CAN MODULAR EXAMPLES AND CONTEXTUAL INTERFERENCE IMPROVE TRANSFER?

A Thesis
Presented to
The Academic Faculty

by

Brian D. Gane

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Psychology

Georgia Institute of Technology
August of 2006

CAN MODULAR EXAMPLES AND CONTEXTUAL INTERFERENCE IMPROVE TRANSFER?

Approved by:

Dr. Richard Catrambone, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Arthur D. Fisk
School of Psychology
Georgia Institute of Technology

Dr. Lawrence James
School of Psychology
Georgia Institute of Technology

Date Approved: May 3, 2006

ACKNOWLEDGEMENTS

Richard Catrambone provided invaluable guidance and pushed me to refine and explore these ideas. Thanks to Dan Fisk and Larry James for helpful comments on both the theoretical and methodological ideas within this paper. Kristie Fisher, Elizabeth Solomon, and Yoko Nakano provided much needed help in running participants. This research was partially funded by the School of Psychology's Graduate Student Development Award.

A special thanks to my family for instilling within me the confidence and motivation to achieve my goals. Finally, this work would not have been possible without the loving support of Megan Reilly, who continuously pushed me to persevere through the numerous challenges that this work presented.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Two instructional design features hypothesized to affect problem solving performance, problem format and contextual interference, were investigated. Problem format was manipulated by altering the format of worked examples to demonstrate a molar or modular solution. Contextual interference was manipulated by randomizing the order in which problem categories were studied. Participants studied worked examples from 5 complex probability categories and solved 11 novel problems. The modular problem format reduced study time and the workload during study and increased performance on the subsequent test. Greater contextual interference increased study time but had no effect on workload or test performance. Additionally, a regression analysis demonstrated that mental workload partially mediated the effect of problem format on test performance. A separate regression analysis did not demonstrate that working memory capacity moderated the effect of problem format on mental workload.

Chapter 1

INTRODUCTION

A cognitive skill is a complex, learned procedure that relies primarily on cognitive processing and not heavily on motor processing (VanLehn, 1996). Cognitive skills, such as solving physics problems and writing computer programs, are developed through extensive practice and study. The two critical components that enable effective cognitive skill performance are *rule application* and *schema development*. Rule application is the process of applying valid operators in order to move a problem from an initial state to a goal state (Sweller, 1999). Consider an example that asks for the product of two fractions, 1/4 and 1/5. The initial state is two fractions, with values 1/4 and 1/5. The goal state is one fraction. To move from the initial state to the goal state, the multiplication rule must be applied. For many, this rule is so well practiced that the rule application has become automated (Sweller, 1999).

In addition to automated rule application, schema robustness guides successful cognitive skill processing. A schema is a hypothetical construct that represents a mental structure which organizes knowledge of a domain (Chi, Glaser, & Rees, 1982; Sweller, 1999). For instance, a schema for manipulating fractions could include how to add, subtract, multiply, and divide fractions. This schema would likely include how to reduce fractions, how to determine the least common denominator, and how to cross-multiply fractions. These concepts may be interconnected in a somewhat hierarchical manner (i.e., multiple levels of abstraction). For instance, one level would hold concepts such as adding, subtracting, multiplying, and dividing fractions. A different level would contain the concept of reducing fractions, which would be connected to subtracting and adding fractions. As new knowledge of a domain is learned, schemas are modified. Schemas can facilitate problem solving when the rule has not yet been automated because they allow declarative knowledge to guide rule application (Chi et al., 1982).

Learning a cognitive skill involves a progression from novice to expert. This progression is marked by three stages of skill acquisition: early, intermediate, and late (Kalyuga, Ayres, Chandler, & Sweller, 2003; Renkl & Atkinson, 2003; VanLehn, 1996). In the early stage learners become familiar with the domain but do not yet apply the knowledge. In the intermediate stage learners begin to apply the knowledge by studying or solving examples; in this stage they are still refining their schemas and usage of rules (VanLehn, 1996). In the final stage learners automate their application of rules, which improves the speed and accuracy of their problem solving ability (Renkl & Atkinson, 2003; VanLehn, 1996). In order to facilitate the progression from novice to expert, instructions should be designed for the learning goals at a given stage (Renkl & Atkinson, 2003). The instructional design manipulations proposed in this study are relevant for learners in the intermediate stage, and are aimed at facilitating learner's rule applications and schema development.

1.1 Novice and Expert Schemas

Research on problem solving has revealed that one critical difference between novices and experts is the schematic representation of their knowledge. Compared to a novice, experts' schemas are more robust. An expert's schema contains more elements and a greater number of interconnections between these elements. These differences in schemas also mediate the representation of the problem, facilitating solutions to novel problems (Chi et al., 1982).

Additionally, experts have schemas for particular problem types. For instance, a statistician would likely have a schema for calculating permutations. Experts in a domain are more likely to have schemas that are organized by structural features of the domain, rather than surface features (Chi et al., 1982). This allows experts to solve problems that seem superficially different but yet are structurally identical to previously encountered problems (Gick & Holyoak, 1983). This underscores an important distinction in the literature on problem solving and schema acquisition. A fair amount of research has focused on problem-type schemas and investigated how to improve a learner's ability to categorize problems according to a problem-type schema (e.g., Fuchs, Fuchs, Prentice, Hamlett, Finelli, & Courey, 2004). However, emphasizing the acquisition of problem-type schemas does not improve (and might hinder) a learner's ability to solve novel problems. To avoid this reliance on problem-type schemas, I will manipulate two instructional features that are hypothesized to facilitate domain-level schemas. I believe these domain-level schemas will allow learners to transfer the knowledge gained during training to the solution of novel problems.

1.2 Instructional Design and Promoting Transfer

Research on problem solving has revealed that novices in a domain have difficulty solving problems that differ from examples they have previously studied (Catrambone, 1998). In contrast, experts in a domain can often solve these novel problems, although the solution is not usually automatic. Problems can be ranked along a continuum of novelty, ranging from isomorphic to far transfer. An *isomorphic problem* has the same structural features as an example which was studied but has different values. For instance, consider the following two isomorphic problems:

> (a) An urn contains five marbles, each a different color - white, yellow, red, green, and blue. Two marbles are taken out, one by one, and are not put back. What is the probability of first taking out the blue marble, and then the white marble?
>
> (b) An urn contains six marbles, each a different color - white, yellow, red, green, blue, and orange. Three marbles are taken out, one by one, and are not put back. What is the probability of first taking out the blue marble, then the white marble, and then the red marble?

The underlying structure of the problem has not changed, only the numbers. In fact, the surface features have not changed either. In contrast, consider the following *near transfer problem*, in which the cover story has changed but the underlying structure is the same:

> (c) At the Olympics 7 sprinters participate in the 100m-sprint. What is the probability of correctly guessing the winner of the gold, the silver, and the bronze medals?

Novices typically have some difficulty in solving these near transfer problems. In order to alleviate this, research on instructional design has focused on methods to improve novices' ability to identify the structural features and thus retrieve the appropriate problem-type schema (Chi, de Leeuw, Chiu, & LaVancher, 1994; Fuchs et al., 2004; Gick & Holyoak, 1983). However, consider a third type of problem, a *far transfer problem*, in which the surface *and* structural features have changed:

(d) In a car race 12 different European countries participate with one driver per country. There are 5 prizes for the participants: The winner receives \$10,000, the second place finisher gets \$5,000 and the third place finisher receives \$1,000. The drivers of the cars who finish fourth and fifth will each win \$500. What is the probability that the Italian driver wins \$10,000, the German \$5,000, the Swedish \$1,000, and that the French and Danish drivers each win \$500?

Novices have an extremely difficult time solving problems on which they have not been trained, although they may have relevant knowledge. That is, although they might have the appropriate prerequisite knowledge to solve the problem, they are not able to combine this knowledge in a unique way (Catrambone, 1996; Catrambone, 1998). In contrast, experts typically can solve these novel problems. I believe that this difference between novices and experts can be explained in terms of the robustness of their schemas. That is, an expert's ability to solve these far transfer problems "lies in the rich internal representation that the expert has generated, a representation that permits many appropriate inferences to be drawn so that the problem can be simplified and reduced" (Chi et al., 1982, p. 20). Therefore, I believe research should continue to explore the means by which instructional design can facilitate the development of a robust, interconnected representation of the problem domain.

1.3 Working Memory's Role in Skill Acquisition

Working memory can be a bottleneck in the process of refining and modifying schemas. Instructions will promote schema development most effectively if instructional designers utilize psychological theories of memory and attention (Sweller, 1999). Although current theories of memory differ in their characterization of working memory/short term memory, they converge on a capacity-limited store (Baddeley, 2003; Engle, 2003). Moreover, individuals differ in their ability to regulate attention, which results in differences in the number of items that can be simultaneously stored in working memory (Engle, 2003). The implication for theories of learning is clear. As the requirements of the learning task approach the capacity of working memory, the ability to learn the material will degrade (Hambrick & Engle, 2003; Sweller, 1999). Interpreted within the context of schema development, as more elements must be simultaneously integrated into a schema the probability of successfully building or refining that schema will decline.

For instance, in order for a learner to solve a probability problem, that problem must be represented in working memory (Chi et al., 1982). If a learner must use all of his or her working memory in order to understand the instruction, then there will not be free capacity to develop schemas. However, if instructional design is altered to reduce the working memory burden, then the unused capacity can be devoted to reasoning about the problem and therefore to refine and extend an existing schema (Gerjets, Scheiter, & Catrambone, 2004; Paas, 1992; Sweller, 1999).

1.4 Workload

The capacity limitation of working memory and its effect on cognitive skill acquisition underscores an additional consideration for instructional design research, the *workload* imposed by the instructions. Workload is a hypothetical construct that is multidimensional. It is defined here as the cost in internal resources that arises from performing a task, such as studying an example or solving a problem. Therefore workload can affect the ability to maintain or reach a certain level of task performance

(Hart & Staveland, 1988; Xie & Salvendy, 2000). Empirical research has measured workload via subjective ratings, physiological measures, and task performance (Xie & Salvendy, 2000). Of these, subjective ratings are typically favored because of their ease of use, face validity, and concurrent validity (Tsang & Wilson, 1997).

Because a task extends through time, workload is subdivided along a temporal dimension: instantaneous workload, peak workload, average workload, accumulated workload, and overall workload. Subjective ratings collected after completion of the task measure overall workload, which is influenced by average and accumulated workload (Xie & Salvendy, 2000). This overall workload is an interaction of task demands and characteristics of the individual. Although individuals (i.e., operators, learners, etc.) do not have difficulty understanding the concept of workload, they vary in which characteristics of a task they extract as an indicator of workload. For instance, on the same task one individual might characterize workload primarily as the time demands of the task while another individual might characterize workload as the amount of effort expended. Therefore, a multidimensional subjective workload scale can best capture each individual's different conceptualizations of workload. The NASA Task Load Index (NASA-TLX) is a multidimensional scale that yields an overall workload score; this score is sensitive to the demands of a task and minimizes between-subject variability (Hart & Staveland, 1988). Although the NASA-TLX is not commonly used in educational psychology research, derivations of it are frequently used (for a review see Paas, Tuovinen, Tabbers & Van Gerven, 2003).

Each of the multidimensional aspects of workload can be further divided into effective and ineffective workload. The demand incurred by successful performance on a task is effective workload. In contrast, ineffective workload does not contribute to positive task performance (Xie & Salvendy, 2000). An elaboration of this distinction between effective and ineffective workload has been championed in instructional design research as *cognitive load theory* (Sweller, van Merriënboer, & Paas, 1998).

1.5 Cognitive Load Theory

Sweller's cognitive load theory (CLT) divides the workload imposed by an instructional design into *intrinsic load*, *extraneous load*, and *germane load* (Sweller et al., 1998; Sweller, 1999). Intrinsic load is based on the number of interacting elements in the learning task and has conventionally been described as immune to instructional design changes (Sweller, 1999; but see Gerjets et al., 2004 for an alternate view). For instance, learning how to solve chemical equations would have a higher intrinsic load compared to learning the chemical symbol for hydrogen (Sweller, 1999). Extraneous load is closely related to ineffective workload; it is workload that does not contribute to learning. For instance, if a learner is studying instructions presented in a hypertext environment, the overhead associated with navigating between pages is extraneous to the learning task. Germane load is incurred when a learner engages in an activity that contributes to schema development, such as self-explanation (Chi et al., 1994).

CLT has been used to explain a wide range of instructional design research findings, such as: the split-attention effect (instructions should be spatially and temporally contiguous; Mayer & Moreno, 1998), the expertise-reversal effect (the most efficient training method varies as a function of the learner's expertise; Kalyuga et al., 2003), the benefit of goal-free problems (Sweller, 1999), and worked examples (Paas, 1992). However, across these findings CLT has been used more as an informal model than a formal model (Dosher, 1998). The inability of CLT to make a priori predictions about the effect of instructional design decisions is partly because researchers often fail to measure the workload associated with their instructional design manipulation. When researchers do measure workload they still face two significant challenges: (1) how to divide the measured workload into the distinctions proposed by CLT, and (2) how to empirically test the hypothesized role that workload plays in learning from instructional design.

1.5.1 Dividing Workload

Because the workload imposed by a learning task can be considered intrinsic, extraneous, or germane, researchers must decide how to measure these types of workload. This is a challenge because measures of workload typically yield a unitary workload score. Although multidimensional workload rating scales, such as the NASA-TLX, record scores on individual scales, these scales do not correspond to the divisions hypothesized by CLT. For instance, although the NASA-TLX has separate scales for physical demand and mental demand, it has no way to isolate whether the mental demand positively contributes to schema development (i.e., germane load). Thus, the three divisions of CLT workload cannot be measured independently.

In an attempt to address this limitation Paas and van Merriënboer (1993) have argued CLT researchers should calculate the *instructional efficiency* (i.e., a ratio of the amount of effort expended relative to the level of performance achieved) of their manipulations. Unfortunately, the lack of instructional efficiency calculations by other researchers has prevented comparisons of instructional efficiency effect sizes across research programs. While Paas & van Merriënboer (1993) have the correct goal in mind, a simpler and more potentially revealing analysis is also possible. For a manipulation that is hypothesized to incur a given cognitive load factor, its effect size should be calculated (e.g., Atkinson, Catrambone & Merrill, 2003). In a factorial design, the simple effects of each hypothesized workload should also be calculated and should then be tested for a significant difference between the levels of that factor (e.g., Klauer & Zhao, 2004).

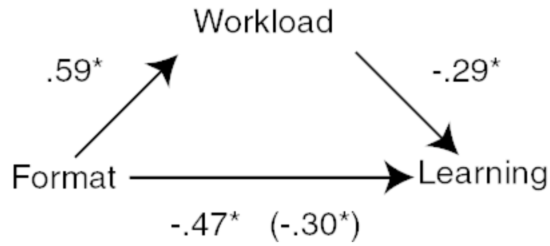
1.5.2 The Role of Workload on Schema Development

In CLT research workload is discussed as though it mediates[1] the effect of learning on performance (e.g., Paas & van Merriënboer, 1994; also, see Figure 1). However, no test of this assumption has been conducted. I hypothesize that the effect of instructional design on learning is partially mediated by workload. Under conditions of high workload there will be fewer resources available to modify schemas; as a result, learning will suffer (Sweller, 1999). However, some features of the instructional design might vary in their effectiveness of promoting schema development while eliciting equivalent amounts of workload. Therefore, I expect workload will act as a partial mediator (Baron & Kenny, 1986).

Furthermore, working memory capacity is proposed to moderate[2] the effect of instructional design on workload (see Figure 2). A higher working memory capacity

---

[1] A mediating variable describes the means by which the antecedent variably causally affects the consequent variable (Baron & Kenny, 1986; James & Brett, 1984). For instance, hours studying could mediate the effects of a student's level of motivation on his or her performance on an academic test. In this example high levels of motivation results in better performance *because* motivation increases the time spent studying.
[2] A moderating variable changes the extent of an independent variable's effect on a dependent variable (Baron & Kenny, 1986). For instance, test anxiety could moderate the effects of a student's knowledge on his or her performance on an academic test. High



*p < .05

Figure 1. Hypothesized model with workload mediating the effects of problem format on learning. The standardized regression coefficients found in this experiment are included. The standardized regression coefficient between problem format and learning, controlling for workload, is in parentheses.

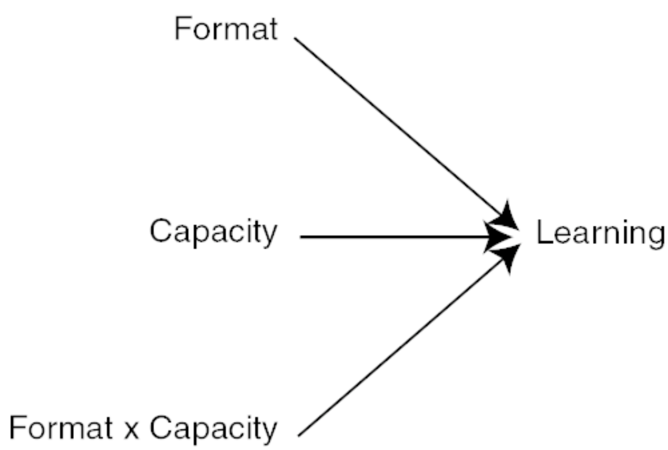


Figure 2. Hypothesized model with working memory capacity moderating the effect of problem format on learning.

levels of test anxiety could weaken the relationship between knowledge of the subject matter and test performance.

indicates more available resources; these resources can be used to offset additional task demands. Because workload is directly related to the resources available to devote to task demands (Xie & Salvendy, 2000), a higher working memory capacity could stabilize workload under varying task demands. Thus, I expect that a low working memory capacity will magnify the effect of the instructional design. If a learner has a low capacity, then any small reduction in workload imposed by the task should improve learning. To date, the moderating role of working memory has not been empirically tested in CLT research.

In this experiment, I will manipulate two features of the instructions: *problem format* and *contextual interference*. I will conduct simple effects analyses of these manipulations in order to measure intrinsic load (presumably affected by problem format) and germane load (presumably affected by contextual interference) independently. The data will also be used to test the hypothesis that workload is a mediating variable and working memory capacity is a moderating variable.

1.6 Problem Format Manipulation

One instructional design that has proven effective in allowing students to solve novel problems is manipulating the worked example format (Gerjets et al., 2004). Traditional training strategies teach learners how to solve a problem using a *molar* strategy. For instance, novices taught to solve complex probability problems learn to categorize the problem in terms of a permutation or combination and then choose the corresponding formula. This molar approach emphasizes the development of a problem-type schema, but does not emphasize the development of a robust, domain-level schema. In contrast, learners can be trained with *modular* examples that do not require classifying the problem

Molar Format

An urn contains five marbles, each a different color - white, yellow, red, green, and blue. Two marbles are taken out, one by one, and are not put back. What is the probability of first taking out the blue marble, and then the white marble?

Identify task features:	permutation without replacement: order of selection is important; no replacement of selected elements
Apply formula:	$A = \dfrac{n!}{(n-k)!}$
Insert values:	5! / (5 - 2)! = 20 permutations
Calculate probability:	1/20

Modular Format

An urn contains five marbles, each a different color - white, yellow, red, green, and blue. Two marbles are taken out, one by one, and are not put back. What is the probability of first taking out the blue marble, and then the white marble?

Find 1st event probability:	1/5
Find 2nd event probability:	1/4
Calculate the overall probability:	1/5 * 1/4 = 1/20

Figure 3. Molar worked example format and modular worked example format.

into a category prior to solving. This modular approach allows problems to be solved with a set of subprocedures that are independent of the problem category (Gerjets, Scheiter, & Catrambone, 2003). In the probability domain, this involves calculating simple event probabilities and then combining them to get the complex event probability. Assessing the category membership of the problem as a whole is unnecessary; assessing the probability of component parts (modules) of the problem is the only requirement. This has two hypothesized benefits: (1) it reduces the number of elements held in working memory simultaneously, and (2) it provides the basis for internal relationships between elements in a schema (Gerjets et al., 2004).

One of the strengths of the modular example format is the reduction of working memory load. In this way, it is consistent with CLT. Because learners do not have to keep a large number of problem features simultaneously activated in working memory, they will have more resources that they can devote to reasoning about the problem. This reduction in workload will allow learners to allocate resources for schema modifications. Consider the permutation problem in Figure 3 (molar). In order to understand why the appropriate formula is $A = \frac{n!}{(n-k)!}$ the learner must keep the following problem characteristics in mind:

(1) the goal state: 2 elements are selected from a set of 5

(2) the order the elements are selected is important

(3) the elements are not replaced after being selected

(4) the same selection constraints apply to both elements (i.e., order without replacement)

In order to understand how to apply the correct formula, the learner must also keep the following elements in mind:

(4) the formula: $A = \frac{n!}{(n-k)!}$

(5) A: total number of possible events

(6) n: number of elements in the set

(7) k: number of elements in the subset of interest

(8) how to use the factorial operator (!)

(9) probability of interest = $\frac{1}{A}$

Beyond the number of elements that must be represented in working memory, there are interrelationships between the elements. Thus, (4) also requires (5), (6), (7), and (8) to be understood. Although some of these elements can be relegated to external memory (e.g., writing notes) the interrelated elements must be simultaneously activated in working memory in order to develop or modify a schema. When the number of elements in working memory is greater than three there are substantial demands on working memory (Sweller, 1999).

Now consider the same permutation problem, but with a different worked example format (Figure 3, modular). To understand the first step, the learner must keep the following characteristics in mind:

(1) the goal of this step: calculate the probability of one event

(2) the formula: $\frac{1}{A}$

(3) A: total number of possible events

To understand the overall process of calculating this complex event the learner must also keep two more characteristics in mind:

(4) probability of interest = first probability * second probability

(5) how to use the multiplier operator (*) with fractions

Note that the number of interacting elements (as well as the total number of elements) has been substantially reduced. In order to understand the solution for the entire problem, the process by which the individual events are calculated does not need to be held in working memory simultaneously. Likewise, in order to understand the solution for the individual event probability, the formula for the complex event probability does not need to be held in working memory simultaneously. I hypothesize that this reduction in working memory demands would allow the learner to devote more resources to reasoning about the steps and could benefit schema development.

I believe that the schemas developed by learners studying these modular examples will be more robust because they will reflect the process by which probability events are calculated, along with the reasoning underlying the process. Because of this, learners' schemas will be better suited to solve a probability problem that has a novel structure. In contrast, learners in the molar condition will have developed schemas related to a specific problem category, and therefore would find it difficult to solve a problem for which they do not have a schema. Although they might know how to apply the formula they have learned to a new problem, they will likely find it difficult to modify the formula to solve a novel problem. Past studies of modular training are consistent with this hypothesis (Gerjets et al., 2004). This study will explore the utility of instructing with modular versus molar examples. I hypothesize that learners in the modular condition will perform better on the transfer test and will rate workload (i.e., intrinsic load) lower during study and testing.

1.7 Contextual Interference Manipulation

In addition to the practice format changes, I am interested in encouraging learners to make inter-example comparisons in order to promote reasoning about the examples. One method by which this might be accomplished is by varying the level of *contextual interference* that occurs during the training session. Contextual interference occurs when there are different variations of the task and these variations are practiced at the same time. For example, interference will occur when one is learning how to solve complex probability problems that require a different formula for each problem category. Contextual interference will be greater if the order in which the task variants are practiced is random. Evidence in other domains, such as paired-associate learning and motor skill learning, has indicated that this interference often causes worse performance during initial acquisition but has better retention and transfer performance (for review see Magill & Hall, 1990). Two explanations of the benefit of contextual interference on transfer performance have been proposed. First, learners might explicitly compare task variants, and this deep processing could improve schema development. Alternatively, the benefit might arise because on each trial learners have to retrieve the appropriate schema from long-term memory, and this frequent retrieval improves learning (Magill & Hall, 1990).

Although training with high contextual interference improves transfer, it also increases the subjective rating of workload of the learners (Paas & van Merriënboer, 1994; van Merriënboer, Schuurman, de Croock, & Paas, 2002). However, given that contextual interference has been shown to improve transfer and retention, this increase in workload should be considered germane load. That is, I believe that the added workload will contribute to schema development. I hypothesize that the contextual interference will elicit comparisons between studied examples, prompting learners to educe the similarities and differences between examples. This type of reasoning allows learners to elaborate on the instructional material and promotes schema development (Atkinson et al., 2003; Atkinson, Renkl, & Merrill, 2003; Chi, Bassok, Lewis, Reimann, & Glasser, 1989; Chi et al., 1994; Markman & Gentner, 2001).

However, this increased workload can be beneficial only if the learner is not already at capacity. Given that the probability material is complicated, I believe that the germane load associated with contextual interference will appear only when the intrinsic load has been reduced by modular examples. Therefore, I predict an interaction between the problem format and contextual interference. I expect this interaction to manifest itself in measures of workload, study time, and transfer performance.

In summation, this research proposes examining the effects of problem format and contextual interference on the acquisition of a cognitive skill, solving complex probability problems. A modular problem format is hypothesized to improve learning due to improved schema development. I believe the modular group will form more robust schemas because the workload during study will be lower, and because the modular presentation will facilitate connections between schema elements. High contextual interference is hypothesized to increase the workload during study. However, I believe this increased workload will arise from learners comparing examples and will thus enhance schema development, at least for learners with sufficient working memory capacity.

CHAPTER 2

METHOD

2.1 Participants

One hundred and thirty-three participants were recruited from Georgia Institute of Technology (n = 114) and Georgia State University (n = 19). Participants who had prior exposure to probability in a college level mathematics course were excluded from the experiment. Despite the prerequisite limit on prior exposure to probability in college, some participants who had taken a course registered for the study. Participants who indicated that they had exposure to probability in a prior college level course (n = 16) were excluded from all data analysis, leaving data from 117 participants. Participants from Georgia Tech were compensated with course credit; participants from Georgia State were compensated with $10 per hour.

2.2 Materials

2.2.1 Learning Materials

The learning materials (pretest and posttest, worked examples, and problem solving test) were similar to the materials of Gerjets et al. (2003). The pretest and posttest contained conceptual questions such as how to manipulate probabilities (simple and complex) and some simple probability calculations (Appendix A).

The worked example training materials had two changes from Gerjets et al. (2003). First, one additional problem type was added. Learners studied an additional problem category that explained how to multiply probabilities (Appendix E, problems 1 & 2). This problem category was included to rule out the possibility that the modular group performed better on the problem solving test due to explicit training on combining two probability events via multiplication. Second, the cover stories of three of the original eight training examples were altered (Appendix E, problems 3, 5, & 9). Gerjets et al. (2003) had four worked example categories: permutation with replacement, permutation without replacement, combination with replacement, and combination without replacement. Each of these categories had a pair of worked examples. The first example of the pair always had the same cover story (removing marbles from an urn). The second example of the pair had a different cover story for each category. In this study, the contextual interference manipulation necessitated changing the cover story of each pair's first example. Using the same cover story for each problem category could cue the learner to the problem's category, therefore limiting the interference in the high contextual interference condition (HighCI). Therefore I created new cover stories for those problems; each of the 10 examples had a unique cover story. To maintain consistency between conditions, I changed the low contextual interference condition's (LowCI) cover stories as well.

2.2.2 Problem Solving Test

The problem solving test was not changed from the Gerjets et al. (2003); it had 5 isomorphic items, 3 near transfer items, and 3 far transfer items (Appendix B). Participants were given a formula sheet for use during the problem solving test. This formula sheet contained the formulas they studied during the training but had only symbolic values and did not define the symbols (Appendix C). The NASA-TLX (Appendix D) was used to collect subjective ratings of workload (Hart & Staveland, 1988).

2.2.3 Automated Operation Span (Aospan) Task

The Aospan (Unsworth, Heitz, Schrock, & Engle, 2005) task is used to give a measure of working memory capacity. A trial in the Aospan task gives participants a series of letters to remember (between 2 letters and 7 letters) while they solve simple arithmetic problems (e.g., “2 / 2 + 1 = 3; True or False?”). Interspersed between each math problem is a new letter to remember. Therefore, if a trial has five math problems there will also be five letters to remember. After completing all the math problems in a trial the participant reports the letters that were presented, in the order they were presented. Scoring can be calculated as partial credit (e.g., remembering 3 out of 5 letters = 3 points; 5 out of 5 letters = 5 points), or as all-or-nothing credit (e.g., remembering 3 out of 5 letters = 0 points; 5 out of 5 letters = 1 point). All-or-nothing credit is most often used by researchers (Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005).

Participants are instructed to maintain their math accuracy above 85%; on each trial their total math accuracy is given in the right-hand corner of the screen (Unsworth et al., 2005). Low math accuracy scores indicate that the participant might have been concentrating on the secondary task (remembering letters) at the expense of the primary task (solving math problems), and as a result their performance on the secondary task might not be valid. Data from participants whose math accuracy is below some cut-off point (e.g., 85%) are typically removed (Conway et al., 2005).

2.3 Design and Assignment to Conditions

I used a 2 x 2 (Problem format x Contextual interference) factorial design. Participants’ score (all-or-nothing score) on the Aospan was used to classify them into one of four groups, based on previously reported Aospan quartile values: 0-26, 27-38, 39-

51, 52-75 (Unsworth et al., 2005). Participants were randomly assigned to condition, with the constraint of keeping the number of participants equal across experimental groups and Aospan quartiles.

In order to manipulate problem format, participants studied worked examples that were in either a modular or molar format. The contextual interference manipulation varied the level of interference between LowCI and HighCI by varying the order in which examples were presented. Participants in the LowCI condition studied worked examples that were ordered by problem category. Each problem categories' pair of worked examples was presented in sequence. Participants in the HighCI condition studied worked examples in what appeared to be a random order of problem categories. A worked example from one problem category was never preceded by the second worked example from the same category (see Table 1).

Table 1. Contextual interference was manipulated by presenting problem categories in a sequential order (LowCI) or a random order (HighCI)

Presentation order	LowCI	HighCI
1	multiplying probability	multiplying probability
2	multiplying probability	permutation without replacement
3	permutation without replacement	combination with replacement
4	permutation without replacement	multiplying probability
5	permutation with replacement	permutation with replacement
6	permutation with replacement	combination without replacement
7	combination without replacement	combination with replacement
8	combination without replacement	permutation without replacement
9	combination with replacement	combination without replacement
10	combination with replacement	permutation with replacement

2.4 Procedure

Participants were introduced to the overall format of the experiment and then completed a demographic questionnaire. Next they completed the Aospan task (Unsworth et al., 2005). Participants then completed the probability pretest. After completing the probability pretest, participants were given instructions on how to complete the NASA-TLX (Appendix F) and completed the NASA-TLX. The probability pretest also served as a reference task that participants could use when evaluating the workload of subsequent experimental tasks; the use of a reference task can reduce between-subject variability in workload estimation (Hart & Staveland, 1988).

Following the pretest, participants were given the worked example training materials to study. They marked the time they started studying the materials and the time they finished studying. Participants in all conditions were allowed to go back to previously studied examples. There was not a time limit for studying. When done studying, participants completed another copy of the NASA-TLX, rating their workload during study.

Participants then completed the problem solving test and recorded the time they started and stopped the test. Participants had 35 minutes to complete the test. Participants were required to monitor their own time, and the experimenter monitored their time also. If the 35 minutes elapsed and the participant was still working, the experimenter told him or her to reach a stopping point and turn in the test. During the problem solving test participants were not able to review the worked examples, but they were given a formula sheet. This minimized the possibility that differences between molar and modular groups would arise from differences in the number of formulas they had to memorize. Participants then completed the NASA-TLX again. Finally, they completed the probability posttest and another NASA-TLX.

CHAPTER 3

RESULTS

3.1 Incomplete and Unequal Data

Learners were instructed to take no more than 35 minutes to complete the problem solving test. Due to experimenter oversight, some participants worked past the 35 minutes. Additionally, a few participants ran out of time to complete the test (i.e., the two hours allotted for the experiment had elapsed, but the 35 minutes for the test had not). Participants who took more than 38 minutes to complete the test ($n = 4$) and participants who did not have the full time to complete the test ($n = 4$) were excluded from all analyses.[3] The actual sample sizes for each cell are shown in Table 2.

All analyses were conducted with $\alpha = .05$.

Table 2. Cell Sizes

Contextual Interference	Molar	Modular	Total
LowCI	30	29	59
HighCI	22	27	49
Total	52	56	

[3] Of the four participants who worked longer than 38 minutes, one belonged to the modular + LowCI group, one belonged to the modular + HighCI group, and two belonged to the Molar + HighCI group. Of the four participants who did not have the full 35 minutes, one belonged to the molar + LowCI group and three belonged to the molar + HighCI group.

Table 3. Correlation Matrix

Variable	1	2	3	4	5	6	7	8
1. Format	--	-.06	.24*	.67**	.59**	.44**	-.47**	.48**
2. CI		--	.02	.10	-.07	-.00	.11	-.12
3. Study time: Intro			--	.34**	.26**	.33**	-.28**	.24*
4. Study time: WE				--	.48**	.49**	-.34**	.31**
5. Study TLX					--	.27**	-.47**	.65**
6. Test time						--	-.27**	.27**
7. Test score							--	-.60**
8. Test TLX								--
9. Pretest time								
10. Pretest score								
11. Posttest time								
12. Posttest score								
13. Aospan score								
14. Aospan total								
15. SAT composite								

*p < .05
**p < .01

Table 3 Continued

9	10	11	12	13	14	15
-.03	-.09	-.15	.09	-.03	-.04	.13
.06	.10	-.17	.06	.03	-.04	-.08
.33**	-.24*	.37**	-.10	-.01	.01	-.19
.22*	-.10	.07	.01	-.09	-.12	-.11
-.04	-.30**	-.06	-.23*	-.10	-.07	-.13
.32**	-.04	.16	.07	-.08	-.06	-.22*
-.07	.37**	.07	.30**	.15	.08	.26*
-.04	-.32**	-.11	-.25**	-.04	.04	-.17
--	-.17	.46**	-.07	-.06	-.10	-.35**
	--	-.15	.58**	.08	.04	.43**
		--	.00	-.03	-.04	-.25*
			--	.09	.07	.35**
				--	.92**	.21*
					--	.16
						--

3.2 Study Phase

3.2.1 Study Workload

Two indicators (total study time and TLX score for the study session) were used to assess the effects of the instructional design manipulations on the hypothetical construct of mental workload. If the mental workload construct manifests itself in both time and subjective workload, then one would expect these dependent variables to be correlated. This hypothesized relationship is present in the observed data (see Table 3).

A MANOVA was used to analyze the effects of the instructional design manipulations on the mental workload construct. The multivariate test statistic Wilk's lambda was used to evaluate the F ratio. The problem format manipulation had a significant effect on workload, multivariate $F(2, 102) = 53.13$, $p < .005$. The contextual interference manipulation did not have a significant effect on workload, multivariate $F(2, 102) = 1.73$, $p = .18$. Additionally, there was not a significant interaction between problem format and contextual interference, multivariate $F(2, 102) = 2.21$, $p = .12$. To follow-up the MANOVA, univariate ANOVAs were conducted for each dependent variable (time and TLX score); this allows one to specify which dependent variable (or variables) were effected and the strength of the effect on each dependent variable.[4]

---

[4] The mental workload construct is hypothesized to demonstrate itself in two variables (time and subjective workload ratings) during study and three variables (time, subjective workload ratings, and performance) during the test. A correlation between these two study variables is a necessary, but not sufficient, condition to demonstrate the presence of the mental workload construct (likewise for the three test variables). It is not a sufficient condition because another, unspecified variable could be causing the relation. For instance, suppose that the problem format manipulation changes the problem solving *strategy* that participants use. This strategy (e.g., a molar approach) might then effect time, workload, and performance. If this occurs, a significant F value in the MANOVA would indicate only that the dependent variables consistently change as a result of the independent variables. It does not, however, resolve the question of why the variables

3.2.2 Study Time

Study time was self-paced and did not have a time limit. Participants recorded their time studying both portions of the learning materials (the introduction and the worked examples). The introductory materials differed only between the molar and modular group; they did not differ between LowCI and HighCI groups. The worked example materials differed between all four groups. Therefore, the time to study the introduction and the worked examples was analyzed separately. This allows one to determine if the order of the worked examples (contextual interference) affected the time participants studied those worked examples.

A one-way ANOVA was used to analyze study time on the introductory packet (see Table 4). The molar group studied the introduction longer than the modular group, $F(1, 106) = 6.31$, $MSE = 11.57$, $p < .05$. Cohen's *f* statistic for these data yields an effect size estimate of 0.24, which corresponds to a medium effect size.

A two-way ANOVA was used to analyze study time on the worked example packet (see Table 4).[5] The molar group studied the learning materials longer than the modular group, $F(1, 103) = 96.58$, $MSE = 20.11$, $p < .005$, $f = 0.97$ (large effect). The HighCI group had a longer study time than the LowCI group, $F(1, 103) = 4.83$, $p < .05$, $f= 0.21$ (small effect). In addition to these two main effects, there was a significant interaction between problem format and contextual interference, $F(1, 103) = 4.59$, $p < .05$, $f = 0.22$ (small effect). A simple effects analysis indicated that contextual

---

change (i.e., what hypothetical construct(s) are responsible). Therefore, interpreting the univariate ANOVAs individually clarifies that my focus is the relation between the independent variables and the dependent variables, not the relation between a hypothetical construct and the dependent variables.

Table 4. Study Times and Test Time

Contextual Interference	Molar	Modular	Avg.
		Learning introduction	
LowCI	8.17 (3.06)	7.03 (3.02)	7.61 (3.07)
HighCI	9.05 (4.09)	6.74 (3.60)	7.78 (3.96)
Avg.	8.54 (3.52)	6.89 (3.29)	7.69 (3.49)
		Learning worked examples	
LowCI	13.87 (5.06)	7.14 (2.85)	10.56 (5.31)
HighCI	17.67 (6.66)	7.19 (2.88)	11.77 (7.15)
Avg.	15.43 (6.01)	7.16 (2.84)	11.10 (6.20)
		Test	
LowCI	31.41 (4.81)	26.62 (4.87)	29.02 (5.37)
HighCI	31.55 (5.17)	26.93 (4.86)	29.00 (5.47)
Avg.	31.47 (4.92)	26.77 (4.82)	29.01 (5.39)

Note. Time is reported in minutes. Standard deviations are in parenthesis.

[5] One participant incompletely recorded worked example study time. This participant was excluded from the analysis of study time.

Table 5. Weighted Overall NASA-TLX Scores: Ratings During Study Phase and Test Phase

Contextual Interference	Molar	Modular	Avg.
		Study	
LowCI	49.18 (17.37)	30.21 (13.88)	39.85 (18.31)
HighCI	52.58 (16.96)	24.40 (15.66)	37.05 (21.43)
Avg.	50.62 (17.11)	27.40 (14.92)	38.58 (19.74)
		Test	
LowCI	72.27 (16.34)	55.39 (17.76)	63.97 (18.92)
HighCI	70.02 (14.12)	50.57 (17.78)	59.30 (18.81)
Avg.	71.31 (15.33)	53.07 (17.77)	61.85 (18.93)

Note. Overall TLX score ranges from 0 to 100; a higher value indicates higher workload. Standard deviations are in parenthesis.

interference (HighCI) increased study time in the molar group ($F(1, 103) = 8.87$, $p < .005$) but not the modular group ($F < 1$).

3.2.3 Workload Assessment

Overall TLX scores were also analyzed with a univariate ANOVA. The means and standard deviations for each group are presented in Table 5. Problem format had a significant main effect on score; the molar group rated their study workload higher than the modular group, $F(1, 104) = 57.92$, $MSE = 255.36$, $p < .005$, $f = 0.75$ (large effect). Contextual interference did not have an effect on TLX score, $F < 1$. Additionally, there was not an interaction between problem format and contextual interference, $F(1, 104) = 2.21$, $p = .14$.

3.2.4 Planned Comparisons: Time and Workload

Planned comparisons were conducted to compare (1) the difference between the modular + LowCI and the molar + LowCI group, (2) the difference between the modular + LowCI group and the modular + HighCI group, and (3) the difference between the molar + LowCI group and the molar + HighCI group. The problem format manipulation was hypothesized to alter the intrinsic load of the task. In contrast, the contextual interference manipulation was hypothesized to alter the germane load of the task. If problem format affects intrinsic load and contextual interference affects germane load, then these comparisons determine (1) if intrinsic load affected the dependent variables, (2) if germane load affected the dependent variables when intrinsic load was hypothesized to be the lowest, and (3) if germane load affected the dependent variables when intrinsic load was hypothesized to be the highest.

Table 6 presents the results of the set of planned comparisons for both dependent variables: study time and study TLX scores. Intrinsic load affected study time (comparison 1). Germane load had a significant effect on study time when intrinsic load was the highest (comparison 3), but no significant effect on study time when intrinsic load was the lowest (comparison 2). Additionally, intrinsic load affected TLX scores during study (comparison 1). Germane load had no significant effect on TLX scores; this was true at both low and high levels of intrinsic load (comparisons 2 and 3).

3.3 Test Phase

3.3.1 Test Workload

The hypothetical construct of workload was assumed to be measurable by three indicators: test time, test performance (total score), and subjective workload during test (NASA-TLX). Therefore, these three dependent variables were analyzed using a 2 x 2 MANOVA with problem format and contextual interference as the two factors. Wilk's lambda was used to evaluate the F ratio. Problem format had a significant effect on the three dependent variables, multivariate $F(3, 101) = 18.99$, $p < .005$. Contextual interference did not have a significant effect, multivariate $F < 1$. Additionally, the interaction between problem format and contextual interference was not significant, multivariate $F < 1$. The MANOVA was followed-up with univariate ANOVAs for each dependent variable.

Table 6. Planned Comparisons on Study Time and Study TLX

Comparison	df	*F*	Cohen's *f*	*p*
			Time	
(1) Modular + LowCI vs. molar + LowCI	1	33.19**	0.57	< .005
(2) Modular + LowCI vs. modular + HighCI	1	< 1	--	.97
(3) Molar + LowCI vs. molar + HighCI	1	8.87**	0.29	< .005
Error	103	(20.11)		
			TLX	
(1) Modular + LowCI vs. molar + LowCI	1	20.78**	0.45	< .005
(2) Modular + LowCI vs. modular + HighCI	1	1.85	--	.18
(3) Molar + LowCI vs. molar + HighCI	1	< 1	--	.45
Error	104	(255.36)		

Note. MSE are in parenthesis.

**p < .005

3.3.2 Test Time

A 2-way ANOVA was used to assess the effects of problem format and contextual interference on time to complete the test.[6] The molar group took a longer time to complete the test than the modular group (see Table 4), $F(1, 103) = 24.25$, MSE = 24.13, $p < .005$, $f = 0.49$ (large effect). There was not a main effect of contextual interference, $F < 1$, nor was there an interaction between problem format and contextual interference, $F < 1$.

3.3.3 Test Performance

The problem solving test had three item types: isomorphic, near transfer, and far transfer. Collapsing across all conditions, participants, on average, correctly answered 46% of the isomorphic items, 44% of the near transfer items and 25% of the far transfer items. Mean performance for each group on each item type is presented in Table 7. For each item type a 2-way ANOVA was conducted. For each item type, only the problem format manipulation significantly affected performance (see Table 8); in each case the modular group performed better than the molar group.

3.3.4 TLX Score

A 2-way ANOVA was used to assess the effects of problem format and contextual interference on subjective ratings of workload during the test. Each participant's overall NASA-TLX score was used. The modular group experienced lower workload when completing the test than the molar group, $F(1, 104) = 31.50$, MSE = 278.65, $p < .005$, $f = 0.55$ (large effect). There was no significant effect of contextual

---

[6] One participant incompletely recorded test time. This participant was excluded from the analysis of test time.

Table 7. Problem Solving Test Performance as a Function of Problem Format, Contextual Interference, and Item Type

Contextual Interference	Molar	Modular	Avg.
		Isomorphic	
LowCI	1.70 (1.47)	2.69 (1.63)	2.19 (1.61)
HighCI	1.45 (1.41)	3.19 (1.57)	2.41 (1.72)
Avg.	1.60 (1.43)	2.93 (1.61)	2.29 (1.66)
		Near transfer	
LowCI	0.77 (0.86)	1.69 (0.81)	1.22 (0.95)
HighCI	1.23 (1.02)	1.63 (0.93)	1.45 (0.98)
Avg.	0.96 (0.95)	1.66 (0.86)	1.32 (0.97)
		Far transfer	
LowCI	0.20 (0.48)	1.10 (1.01)	0.64 (0.91)
HighCI	0.41 (0.73)	1.26 (1.10)	0.88 (1.03)
Avg.	0.29 (0.61)	1.18 (1.05)	0.75 (0.97)

Note. Mean accuracy. Standard deviations are in parenthesis. Test contained 11 items: 5 isomorphic, 3 near transfer, 3 far transfer.

Table 8. Problem Solving Test Performance ANOVAs, as a Function of Problem Format, Contextual Interference, and Item Type

Source	df	*F*	Cohen's *f*	*p*
		Isomorphic		
Format	1	21.16**	0.45	< .005
CI	1	< 1	--	.67
CI * Format	1	1.57	--	.21
Error	104	(2.33)		
		Near transfer		
Format	1	14.52**	0.37	< .005
CI	1	1.33	--	.25
CI * Format	1	2.24	--	.14
Error	104	(0.81)		
		Far transfer		
Format	1	27.28**	0.51	< .005
CI	1	1.18	--	.28
CI * Format	1	< 1	--	.87
Error	104	(0.75)		

Note. MSE in parenthesis.

**p < .005

interference, $F(1, 104) = 1.20$, $p = .28$. There was also no significant interaction between problem format and contextual interference, $F < 1$.

3.3.5 Planned Comparisons: Time, Performance, and Workload

The same planned comparisons that were conducted on the dependent variables collected during the study phase were conducted on the dependent variables collected during the test phase. The results are shown in Table 9. Differences in intrinsic load during study (comparison 1) affected test time, test performance, and test workload. Germane load during study did not affect test time, test performance, or test workload at either the low level of intrinsic load (comparison 2) or the high level of intrinsic load (comparison 3).

3.4 Pretest-posttest Difference Scores

A pretest-posttest difference score was calculated for each participant and compared across groups using a 2 x 2 ANOVA. Four subjects were not given the posttest because the time allocated for the experiment had elapsed before they could be given the posttest; three of those participants were in the molar + HighCI group and one participant was in the molar + LowCI group. These four participants were excluded from this analysis.

The molar group had a higher mean difference score ($M = 1.56$, $SD = 1.74$) than the modular group ($M = 0.86$, $SD = 1.80$), but this difference was not significant, $F(1, 100) = 3.44$, $MSE = 3.18$, $p = .07$. There was not a main effect of contextual interference, $F < 1$. Additionally there was not an interaction between problem format and contextual interference, $F < 1$. One possible explanation for the slightly higher scores in the molar group, relative to the modular, is that one question asked participants about the definition

Table 9. Planned Comparisons on Test Time, Test Score, and Test TLX

Comparison	df	*F*	Cohen's *f*	*p*
		Time		
(1) Modular + LowCI vs. molar + LowCI	1	13.81**	0.37	< .005
(2) Modular + LowCI vs. modular + HighCI	1	< 1	--	.82
(3) Molar + LowCI vs. molar + HighCI	1	< 1	--	.93
Error	103	(24.13)		
		Test score		
(1) Modular + LowCI vs. molar + LowCI	1	15.06**	0.38	< .005
(2) Modular + LowCI vs. modular + HighCI	1	< 1	--	.43
(3) Molar + LowCI vs. molar + HighCI	1	< 1	--	.59
Error	104	(7.77)		
		TLX		
(1) Modular + LowCI vs. molar + LowCI	1	15.07**	0.38	< .005
(2) Modular + LowCI vs. modular + HighCI	1	1.17	--	.28
(3) Molar + LowCI vs. molar + HighCI	1	< 1	--	.63
Error	104	(278.65)		

Note. MSE in parenthesis.

**p < .005

of the ! sign (factorial operator). The molar group's learning materials defined the ! sign, whereas the modular group's learning materials did not (because the modular instructional approach does not use the ! sign).

3.5 Instructional Efficiency

Because measures of workload themselves cannot indicate a difference between effective and ineffective workload (or a difference between intrinsic, extraneous, and germane load), Paas and van Merriënboer (1993) have suggested the use of a measure of *instructional efficiency*. An instructional design is said to be efficient if it has low workload but high performance; conversely, if it has high workload but low performance it is inefficient. To calculate instructional efficiency one converts performance and workload raw scores to z-scores and divides the difference between the two scores by $\sqrt{2}$. This allows each condition's instructional efficiency to be compared to a theoretical baseline efficiency of 0 (Paas & van Merriënboer, 1993).

Instructional efficiency was calculated for each group, using subjective workload ratings (overall TLX score) during study and total test score. The conditions, ordered from most efficient to least efficient, are modular + HighCI (E = 1.06), modular + LowCI (E = 0.81), molar + LowCI (E = -1.04), and molar + HighCI (E = -1.09).

3.6 Working Memory Capacity as a Moderator and Workload as a Mediator

Two hypothesis were tested: (a) working memory capacity moderates the effect of problem format on workload, and (b) the effect of problem format on problem solving performance is mediated by workload (see Figure 4).[7] The model in Figure 4 can be

[7] Hypotheses a and b could be generalized to both instructional design manipulations, but here I present them in terms of the problem format manipulation only. Hierarchal regression models with problem format and contextual interference were evaluated for

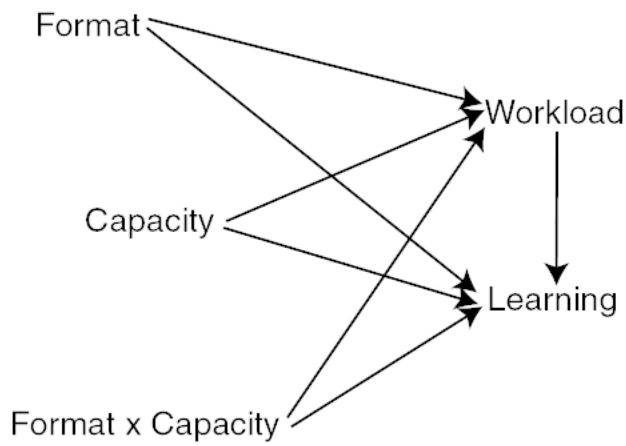


Figure 4. Hypothesized model of problem format (format), working memory capacity (capacity), workload, and learning.

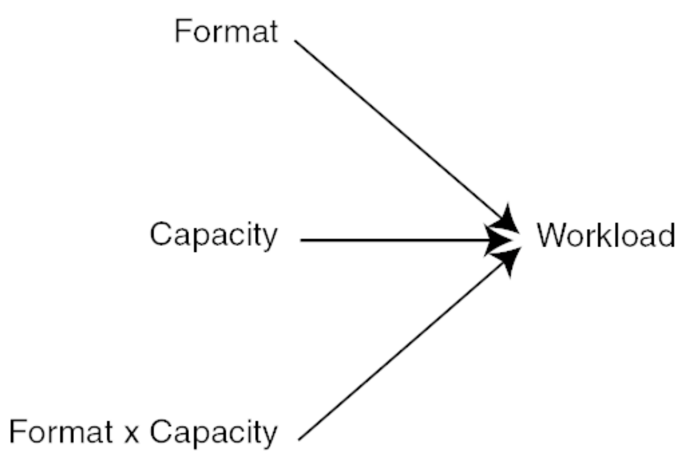


Figure 5. Hypothesized model with working memory capacity moderating the effect of problem format on workload.

---

both workload and total score; in both models contextual interference was not a significant predictor and did not moderate the effect of problem format. Therefore, the contextual interference variable was not included in the subsequently reported analyses.

tested with regression equations that form a set of ordered analyses to determine (1) if problem format and working memory capacity[8] affect problem solving performance, (2) if format and capacity affect workload and if workload affects problem solving performance, and (3) if workload mediates the effect of problem format and working memory capacity on problem solving performance (Baron & Kenny, 1986).

Step 1 of the ordered analyses tested the model in Figure 2. Problem format was a significant predictor of total score (see Table 10). However, working memory capacity was not a significant predictor of total score. Furthermore, working memory capacity did not moderate the effect of problem format on total score (see Table 10).

Step 2 of the ordered analyses tested the model in Figure 5. In this step subjective mental workload rating (NASA-TLX score) was the criterion variable, rather than total score. Problem format was a significant predictor of workload score (see Table 11). Working memory capacity was not a significant predictor of workload score, nor did it moderate the effect of problem format on workload score (Step 1 and 2, respectively, in Table 11).

The first two steps of the ordered analysis indicated that working memory capacity did not predict subjective ratings of workload nor problem solving performance. Additionally, it did not moderate the effects of problem format on either variable. Therefore, the hypothesized moderating role of working memory was removed from the original model, resulting in a partial mediation model (see Figure 1). This model states that the effect of problem format on total score is partially mediated by workload.

---

[8] All regression analyses used participants' Aospan total score, which uses partial credit scoring, as the measure of working memory capacity (see Conway et al., 2005).

Table 10. Summary of Hierarchical Regression Analysis for Variables Predicting Problem Solving Test Performance (N = 107)

Variable	b	SE b	*ß*
Step 1			
Format	-2.96	0.54	-.47*
WMC	-0.02	0.02	.06
Step 2			
Format X WMC	-0.07	0.05	-.68

Note. R^2 = .23 for Step 1; ΔR^2 = .02 for Step 2 (p > .05). WMC was measured using Aospan total score.
*p < .05

Table 11. Summary of Hierarchical Regression Analysis for Variables Predicting Subjective Workload Ratings (NASA-TLX; N = 107)

Variable	b	SE b	*ß*
Step 1			
Format	22.83	3.11	.58*
WMC	-0.08	0.13	-.05
Step 2			
Format X WMC	-0.14	0.27	-.22

Note. R^2 = .35 for Step 1; ΔR^2 < .01 for Step 2 (p > .05). WMC was measured using Aospan total score.
*p < .05

A series of three regression equations were used to test the partial mediation model (see Baron & Kenny, 1986). Equation 1 regresses workload on problem format, $R2 = .35$, $F(1, 106) = 56.66$, $MSE = 256.39$, $p < .005$. Equation 2 regresses total score on problem format, $R2 = .22$, $F(1, 106) = 29.95$, $MSE = 7.69$, $p < .005$. Equation 3 regresses total score on problem format and workload, $R2 = .28$, $F(2, 105) = 19.96$, $MSE = 7.21$, $p < .005$. Figure 1 shows the model with the standardized regression coefficients from the regression equations. Problem format and workload are significant predictors of total score. Additionally, the effect of problem format on total score is partially mediated by workload. When workload is held constant the problem format ß drops from -.47 to -.30.

CHAPTER 4

DISCUSSION

4.1 Problem Format

Manipulating the problem format significantly altered learners' behavior during study and test. Participants who were taught a modular approach, in which they learned to decompose complex probability problems and apply a general, modular solution, spent less time studying worked examples and rated their mental workload lower. This reduced time and effort during study and improved performance on a subsequent problem solving test. This improvement was seen across problem types: isomorphic, near transfer, and far transfer items had significantly more correct answers for the modular group. Additionally, participants took less time to solve the problems and rated their mental workload lower. These results are consistent with previous work in this domain (Gerjets et al., 2004).

This study extends Gerjets et al. (2004) by ruling out the possibility that one cause of the different performances between problem approach groups arose because learners did not know how to combine probability events. To correctly answer both the near and far transfer items one must combine probabilities via multiplication. In previous studies the modular group studied worked examples that illustrated multiplying probabilities, but the molar group did not (Gerjets et al., 2004). In this experiment both groups studied two worked examples on multiplying probabilities. Although the molar group was explicitly instructed on this method, they were still not able to transfer that knowledge to novel problems.

This combination of evidence suggests that a general, modular approach can help novices learn to perform the task and transfer that knowledge to novel tasks. This claim is also supported by research in a domain outside of probability. Frederiksen & White (1989) taught participants to pilot a ship in a video game. Although many other variables were simultaneously manipulated in their training program, they found evidence that instructing learners on general ship movement commands, rather than specific ship movement strategies, was helpful (Frederiksen & White, 1989). This finding demonstrates the promise of using general, modular instructions. Nevertheless, subsequent work outside of the probability domain needs to be conducted.

The problem format manipulation has strong implications for CLT (Sweller et al., 1998). The modular problem format reduced the intrinsic load of the learning task. This finding is counter to many conceptions of CLT, which claim that intrinsic load is inherent to the domain and is not mutable by instruction (e.g., Sweller et al., 1998). Recently, however, research has shown that intrinsic load might be a function of multiple variables. For instance, the amount of prior knowledge of the domain that the learner has can affect the efficacy of instructional design manipulations (Kalyuga et al., 2003). In terms of intrinsic load, one could argue that learners with higher domain knowledge have more robust schemas that allow them to chunk more information, thereby decreasing the amount of intrinsic load. The effects of problem format in this study and prior studies (Gerjets et al., 2004), combined with findings about prior knowledge (Kalyuga et al., 2003) suggest that when reasoning about intrinsic load differences, the researcher must consider how multiple variables will affect intrinsic load.

Although the problem format improved scores on the problem solving test, it did not affect the change in scores on conceptual questions from pretest to posttest. The content of the questions on these tests centered around basic probability ideas (see Appendix A). Most of this knowledge was contained in the introduction of the learning materials that all participants studied. This might explain why there was no difference between groups.

4.2 Contextual Interference

The contextual interference results were not expected. Higher levels of contextual interference were hypothesized to increase workload regardless of problem format. Because of differences in intrinsic load, however, contextual interference was hypothesized to improve problem solving performance for the modular group only. Instead, the contextual interference manipulation increased time (but not workload) during study for the molar group, and had no effect on problem solving performance in either the molar or modular group. This study time difference suggests that contextual interference did change the molar group's learning behavior. Given the data collected, I cannot determine whether the increased time arose because learners explicitly compared the current example to previous examples (e.g., revisiting previous examples), or because learners spent more time studying each example (e.g., retrieving information about the current problem type from long-term memory). Future work could answer this question.

It is surprising that there were no differences in subjective workload ratings between the molar + LowCI and molar + HighCI groups. I do not yet have a compelling explanation for why study time increased but workload ratings did not. One possibility is that the NASA-TLX was not sensitive enough to pick up differences in workload caused by the different ordering of worked examples.

The contextual interference manipulation did not affect scores on the problem solving test. For the molar group, study time increased from low to high contextual interference, but performance did not change. Thus, for this group it appears that the detrimental effect of contextual interference (increased learning time) occurred, but the beneficial effect did not occur (Magill & Hall, 1990). One explanation might be provided by Shea, Kohl, & Indermill (1990). They compared blocked and random practice (LowCI and HighCI, respectively) on a force production task across varying numbers of practice trials (50, 200, and 400) and across blocked and random retention tests. The random practice trials reduced errors on the random retention test, but only when the number of practice trials was 200 or 400. Interestingly, the random practice group was also better than the blocked practice group test on the blocked retention test, but only after 400 practice trials. Shea and colleagues (1990) argue that a novice who is first learning the task domain will not benefit from random practice trials. They argue the learner needs to move from the beginning stage to the intermediate stage of skill acquisition before he or she can gain the benefits of random practice (Shea et al., 1990; see also VanLehn, 1996). It might be that in the present experiment the contextual interference manipulation began too early in the process of skill acquisition. In the molar condition, the learners might have been overwhelmed by the random order at a time when they were still trying to understand and integrate the different features of the problem domain.

It is also surprising that no effects of the contextual interference manipulation were found in the modular group. From a CLT perspective (Sweller et al., 1998), the modular problem format should have reduced intrinsic load, allowing the learner to engage in germane processing, which should improve performance (van Merriënboer et al., 2002). No evidence for increased germane processing was found (as measured by TLX scores), and no evidence for improved performance was found (as measured by problem solving score). Because this pattern of findings is different from what CLT would predict, it might suggest revisions to CLT are necessary. Alternatively, it is possible that a feature unique to the modular group's problem format made the manipulation ineffective. For instance, when using the modular format, one of the learner's critical decision is to identify the number of simple events in the problem. The implementation of the contextual interference manipulation, however, did not systematically vary the number of simple events from problem to problem.

4.3 Implications for CLT Methodology

The measure of instructional efficiency (Paas & van Merrienboer, 1992) supports the use of modular instruction with contextual interference. For the modular group, the contextual interference manipulation's effect on efficiency is not dramatic. Both modular groups are substantially better than the molar groups, however. Unfortunately the measure of instructional efficiency does not allow one to distinguish between the different types of cognitive load. For instance, the group with the highest instructional efficiency rating had lower mental workload and higher scores. This lower mental workload could have arisen from a reduction in intrinsic load or a reduction in extraneous load. Currently, there is no empirical way to determine which type of cognitive load was reduced. The data (and the instructional efficiency measure) only support the distinction between effective and ineffective workload (e.g., Xie & Salvendy, 2000), not the distinction between intrinsic, extraneous, and germane. Additionally, the effective/ineffective distinction is more parsimonious because it uses two variables, rather than three.

Instead of using instructional efficiency, the assessment strategy used in this study might be a more useful method to assess mental workload within a CLT framework. In addition to reporting significance values of the subjective workload differences between condition, I report the effect sizes. Therefore, one can see the size of the workload effect for a given manipulation. When this is supplemented with performance data, one can determine if the workload was effective or ineffective and has a point estimate of the workload, which is not dependent on the measure used. Other researchers can then make a priori predictions about the direction and strength of effect on workload for a given manipulation. The method is a step towards helping the CLT community to quantify the load imposed by the instructional design manipulations. These estimates of workload could be combined with models of learner characteristics (e.g., domain knowledge) and task characteristics (e.g., elements in working memory). As such quantification becomes more common, formal models based on a CLT framework will become possible.

4.4 Working Memory Capacity as a Moderator and Workload as a Mediator

Working memory capacity was hypothesized to moderate the effects of problem format on subjective ratings of workload. In this paper I define workload as the utilization of internal resources. Working memory capacity is assumed to reflect the amount of mental resources available or the ability to focus those resources on the task (e.g., Engle, 2002). One might expect that as working memory capacity increases, the greater amount of mental resources reduces the task demands, and thus lowers subjective ratings of workload. The regression results did not support the hypothesis that working memory capacity moderates workload, however. Moreover, working memory capacity did not moderate the effect of problem format on performance.

It is possible that a restriction of range prevented finding effects of working memory. The current sample had a median score of 47.0, which is higher than previously reported distributions (median = 39.2 in Unsworth et al., 2005). Additionally, this sample was more negatively skewed (skew = -.47) than previously reported (skew = -.02 in Unsworth et al., 2005). This indicates the working memory capacity of the sample was shifted toward the higher end of the distribution, relative to a previous study with participants from the Atlanta community (e.g., Unsworth et al., 2005). It might be that the large proportion of Georgia Tech students in the sample increased mean working memory capacity, and reduced the ability to detect correlations between working memory capacity and workload and between working memory capacity and total score.

Workload was hypothesized to mediate the effect of problem format on learning. The set of regression equations demonstrated that workload partially mediates the effect. There was a significant indirect effect of problem format on total score (via workload), and a significant direct effect of problem format on total score. The standardized correlation coefficient of workload is negative, indicating that higher workload during learning is associated with lower test score. This pattern suggests that some of the effect of the problem format is carried by ineffective workload; the molar group experienced a high level of workload during study, and this workload did not translate into better performance. This finding is important for cognitive load theory. The mediating role of workload has been assumed (Sweller et al., 1998), but never empirically verified.

Also, this pattern of results shows that workload is not the only explanation for the differences in test scores; the problem format significantly predicted test score when workload was held constant. This finding demonstrates that although a CLT framework might be useful to interpret the effect of manipulating instructional design, differences in workload alone might not be a sufficient causal explanation. The cognitive task analysis outlined in the introduction can inform plausible explanations for the significant direct effect of problem format on total score. When the modular group solves transfer problems they do not have to use a different problem solving process then they have been using. When the molar group solves transfer problems they must recognize that there are two different sub-problems within the problem and that they must be solved separately and then combined. This explanation is not dependent on workload differences and can account for differences in test scores. In this case the cognitive explanation provides further support for the use of a general approach that is adaptable to a variety of problems.

4.5 Future Work

These results suggest follow-up studies along two research paths. First, the problem format manipulation should be replicated in a different domain. Second, the contextual interference manipulation should be further explored. One promising direction is to alter the contextual interference at different stages of skill acquisition; this would extend the findings of Shea et al. (1990). Instead of training participants with a different number of practice trials and comparing blocked to random practice (Shea et al., 1990), one could introduce the random practice after different amounts of blocked practice. Additionally, it might be useful to use traditional problem solving problems rather than worked examples. This would provide timing and accuracy data on each item during the acquisition phase, which would allow one to measure the rate of learning. Based on the rate of learning, the optimal point at which to switch from blocked practice (LowCI) to random practice (HighCI) could be determined.

APPENDIX A

PROBABILITY PRETEST/POSTTEST

1. What is a random experiment?
(a) An experiment where the probability of getting a particular outcome is less than chance
(b) An experiment that always leads to a different outcome when being repeated under essentially the same conditions
(c) An experiment whose outcome can not be predicted with certainty and that can be indefinitely repeated under essentially the same conditions
(d) An experiment in which the set of possible outcomes is not known

2. How are individual events related to complex events?
(a) Individual events have a lower probability of occurrence than complex events.
(b) Complex events are constituted by a number of individual events that result from conducting a random experiment repeatedly.
(c) Complex events result when selecting elements from two different sets of elements, whereas individual events always result when there is only one set of elements to choose from.
(d) Individual events can be considered complex events as long as the number of acceptable outcomes can not be determined in advance.

3. What is meant in probability theory by "selecting" elements from a set?
(a) Determining specific elements in a set of elements by taking into account their specific features (e.g., their value)
(b) Randomly choosing elements from a set of elements
(c) Choosing elements that differ from the other elements in the set
(d) Choosing elements from the set and returning them in case they belong to a specific category of elements

4. Imagine you select elements from a set by chance. What is meant by "with replacement" in this case?
(a) The number of different elements that are eligible for selection is infinite.
(b) Elements that were once selected are not eligible for being selected again.
(c) Elements that were once selected are eligible for being selected again.
(d) Each element that is selected is then replaced by another element that is at least similar to the selected element.

5. Imagine that you randomly select two cards from a card game. In which case is the probability of selecting a specific card as the second card independent of selecting the first card?
   (a) If this specific second card occurs twice in the card game
   (b) If the first card is replaced after selection
   (c) If the first card is not replaced after selection
   (d) If this specific second card is different from the first card

6. What is the meaning of the term n! ?
   (a) n! = n * (n-1) * (n-2) * ... * 2 * 1
   (b) n! = n + (n-1) + (n-2) + ... + 2 + 1
   (c) n! = n * (n+1) * (n+2) * ... * (n+n)
   (d) n! = (n*1) + (n*2) + ... + (n*n)

7. What is meant by "possible outcomes" in probability theory?
   (a) Possible outcomes are those events whose probability of occurrence has to be calculated.
   (b) Possible outcomes are all events that can occur when conducting a random experiment.
   (c) Possible outcomes are those events that occur with a lower probability than the other outcomes.
   (d) Possible outcomes are those events that occur with a higher probability than the other outcomes.

8. What is less probable: Tossing a "6" three times in succession or tossing a "2", a "3", and finally tossing a "4" in succession?
   (a) Tossing "2", "3", and "4" in succession is less probable.
   (b) Tossing a "6" three times in succession is less probable.
   (c) Both events occur with the same probability.
   (d) There is no systematic relation between the two probabilities.

9. How do you calculate the probability of an event?
   (a) I divide the number of outcomes that meet some particular criteria by the number of possible outcomes.
   (b) I multiply the number of outcomes that meet some particular criteria by the number of possible outcomes.
   (c) I divide the number of possible outcomes by the number of outcomes that meet some particular criteria.
   (d) I subtract the number of outcomes that do not meet some particular criteria from the number of possible outcomes.

10. Imagine that you know the probability (p1) of randomly selecting a red, a yellow, and a green marble in succession from a bowl with many marbles with a variety of colors. Now you are asked to determine the probability (p2) of selecting a red, a yellow, and a green marble, however, with the order of selection being not important this time. How are the two probabilities related to each other?

(a) The probability p1 is higher than p2.

(b) They are both the same.

(c) The probability p1 is lower than p2.

(d) There is no systematic relation between the two probabilities.

11. What happens to the probability of selecting a specific element from one selection to the next if each element is put back after having been selected?

(a) The probability of selecting a specific element decreases from one selection to the next.

(b) The probability of selecting a specific element increases from one selection to the next.

(c) The probability of selecting a specific element remains the same for each selection.

(d) There is no systematic relation between the two probabilities.

APPENDIX B

PROBLEM SOLVING TEST ITEMS

1. Angler problem

A club of vegetarian anglers has 4 members. All vegetarian anglers have committed themselves to throw the fish they catch directly back into the lake. One day the club members, one after another, go fishing at a lake that is 8 square meters in size and has 5 fish in it: one pike-perch, one salmon, one trout, one pike, and one carp. In the order of their age all club members catch one fish and then immediately throw their fish back into the lake.
What is the probability of the oldest angler catching the salmon and the second oldest catching the trout?

2. Dog problem

An animal home currently hosts 11 dogs. Four of them are terriers, the remaining are half-breeds. Two blond and 4 brunette children come to the animal home wanting dogs as pets. Because the brunette children arrived first, they are allowed to choose dogs before the blond children. Each brunette child chooses a dog randomly.
What is the probability of every brunette child getting a terrier?

3. Knight problem

Ten knights participate in the "9th King's Tournament". The king provides the tournament with 12 horses. The knights have to pick their horses blindfold. The heaviest knight gets to pick first, then the second heaviest, and so on.
What is the probability of the heaviest knight getting the biggest horse, the second heaviest knight getting the second biggest horse, and the third heaviest knight getting the third biggest horse?

4. Picnic problem

John has a spontaneous picnic with his friends. As the shops are already closed he decides to buy sweets and drinks from two vending machines. The first vending machine contains 15 different sweets whereas the second machine offers 10 different drinks. John randomly chooses the items from each machine, 4 from the candy machine and 3 from the drink machine.
What is the probability of John getting one Coke, one Seven-Up, and one Sprite from the drink machine (in that order), and one Twix, one Snickers, one Milky Way, and one Kit Kat from the candy machine (in that order)?

5. Soccer problem

At a soccer game there are two dressing rooms for the two teams. The 11 players from Oxford wear T-shirts with odd numbers from 1 to 21 and the 11 players from Manchester have even numbers from 2 to 22. Because the aisle from the dressing rooms is very narrow only one player at a time can enter the field. The players of the two teams leave their rooms alternately with a player from Oxford going first.
What is the probability of the first five players entering the field having the numbers five, two, thirteen, eight, and one (i.e., the first has the number five, the second has the number two, and so on)?

6. Tennis problem

A tennis club has 20 members, 9 women and 11 men, all of them with different last names. For a friendly game against another club a team has to be organized that consists of 2 women and 3 men. The tennis players are chosen by chance (2 from the female members and 3 from the male members).
What is the probability of building up a team that consists of Mrs. Miller, Mrs. Jackson, Mr. Byrne, Mr. Thomson and Mr. Myles?

7. Car race problem

In a car race 12 different European countries participate with one driver per country. There are 5 prizes for the participants: The winner receives $10,000, the second place finisher gets $5,000 and the third place finisher receives $1,000. The drivers of the cars who finish fourth and fifth will each win $500.
What is the probability that the Italian driver wins $10,000, the German $5,000, the Swedish $1,000, and that the French and Danish drivers each win $500?

8. Cocktail problem

In New York's famous cocktail bar "The Shark" there is a special Happy-Hour offer on Wednesday evenings. The bartenders have 16 bottles of different liquors and soft drinks available for mixing cocktails. In the bar's "Happy Hour" the bartenders randomly choose 5 of those liquids and empty the bottles into a large bucket in order to mix their special Shark-Bite cocktail. Additionally, every guest receives his personal White-Shark cocktail that is made up of 3 lacings of the remaining liquids according to the following procedure: After having selected the first liquid randomly, the bottle is put back into the shelf and a second liquid is randomly chosen. A third liquid is chosen in the same way.
What is the probability of you getting a Shark-Bite cocktail with gin, vodka, tequila, orange juice, and lime juice - and a subsequent White-Shark cocktail that is made up of first pouring a lacing of pineapple juice into the glass, then some white rum, and finally pineapple juice again?

9. Clairvoyant problem

A clairvoyant at a carnival claims to be able to recognize 8-digit numbers that are covertly written on a sheet of paper. He instructs you to first write down your 3 favorite digits in a sequence and then to use some of the remaining 7 digits in order to additionally construct your "magic number" which consists of five digits. In the "magic number" the same digit may occur more than once. You try to test the clairvoyant and write down 8 digits. First you write down your favorite digit (seven), then your second favorite digit (five) and then your third favorite digit (three). Then you use some of the remaining 7 digits to write down your five-digit "magic number" (66099).
What is the probability of the clairvoyant correctly guessing your eight-digit number by chance?

10. Lighthouse problem

A lighthouse can create flares using 4 different colors (red, yellow, green, blue). Colors are randomly chosen to form a flare. Each flare is created by mixing 2 colors. Because the colors that constitute a flare are randomly determined it may happen that the same color is used twice in one flare (e.g. "mixing" yellow with yellow).
What is the probability of the lighthouse creating a flare by mixing red and yellow (regardless of order)?

11. Car collision problem

There are 8 learners who practice car driving at a parking lot next to their driving school. Each learner is driving a car with a unique color.
What is the probability that the red car and the yellow car randomly collide?

APPENDIX C

FORMULA SHEET: MOLAR

$$A = \frac{n!}{(n-k)!}$$

$$A = n^k$$

$$A = \frac{n!}{(n-k)!k!}$$

$$A = \frac{(n+k-1)!}{(n-1)!k!}$$

$$p(O) = \frac{1}{A}$$

$$p(O) = a_1 * a_2$$

FORMULA SHEET: MODULAR

$$p(O) = a_1 * a_2$$

APPENDIX D

NASA-TLX

NASA Task Load Index

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

PHYSICAL DEMAND	MENTAL DEMAND
TEMPORAL DEMAND	MENTAL DEMAND
PERFORMANCE	MENTAL DEMAND
FRUSTRATION	MENTAL DEMAND
EFFORT	MENTAL DEMAND
TEMPORAL DEMAND	PHYSICAL DEMAND
PERFORMANCE	PHYSICAL DEMAND
FRUSTRATION	PHYSICAL DEMAND
EFFORT	PHYSICAL DEMAND
TEMPORAL DEMAND	PERFORMANCE
TEMPORAL DEMAND	FRUSTRATION
TEMPORAL DEMAND	EFFORT
PERFORMANCE	FRUSTRATION
PERFORMANCE	EFFORT
EFFORT	FRUSTRATION

APPENDIX E

COVER STORIES FOR PRACTICE PROBLEMS

1. Multiplying simple event probability

You are in the audience of a game show. In order to compete in the game show you must first be selected from the audience to play the first round. The probability that you are selected to play the first round is 1/100. However, once in the first round, the probability that you will advance is 1/5.
What is the probability that you will be picked from the audience and advance past the first round?

2. Multiplying simple event probability

At Harrison college, the probability that a student will be an art student is 1/80. For any given student at Harrison, there is a 1/3 probability that he or she lives on campus.
What is the probability that a randomly selected student will be an art student who lives on campus?

3. Permutation without replacement

Joni is putting her groceries on the conveyer belt in the supermarket. She has a potato, an onion, a tomato, a squash, and broccoli in her basket.
What is the probability that she will put the tomato on the conveyer belt first and the squash on the belt second (assuming that she randomly picks produce from the basket)?

4. Permutation without replacement

At the Olympics 7 sprinters participate in the 100m-sprint.
What is the probability of correctly guessing the winner of the gold, the silver, and the bronze medals?

5. Permutation with replacement

Les has an 8-sided die; its sides are numbered from 1 to 8.
What is the probability that he will get an 8 on his first role and a 2 on his second role?

6. Permutation with replacement

A bank distributes a random four-digit secret code as a personal identification number (PIN) for its credit cards. Suppose one credit card has been lost.
What is the probability that anybody finding the card and trying to get money with it will guess the correct secret code on the first try?

7. Combination without replacement

An urn contains five marbles, each a different color - white, yellow, red, blue, and green. Two marbles are taken out, one by one, and are not put back.
What is the probability of taking out the red marble one of the times and the blue marble the other time?

8. Combination without replacement

You are playing cards with your friends. The card game contains 52 cards and each person gets 4 cards. First, you receive all your cards and then your friends receive their cards.
What is the probability that you get all four aces?

9. Combination with replacement

Andre has a cell phone with each of the buttons from 1 to 9 programmed to speed dial a different friend. Unfortunately the buttons are on the face of the phone and when the phone is in his pocket he often accidentally calls his friends. Last night his phone placed 2 calls unintentionally.
What is the probability that he called friend #3 and friend #4?

10. Combination with replacement

A car rental service owns 10 cars, each of which has a unique color. Within 14 days a person rents a car twice and chooses the car randomly.
What is the probability of getting a red car one of the times and a blue one the other time?

APPENDIX F

NASA-TLX INSTRUCTIONS[9]

Scale Rating Evaluation

We are not only interested in assessing your performance but also the experiences you had during the different task conditions. Right now we are going to describe the technique that will be used to examine your experiences. In the most general sense we are examining the “Workload" you experienced. Workload is a difficult concept to define precisely, but a simple one to understand generally. The factors that influence your experience of workload may come from the task itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you felt, The workload contributed by different task elements may change as you get more familiar with a task, perform easier or harder versions of it, or move from one task to another. Physical components of workload are relatively easy to conceptualize and evaluate. However, the mental components of workload may be more difficult to measure.

Because workload is something that is experienced individually by each person, there are no effective "rulers" that can be used to estimate the workload of different activities. One way to find out about workload is to ask people to describe the feelings they experienced. Because workload may be caused by many different factors, we would like you to evaluate several of them individually rather than lumping them into a single global evaluation of overall workload. This set of six rating scales was developed for you to use in evaluating your experiences during different tasks. Please read the descriptions of the scales carefully. If you have a question about any of the scales in the table, please ask me about it. It is extremely important that they be clear to you. You may keep the descriptions with you for reference during the experiment.

---

[9] Instructions adapted from: Human Performance Research Group (n.d.) NASA TASK LOAD INDEX (TLX) v. 1.0: Computerized Version [Computer software manual].

After performing each task, you will be presented with six rating scales. You will evaluate the task by marking each scale at the point which matches your experience. Each line has two endpoint descriptors that describe the scale. Note that "own performance" goes from “good" on the left to “bad" on the right. This order has been confusing for some people. Please consider your responses carefully and consider each scale individually. Your ratings will play an important role in the evaluation being conducted, thus, your active participation is essential to the success of this experiment, and is greatly appreciated.

Sources of Workload Evaluation

Throughout this experiment the rating scales are used to assess your experiences in the different task conditions. Scales of this sort are extremely useful, but their utility suffers from the tendency people have to interpret them in individual ways. For example, some people feel that mental or temporal demands are the essential aspects of workload regardless of the effort they expended or the performance they achieved. Others feel that if they performed well the workload must have been low, and vice versa. Yet others feel that effort or feelings of frustration are the most important factors in workload; and so on. The results of previous studies have found every conceivable pattern of values. In addition, the factors that create levels of workload differ depending on the task. For example, some tasks might be difficult because they must be completed very quickly. Others may seem easy or hard because of the intensity of mental or physical effort required. Yet others feel difficult because they cannot be performed well, no matter how much effort is expended.

The evaluation you are about to perform is a technique that has been developed by NASA to assess the relative importance of six factors in determining how much workload you experienced. The procedure is simple: You will be presented with a series of pairs of rating scale titles (for example, Effort vs. Mental Demands) and asked to choose which of the items was more important to <u>your</u> experience of workload in the task that you just performed. Each pair of scale titles will appear separately on the screen. <u>Select the Scale</u>

Title that represents the more important contributor to workload for the specific task(s) in this experiment.

After you have finished the entire series we will be able to use the pattern of your choices to create a weighted combination of the ratings from that task into a summary workload score. Please consider your choices carefully and make them consistent with how you used the rating scales during the particular task you were asked to evaluate. Don't think that there is any *correct* pattern; we are only interested in your opinions. If you have any questions, please ask them now. Thank you for your participation.

REFERENCES

Atkinson, R. K., Catrambone, R., & Merrill, M. M. (2003) Aiding transfer in statistics: Examining the use of conceptually oriented equations and elaborations during subgoal learning. *Journal of Educational Psychology 95*, 762-773.

Atkinson, R. K., Renkl A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*, 774-783.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Neuroscience, 4,* 829-839.

Baron R. M. & Kenny D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Catrambone R. (1996). Generalizing solution procedures learned from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1020-1031.

Catrambone, R. (1998). The subgoal learning model: Creating better examples to improve transfer to novel problems. *Journal of Experimental Psychology: Genera1, 27*, 355-376.

Chi, M. T. H., Glaser, R. & Rees, E. (1982). Expertise in problem solving. In: R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, vol. 1* (pp. 7-75). Hillsdale, NJ: Erlbaum.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Chi, M. T. H., de Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769-786.

Dosher, B. A., (1998). Models of visual search: Finding a face in the crowd. In D. Scarborough & S. Sternberg (Eds.), *Methods, models, & conceptual Issues: An invitation to cognitive science, vol. 4* (pp. 455-521). Cambridge, MA: MIT Press.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19-23.

Fuchs, L. S., Fuchs, D., Prentice, K., Hamlett, C. L., Finelli, R., & Courey, S. J. (2004). Enhancing mathematical problem solving among third-grade students with schema-based instruction. *Journal of Educational Psychology, 96*, 635-647.

Gerjets, P., Scheiter, K., & Catrambone, R. (2003) Reducing cognitive load and fostering cognitive skill acquisition: Benefits of category-avoiding examples. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science, 32*, 33-58.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction in analogical transfer. *Cognitive Psychology, 15*, 1-38.

Frederiksen, J. R., & White, B. Y. (1989). An approach to training based upon principled task decomposition. *Acta Psychologica, 71*, 89-146.

Hambrick, D. Z., & Engle, R. W. (2003). The role of working memory in problem solving. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 176-206). London: Cambridge Press.

Hart, S. G., & Staveland, L. E. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

James, L. R. & Brett J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology, 69*, 307-321.

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

Klauer, K. C. & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General, 133*, 355-381.

Magill, R. A., & Hall, K. G. (1991). A review of the contextual interference effect in motor skill acquisition. In R. B. Wilberg (Ed.), *The learning, memory, and perception of perceptual-motor skills* (pp. 35-83). Amsterdam: Elsevier.

Markman A. B. & Gentner D. (2001). Thinking. *Annual Review of Psychology, 52*, 223-247.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312-320.

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

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

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

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

Shea, C. H., Kohl, R., & Indermill, C. (1990). Contextual interference: Contributions of practice. *Acta Psychologica, 73*, 145-157.

Sweller, J. (1999). *Instructional design in technical areas*. Melbourne, Australia: ACER Press.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-295.

Tsang, P., & Wilson, G. F. (1997). Mental workload. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 417-449). New York: Wiley-Interscience.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37, 498-505.

van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. G. W. C.(2002). Redirecting learners' attention during training: effects on cognitive load, transfer test performance and training efficiency. *Learning & Instruction, 12*, 11-37.

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

Xie, B., & Salvendy, G. (2000). Prediction of mental workload in single and multiple task environments. *International Journal of Cognitive Ergonomics, 4*, 213–242.