidance procedure corrects errors as the errors occur. In contrast, use of example-problem pairs results in learners committing more errors, because of the abrupt change from studying a worked example to solving a whole problem, and does not immediately correct errors, because feedback is only provided after the entire problem is solved (Renkl et al., 2002).

Therefore, the fading guidance procedure allows participants to gradually learn to solve a problem with appropriate scaffolding and feedback support. Consistent with CLT, a smooth transition of scaffolding support should maintain cognitive load within acceptable levels (Renkl et al., 2002). In contrast, the use of example-problem pairs does not provide a smooth transition of scaffolding support or provide immediate feedback and therefore should require more mental effort and impose a higher cognitive load than a fading guidance procedure. As a result, Renkl et al. (2002) found that groups employing a fading guidance procedure scored higher on posttests than a group employing an example-problem pair procedure.

Speculated Impact on Cognitive Load of Repeating Example-Problem Pairs

Repeating example-problem pairs in the Kalyuga (2006a) study may have contributed to the higher instructional efficiency score of the mental efficiency group compared to the non-adapted control group. The mental efficiency group achieved a significantly higher instructional efficiency score (10.43) than the non-adapted control group (3.52). Instructional efficiency was calculated as average knowledge gain divided by average mental effort rating of the training session. The mental efficiency and non-adapted control groups achieved knowledge gains of 23.00 and 18.31, respectively, but the difference in knowledge gains was statistically non-significant. The mental efficiency and non-adapted control groups rated the mental effort of the training session 3.21 and 5.77, respectively, and the difference in mental effort ratings was significant. Therefore, the knowledge gains did not influence the instructional efficiency calculation as much as the mental effort ratings. The mental efficiency group achieved a higher average instructional efficiency score than the non-adapted control group because the mental efficiency group rated the mental effort of the training session lower.

Since the lower mental effort rating by the mental efficiency group contributed to the higher instructional efficiency score, causes of the lower mental effort rating should be explored. Because the present study, when using a faded guidance procedure, did not find significant differences in mental effort between the adapted groups and non-adapted control group, and the Kalyuga study, when using example-problem pairs, did find significant differences, it is suggested that the repetition of example-problem pairs by the adapted groups may have reduced the cognitive load of the adapted groups and consequently led to the adapted groups rating the mental effort of the training session lower.

The use of example-problem pairs may place high cognitive load on all participants but may be reduced for the adapted group participants through the repetition of the example-problem pairs. Consequently, using example-problem pairs may have placed the non-adapted control group at a disadvantage because the non-adapted control group could not repeat the example-problem pair. Use of example-problem pairs places high cognitive load on the learner when switching from the example to the problem (Renkl et al., 2002) and may result in the learner committing errors the first time the problem was solved. Repeating the example-problem pair may reduce the cognitive load and reduce the number of errors committed with each repetition of the problem.

To take an example from the Kalyuga study, consider Stage 1 of the training session, in which the learner must first study a worked example with all the solution steps displayed and then solve a conventional problem with no solution steps displayed. After one minute, the solution to the conventional problem is provided to the learner, and the learner takes a rapid assessment (and the mental efficiency group also rates the mental effort of the assessment). The non-adapted control group participant automatically must advance to the next training session stage. There is no chance to try the problem again and reduce the errors committed. In contrast, the adapted group participant must repeat the example-problem pair if the participant "fails" the rapid assessment (either does not correctly verify the solution steps to the assessment problem or, for the mental efficiency group, rates the mental effort too high). Therefore, some adapted group participants must repeat the conventional problem in the example-problem pair. Because the solution to the problem is provided after the previous try, some portion of the solution step should remain in memory for the adapted group participant. Each time the problem is repeated, more solution steps should remain in memory and be chunked together into a more integrated whole (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). By repeating the same conventional problem, the participant may commit fewer errors and may feel that the problem becomes easier to solve because the participant is able to successively construct solution steps in memory. Consequently, the adapted group participants may rate the mental effort of the training lower than the non-adapted control group. However, further research is necessary to verify this claim.

Of the two adapted groups, the mental efficiency group should repeat more example-problem pairs than the performance group because the mental efficiency group must not only correctly verify the solution steps to the assessment problem but rate the mental effort of the assessment a four or below out of a nine-point scale. Data from the present study computer log files indeed confirm that a higher percentage of mental efficiency group participants repeated each training session stage compared to performance group participants. Although similar data were not provided in the Kalyuga study, assuming that the mental efficiency group of the Kalyuga study also repeated more training session stages than the performance group, it is possible that the mental efficiency group would rate the mental effort of the training session lower than the performance group (who would repeat fewer example-problem pairs) and non-adapted control group (who would not repeat any example-problem pairs). The mental efficiency group of the Kalyuga study rated the mental effort of the training session lower than the non-adapted control group ($p < 0.05$), and the performance group rated the mental effort lower then the non-adapted control group (but somewhat higher than the mental efficiency group) ($p < 0.10$). The mental efficiency group rated the mental effort lower than the performance group and non-adapted control group, but the difference between the mental efficiency and performance groups was not statistically significant. Although the data cited above do not provide conclusive evidence, the pattern of the data and the logic previously discussed provide some evidence that repeating a conventional problem may contribute to the participant rating the mental effort of the training session lower, but additional research is necessary to confirm this effect.

Speculated Impact on Cognitive Load of Repeating a Fading Guidance Procedure

Repeating a fading guidance procedure may not have had as great an impact on participants in the present study as repeating example-problem pairs had in the Kalyuga (2006a) study. It is possible the non-adapted control group of the present study was not placed in as disadvantaged a position as the non-adapted control group of the Kalyuga study. Because a fading guidance procedure facilitates successive approximations to a target skill, repeating a faded completion problem may not lower cognitive load and improve skill acquisition as much as repeating a conventional problem.

To take an example from the present study, consider Stage 1 of the training session, which employed a fading guidance procedure. The learner must first study a worked example with all the solution steps displayed and then solve a faded completion problem with the last solution step omitted but the previous steps displayed. The scaffolding provided by the display of the previous steps reduces cognitive load and allows the learner to devote cognitive resources to solve the single step omitted. Repeating the completion problem merely provides redundant information because all of the solution steps, with the exception of the last, were originally provided, and the repetition provides little new information. Therefore, it is speculated that repeating a completion problem may not provide the same benefit as repeating a conventional problem, but additional research is needed to verify this claim.

Speculated Failure of Current Adaptation Methods to Improve Fading Guidance Procedure

It is possible that the adaptation methods used in the present and Kalyuga (2006a) studies do not improve upon a fading guidance procedure. However, additional research is needed to verify this claim. The Kalyuga (2006a) study, when employing an example-problem pair procedure, found significant differences in mental effort and instructional efficiency between one adapted group (mental efficiency group) and the non-adapted control group. In contrast, the present study, when employing a fading guidance procedure, found no significant differences in mental effort and instructional efficiency among any of the three treatment groups. Since a fading guidance procedure is designed to gradually reduce instructional support at each training session stage and to provide a smooth transition from scaffolded problem solving to independent problem solving, adapting instruction (by repeating or bypassing training session stages) using a fading guidance procedure does not seem to reduce cognitive load and increase instructional efficiency.

Summary: Instructional Procedure Difference Between Present and Kalyuga (2006a) Study

The present study found no significant differences in any dependent variable among the three treatment groups. This contrasts with the Kalyuga (2006a) study, which found significant differences in mental effort ratings and instructional efficiency scores between the mental efficiency group and non-adapted control group. The non-significant differences of the present study may be due to a change in one instructional procedure made in the present study from the Kalyuga (2006a) study. The present study substituted faded completion problems for conventional problems, using faded worked example-faded completion problem pairs at each training session stage, compared to the Kalyuga (2006a) study, which used faded worked example-conventional problem pairs at each training session stage. Since the non-adapted control group of the present study proceeded through the training stage-by-stage, neither bypassing nor repeating any stage, the non-adapted control group essentially followed a fading guidance procedure, receiving feedback after each step supplied by the participant for the faded completion problems.

Renkl et al. (2002) found that a faded guidance procedure produced higher post test scores than an example-problem pair procedure because the example-problem pair procedure placed excessive demands on the learner, raising cognitive load and impairing schema construction. The non-adapted control group of the Kalyuga (2006a) study was not allowed to repeat example-problem pairs and may have been placed in a disadvantaged position. While the adapted groups could repeat the example-problem pairs and possibly reduce the mental effort of solving the conventional problems, the non-adapted control group could not. In contrast, the non-adapted control group of the present study was not allowed to repeat a fading guidance procedure, but this may not have placed the non-adapted control group in a disadvantaged position. Repeating a completion problem in a fading guidance procedure, compared to repeating a conventional problem in an example-problem pair procedure, may not provide a comparable reduction in mental effort. The completion problem only requires the participant to supply a single step at a time, with correct answer feedback provided after each step is submitted, but the conventional problem requires the participant to supply all five steps, with correct answer feedback provided after all steps have been submitted. Because the completion problem provides more scaffolding, repeating the completion problem may not lower cognitive load as much as repeating a conventional problem, which provides little scaffolding. The non-significant results of the present study may indicate that the adaptation methods used in both the present and Kalyuga (2006a) studies do not improve upon a fading guidance procedure.

Implication of Findings for Adapting Instruction Using Mental Efficiency Measures

The results of the present study may indicate that the adaptive method advocated by Kalyuga (2006a) may not work effectively with a fading guidance procedure advocated by Renkl & Atkinson (2003) and Renkl et al. (2002). The following question is posed: Will a training session that requires participants to bypass or repeat stages based upon a measure of performance or a measure of mental efficiency provide more efficient and effective instruction than a training session that employs a fading guidance procedure, which does not require participants to bypass or repeat stages? It is true that adaptive instruction that employs example-problem pairs seems to be more efficient and effective than non-adaptive instruction (Kalyuga, 2006a). However, adaptive instruction that employs a fading guidance procedure, such as that used in the present study, seems to *not* be more efficient or effective than non-adaptive instruction.

Two other adapted instruction studies (Kalyuga & Sweller, 2004; 2005) did employ a modified completion strategy, but not with a fading guidance procedure. However, both of the Kalyuga and Sweller studies did add a conventional problem to the series of completion problems provided during the training session. The Kalyuga and Sweller (2004) study found significantly higher knowledge gains for the mental efficiency group (the only adapted group) compared to the non-adapted control group (instructional efficiency was not computed), and the Kalyuga and Sweller (2005) study found significantly higher instructional efficiency scores for the mental efficiency group (the only adapted group) compared to the non-adapted control group. However, the studies employed a yoked design, where the non-adapted control group was initially placed in the same training session stage as the adapted group counterpart and repeated the same training session stages as the adapted group counterpart. It is possible the non-adapted control group was not able to benefit from the smooth transition provided by a completion strategy, which started at the first training session stage and progressed stage-by-stage to the last stage.

Therefore, none of the previous CLT adaptive instructional research employed a fading guidance procedure, such as that employed by the present study, leading to the conjecture that the adaptive methods used in the other studies may not work effectively with a fading guidance procedure. As will be discussed next, the fading guidance procedure is more effective promoting both near- and far-transfer learning compared to an example-problem pair procedure, but the CLT adaptive instruction research so far has not provided an adaptation method that is more effective promoting far-transfer learning than a non-adaptation method.

The present study and studies previously discussed (Kalyuga, 2006a; Kalyuga & Sweller, 2004, 2005) measured near-transfer performance on a posttest, but did not measure far-transfer performance. Other adaptive instructional studies (Camp et al., 2001; Salden et al., 2004) measured both near- and far-transfer performance on a posttest but found no significant differences for far-transfer test performance between the adapted groups and non-adapted control groups. In contrast, Renkl et al. (2002) found that a backward fading procedure resulted in significantly higher far transfer posttest scores than a example-problem pair procedure. Furthermore, adding self-explanation prompts to a backward fading procedure resulted in significantly higher near- and far-transfer posttest scores compared to a backward fading procedure without prompts (Renkl & Atkinson, 2003). A self-explanation is a mental dialog that a learner conducts with his or herself while studying a worked example (or worked out steps to a completion problem). Self-explanations help the learner to understand worked examples and construct solution schemas (Renkl, 1997).

Renkl and Atkinson (2003) summarized their findings as follows: a) fading provides a bridge between early and later phases of cognitive skill development; b) fading is effective for learning near-transfer problem solving rules; c) providing step-by-step feedback reduces errors and is affective for learning near-transfer problem solving rules; d) fading in a backward manner is better than fading in a forward manner; e) adding self-explanatory prompting at worked-out steps fosters both near- and far-transfer. Renkl et al. (2002) and Renkl and Atkinson (2003) have demonstrated the benefits of using a fading guidance procedure compared to an example-problem pair procedure. Kalyuga (2006a) has demonstrated the benefits of adapting instruction using an example-problem pair procedure. However, the present study has not demonstrated the benefit of adapting instruction using a fading guidance procedure. The present study results may indicate that the adapted methods advocated by Kalyuga (2006a) do not work with a fading guidance procedure. Therefore, more research is required.

Recommendations

The present study indicates that further research is needed in the following areas: 1) The present study should be replicated with a longer training session time of perhaps 2 hours (as followed by the Camp et al. (2001) and Salden et al. (2004) studies), rather than an average training session time of 45 minutes, followed by the present study. 2) Adapting instruction using a fading guidance procedure should be investigated to determine if it is more effective and efficient than adapting instruction using an example-problem pair procedure (as followed by the Kalyuga (2006a) study). 3) A fading guidance procedure without adaptation (as followed by the Renkl et al. (2002) study) should be investigated to determine if it is more effective and efficient than adapting instruction with an example-problem pair procedure (as followed by the Kalyuga (2006a) study).

In addition, Renkl and Atkinson (2003) call for research on adaptive instruction that explores two approaches: 1) externally determined adaptation, in which a learner's prior knowledge is diagnosed to determine which problem solving steps a learner cannot solve and to initially place the learner in a training session stage that employs a fading guidance procedure that starts with the step that the learner cannot solve and 2) internally determined adaptation, in which learners are a) trained to self-explain, b) required to anticipate a solution step, and c) required to look up worked-out steps for incorrect answers and self explain why the worked out steps are correct.

Limitations of the Study

There were limitations to this study that should be considered when interpreting the results, including the study's participants, environment, domain, sample sizes, duration, and validity of instruments. Perhaps different results will be generated by others because of the following factors that existed with the present study: 1) the participants for the study were college students at a regional university and may possess unique demographic profiles, mathematical abilities, and prior knowledge of business and accounting related subjects not found elsewhere; 2) the experiment was conducted in a realistic class environment with unique characteristics that may not be found elsewhere; 3) the experiment was conducted in the domain of accounting requiring unique problem solving skills that may not be found in other domains; 4) the sample sizes of the groups used in the study ranged from 16 to 19 participants, which may have been too small; 5) the experiment was conducted during one 1.5 hour class period, and the training session lasted 45 minutes, which may have been too short; 6) the paper-based pretest and rapid diagnostic tests did not provide very high reliability or validity measures, which may have resulted in the assessments not really assessing what they purported to assess.

APPENDICES

APPENDIX A
Paper-Based Pretest

Name ____

Paper-based Pretest – 10 Minutes

Multiple Choice. Circle the best response.

1) Costs that remain constant in total dollar amount as the level of activity changes are called:
   a) Fixed costs
   b) Mixed costs
   c) Opportunity costs
   d) Variable costs

2) Which of the following describes the behavior of the fixed cost per unit?
   a) Decreases with increasing production
   b) Decreases with decreasing production
   c) Remains constant with changes in production
   d) Increases with increasing production

3) Costs that vary in total in direct proportion to changes in an activity level are called:
   a) Fixed costs
   b) Sunk costs
   c) Variable costs
   d) Differential costs

4) Which of the following describes the behavior of the variable cost per unit?
   a) Varies in increasing proportion with changes in the activity level
   b) Varies in decreasing proportion with changes in the activity level
   c) Remains constant with changes in the activity level
   d) Varies in direct proportion with the activity level

5) Contribution margin is
   a) The excess of sales revenue over variable cost
   b) Another term for volume in the "cost-volume-profit" analysis
   c) Profit
   d) The same as sales revenue

6) The cost-volume-profit equation is:
   a) (Unit contribution margin + Fixed costs)/Target profit
   b) (Unit contribution margin + Fixed costs)/Net income
   c) (Fixed costs + Target profit)/Unit contribution margin
   d) (Fixed costs + Net income)/Sales price

7) Which of the following is true:
a) Operating costs are 100% variable costs
b) Gross margin is sales minus operating costs
c) Cost of goods sold contains variable and fixed costs
d) Cost of goods sold are 100% variable costs

8) If variable costs per unit increased because of an increase in hourly wages rates, the breakeven point would:
a) decrease
b) increase
c) remain the same
d) increase or decrease, depending upon the percentage increase in wage rates

9) If fixed costs decreased because of a decrease in insurance costs, the breakeven point would
a) decrease
b) increase
c) remain the same
d) increase or decrease, depending upon the percentage increase in insurance costs

10) Which of the following conditions would cause the breakeven point to decrease?
a) Total fixed costs increase
b) Unit selling price decreases
c) Unit variable cost decreases
d) Unit variable cost increases

11) If fixed costs are $350,000, the unit selling price is $75, and the unit variable costs are $25, what is the breakeven sales in units?
a) 3,500 units
b) 4,667 units
c) 14,000 units
d) 7,000 units

12) If fixed costs are $500,000, the unit selling price is $40, and the unit variable costs are $32, what is the breakeven sales in units if the sales price is increased by $2?
a) 50,000 units
b) 12,500 units
c) 62,500 units
d) 83,333 units

13) If fixed costs are $600,000, the unit selling price is $100, and the unit variable costs are $60, what is the amount of sales required to realized an operating income of $100,000?

a) 10,000 units
b) 11,667 units
c) 17,500 units
d) 15,000 units

14) If fixed costs are $500,000, the unit selling price is $40, and the unit variable costs are $32, what is the breakeven sales in units if fixed costs are reduced by $80,000?

a) 60,500 units
b) 52,500 units
c) 62,500 units
d) 64,500 units

15) If fixed costs are $500,000, the unit selling price is $40, and the unit variable costs are $32, what is the breakeven sales in units if the variable costs are decreased by $2?

a) 50,000 units
b) 12,500 units
c) 62,500 units
d) 83,333 units

16) Problem: Please solve the following problem and write your complete solution below, including your solution steps. Given the information below, calculate the number of units the company must sell to earn a target profit of $40,000?

Contribution Margin Income Statement	
Sales (10,000 units)	$400,000
Variable costs	300,000
Contribution margin	100,000
Fixed costs	80,000
Net income	$20,000

APPENDIX B

Initial Diagnostic Test Example: Problem 2

PRE-TEST

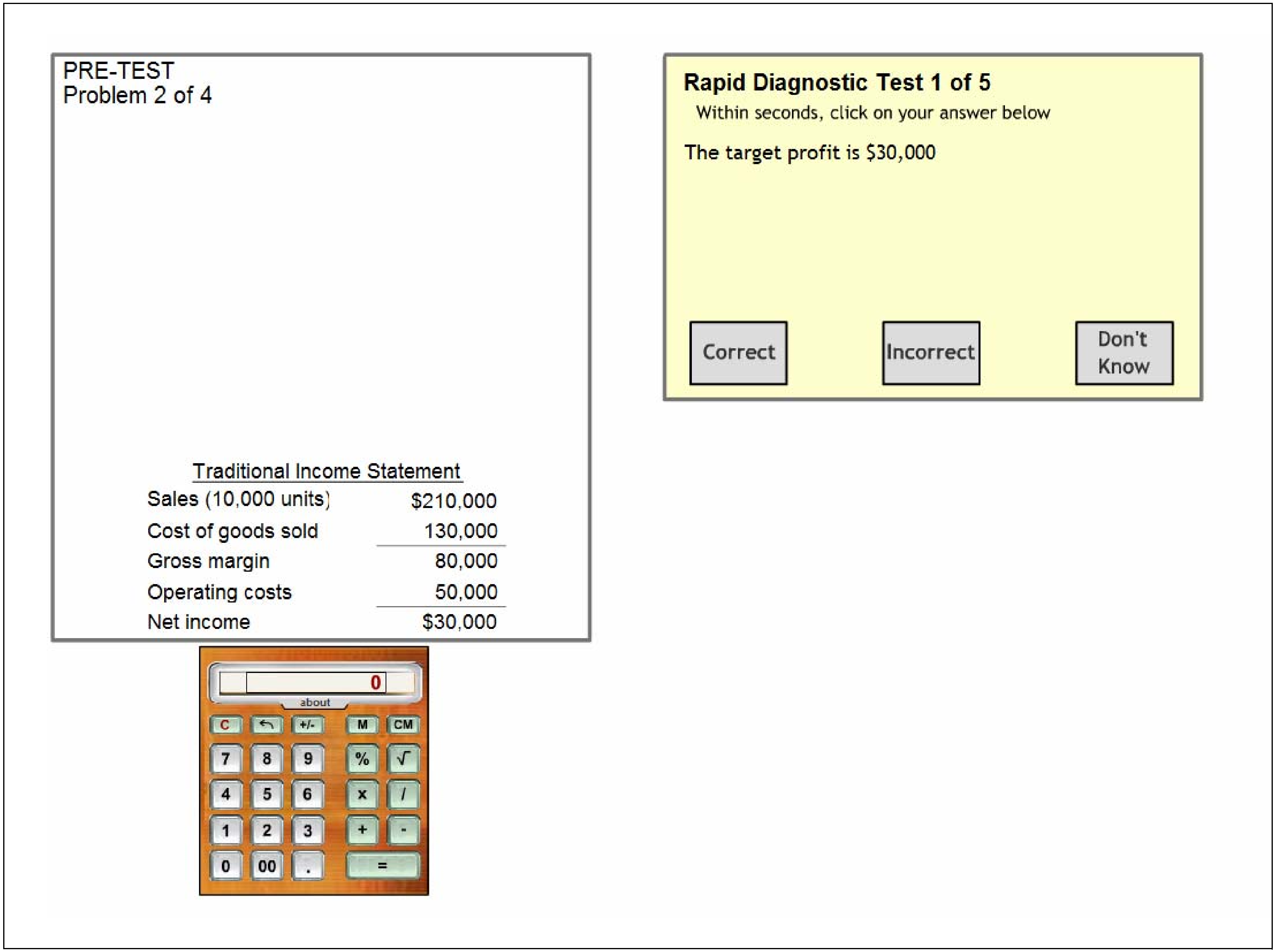
Problem 2 of 4

Camping Outfitters, Inc. manufactures and sells a camping lantern. A traditional annual income statement appears below. The company sells 10,000 camping lanterns per year for $21.00 per lantern. Of the $180,000 total costs, $110,000 are variable, and the remaining costs are fixed. When you are ready to take five rapid diagnostic tests on this problem, click *Next*. **This text will disappear, so remember the key variables**.

Required: Calculate the number of units the company must sell per year to earn a profit of $60,000.

Traditional Income Statement	
Sales (10,000 units)	$210,000
Cost of goods sold	130,000
Gross margin	80,000
Operating costs	50,000
Net income	$30,000

Next


PRE-TEST
Problem 2 of 4
Traditional Income Statement
Sales (10,000 units) $210,000
Cost of goods sold 130,000
Gross margin 80,000
Operating costs 50,000
Net income $30,000
Rapid Diagnostic Test 1 of 5
Within seconds, click on your answer below
The target profit is $30,000
Correct
Incorrect
Don't Know

PRE-TEST
Problem 2 of 4

Traditional Income Statement

Sales (10,000 units)	$210,000
Cost of goods sold	130,000
Gross margin	80,000
Operating costs	50,000
Net income	$30,000

Rapid Diagnostic Test 1 of 5 Mental Effort

How easy or difficult was this rapid verification step?
Click one of the nine boxes below on the scale from "Extremely easy" to "Extremely difficult".

Extremely easy				Neither easy nor difficult				Extremely difficult

PRE-TEST
Problem 2 of 4

Traditional Income Statement	
Sales (10,000 units)	$210,000
Cost of goods sold	130,000
Gross margin	80,000
Operating costs	50,000
Net income	$30,000

Rapid Diagnostic Test 2 of 5

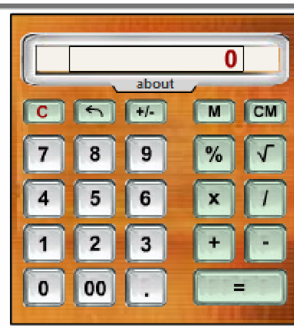
Within seconds, click on your answer below

The target profit is $60,000
The fixed costs are $70,000

Correct | Incorrect | Don't Know

PRE-TEST
Problem 2 of 4

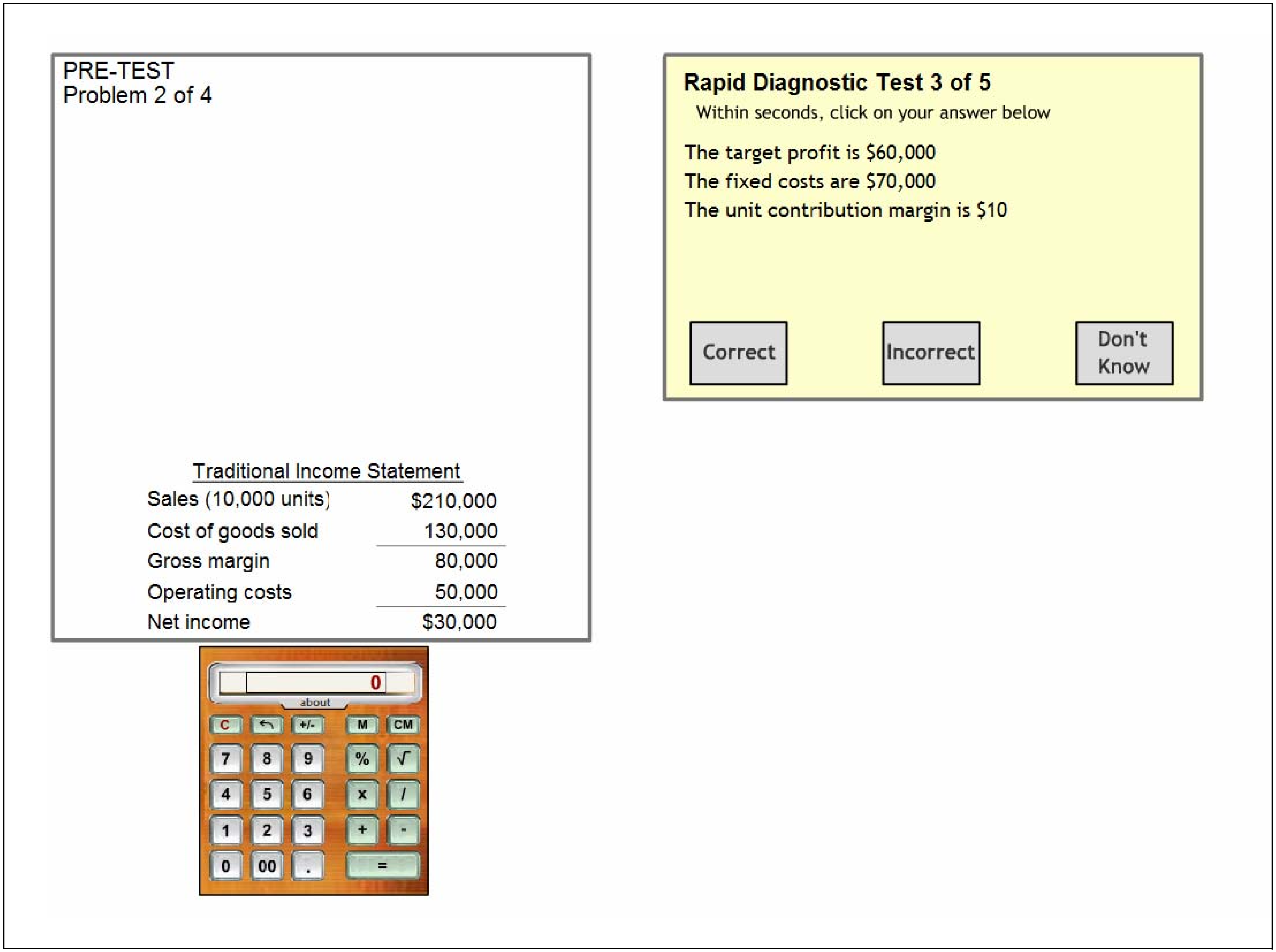
Traditional Income Statement	
Sales (10,000 units)	$210,000
Cost of goods sold	130,000
Gross margin	80,000
Operating costs	50,000
Net income	$30,000



Rapid Diagnostic Test 2 of 5 Mental Effort

How easy or difficult was this rapid verification step?
Click one of the nine boxes below on the scale from "Extremely easy" to "Extremely difficult".

Extremely easy				Neither easy nor difficult				Extremely difficult


PRE-TEST
Problem 2 of 4
Traditional Income Statement
Sales (10,000 units) $210,000
Cost of goods sold 130,000
Gross margin 80,000
Operating costs 50,000
Net income $30,000
Rapid Diagnostic Test 3 of 5
Within seconds, click on your answer below
The target profit is $60,000
The fixed costs are $70,000
The unit contribution margin is $10
Correct
Incorrect
Don't Know

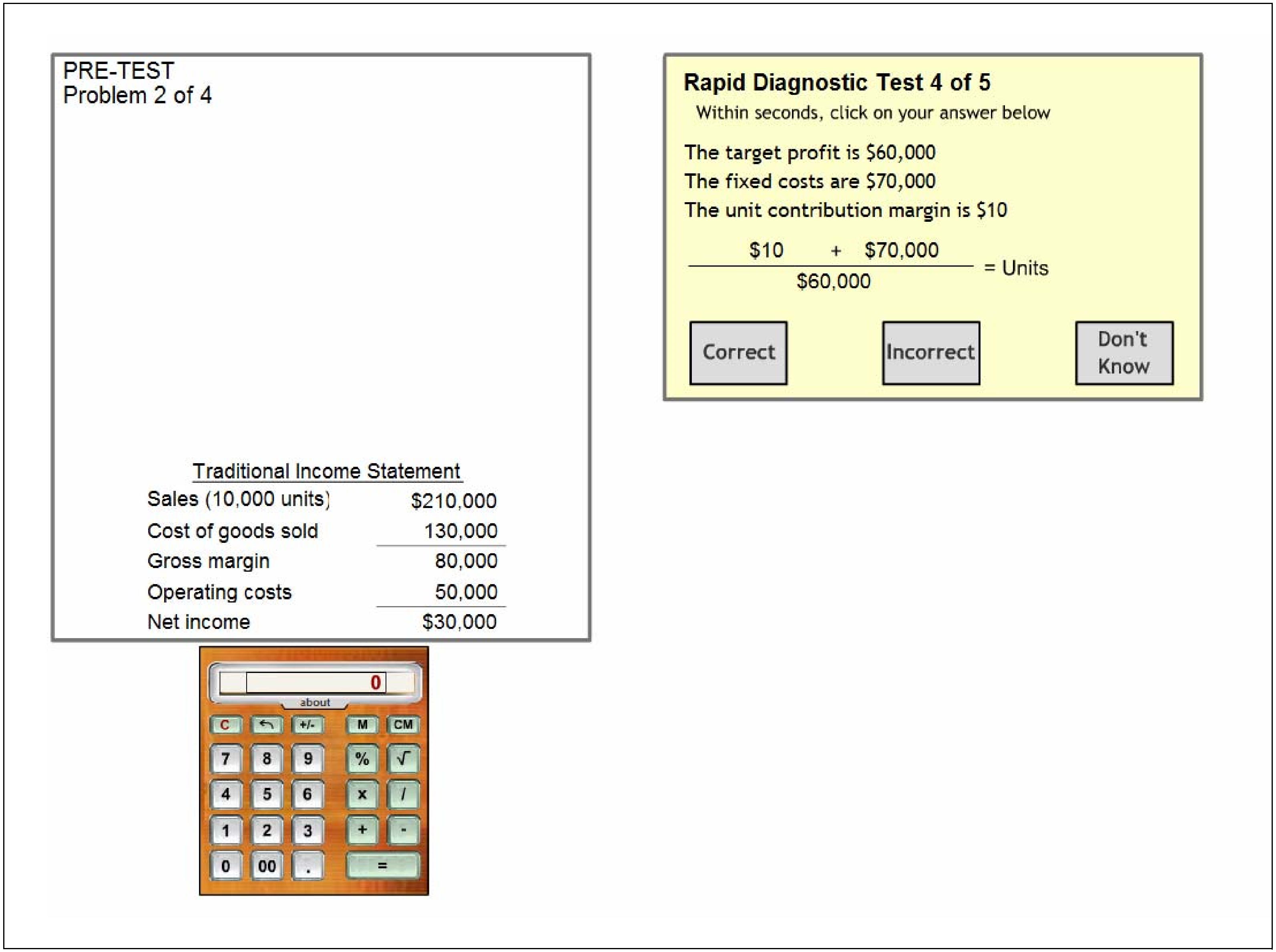
PRE-TEST
Problem 2 of 4

Traditional Income Statement	
Sales (10,000 units)	$210,000
Cost of goods sold	130,000
Gross margin	80,000
Operating costs	50,000
Net income	$30,000

Rapid Diagnostic Test 3 of 5 Mental Effort

How easy or difficult was this rapid verification step?
Click one of the nine boxes below on the scale from "Extremely easy" to "Extremely difficult".

Extremely easy				Neither easy nor difficult				Extremely difficult


PRE-TEST
Problem 2 of 4
Traditional Income Statement
Sales (10,000 units) $210,000
Cost of goods sold 130,000
Gross margin 80,000
Operating costs 50,000
Net income $30,000
Rapid Diagnostic Test 4 of 5
Within seconds, click on your answer below
The target profit is $60,000
The fixed costs are $70,000
The unit contribution margin is $10
$10 + $70,000
$60,000
= Units
Correct
Incorrect
Don't Know

PRE-TEST
Problem 2 of 4

Traditional Income Statement

Sales (10,000 units)	$210,000
Cost of goods sold	130,000
Gross margin	80,000
Operating costs	50,000
Net income	$30,000

Rapid Diagnostic Test 4 of 5 Mental Effort

How easy or difficult was this rapid verification step?
Click one of the nine boxes below on the scale from "Extremely easy" to "Extremely difficult".

Extremely easy	Neither easy nor difficult	Extremely difficult

PRE-TEST
Problem 2 of 4

Traditional Income Statement

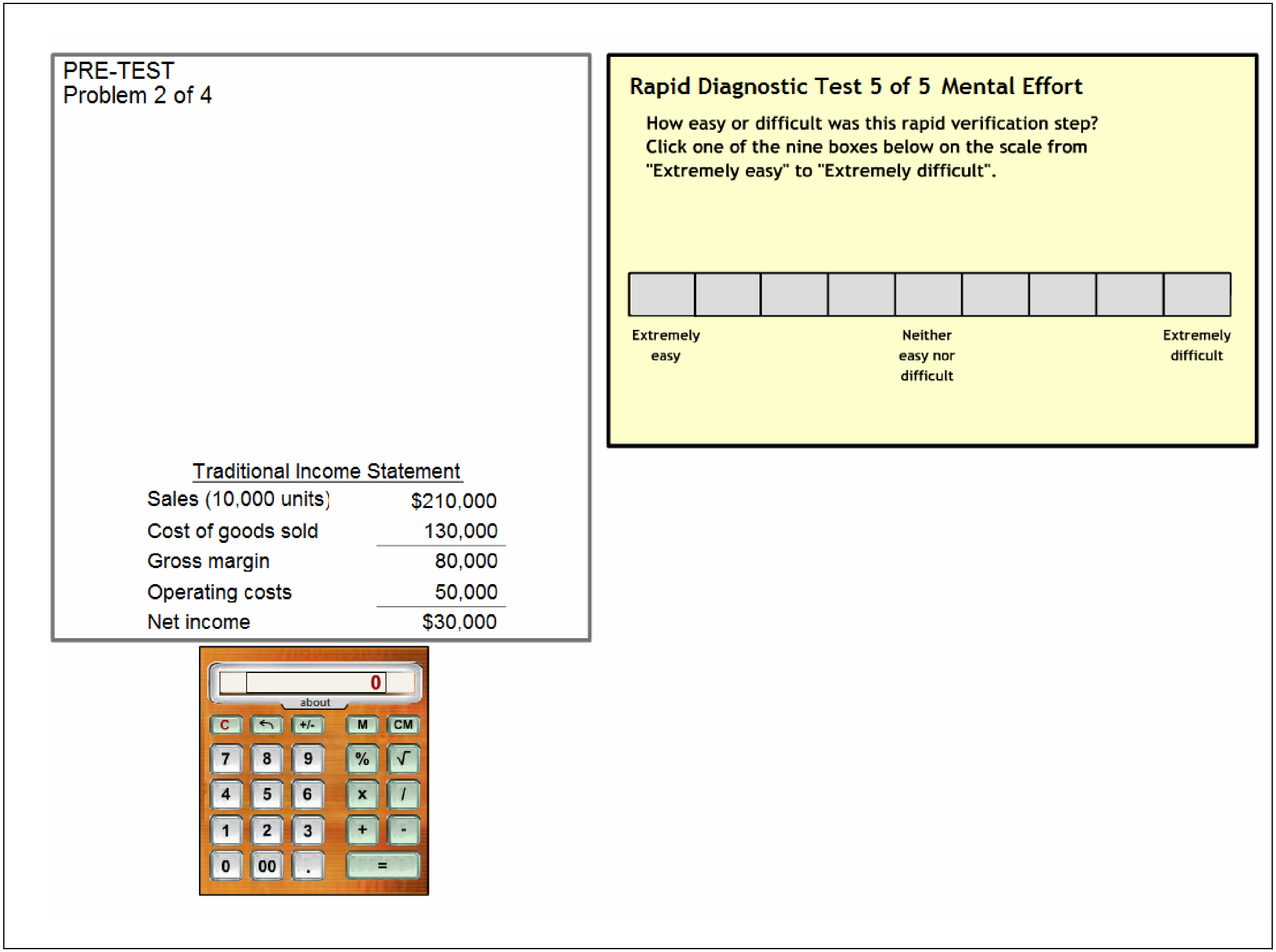
Sales (10,000 units)	$210,000
Cost of goods sold	130,000
Gross margin	80,000
Operating costs	50,000
Net income	$30,000

Rapid Diagnostic Test 5 of 5

Within seconds, click on your answer below

The answer is 13,000 units

Correct | Incorrect | Don't Know


PRE-TEST
Problem 2 of 4
Traditional Income Statement
Sales (10,000 units) $210,000
Cost of goods sold 130,000
Gross margin 80,000
Operating costs 50,000
Net income $30,000
Rapid Diagnostic Test 5 of 5 Mental Effort
How easy or difficult was this rapid verification step?
Click one of the nine boxes below on the scale from
"Extremely easy" to "Extremely difficult".
Extremely easy
Neither easy nor difficult
Extremely difficult

APPENDIX C

Training Session Example: Difficulty Level 2, Stage 2

TRAINING LEVEL 2 of 4

Self-Study Problem 2-2

Elegance Ties, Inc. manufactures and sells men's ties. A traditional annual income statement appears below. The company sells 10,000 ties per year for $25 per tie. Of the $220,000 total costs, $150,000 are variable, and the remaining costs are fixed. Study the solution to the right. (The first step has been skipped.) When you are finished, click *Next*.

Required: Calculate the number of units the company must sell per year to earn a profit of $60,000.

Traditional Income Statement

Sales (10,000 units)	$250,000
Cost of goods sold	150,000
Gross margin	100,000
Operating costs	70,000
Net income	$30,000

Please Study this Solution

You may take as long as you like.
When you are finished, click Next.

Determine target profit step skipped

The fixed costs are
$70,000 (Total costs - Variable cost = Fixed cost. $220,000 - $150,000 = $70,000)

The unit contribution margin is
$10 (Total sales - Total variable cost = Total contribution margin. Total contribution margin/Total units = Unit contribution margin. $250,000 - $150,000 = $100,000. $100,000/10,000 = $10)

$$\frac{\text{Target profit + Fixed cost}}{\text{Unit contribution margin}} = \text{Units}$$

$$\frac{\$60{,}000 \quad + \quad \$70{,}000}{\$10} = \text{Units}$$

The answer is **13,000**

Proof:
(13,000 x $10) - $70,000 = $60,000

Next

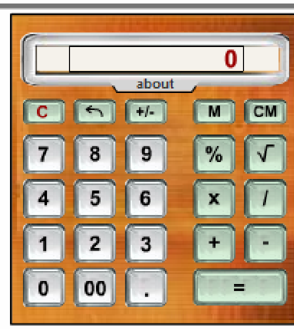
TRAINING LEVEL 2 of 4
Completion Problem 2-2

Leather Goods Co. manufactures and sells wallets. A traditional annual income statement appears below. The company sells 10,000 wallets per year for $35 per wallet. Of the $310,000 total costs, $250,000 are variable, and the remaining costs are fixed. Complete the steps to the completion problem to the right.

Required: Calculate the number of units the company must sell per year to earn a profit of $80,000.

Traditional Income Statement

Sales (10,000 units)	$350,000
Cost of goods sold	210,000
Gross margin	140,000
Operating costs	100,000
Net income	$40,000



Please enter your answer below and click *Submit*.

You have three tries, after which the answer will be provided. If an answer is zero, please enter 0 and don't leave the field blank. Also, just enter dollar figures with no cents.

The target profit is $80,000
The fixed costs are $60,000
The unit contribution margin is $10

$\frac{\boxed{} + \boxed{}}{\boxed{}}$ = Units

Submit

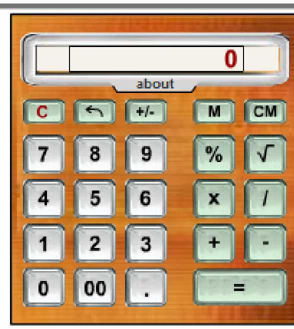
TRAINING LEVEL 2 of 4

Completion Problem 2-2

Leather Goods Co. manufactures and sells wallets. A traditional annual income statement appears below. The company sells 10,000 wallets per year for $35 per wallet. Of the $310,000 total costs, $250,000 are variable, and the remaining costs are fixed. Complete the steps to the completion problem to the right.

Required: Calculate the number of units the company must sell per year to earn a profit of $80,000.

Traditional Income Statement	
Sales (10,000 units)	$350,000
Cost of goods sold	210,000
Gross margin	140,000
Operating costs	100,000
Net income	$40,000



Please enter your answer below and click *Submit*.

You have three tries, after which the answer will be provided. If an answer is zero, please enter 0 and don't leave the field blank. Also, just enter dollar figures with no cents.

The target profit is $80,000

The fixed costs are $60,000

The unit contribution margin is $10

$$\frac{\$80{,}000 + \$60{,}000}{\$10} = \text{Units}$$

OK, here is the answer.

The answer is ______ Units

Submit

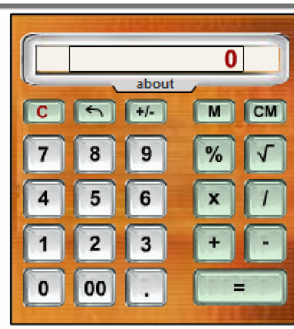
TRAINING LEVEL 2 of 4
Completion Problem 2-2

Leather Goods Co. manufactures and sells wallets. A traditional annual income statement appears below. The company sells 10,000 wallets per year for $35 per wallet. Of the $310,000 total costs, $250,000 are variable, and the remaining costs are fixed. Complete the steps to the completion problem to the right.

Required: Calculate the number of units the company must sell per year to earn a profit of $80,000.

Traditional Income Statement

Sales (10,000 units)	$350,000
Cost of goods sold	210,000
Gross margin	140,000
Operating costs	100,000
Net income	$40,000



Please enter your answer below and click *Submit*.

You have three tries, after which the answer will be provided. If an answer is zero, please enter 0 and don't leave the field blank. Also, just enter dollar figures with no cents.

The target profit is $80,000
The fixed costs are $60,000

The unit contribution margin is $10

$$\frac{\$80{,}000 \quad + \quad \$60{,}000}{\$10} = \text{Units}$$

The answer is 14,000 Units OK, here is the answer.

Proof:
(14,000 x $10) - $60,000 = $80,000

Next

TRAINING LEVEL 2 of 4

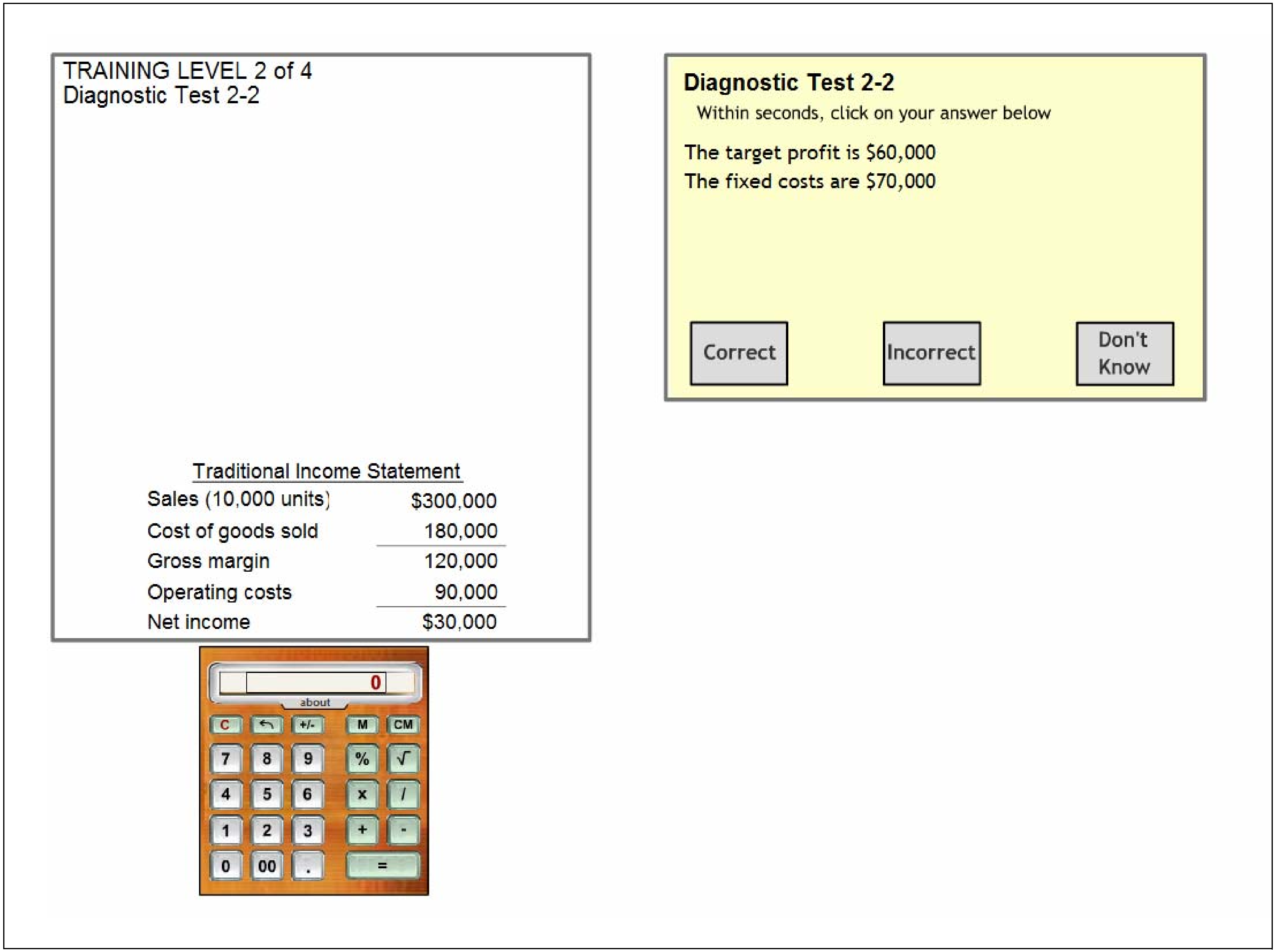
Diagnostic Test 2-2

Carpenter's Helper Co. manufactures and sells hammers. A traditional annual income statement appears below. The company sells 10,000 hammers per year for $30 per hammer. Of the $270,000 total costs, $200,000 are variable, and the remaining costs are fixed. When you are ready to take a single rapid diagnostic test on this problem, click *Next*. This text will disappear, so remember the key variables.

Required: Calculate the number of units the company must sell per year to earn a profit of $60,000.

Traditional Income Statement	
Sales (10,000 units)	$300,000
Cost of goods sold	180,000
Gross margin	120,000
Operating costs	90,000
Net income	$30,000

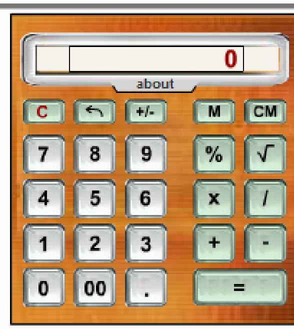
Next


TRAINING LEVEL 2 of 4
Diagnostic Test 2-2
Traditional Income Statement
Sales (10,000 units) $300,000
Cost of goods sold 180,000
Gross margin 120,000
Operating costs 90,000
Net income $30,000
Diagnostic Test 2-2
Within seconds, click on your answer below
The target profit is $60,000
The fixed costs are $70,000
Correct
Incorrect
Don't Know

TRAINING LEVEL 2 of 4
Diagnostic Test 2-2

Traditional Income Statement

Sales (10,000 units)	$300,000
Cost of goods sold	180,000
Gross margin	120,000
Operating costs	90,000
Net income	$30,000



Diagnostic Test 2-2 **Mental Effort**

How easy or difficult was this rapid verification step?
Click one of the nine boxes below on the scale from "Extremely easy" to "Extremely difficult".

Extremely easy				Neither easy nor difficult				Extremely difficult

REFERENCES

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*, 167-207.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology, 23*, 932-945.

Atkinson, R., Renkl, A., & Merrill, M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*, 783.

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

Blessing, S. B., & Anderson, J. R. (1996). How people learn to skip steps. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 576-598.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(3-16).

Brunken, R., Plass, J., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*, 53-61.

Burns, M., MacQuarrie, L., & Campbell, D. (1999). *The difference between curriculum-based assessment and curriculum-based measurement: A focus on purpose and result*. Retrieved 3/13/2006, 2006, from http://www.nasponline.org/publications/cq276cba.html

Camp, G., Paas, F., Rikers, R., & van Merriënboer, J. (2001). Dynamic problem selection in air traffic control training: a comparison between performance, mental effort and mental efficiency. *Computers in Human Behavior, 17*, 575-595.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332.

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Clark, R. C., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. San Francisco: Pfeiffer.

Cohen, L., Manion, L., & Morrison, K. (2003). *Research methods in education* (5th ed.). London & New York: Routledge Falmer.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction* (pp. 453-494). Hillsdale, NJ: Erlbaum.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem solving transfer. *Journal of Educational Psychology, 79*, 347-362.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245.

Fletcher, J. D. (2003). Evidence for learning from technology-assisted instruction. In J. Harold F. O'Neil & R. S. Perez (Eds.), *Technology Applications in Education, A Learning View* (pp. 79-99). Mahway, NJ: Lawrence Erlbaum Associates.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*, 488-500.

Jonassen, D. (1988). Adaptive designs for courseware. In D. H. Jonassen (Ed.), *Instructional designs for microcomputer courseware*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Jonassen, D., & Grabowski, B. L. (1993). *Handbook of individual differences: Learning & instruction*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Kalyuga, S. (2005). *When less is more in cognitive diagnosis.* Paper presented at the CogSci 2005 Annual Conference of the Cognitive Science Society, Stresa, Italy.

Kalyuga, S. (2006a). Assessment of learners' organized knowledge structures in adaptive learning environments. *Applied Cognitive Psychology, 20*, 333-342.

Kalyuga, S. (2006b). *Instructing and testing advanced learners: A cognitive load approach*. Hauppauge NY: Nova Science Publishers, Inc.

Kalyuga, S. (2006c). Rapid assessment of learners' proficiency: A cognitive load approach. *Educational Psychology, 26*, 613-627.

Kalyuga, S. (2006d). Rapid cognitive assessment of learners' knowledge structures. *Learning and Instruction, 16*, 1-11.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-31.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*, 1-17.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579-588.

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96*, 558-568.

Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development, 53*(3), 83-93.

Kirschner, P., Sweller, J., & Clark, R. A. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75-86.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from the Tower of Hanoi. *Cognitive Psychology, 17*, 248-294.

Larkin, H., McDermott, J., Simon, D., & Simon, H. (1980). Models of competence in solving physics problems. *Cognitive Science, 4*, 317-345.

Marshall, S. (1995). *Schemas In problem solving*. New York: Cambridge University Press.

Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: The Guilford Press.

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

Mislevy, R. (1993). Foundations of a new test theory. In N. Fredericksen, R. Mislevy & I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 19-39). Hillsdale, NJ: Lawrence Erlbaum Associates.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*, 429-434.

Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63-71.

Paas, F., & van Merriënboer, J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*, 737-743.

Paas, F., & van Merriënboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive load approach. *Journal of Educational Psychology, 86*, 122-133.

Paas, F., van Merriënboer, J., & Adam, J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*, 419-430.

Park, O., & Lee, J. (2004). Adaptive instructional systems. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd ed., pp. 651-684). Mahwah, N.J.: Lawrence Erlbaum.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 1*, 1-29.

Renkl, A., & Atkinson, R. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*, 15-22.

Renkl, A., Atkinson, R., Maier, U., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *The Journal of Experimental Education, 70*, 315.

Rikers, R. (2006). A critical reflection on emerging topics in cognitive load research. *Applied Cognitive Psychology, 20*, 359-364.

Salden, R., Paas, F., Broers, N., & van Merriënboer, J. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in air traffic control training. *Instructional Science, 32*, 153-172.

Schneider, W., & Shiffrin, R. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1-67.

Shiffrin, R., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*, 127-190.

Shinn, M., & Bamonto, S. (1998). Advanced applications of curriculum-based measurement: "big ideas" and avoiding confusion. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 1-31). New York: The Guilford Press.

Simon, S., & Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology, 5*, 29-46.

Singley, M., & Bennett, R. (2002). Item generation and beyond: Applications of schema theory to mathematics assessment. In S. Irvine & P. Kyllonen (Eds.), *Item generation of test development* (pp. 361-384). Mahwah, NJ: Lawrence Erlbaum Associates.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4*, 495-312.

Sweller, J. (2005). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press.

Sweller, J. (2006). Discussion of emerging topics in cognitive load research: Using learner and information characteristics in the design of powerful learning environments. *Applied Cognitive Psychology, 20*, 353-357.

Sweller, J., Chandler, P., Tierner, P., & Cooper, G. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General, 119*, 176-192.

Sweller, J., & Cooper, G. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*, 59-89.

Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

Tuovinen, J., & Paas, F. (2004). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional Science, 32*, 133-152.

Tuovinen, J., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*, 334-341.

Tyack, D., & Cuban, L. (2000). Teaching by machine. In R. Pea (Ed.), *The Jossey-Bass reader on technology and learning*. San Francisco: Jossey-Bass.

Van Gerven, P., Paas, F., van Merriënboer, J., & Schmidt, H. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology, 41*, 167-174.

van Merriënboer, J. (1990). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of Educational Computing Research, 6*, 265-285.

van Merriënboer, J. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.

van Merriënboer, J., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development, 53*(3), 5-13.

van Merriënboer, J., & de Croock, M. (1992). Strategies for computer-based programming instruction: Program completion vs. program generation. *Journal of Educational Computing Research, 8*, 365-394.

van Merriënboer, J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*, 147-177.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47*, 513-539.

Zhu, X., & Simon, H. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction, 4*, 137-166.