Learner Control, Expertise, and Self-Regulation:

Implications for Web-Based Statistics Tutorials

A final project submitted to the Faculty of Claremont Graduate University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Psychology

by

Amanda T. Saw

Claremont Graduate University

2011

APPROVAL OF THE REVIEW COMMITTEE

This dissertation has been duly read, reviewed, and critiqued by the Committee listed below, which hereby approves the manuscript of Amanda T. Saw as fulfilling the scope and quality requirements for meriting the degree of Doctor of Philosophy in Psychology.

Dale E. Berger, Chair
Claremont Graduate University
Professor

Diane Halpern
Claremont McKenna College
Professor

Kathy Pezdek
Claremont Graduate University
Professor

Richard Mayer
Visiting Examiner
University of California, Santa Barbara
Professor

Abstract

Learner Control, Expertise, and Self-Regulation:

Implications for Web-Based Statistics Tutorials

by

Amanda T. Saw

Claremont Graduate University: 2011

Many statistics students have only a rudimentary understanding of distributions and variability, fundamental concepts in statistical inference. Computer-based instruction can improve understanding, but its effectiveness may vary due to interactions between instructional features and learner characteristics such as domain expertise and self-regulated learning ability (planning, monitoring, and evaluating one's own learning). Especially in computer-based instruction, learning can depend upon the learner's control over instructional processes.

The current study manipulated levels of learner control over exposure to feedback and supplementary questions in a Web-based tutorial on standard deviation. The study examined how learner control and learning are related to domain expertise, self-regulation of learning, self-efficacy (belief that one will succeed on the tutorial), and task value (importance of learning about standard deviation).

Although the tutorial significantly improved understanding of standard deviation for all learners, $t(200) = 6.75$, $p < .001$, $d = .42$, experts (who had

completed one or more statistics courses) benefited more from learner control than did novices (who had not completed their first statistics course). In contrast, novices benefited from greater control exercised by the program and suffered from greater learner control, as reflected by impaired learning and increases in reported frustration and difficulty with the tutorial. Experts, who experienced less cognitive load overall, learned equally well with either level of control.

However, the prediction that program control would be more beneficial for low self-regulating learners than high self-regulating learners was not supported. Self-regulation of learning, self-efficacy, and task value (all self-reported) were positively and significantly associated with learning; however, when expertise was statistically controlled, these predictors were no longer significant. Perceived cognitive load was negatively associated with learning.

Supporting Cognitive Load Theory, these results have implications for the design of computer-based instruction. Learner expertise must be considered so that cognitive load can be managed via instructional control that enables learners to focus on essential material and make connections with prior knowledge. A high level of learner control that allows experienced learners to exercise efficient learning, may be detrimental to novices, who possess limited domain expertise and may not effectively self-regulate their learning.

Dedication

To the ones who made me (me),

including Deb, Dad, and D. E. B.

Acknowledgements

This dissertation took the shape (and range) it did mainly due to my advisor, mentor, and statistics guru, Dale E. Berger. I also am indebted to the rest of my dissertation committee for their incredible support and helpful suggestions on dissertation drafts: Diane Halpern, Kathy Pezdek, and Richard Mayer. The rest of the WISE team, Justin Mary and Giovanni Sosa, also deserve my thanks for keeping me grounded. Thanks also to my pilot testers: En-Ling Chiao, Thipnapa Huansuriya, Vera Mauk, and Johan Sandqvist.

I would like to thank the instructors who assisted in me recruiting participants and their students: Christopher Aberson, Dale Berger, Bettina Casad, Andrew Lac, Kevin Matlock, Richard Mayer, and Michelle Samuel. Without their students completing the Standard Deviation tutorial, this dissertation would not have been possible.

Thanks are also due to friends and family who supported and believed in me, in particular my parents, Sophia, Sonny, Karen, and Helen, the other statistics guru in my life.

I was very much inspired by the cutting-edge statistics education research conducted by individuals such as Robert delMas, Beth Chance, and Joan Garfield. I hope to build upon their work in illuminating the statistical understanding—and misconceptions—demonstrated by statistics students.

This dissertation was funded by the Claremont Graduate University Dissertation Grant.

Table of Contents

Chapter 2: Method

Chapter 3: Results

Chapter 4: Discussion

List of Tables

List of Figures

List of Appendices

Chapter 1: Introduction

A large body of statistical education research shows that students often have fundamental misconceptions about inferential statistical concepts, even after completing relevant statistical exercises and activities. These conceptual weaknesses include difficulty specifying null and alternative hypotheses (Aquilonius, 2005); poor understanding of the fundamental concepts of randomness and sampling (Kahneman, Slovic, & Tversky, 1982); failure to consider the role of sample size in interpreting statistical findings, such as how sample size affects the sampling distribution (Well, Pollatsek, & Boyce, 1990); failure to differentiate between the population, sample, and sampling distributions (Saldanha & Thompson, 2003); and erroneous reasoning and interpretations regarding *p*-values (Lane-Getaz, 2007; Saw, Berger, Mary, & Sosa, 2009; Sosa, Berger, Saw, & Mary, 2009). These related conceptual weaknesses may be due to an incomplete understanding of distributions and variability that form the basis for comprehending inferential statistics and more advanced statistical topics. Incomplete knowledge of fundamental concepts may prevent effective learning of more advanced topics.

Learning can be conceptualized as a change in long-term memory due to the integration of new schemata with existing knowledge (e.g., Kirschner, Sweller, & Clark, 2006). Consistent with this cognitive perspective, a constructivist framework of learning emphasizes the role of prior knowledge in acquiring new information and the role of the learner as an active participant seeking knowledge (Bransford, Brown, & Cocking, 1999). A constructivist approach, in which the learner actively builds new knowledge upon prior knowledge, has been advocated by many researchers specifically in the teaching of statistics (e.g., Franklin & Garfield, 2006; Garfield & Ahlgren, 1988; Lovett & Greenhouse, 2000; Mills, 2002; Romero, Berger, Healy, & Aberson, 2000). Using such a framework, active learners are expected to interact with the learning environment and select tasks or activities that may develop their understanding of concepts, emphasizing the need to self-regulate their learning processes and to manage cognitive load (dedicating cognitive resources to certain subtasks), especially when using computer-based instruction (Kostons, Van Gog, & Paas, 2009; Kostons, Van Gog, & Paas, 2010; Scheiter & Gerjets, 2007).

In computer-based instruction, the benefits of giving the learner control over which tasks to do, including pacing and sequencing, may be moderated by the learner characteristics, including their prior knowledge and ability to self-regulate their learning (Gerjets et al., 2009; Lunts, 2002; Vovides, Sanchez-Alonso, Mitropoulou, & Nickmans, 2007). Having too much control over learning processes may impair learning for learners who cannot effectively self-regulate their learning or who possess a limited amount of prior knowledge or expertise in a particular domain. Motivation, such as self-efficacy (i.e., the belief that one will succeed on a certain task), is another important factor in learning, a concept intricately related to self-regulation of learning (Artino, 2008; Zimmerman, 2000).

In addition to prior knowledge, learning abilities, and motivation, general human cognitive architecture must also be considered when designing instruction (Kirschner et al., 2006). One such aspect is the limited amount of cognitive resources an individual has available to process and integrate information in working memory. When cognitive resources are overburdened, the learner experiences cognitive overload, and learning is disrupted (Moreno & Mayer, 2007).

Active or discovery learning, which is central to some constructivist paradigms, may cause learners to be overwhelmed by information, leaving them unable to integrate information efficiently (van Merriënboer & Sweller, 2005) or to select subsequent learning activities that enhance their understanding (Kostons et al., 2009). Mayer (2004) reviewed three decades of discovery learning research from the 1960s to the 1980s in the domains of problem solving, Piagetian conservation strategies, and computer programming. He concluded that while there is merit in constructivist approaches, pure discovery learning, in particular, has proved to be less effective than guided discovery.

Guided instruction can make use of advance organizers, which are defined as "appropriately relevant and inclusive introductory materials" that are "presented at a higher level of abstraction, generality, and inclusiveness" than the target passage (Ausubel, 1968, p. 148). An advance organizer serves "to provide ideational scaffolding for the stable incorporation and retention of the more detailed and differentiated material that follows" (Ausubel, 1968, p. 148). An overview of topics to be learned, or a set of learning goals, could also serve as an advance organizer designed to enhance learning. Beneficial effects of advance organizers were first reported by Ausubel (1960) in a study of students' understanding of an unfamiliar topic, metallurgy. Students who were presented with an advance organizer performed significantly better on post-tests and retention tests than those presented with a historical background.

Initial reviews of advance organizers concluded that they are ineffective (Barnes & Clawson, 1975) or questioned their theoretical and practical usefulness (Hartley & Davies, 1976). However, subsequent meta-analyses and reviews (Ausubel, 1978; Luiten, Ames, & Ackerson, 1980; Mayer, 1979a; Mayer, 1979b; Stone, 1983) have reported overall positive learning effects of advance organizers. Several of these subsequent reviews also challenged the logic of the initial negative reviews that discounted the utility of advance organizers without considering other factors. For instance, Barnes and Clawson (1975) failed to separate conditions where advance organizers are effective from those where they are ineffective (Mayer, 1979a). The effectiveness of advance organizers may depend upon the ability and knowledge of learners. Luiten et al. (1980) concluded that advance organizers are particularly effective with high-ability learners, whereas Mayer (1979b) suggested that advance organizers were more helpful to low-knowledge learners. This debate concerning the efficacy of advance organizers parallels the debate concerning other instructional features such as how much control a learner should have over their learning.

In fact, several lines of research, including statistics education research, have demonstrated that while scaffolding (i.e., instructional supports) beyond advance organizers and guided instruction in general may enhance learning, the effectiveness of scaffolds may be affected by various learner characteristics. These characteristics include the learner's prior knowledge (Lipson, Kokonis, & Francis 2003) and how well individuals self-regulate their own learning process (Moos & Azevedo, 2008b). Especially in a hypermedia instructional setting, the learner's ability to self-regulate their learning impacts how well they process and retain the material (Azevedo & Hadwin, 2005; Chen, Fan, & Macredie, 2006). Winters, Greene, and Costich (2008) identified 33 computer-based learning environment studies that explicitly cited self-regulated learning as a key construct. However, nearly a third of these studies did not report any objective measure of learning. In addition, self-efficacy, which is the belief that one will successfully perform a task, can also impact the learning process (Zimmerman, 2000). It is crucial to relate learning outcomes to both cognitive and motivational processes (Pintrich, 1999). The current study examined the effects of prior knowledge, self-efficacy, self-regulation of learning strategies, cognitive load, and learner control on a Web-based tutorial about variability (in the context of standard deviation).

1.1 *Effectiveness of Technology in Teaching Statistical Variability*

Although several meta-analyses have demonstrated that instructional technology contributes to better learning in general educational domains (Kulik & Kulik, 1991; Kulik & Kulik, 1986; Kulik, Kulik, & Cohen, 1980) and in statistics education (Hsu, 2003; Schenker, 2007; Sosa, Berger, Saw, & Mary, 2011), the evidence is mixed regarding the effects of technology on comprehension and mastery of the specific topics of distributions and statistical variability. In fact, some applications of technology can introduce new misconceptions. For instance, when using software that simulated the statistical technique of repeated sampling, some students erroneously concluded that multiple samples are needed for conducting inferential statistics (Hodgson & Burke, 2000).

Other studies have revealed that students have many misconceptions of basic statistical concepts even after using educational technological resources designed to correct them. For example, deficiencies in understanding concepts such as the Central Limit Theorem and the sampling distribution of the mean persisted even after using a computer program that simulated sampling to illustrate the effect of sample size on sampling variability (Well, Pollatsek, & Boyce, 1990). After using another similar program, students still found it difficult to differentiate between the sample, population, and sampling distributions (Saldanha & Thompson, 2003). However, students' understanding of the sampling distribution after attending a traditional lecture on the topic was found by Aberson et al. (2000) to be no better than after using a Web-based tutorial.

Lipson et al. (2003) demonstrated the importance of highlighting key features in a computer simulation to facilitate learning. They tracked the development of eight students' statistical reasoning as the students completed a dynamic simulation software program to explore sampling distributions. In the simulation activity, the students assessed the veracity of a postal carrier's claim that at least 96% of letters were delivered on time, which conflicted with a journalist's finding that 88% of letters were delivered on time in his sample. Only after repeated use of the simulation program, did the students gradually recognize different aspects of the simulation display and distinguish between samples and sampling distributions of means. Initially they favored practical or motivational explanations (e.g., the journalist did something incorrectly), rather than statistical explanations for simulation outcomes, demonstrating the role of prior knowledge. Only when probed by the interviewer, did students offer statistical explanations.

Further highlighting the importance of prior knowledge in learning statistics, Chance, delMas, and Garfield (2004) identified four concepts that are prerequisites to understanding sampling distributions, based upon conceptual analyses of classroom observations, colleagues' contributions, and performance on items assessing statistical comprehension. These concepts are variability, distribution, normal distribution, and sampling (see Table 1.1). An implication is that learners of the sampling distributions should understand how observations vary and be able to describe and compare distributions, interpret graphs, and distinguish between samples and population. Chance et al. noted that as she and her colleagues continued to conduct statistics education research, they found that they needed to explore students' understanding of even more basic concepts (e.g., distributions and variability) than those being empirically examined (e.g., sampling distributions).

Because students find it difficult to differentiate between population and sampling distributions (Saldanha & Thompson, 2003), a sampling simulation program such as by Lane and Tang (2000) may be beneficial in helping students untangle these concepts. The Lane and Tang program (found online at: http://onlinestatbook.com/stat_sim/) graphically displays three separate histograms showing the population distribution, individual scores from a

Table 1.1

Prerequisite Knowledge to Learning about Sampling Distributions (from Chance, delMas, & Garfield, 2004, p. 300).

Concepts	Description
Variability	What is a variable? What does it mean to say observations vary? Students need an understanding of the spread of a distribution in contrast to common misconceptions of smoothness or variety.
Distributions	Students should be able to read and interpret graphical displays of quantitative data and describe the overall pattern of variation. This includes being able to describe distributions of data; characterizing their shape, center, and spread; and being able to compare different distributions on these characteristics. Students should be able to see between the data and describe the overall shape of the distribution, and be familiar with common shapes of distributions, such as normal, skewed, uniform, and bimodal.
Normal distribution	This includes properties of the normal distribution and how a normal distribution may look different due to changes in variability and center. Students should also be familiar with the idea of area under a density curve and how the area represents the likelihood of outcomes
Sampling	This includes random samples and how they are representative of the population. Students should be comfortable distinguishing between a sample statistic and a population parameter. Students should have begun considering or be able to consider how sample statistics vary from sample to sample but follow a predictable pattern.

sample, and the resulting distribution of sample means from repeated sampling. Lane and Tang tested the instructional effectiveness of the simulation program in a 30-minute demonstration led by an experimenter, which contrasted sampling distributions of the mean obtained from sampling two different sample sizes. Compared to students who read a text description of this sampling process, students who viewed the simulation did significantly better on problem solving items regarding sampling. Prompting students beforehand with "specific" questions about the sampling simulation outcomes—rather than general questions — demonstrated a trend for improved learning (although this difference was not statistically significant, p = .061), which suggests the utility of guided instruction and advance organizers. Although motivation was not explicitly measured, Lane and Tang observed that the students viewing the simulation appeared more engaged during the training.

Learning about standard deviations encompasses examining both variability and distributions. DelMas and Liu (2003; 2005; 2007) examined students' conceptual understanding of standard deviation using an interactive game-like computer program, in which students manipulated bars of observations (i.e., observations of the same value) in a histogram to understand how these changes impact standard deviation. In five games progressing from histograms with two bars to five bars of equal or unequal frequency, the students individually had to manipulate the configuration of bars to produce two different configurations of the largest standard deviation possible and three different configurations of the smallest standard deviation possible (see Figure 1.1 for possible solutions to a four-bar histogram, illustrating largest and smallest standard deviations). For each game, the student therefore produced five different configurations (for a total of 25 configurations), and verbally justified each of their answers to an interviewer. The program illustrated a mean-centered conception of standard deviation by highlighting the sample mean and how much each observation deviated from the mean. It also demonstrated how the shape of the distribution (e.g., bell-shaped vs. U-shaped) and its range impacted standard deviation.

By the end of this one-hour training, all 12 students seemed to understand that a mirror image of the configuration of bars produced the same standard deviation and that the relative position of bars to the sample mean, not the

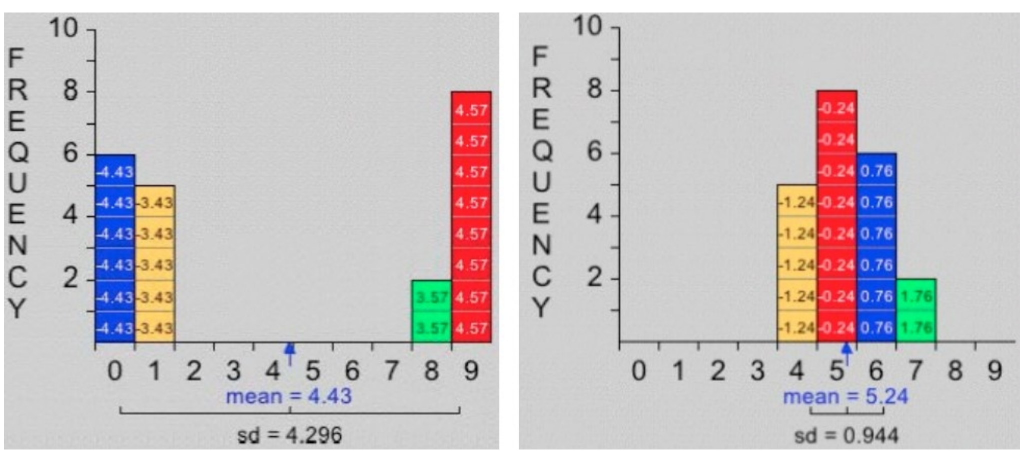


Figure 1.1. Item illustrating possible solutions for largest and smallest standard deviation for a four-bar histogram (from delMas & Liu, 2003).

absolute position on the histogram scale, determined the standard deviation (delMas & Liu, 2007). However, the justifications the students provided for why the standard deviation was larger or smaller were not always completely comprehensive (e.g., the standard deviation is smaller when the bars are contiguous or when the sample mean is in the middle of configuration of bars) or plainly wrong (e.g., a bigger sample means a larger standard deviation). Sometimes students neglected to mention the role of the mean in defining standard deviation, and relied on explanations such as bars are “spread out” or “equally spread out” to justify higher variability. On the 10-item post-test (on which the student identified which histogram of two had the greater standard deviation, see Figure 1.2), nine students got nine items right and three students got seven items right, for an average of 8.5 out of 10 correct. This exploratory study was a post-test-only design, thus precluding a pre-test comparison.

Performance on the post-test items reveals that the students did not always integrate information about shape and spread to make judgments about standard deviation (delMas & Liu, 2005). Test items 5, 7, and 9 (see Figure 1.2) tested students’ knowledge of how gaps in the distribution affected standard deviation. While all 12 students answered items 7 and 9 correctly, two students overlooked the gaps in item 5 and responded that the distributions had equal standard deviations, indicating an over-reliance on the shape of the distribution rather than spread. Test items 8 and 10 challenged the notion that symmetric, bell-shaped

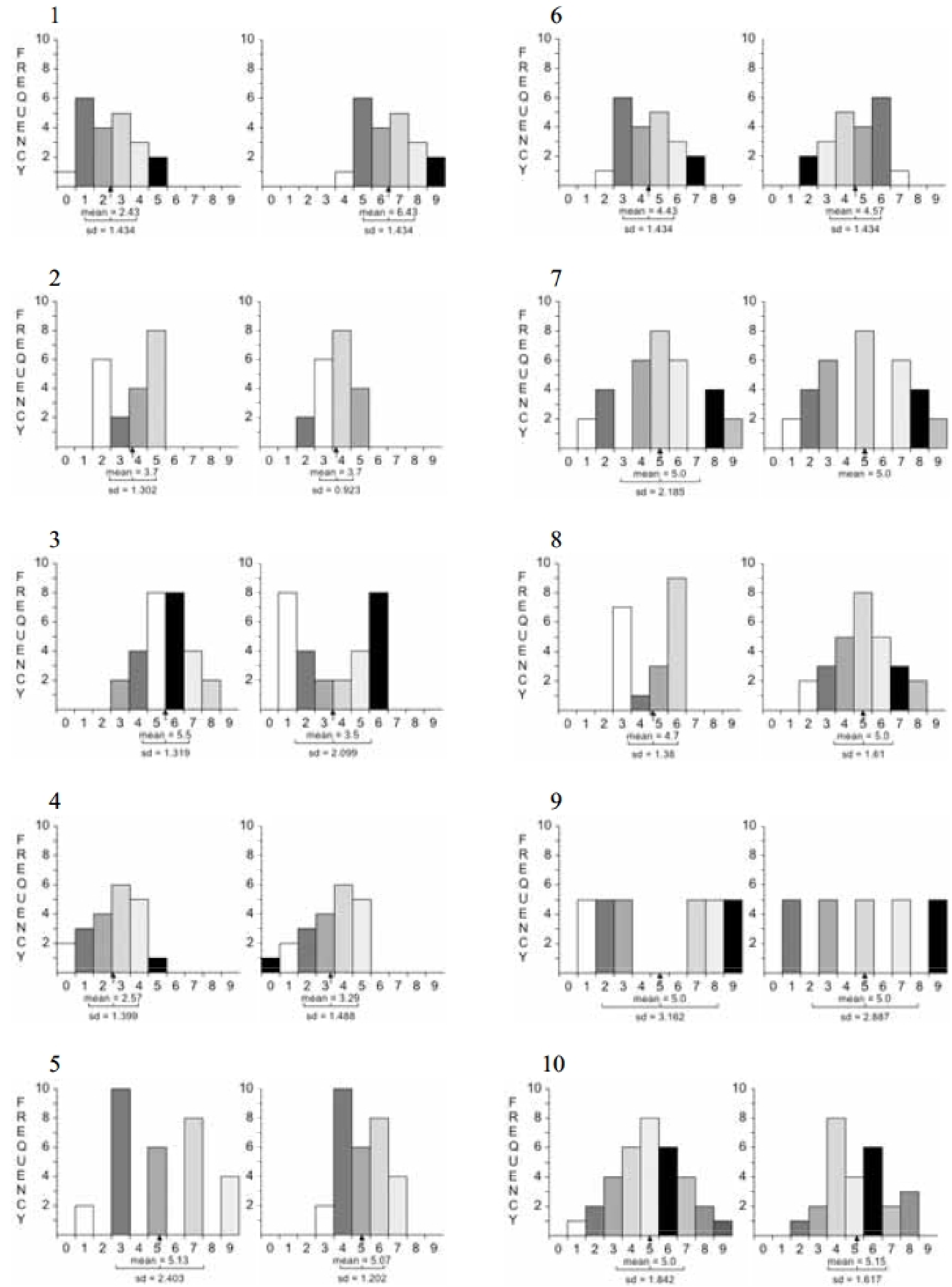


Figure 1.2. Test items assessing understanding of standard deviation (from delMas & Liu, 2005). Sample means were shown while standard deviations were not shown.

distributions have smaller standard deviations than non-bell-shaped distributions. Test item 8 was the most difficult item; only one student answered it correctly. Nine students correctly answered item 10, primarily through calculations. Students received verbal feedback from an interviewer on each of their responses as they completed the post-test. The one student who correctly answered item 8 and thus did not receive guidance on looking beyond the shape of the distribution, as did the other nine students, was the only one who answered item 10 incorrectly. This highlights both the importance of feedback to correct possible misconceptions as well as the possibility that a student may not have valid conceptual knowledge despite doing well on an assessment item.

1.2 *Scaffolding and Learner Control*

The studies on statistics education reviewed here suggest the need for guided and structured activities for effective learning through the use of "scaffolding." Scaffolding is support given to learning in the initial phases by a more knowledgeable other who operates within the learner's "zone of proximal development" to build upon knowledge (Vygotsky, 1978), but once mastery is achieved, this support is "faded" out (Lajoie, 2005). In this way, scaffolding can be thought of as bridging prior knowledge and new knowledge. In computer-based instruction, scaffolding comes in a variety of forms, including corrective feedback and prompts to learners, when appropriate, to produce explanations to facilitate their understanding of a concept. Scaffolding can also provide structure and emphasis to relevant information in a complex learning situation and thus reduce cognitive load by focusing the learner's cognitive resources on the most relevant aspects of a task (Kirschner et al., 2006). Prior knowledge becomes even more important when the learning environment is self-regulated, such as in a Web-based tutorial (Lajoie, 2005; Shapiro & Niederhauser, 2004).

Effective use of scaffolds is dependent upon accurate and frequent assessment of the learner's understanding as the learning process progresses. Assessing new knowledge on an ongoing basis is known as dynamic assessment. Dynamic assessment provides information needed for the instructional program to give appropriate feedback, explanations, and prompts, as well as structure the sequencing of learning activities. Lajoie (2005) described this process as follows:

> Dynamic assessment implies that human or computer tutors can evaluate transitions in knowledge representations and performance while learners are in the process of solving problems, rather than after they have completed a problem. Immediate feedback in the form of scaffolding can then be provided to learners during problem solving, when and where they need assistance. The purpose of assessment in these situations is to improve learning in the context of problem solving, while the task is carried out. (p. 545)

Using such an approach, a computer-based tutorial would provide guided instruction and feedback tailored to a learner's knowledge and misconceptions.

However, the usefulness of dynamic assessment depends upon the learner's ability to use the information presented by the feedback, which can be moderated by learner's characteristics including their prior knowledge, accurate assessment of what constitutes good performance, and the ability to process self-assessment information in addition to the content to be learned (Kostons et al., 2010). The ability to use feedback to select appropriate subsequent learning tasks to enhance learning is one aspect of the learner's ability to self-regulate their learning processes (Kostons et al., 2009).

One way to support computer-based learning is to limit how much control the learner has over learning processes in favor of program (or computer) control. For instance, learners using program-control instruction may be required to complete integrative, review questions before proceeding. In contrast, learners using learner-control instruction could choose whether to use or to skip these tasks. Learner control has at least three dimensions: controlling the order (sequencing) of information, selecting content to access, and pacing how fast the material is presented (Lunts, 2002; Milheim & Martin, 1991; Scheiter & Gerjets, 2007). Giving the learner more control can lead to more positive attitudes regarding the instructional program (Burke, Etnier, & Sullivan, 1998; Hannafin & Sullivan, 1995). Yet the effectiveness of learner control also depends upon the learner's self-regulation abilities (Vovides et al., 2007). In the absence of guidance from a computer program, the learner must depend upon self-evaluation to monitor their own performance and to make decisions regarding learning activities and feedback. At the same time, computer-based instruction can enhance self-regulation of learning by providing the cognitive tools to support self-monitoring (Lajoie, 2008).

Besides distinguishing between interactivity and learner control, Scheiter and Gerjets (2007) also made a distinction between multimedia and hypermedia learning. Hypermedia learning involves the use of hyptertext that links to other informational screens, and may include multimedia presentations. Unlike multimedia learning, which tends to be system-controlled and linear, hypermedia learning is more interactive and requires more user response/input. Although both deal with how users may manipulate how content information is represented, interactivity is not as multi-dimensional as learner control; interactivity usually refers to instances of manipulating single instances of representations. In contrast, learner control, which characterizes most hypermedia environments, reflects a broader perspective on how the learner interacts with the learning environment, including how information is represented and sequenced and which activities are selected and pursued. Thus, effective hypermedia learning may require more self-regulated learning processes from the user. Scheiter and Gerjets (2007) cited several reasons why hypermedia may be effective:

1. Like the mind, hypermedia/hypertext reflects nodes and the interconnected structure of information.
2. Hypermedia promotes motivation and interest (self-efficacy).
3. Interactivity is adaptive and subject to learner control to fit learner's needs, including prior knowledge.
4. Hypermedia instruction forces learners to constantly evaluate their learning goals and processes.
5. Hypermedia instruction facilitates deeper processing of information and self-regulation of learning.

On the other hand, there are potential problems with hypermedia learning, including disorientation with where one is in the learning process (Chen et al., 2006) and cognitive overload (Gerjets et al., 2009). In its infancy, compared to non-hypertext instruction, hypertext instruction has been shown to have a medium effect in promoting learning (Chen & Rada, 1996). In addition, early hypermedia instruction has been shown to be most effective for learning that is drill-and-practice instruction, and learner control may be most beneficial for high-ability learners (Dillon & Gabbard, 1998). However, a limitation of the early studies that evaluated hypermedia learning is that they usually had small sample sizes and confounded variables in their experimental manipulations (Scheiter & Gerjets, 2007).

Learner control can be further broken down into full vs. lean versions of computer-based learning programs, as it was in Hannafin and Sullivan's (1995) study of geometry students using a computer-based mathematics program. In the full version, learners were given the complete set of instructional content that was given to the program-control group, but with the option of bypassing or "skipping" sections of instruction. In the lean version, learners could optionally choose to do these same sections, reframed as being "supplemental." Using a 2 (version: full vs. lean) x 2 (instructional control: program vs. learner) design, Hannafin and Sullivan compared these two versions of learner-control instruction to comparable versions of program-control instruction. The program-control full version contained basic information along with examples, practice problems, and review; program-control lean version contained the same basic information but no examples, practice problems, or review. The learner-control versions contained the same basic information as the program-control version, but students could either skip (full version) or supplement instruction with (lean version) the optional examples, practice problems, and review.

Students using the learner-control versions reported liking the program more than those using the program-control versions (Hannafin & Sullivan, 1995). Furthermore, students using the full versions reported liking the option to skip instructional sections more than those using the lean versions reported liking the option to do supplemental sections. More importantly, students using the learner-control versions scored significantly higher on a 30-item post-test (M = 14.97) than those in the program-control condition (M = 13.69). The interaction between instructional control and version was not significant.

1.3 *Prior Knowledge and Cognitive Load*

A novice may not know what features of a presentation to attend to when using a computer program, therefore hampering the learning process (Lipson et al., 2003). Thus, it may be especially helpful to orient users to relevant features before the main learning activity. In particular, pre-instructional activities can improve the effectiveness of computer-instruction. For instance, not so different from advance organizers, pretraining is prior instruction that introduces the components in the system that is the focus of instruction. Pretraining is based on the assumption that activation of relevant prior knowledge before instruction helps to focus cognitive resources and to integrate new knowledge (Mayer & Moreno, 2003; Moreno & Mayer, 2007). In a variant of pretraining, delMas, Garfield and Chance (1999) demonstrated that having students make predictions and then test their predictions using simulation software can benefit learning. Furthermore, this method was most effective when students were required to confront their misconceptions. Pre-instructional activities may also involve the use of advance organizers that are short text passages that help connect prior knowledge with incoming knowledge (McManus, 2000).

Aside from activating prior knowledge schemata to facilitate the integration of new information, these pre-instructional activities may also enhance learning by helping learners focus on relevant information, reducing cognitive resources allocated to less relevant information. According to Cognitive Load Theory, cognitive load can be classified into three different types: germane, intrinsic, and extraneous (Sweller, van Merriënboer, & Paas, 1998). *Germane* cognitive load is necessary to the construction of schemata and their storage into long-term memory, which is essential to learning (van Merriënboer & Sweller, 2005). *Intrinsic* load is determined by the interaction between complexity of the learning task and learner's prior knowledge. Traditionally, it is assumed that intrinsic load cannot be changed for a given learning task. In contrast to both germane and intrinsic load, *extraneous* load is not related to the learning process and actually interferes with schemata acquisition. Optimal instructional design maximizes germane load (by encouraging elaboration of information to facilitate schemata integration) while minimizing extraneous load (Gerjets, Scheiter, & Catrambone, 2006; Zumbach, 2006).

Prior knowledge in the form of expertise can influence the effectiveness of instructional scaffolding. In numerous examples of the *expertise reversal effect*, scaffolding has been shown to impair the performance of expert learners who have high prior knowledge of a domain (Kalyuga, 2007; Kalyuga et al., 2003). Cognitive Load Theory has been used to explain the expertise reversal effect (e.g., Kalyuga & Sweller, 2004; Kalyuga & Sweller, 2005; van Merriënboer & Sweller, 2005). This explanation is based upon the assumptions that short-term working memory is limited, whereas long-term memory is virtually unlimited, and that effective use of long-term memory can help overcome the processing limitations of working memory. Domain experts usually have an advantage over novices in acquiring new information because experts can more easily organize knowledge into chunks of long-term memory schemata that place less demand on working memory when integrating new information with prior knowledge. In contrast, novices lack these structures and need to exert more effort in constructing schemata, thus experiencing more cognitive load during learning. Hence novices may benefit more from scaffolding that helps build schemata, such as textual explanations in diagrams. However, such scaffolding may not help, or may even be detrimental to expert learners because such information is redundant with, or possibly organized differently from, what they already know. Processing the new scaffolding to be compatible with existing cognitive structures may actually increase cognitive load and interfere with learning. Thus, what may be beneficial to initial learning may be detrimental to later learning, just as what deters initial learning may result in better long-term learning (Schmidt & Bjork, 1992). This distinction between experts and novices highlights the importance of dynamic assessment and the need to provide differential instruction for low-knowledge and high-knowledge learners.

Further supporting the notion that scaffolds can be detrimental to learning under certain circumstances is a study in which 60 undergraduate and graduate students learned new Japanese words in a 15-minute lexicon hypertext lesson (Tripp & Roby, 1990). The students' learning was scaffolded using either an advance organizer that described the structure of the lexicon, or with a visual metaphor that indicated spatial relations, or both the advance organizer and visual metaphor, or neither. Both scaffolds by themselves provided post-test benefits over having no scaffolds at all; however, when both scaffolds were used, students did worse than when presented with only one scaffold. This suggests that having too much scaffolding material may interfere with learning by contributing to cognitive overload.

Although measuring cognitive load has proved to be challenging, such measurement is crucial to understanding and optimizing the learning process. Paas, van Merriënboer, and Adam (1994) found that self-reported, subjective measures of mental effort were adequate as indicators of cognitive load, whereas cardiovascular measures were less reliable and sensitive. Thus, they concluded that self-reported mental effort can be used as an index of cognitive load. One example of a cognitive load measure is the NASA Task Load Index (NASA-TLX), which is a self-reported multi-dimensional measure of workload (Hart & Staveland, 1988). It consists of six subscales: three subscales focus on the individual (Mental, Physical, and Temporal Demands) and the other three focus on the interaction between the individual and the task (Frustration, Effort, and Performance) (see Table 1.2). Although each of the subscales was originally designed to be weighted to compute an overall workload value, a common modification has been either to compute an overall score or to use each subscale individually (Hart, 2006).

Potential problems with the NASA-TXL scale are that each item is multidimensional, and it may be too extensive to administer in some settings, including learning tasks that are more cognitive rather than physical in nature. The scale was originally developed for aviation use and has been used mostly in studies evaluating interface and human factors design (Hart, 2006). Although it has been used in various studies, including flight simulation and other visual/motor tasks (Cao et al., 2009), it may not be optimal to be used in cognitively-oriented learning studies that involve less physical demands. More specifically, the items are not linked to cognitive load as described by Cognitive Load Theory, namely intrinsic, germane, and extraneous load.

Table 1.2

Subscales of NASA-TLX Rating Scale (Hart, 2006)

Title	Endpoints	Description
Mental Demand	Low/High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Good/Bad	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level	Low/High	How insecure, discouraged, irritated, stressed, and annoyed or secure, gratified, content, relaxed, and complacent did you feel during the task?

A few researchers have attempted to separate the cognitive load types by using different self-reported measures. For instance, Gerjets et al. (2009) examined learner control and hypermedia instruction on probability theory using five measures on a 9-point Likert scale (originally in German) to assess cognitive load. They measured intrinsic cognitive load with one item: "How easy or difficult do you consider probability theory at this moment?" Germane cognitive load, critical to integrating new knowledge with existing schemata, was also measured by one item: "Indicate on the scale the amount of effort you exerted to follow the last example." Extraneous cognitive load, which is detrimental to learning, was assessed via three items: (1) "How easy or difficult is it for you to work with the learning environment?," (2) "How easy or difficult is it for you to distinguish important and unimportant information in the learning environment?," and (3) "How easy or difficult is it for you to collect all the information that you need in the learning environment?" Experiment 1 used six instructional conditions varying in information complexity. However, none of these cognitive load measures significantly varied across instructional conditions, undermining their validity. Thus, the cognitive load measures were eliminated in Experiment 2 which compared a high learner-control program version to the six versions from Experiment 1 aggregated. Experiment 2 results indicated slightly better learning with the high learner-control version, but failed to find any interaction between learner control and prior knowledge.

In contrast, DeLeeuw and Mayer (2008) found better evidence for distinct items measuring intrinsic, extraneous, and germane load in two separate experiments using a 6-minute animated multimedia lesson on the electric motor. Extraneous load was measured by response times to a secondary task, which varied in redundancy (concurrent animation, narration, and on-screen text took longer to process than non-redundant animation and narration only). Intrinsic load, measured by effort ratings ("Please rate your level of mental effort on this part of the lesson" on a 9-point Likert scale) at eight different times during the task, was most sensitive to manipulations of sentence complexity used in the multimedia lesson. Germane load was assessed with another 9-point self-rating on the difficulty of the lesson, and was positively correlated with performance on transfer post-test items, reflecting better transfer with higher germane processing. These three measures of load were not highly correlated with one another (r's were .12 to .33), further supporting the notion that these load types are distinct. However, DeLeeuw and Mayer cautioned that participants in both experiments were of low prior knowledge and that higher prior knowledge participants may exhibit a different pattern of results; that is, show different relationships between these cognitive load measures. It is noteworthy that the "difficulty" ratings in Gerjets et al. (2009) were used to reflect intrinsic load, whereas "effort" ratings were used to measure germane load.

1.4 *Self-Regulation of Learning and Expertise*

In addition to cognitive load and prior knowledge, a learner's metacognitive and self-regulated learning abilities must also be considered when designing computer-based instruction. In fact, several studies have shown that the effect of prior knowledge on learning may be mediated by self-regulated learning strategies, such that low-knowledge learners, who tend not to self-regulate their learning effectively, may require more support or scaffolding (Azevedo, 2005; Chen et al., 2006; Moos & Azevedo, 2008b; Shin, Schallert, & Savenye, 1994; Winne, 1996). Self-regulated learning is a cyclical process in which the learner must plan, manage, and control their learning by setting goals and enacting strategies to achieve those goals (Moos & Azevedo, 2008b; Puustinen & Pulkkinen, 2001; Zimmerman, 2002). The perspective of self-regulated learning has replaced the information processing perspective in learning research, by supplementing cognitive factors with motivational, affective, and social factors (Pintrich, 2004).

Self-regulating learners can be distinguished by both their use and awareness of self-regulation strategies, that is, deliberate and systematic actions aiming to gain knowledge (Zimmerman, 1990). They are proactive and take responsibility for their own learning and continuously self-monitor their learning processes and self-evaluate their performance. Self-regulation of learning is a multi-faceted construct that involves implementing cognitive and behavioral strategies that enhance acquisition of knowledge and skills (Boekaerts, 1999). Pintrich (2004) noted that even the five cognitive strategy subscales (rehearsal, elaboration, organization, metacognition, and critical thinking) on the Motivated Strategies for Learning Questionnaire (MSLQ) do not capture all cognitive aspects of self-regulated learning. Self-regulated learning involves at least three different components: (1) cognitive learning strategies, (2) self-regulatory strategies to control cognition, and (3) resource management strategies (Pintrich, 1999). Monitoring learning processes and outcomes can be a complicated metacognitive activity (Zimmerman, 1990; Winne, 1995). Thus, self-regulating one's learning can be a cognitively intensive and demanding task.

The learner's ability to self-regulate their learning process plays a crucial role in determining what is learned, especially in non-linear hypermedia or multimedia learning (Azevedo & Hadwin, 2005; Vovides et al., 2007; Winters et al., 2008). Difficulties in learning with hypermedia instruction may be due to inappropriate or inadequate self-regulated learning strategies (Azevedo, 2005). Novice learners are especially susceptible to *disorientation* (i.e., getting lost where they are within the instruction and not knowing where to find relevant information) when using hypermedia learning tools (Amadieu, Tricot, & Mariné, 2009; Chen et al., 2006; Zumbach, 2006). In contrast, learners with high prior domain knowledge demonstrate more planning and monitoring behaviors during learning (Moos & Azevedo, 2008b). Regardless of the learner's prior knowledge, disorientation may also occur when the presentation format (e.g., linear vs. non-linear) is not aligned with narration format (e.g., non-ordered encyclopedic presentation of information vs. linear narrative story), creating extraneous cognitive load (Zumbach & Mohraz, 2008). Conversely, conceptual scaffolds, which are learning aids that support the development of domain knowledge, may help increase self-regulated learning strategy usage, such as planning, and thus benefit learning (Moos & Azevedo, 2008a). In addition, scaffolds such as navigational aids may reduce disorientation and increase interactivity and positive attitudes (Burke et al., 1998) and may help learners self-regulate their learning by reducing cognitive load.

As an example of scaffolding, goal-setting in the form of providing a list of conceptual topics to be learned may be helpful especially for low-knowledge learners or low self-regulating learners. In the domain of problem solving, providing conceptually-oriented subgoals, rather than computational subgoals, has been shown to result in better transfer to novel statistical problems (Atkinson, Catrambone, & Merill, 2003). Perhaps this benefit is derived from a set of goals acting as an advance organizer that help learners focus on more relevant ideas presented in the learning environment. Segmenting instruction by subgoals and ordering topics by increasing difficulty can reduce cognitive load but also increase self-management demands (Kalyuga, 2007).

Self-evaluation opportunities can be especially important in enhancing self-regulation of learning in computer-based instructional settings (Kostons et al., 2009; Vovides et al., 2007). Novices may have trouble with self-evaluating their own progress for several reasons. One reason might be that they are already experiencing cognitive overload from doing the learning task and not be able to devote resources to monitor their own performance (Butler & Winne, 1995). Even if they are self-monitoring their progress, they may lack the knowledge or criteria to judge their performance accurately (Kostons et al., 2010). In contrast, domain experts may have more resources available to monitor their own performance accurately and they may possess superior metacognitive abilities to make accurate judgments of learning in their domain (Winne, 1996).

Feedback, an essential component of self-evaluation, is most powerful when it addresses misconceptions rather than deficits in knowledge and when the task complexity is low (Hattie & Timperley, 2007). Feedback is crucial to self-regulation of learning as monitoring is dependent upon internal—and external if available—feedback while engaging in a task, possibly resulting in adjustment of learning goals, strategies, and subsequent activities (Butler & Winne, 1995). Self-regulators are attuned to “self-oriented feedback,” including their motivational and cognitive states (Zimmerman, 1990). There is a distinction between scaffolding and external regulation (e.g., program control); the former is more a collaborative effort with an active learner in the building of knowledge, and the latter places control with the program, which ultimately leaves learners not responsible for their own learning (Boekaerts, 1997).

Whether external feedback is ignored, rejected, or incorporated, depends upon the learner’s prior knowledge and cognitive resources (Butler & Winne, 1995). For instance, a task cue, such as an advance organizer, that is overlooked by the learner has no value and cannot benefit learning. Computer-based instruction itself can provide prompts and cues to encourage learners to think about their learning, reinforcing self-regulation of learning (Vovides et al., 2007). However, the utility of outcome feedback may be limited; cognitive feedback that guides the learner towards cues for learning may be more useful (Butler &

Winner, 1995). Further complicating self-evaluation, learners are susceptible to "illusions of competence," that is, overestimating how much they have learned, based upon cues that are present during learning but absent during testing (Bjork, 1994; Koriat & Bjork, 2005; Koriat & Bjork, 2006).

Even if learners reliably self-evaluate their performance, it is questionable if they can use this self-evaluation to make decisions about their subsequent learning processes and tasks. In an explorative study examining self-assessment and task-selection process, learners worked on a learner-control computerized tutorial on heredity and selected a series of eight learning tasks (Kostons et al., 2010). These learning tasks varied by five levels of complexity, and for each complexity level there were three different problem types: worked examples, completion problems, and conventional problems. After completing each task (and before moving on to the next task), learners self-assessed their performance on seven criteria: solution, approach, time on task, enjoyment, difficulty, mental effort, and overall evaluation. These self-assessment criteria could potentially be used to help the learner decide which learning task to do next. Although effective learners (those who had higher learner gains) tended to be more accurate in self-assessing their performance, based upon analyzing their think-aloud protocol, they did not differ on task selection criteria. Moreover, none of the groups demonstrated perfect accuracy in their self-assessments. Kostons et al. (2010) speculated that the lack of group differences on task selection could be due to both groups being novices and not having adequate criteria to make such judgments. Similar to DeLeeuw and Mayer (2008), Kostons et al. imply that expert learners may have exhibited a different pattern of behavior.

Having a low-knowledge base of the domain may contribute to greater intrinsic cognitive load for novices that interferes with their processing that enhances learning (germane load). In contrast, having more domain knowledge would allow the learner to incorporate more information in fewer schemata units, reducing intrinsic cognitive load and freeing up cognitive resources to do more germane processing critical to knowledge acquisition and to self-regulate their learning. In another study examining self-regulated learning on a computer-based tutorial on heredity, learners' actions on the computer screen and eye movements were recorded and replayed for learners as a cue to remind them of their learning paths (Kostons et al., 2009). Eye movements were thought to be reflective of attention and underlying cognitive processes. As reflected by think-aloud verbalizations, this cue was more helpful to novices than to experts in remembering, but not monitoring and self-assessing, their learning processes. Only the experts benefitted from the cue in terms of monitoring and self-evaluating their learning. Presumably the experts experienced lower cognitive load, freeing cognitive resources that allowed them to be able to engage in supplemental processing.

Self-regulation of learning also has motivational components (Boekaerts, 1997; Lynch & Dempo, 2004; Pintrich, 2004; Zimmerman, 1990). In fact, self-efficacy can help promote self-regulation of learning behaviors (Pintrich, 1999), including persistence on difficult tasks (Pintrich & DeGroot, 1990) and setting learning goals (Winne, 1996).

1.5 *Self-Efficacy in Learning*

Computer-based statistics tools that offer learners control over how material is presented can increase engagement with the content (Larreamendy-Joerns, Leinhardt, & Correador, 2005). Self-efficacy and motivation are two related concepts that also have an important role in hypermedia or online learning. Self-efficacy is a self-judgment of one's performance capabilities on a given task; hence, it is context-specific and future-oriented (Bandura, 1977; 1997). Self-efficacy is conceptually different from perceived control over outcomes (such as grades), expectancies and values concerning outcomes, attributions (perceived causes of outcomes), and the learner's self-concept (Schunk, 1991). For instance, one may believe that one will succeed on a task, but not necessarily value the outcome or believe that one has control over the outcome. In addition to being conceptually distinct, self-efficacy has been shown to have discriminant validity in predicting various learning outcomes (Zimmerman, 2000).

Self-efficacy fosters motivation in at least two ways: persistence on a task and goal-setting (Bandura, 1993). Self-efficacy has been shown to have both a direct effect and an indirect effect on skill acquisition by increasing persistence (Pintrich & DeGroot, 1990; Schunk, 1981). Self-efficacy may also influence learning indirectly by encouraging learners to set high personal goals, which is an aspect of self-regulated learning. Although prior grades are good predictors of future achievement, self-efficacy and goal setting have been shown to add predictive power (Zimmerman, Bandura, & Martinez-Pons, 1992; Zimmerman & Bandura, 1994).

Pintrich and DeGroot (1990) sought to construct and validate an instrument to measure components of both motivation and self-regulation of learning. A sample of 173 seventh-grade students in science and English classes responded to a self-report questionnaire (the Motivated Strategies for Learning Questionnaire, or MSLQ) that included 56 items on student motivation, cognitive strategy use, metacognitive strategy use, and management of effort. Each of these items was on 7-point Likert scale (1 = *not at all true of me* to 7 = *very true of me)*. MSLQ scores were collected in between first and second semester course grades. In addition to semester grades, academic performance was also measured by classwork, exams/quizzes, and reports/essays. Factor analysis revealed three distinct components of motivation: Self-Efficacy (9 items; α = .89); Intrinsic Value (9 items; α = .87); and Test Anxiety *(*4 items; α = .75). In addition, two cognitive scales were constructed: (1) Cognitive Strategy Use (13 items; α =.83) which included rehearsal, elaboration, and organizational strategies; and (2) Self-Regulation (9 items; α =.74) which reflected metacognitive and effort management.

Pintrich and DeGroot (1990) found that prior academic achievement predicted metacognitive self-regulation, but not cognitive strategy use. Students high in self-efficacy reported significantly greater use of cognitive strategies and self-regulation strategies than did students low in self-efficacy. Regression analyses revealed that self-efficacy was not significantly related to performance on seatwork, exams, or essays when controlling for the cognitive and self-regulation variables. Pintrich and DeGroot thus argued that cognitive and self-regulatory strategies have a more direct impact on academic performance and that self-efficacy may play a more supportive role by encouraging use of metacognitive strategies.

Much of the research done on the constructs of self-regulated learning and self-efficacy within online settings has been correlational, attempting to relate these constructs to others like motivation and academic achievement (Artino & Stephens, 2009b). However, these motivational constructs have been limited to variables such as self-efficacy and task value, and not other motivational factors such as course satisfaction, intention to continue with online courses, and frustration or boredom with online instruction. In addition, these self-regulation of learning and motivational variables have not often been subject to experimental manipulations or related to learning outcomes. Most of these studies have used self-reported measures of self-regulated learning and learning strategy usage (e.g., MSLQ subscales), not behavioral measures.

Self-efficacy for self-regulated learning has been positively correlated to academic self-efficacy as well as to self-reported cognitive and self-regulated strategy use among online learners (Joo et al., 2000). Self-efficacy for self-regulated learning deals with the belief in one's ability on tasks such as completing assignments on time, concentrating in class, and being able to find a distraction-free environment for studying. Although self-efficacy for self-regulated learning was not directly related to learning outcomes, it was indirectly related through more specific self-efficacy measures: academic self-efficacy and Internet self-efficacy. Contrary to hypotheses, both cognitive and self-regulated strategy use, as self-reported on the MSLQ subscales, were also not positively correlated with learning outcomes. Joo et al. (2000) conjectured that the lack of significant correlations between the strategy use variables and learning outcomes may be due to the self-reported items referring to general course behaviors, and not reflective of nor accurate in describing what learners may be doing online.

Artino (2008) explored self-efficacy and other motivational variables, including intrinsic task value, among those taking a military self-paced online training course. Controlling for other variables, such as gender and prior online learning experience, both self-efficacy and task value reliably predicted course satisfaction. However, limitations of Artino's study include not relating these affective variables to learning outcomes, only course satisfaction, and that these measures were solely self-reported and not linked to behavioral measures. Additionally, the sample was limited to those in the military and thus was not as heterogenous as the online learning community at large.

Expertise differences could also impact self-efficacy and cognitive strategy use on online learning. In a sample of undergraduate and graduate students taking WebCT-managed online courses, there was a significant difference in critical thinking (on the self-reported MSLQ subscale) favoring graduate students but no significant group differences on task value, self-efficacy, or elaboration as a cognitive strategy (Artino & Stephens, 2009a). However, undergraduates reported more procrastination behaviors (e.g., delaying studying). They also reported having more experiences with online courses and greater motivation to do online courses in the future. These differences between learners of varying expertise levels could have implications for actual online learning processes and outcomes and should also be examined in experimental settings.

To examine specifically online instruction, survey instruments related to self-efficacy and self-regulation of learning in online courses have been developed. These include the Online Learning Value and Self-Efficacy Scale (OLVSES) by Artino and McCoach (2008) and the Online Self-Regulated Learning Questionnaire (OSLQ) by Barnard et al. (2009). Although both of these instruments have been validated using factor analysis and large samples, they deal primarily with online courses and not with individual stand-alone computer-based tutorials. The OSLQ consisting of 24 self-reported items, for instance, has been validated based upon responses from separate samples, those taking online courses and those taking hybrid courses (Barnard et al., 2009). It has six subscales measuring different aspects of self-regulated learning; all subscales have high Cronbach's reliability (at least .67): (1) environment structuring, (2) goal setting, (3) time management, (4) help seeking, (5) task strategies, and (6) self-evaluation. Even though it is psychometrically sound, the OSLQ should also be linked to actual learner behavior and learning outcomes in online instructional settings to support its validity as a measure of self-regulated learning and applicability to online learning, as well as to help establish a causal link between self-reported behavior and actual learning.

The OLVSES (Artino & McCoach, 2008), reflecting the notion that self-efficacy is domain-specific, was designed to measure self-efficacy for self-paced, online learning. Based upon factor analyzing responses from hundreds of U.S. Naval Academy students and military personnel, a 11-item scale was created to measure two factors: (1) task value (e.g., interest in course content) and (2) self-efficacy (e.g., confidence in learning). Among the naval academy undergraduates, both task value and self-efficacy were positively and significantly correlated (*p'*s $< .001$) with two cognitive subscales from the MSLQ: the elaboration subscale (e.g., summarizing), $r = .59$ and $r = .27$, respectively, and the metacognitive self-regulation subscale (e.g., goal setting and evaluating knowledge), $r = .62$ and $r = .20$, respectively. However, when examining the unique contribution of each variable in predicting learning outcomes, only task value was a significant predictor, indicating that it was more influential than self-efficacy on cognitive processes. Similar to the OSLQ, the OLVSES needs to be related to actual learning processes and outcomes, and be used with more heterogenous samples to help establish its validity and utility as an online learning survey instrument.

Web-based or hypermedia learning potentially allows for the learner to take more responsibility for their learning, and could increase self-efficacy and motivation by promoting learning goals and giving timely feedback. The effects of learner control on motivation, however, are so varied that Lunts (2002) concluded that there are three types of studies: those that find no effect, those that find a positive effect, and those that find a negative effect. Niemiec, Sikorski and Walberg (1996) also concluded that the effects of learner control on learning outcomes are inconsistent. A better theoretical framework of learner control is needed to reconcile these inconsistent findings. It is not a question of providing learner control or not, but how much and how to facilitate use of learner control (Chung & Reigeluth, 1992). As with traditional learning, learner characteristics such as expertise, self-efficacy, and self-regulated learning abilities may also impact the effect of learner control in computer-based instructional settings.

1.6 *Learner Control Interacting with Scaffolding and Learner Characteristics*

Several studies have investigated the effects of learner control on computer-based learning and come to various conclusions concerning how learner control is moderated by different program features (e.g., scaffolding) and learner characteristics, including self-regulation of learning and learner expertise in a variety of domains.

For instance, Burke et al. (1998) examined learner control and scaffolding in the form of navigational aids in a study involving fifth graders completing a hypermedia lesson on the solar system. Although there was no difference between the learner-control and program-control versions, nor a main effect of navigational aids on achievement, there was a marginally significant interaction between instructional control and navigational aids, $F(1,85) = 3.43$, $p = .067$. Learners in the learner-control condition benefited from having navigational aids ($M = 23.82$ and $M = 20.55$, respectively); learners in the program-control condition did better with no navigational aids ($M = 22.26$, and $M = 20.14$, respectively). In addition, there were differences in attitudes and navigational paths between groups. Compared to learners in the program-control condition, those in the learner-control condition had more positive attitudes about the instruction. Compared to learners who did not have navigational aids, those who had navigational aids demonstrated more non-linear paths, possibly reflecting more cognitive engagement with the instructional content. Thus, especially for those in the scaffolded learner-control condition, learning may have been enhanced because of both motivational and cognitive reasons. An examination of a learner characteristics such as expertise, cognitive load, motivation, or self-regulation of learning ability would have been helpful in clarifying causal relationships between the experimental manipulations with learning outcomes.

Young's (1996) study more directly related learner control to a learner characteristic, namely self-regulation of learning. The study examined seventh grade students who completed a computer-based tutorial on propaganda techniques in advertisement. Students' self-regulation of learning ability was measured with the Self-Regulatory Skills Measurement Questionnaire (SRSMQ), a 33-item survey adapted from Pintrich and De Groot's (1990) MSLQ and Zimmerman and Martinez-Pons' (1986) Self-Regulated Learning Interview

Schedule. The SRSMQ items were designed to focus entirely on self-regulation of learning strategies and thus eliminated motivational items from the original scales. There was a significant interaction between instructional control and self-regulation of learning, but no significant main effects. Low self-regulators in the learner-control condition had the worst post-test scores of the four conditions; regardless of instructional control, high self-regulators had comparable post-test scores. Limitations to this study include the small sample size ($N = 26$), post-test-only design, and elimination of motivation as a component of self-regulated learning. By eliminating motivation, which is an important aspect of self-regulation of learning (Pintrich & DeGroot, 1990; Zimmerman, 2002) and critical to academic achievement (Zimmerman et al., 1992), an important factor may have been overlooked that might have helped to explain why the low self-regulators in the learner-control condition did worse than the other groups.

In addition to self-regulated learning abilities and learner control (in the form of non-linear topic sequencing), McManus (2000) examined the effects of advance organizers (i.e., scaffolding) in a Web-based hypermedia instruction on using a computer operating system. Half of the tutorials included advance organizers that attempted to connect prior knowledge with the learning content. There were three levels of sequencing linearity that reflected varying levels of learner control. Self-regulated learning, which was classified into three levels, was measured by a modified version of the MSLQ, adapted for computer-based instruction rather than in a traditional course. On a post-test, learners were tested on both procedural knowledge and declarative basic computer knowledge.

McManus (2000) found two interactions that were marginally significant at the .05 alpha level: between non-linearity and self-regulated learning, $F(2, 101) = 2.42$, $p = 0.054$; and between non-linearity and advance organizers, $F(2, 101) = 3.05$, $p = 0.052$. No other interaction or main effect was significant or nearly significant. The low level of statistical power (119 college participants for a 3 x 3 x 2 between-subjects design) may explain the lack of significant findings. However, the trends identified in this study suggest that the effect of non-linearity (learner control) may depend upon the presence of advance organizers and upon the self-regulating attributes of the learner. These findings indicate that scaffolding, in the form of advance organizers, may be more effective in highly and moderately nonlinear instruction (higher learner control), respectively, than in mostly linear instruction (lower learner control). The findings also indicate that lower learner control may impair learning for highly self-regulating learners, whereas, higher learner control may hurt the learning of less self-regulating learners. McManus (2000) conjectured that the lack of a self-regulation of learning main effect may be due to learners not being familiar with an online learning environment, and thus having less than accurate self-assessments of their learning behaviors in that setting.

In a hypermedia study with second-graders learning about food groups, Shin et al. (1994) looked at three variables similar to those examined by

McManus (2000): advisement (provided or not), learner control (free access vs. limited access), and prior knowledge. Advisement, a form of scaffolding, provided optional recommendations that students could use to navigate the sequence of topics, as well as visual aids to help them navigate the tutorial. In this study, learner control could be considered similar to McManus' conceptualization of linear sequencing. Whereas McManus examined self-regulated learning ability, Shin et al. focused on prior knowledge, as measured on a pre-test.

Shin et al. (1994) found that high-knowledge learners had significantly higher post-test scores than did low-knowledge learners. Moreover, there was a significant interaction between learner control and prior knowledge. Low-knowledge students benefited more from the more linear limited-access module than the more non-linear free-access module. In contrast, high-knowledge students benefited equally from both. These findings regarding low-knowledge learners parallel those of McManus (2000), who found that learners with low or moderate self-regulating learning strategies were negatively affected by non-linear sequencing (higher learner control); however, McManus found that high self-regulators were differentially affected by linearity of sequencing, unlike Shin et al.'s findings of a comparable benefit.

Although Shin et al. (1994) found that advisement on sequencing topics did not affect learning gains, advisement interacted with prior knowledge to impact the amount of time to finish the tutorial. The low-knowledge students completed the lesson faster when they received no advisement, 11.9 min, than those receiving advisement, 14.8 min; whereas the high-knowledge students did not differ in completion times, 13.3 min and 12.8 min, respectively. This suggests that low-knowledge students may not be able to correctly gauge their learning and take appropriate steps to fill in any gaps in knowledge they may have.

1.7 *Summary*

There is some evidence that high levels of learner control may be detrimental to learning for some groups, including learners with low self-regulation of learning abilities and low prior knowledge or domain expertise. In addition, computer-based instruction may be the ideal setting for examining and promoting self-regulated learning via scaffolding and giving learners control over their learning processes (Scheiter & Gerjets, 2007). However, only a few studies have assessed the relationship between learner control, self-regulated learning, and self-efficacy during computer-based learning (Moos & Azevedo, 2008a). Furthermore, much of the research done on self-regulated learning and self-efficacy in online learning environments has been correlational, attempting to relate these learner characteristics to motivation and academic achievement (Artino & Stephens, 2009b). Although computer-based instruction in statistics education has been shown to be effective (Hsu, 2003; Schenker, 2007; Sosa et al., 2011), it is unclear how prior knowledge, self-efficacy, and cognitive-strategies differentially impact learning outcomes and how these factors influence one another especially in online learning environments teaching statistical concepts.

The effectiveness of instructional scaffolds, implemented in a variety of forms, advance organizers, advisement, navigational aids, can depend upon the learner's expertise and prior experiences. In terms of cognitive load, the expertise reversal effect may help explain how scaffolding instruction may be useful to novices with low domain knowledge but disrupt learning for experts with higher domain knowledge (Kalyuga & Sweller, 2005; van Merriënboer & Sweller, 2005). Experts bring with them well-established schemata of related concepts and these schemata may be in conflict with models presented by the instruction, placing a burden on their cognitive system to reconcile these conflicting models of information (Kalyuga, 2007). Thus, learners with higher levels of domain expertise or self-regulation of learning ability may require different types of scaffolding or instructional control than novices, and experience different types of cognitive load when using the same instruction. Cognitive load has been measured in a variety of ways, some more suitable for learning tasks than others, and some more discriminating among the different types of load (e.g., DeLeeuw & Mayer, 2008; Gerjets et al., 2009).

Self-regulation of learning is a cyclical process that involves planning, managing, and evaluating one's learning (Puustinen & Pulkkinen, 2001). Self-evaluation is an important aspect of self-regulated learning and can place heavy cognitive demands on learners, especially novices, and may affect learning processes and ultimately outcomes (Kostons et al., 2009). Novices may lack the knowledge of what constitutes good performance and even if they had adequate criteria to self-assess their performance, may lack knowledge about how to improve their learning (Kostons et al., 2010). Scaffolds that are intended to improve self-regulation of learning, such as goal-setting and feedback, may be used differently by experts and novices. Instructional design, such as learner control, also has implications for motivation and self-efficacy, which in turn, also affect self-regulation of learning behaviors (Pintrich & de Groot, 1990; Zimmerman, 2000).

Thus, it is critical to link cognitive and motivational processes with actual learning outcomes. Niemiec et al. (1996) concluded that the research on learner control is mixed, its impact on learning is “neither powerful nor consistent” (pg. 157). Therefore, to clarify the effects of learner control, it is important to examine the link between learner control and learner characteristics, including self-regulated learning abilities, and how this link affects learning performance (Young, 1996). It is also important to examine and validate the mediating effects of learner control on a learner’s cognitive load by manipulating learner control as part of the research design.

1.8 *The Current Study*

One possible reason why many computer-based statistics instruction programs have produced suboptimal results in learning is that they might have been implemented without regard to assessing prior knowledge and domain expertise. As a result, these programs have accordingly failed to provide relevant scaffolding and given learners too much, or too little, control over the instructional process to engage learners. Low-knowledge learners may exhibit less effective learning strategies than high-knowledge learners and may benefit more from scaffolding and less learner control. Furthermore, prior knowledge and metacognition factors that impact learning, self-evaluation, and self-regulation of learning practices, were overlooked by many studies.

In fact, analyzing 45 statistics education studies that compared computer-based instruction to traditional instruction (little or no use of technology) in teaching statistical concepts, Sosa et al.'s (2011) meta-analysis which determined a moderate advantage of computer-based instruction in statistics education, $d = 0.33$, found that only 13 studies employed a pre-test of students' prior knowledge and only 15 studies statistically controlled for pre-existing group differences. Moreover, none of the 45 studies were designed to test specific learner-centered variables. Many of these statistical educational studies were explorative or lacked strong methodological considerations to be able to make causal inferences or theoretical connections. Therefore, it would be particularly valuable to examine how to enhance understanding of basic statistical concepts such as variability using computer-based instructional and an experimental design involving learner characteristics and instructional control.

The current study examined the effects of statistical expertise (in terms of number of statistics courses taken), self-efficacy, self-regulation of learning strategies, task value, and learner control on learning about standard deviation in the context of a Web-based hypermedia tutorial. Two versions of a hypermedia tutorial were created, differing only in the degree of control given to learners regarding the amount of instructional material they were given and the amount of feedback they received on their responses to questions.

Learners in the low learner-control (i.e., program-control) condition were automatically presented explanative feedback on why each of their responses to multiple-choice questions was correct or incorrect. After completing tutorial sections, they were also quizzed on summative true-or-false questions intended to integrate the concepts presented in that section. In contrast, learners in the high learner-control condition could choose to skip these integrative, summative questions at the end of each section and to skip the feedback on why their response was correct or incorrect.

1.9 *Hypotheses and Analyses*

The current study was designed to test the hypothesis that learning about standard deviation would benefit from program-control (vs. learner-control) instruction in the form of automatically providing feedback for each learner response and requiring learners to complete integrative, summary questions at the end of each section before moving on to the next section. That is, participants in a program-control (PC) condition were expected to demonstrate greater learning than those in a learner-control (LC) condition. Moreover, the PC benefit was hypothesized to be greater for learners with a low level of statistical experience than for more experienced learners and for learners who reported using less effective self-regulatory strategies than for highly self-regulating learners. It was also hypothesized that higher levels of self-efficacy would be associated with greater learning, but self-regulating strategies would have an even larger impact on learning. Finally, it was predicted that higher levels of cognitive load would be associated with less learning.

A hierarchical regression analysis on learning outcomes, as measured by post-test performance adjusted for pre-test performance, was conducted with the predictors entered in blocks in the following order: (1) prior knowledge, time spent on tutorial, and statistical expertise (i.e., whether the participant had completed one or more statistics courses); (2) task value (i.e., valuing learning about standard deviation), self-efficacy and self-regulated learning; (3) instructional control; (4) cognitive load; and (5) instructional control x statistical expertise, and instructional control x self-regulated learning. Predictors in the first block were considered to reflect general learner knowledge before the tutorial; predictors in the second block were considered to be learner characteristics more related to the tutorial itself. Instructional control was assumed to have bearing on cognitive load and thus entered the model before cognitive load. Lastly, interaction terms were entered after the main effects. Continuous predictors in the model were centered prior to entry and computing interaction terms to reduce problems associated with multicollinearity, according to procedures described by Aiken and West (1991).

Several predictions were made concerning learning about the standard deviation (as measured by post-test performance) on a hypermedia tutorial, with time spent on the tutorial, prior knowledge, and statistical expertise controlled for. These predictions were as follows:

H1. Greater task value and self-efficacy both will be associated with greater learning.

This will be evidenced by positive and significant beta weights.

H2. Self-regulated learning will have an even larger positive effect than task value and self-efficacy on learning.

This will be evidenced by a significant and even larger positive beta weight for self-regulated learning than for task value and self-efficacy.

H3. Learning will be greater for the PC condition than the LC condition.

This will be tested by examining the significance and direction of the beta weight for instructional control, which is dummy coded as 0 for the PC condition and as 1 for the LC condition. As such, it is expected that the beta weight will be negative, reflecting better learning for the PC condition.

H4. Higher levels of cognitive load will be related to less learning.

This will be reflected by a negative beta weight of cognitive load.

H5. Instructional control was predicted to interact with both statistical expertise and self-regulated learning strategy.

H5a. The benefits of PC instruction compared to LC instruction will be greater for novice learners than for expert learners.

This will be tested by examining the significance and magnitude of the beta weight for this interaction.

H5b. The benefits of PC instruction compared to LC instruction will be greater for low self-regulating learners than for high self-regulating learners.

This will be tested by examining the significance and magnitude of the beta weight for this interaction.

1.10 *Importance and Implications*

As early as 1991, there has been a call to provide a more explicit theoretical framework for learner control research (Milheim & Martin, 1991). Years later, Dinsmore, Alexander, and Loughlin (2008) explored how the constructs of metacognition, self-regulation, and self-regulated learning have been defined, studied, and reported in empirical research. In a review of 255 studies examining these self-regulation of learning constructs, Dinsmore et al. concluded that there is conceptual ambiguity concerning the constructs and a need to sharpen their definitions. Moreover, the literature on learner control would benefit from making more theoretical connections to cognitive processes such as cognitive load (Lunts, 2002; Scheiter & Gerjets, 2007). The current study was designed to disentangle these constructs and help clarify the learning process. In particular, the study examines whether an expertise reversal effect could explain differences between novice and expert learners on both learning processes and outcomes in a computer-based statistics instructional setting.

In addition, the current study provides information on how learners of varying statistical expertise utilize scaffolding and benefit differently from having control over instructional materials (i.e., feedback and additional exercises) when learning about standard deviation, a basic statistical concept dealing with variability and distributions. Many researchers and instructors would agree that understanding variability and distributions is necessary for understanding more complex topics such as statistical inference, *p*-values, and confidence intervals (e.g., Chance et al., 2004). However, it is unclear what is required for understanding distributions and variability and how best to support learning about these concepts for learners with different levels of expertise.

Chapter 2: Method

2.1 *Participants and Design*

Two hundred and ten undergraduate and graduate students were recruited at two public colleges and four private educational institutions, primarily those who were taking or have taken an introductory or more advanced statistics course. For participants currently taking a statistics course, participation in the study was optional or students were given extra credit for participating. Approximately half of the participants were randomly assigned to the program-control (PC) condition (n =106); the rest were assigned to the learner-control (LC) condition (n = 104).

To determine an appropriate sample size to obtain power of .80 with a Bonferroni-adjusted alpha of .005, based upon a nominal alpha of .05 and 10 comparisons, power analyses were conducted using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007). Given a multiple regression design, 183 participants would provide adequate power for a minimum overall effect size of f^2 = 0.15 (a medium effect for overall R^2 according to Cohen, 1988) with a total of eight between-subjects predictors and two interaction predictors. These predictors of learning were time spent on the tutorial, prior knowledge, statistical expertise, task value, self-efficacy, self-regulated learning, instructional control, cognitive load, and two-way interactions between instructional control with prior knowledge and self-regulated learning.

2.2 *Materials and Procedure*

In this hypermedia study, an informed consent form (Appendix A) was presented to participants, followed by demographic questions (Appendix B). Demographic questions concerned whether participants have learned about standard deviation before as well as their experience with statistics, number of statistics courses taken, educational level, educational field, age category, gender, and institutional affiliation.

Participants then rated themselves on twelve items on self-regulated learning behaviors, self-efficacy and task value. Several of these self-rating items were adapted from the self-efficacy and self-regulation subscales on the Motivated Strategies for Learning Questionnaire (MSLQ, Duncan & McKeachie, 2005, see Appendix C), as well as the Online Self-Regulated Learning Questionnaire (OSLQ, Barnard et al., 2008). For the current study, three self-efficacy items were adapted to focus specifically on learning about standard deviation on the online tutorial; whereas, the seven self-reported ratings on self-regulation of learning were modified to reflect more general learning strategies (see Appendix D). These self-regulation of learning items were designed to capture aspects of goal-setting, strategy usage, and self-evaluation applicable to online learning. Two additional self-ratings were to measure the task value of doing well on the tutorial and learning about standard deviation. On these twelve items, the learner rated how true these statements were of themselves on a 1-7 scale, from “Not true at all” to “Very true of me.”

Following these self-reported items, participants were introduced to the tutorial: "The goal of this tutorial is to provide a foundation for understanding the variability of observed scores. Variability is a key concept for basic statistics and for many advanced statistical techniques you may encounter." Participants were then presented the learning goals for the tutorial:

1. How variability is related to the shape of a distribution.
2. How standard deviation is used as a measure of variability.
3. What makes a standard deviation larger or smaller.

These goals were presented to help participants focus their learning and to give them some criteria to self-evaluate their performance at later points in the tutorial. After this brief introduction, participants completed a 15-item pre-test (Section 1 of 4) assessing baseline statistical knowledge, or prior knowledge, of interpreting histograms and comparing standard deviations in pairs of histograms. On five items that assessed understanding of distributions, the learner needed to interpret and match histograms with descriptions of various situations, such as scores on a very easy quiz (see Appendix E). These five items were drawn from a nationally validated test called the CAOS (Comprehensive Assessment of Outcomes in Statistics) Test, which was developed by delMas, Garfield, Ooms, and Chance (2007) to assess concepts that introductory statistics students should master. On 10 items adapted from those used by delMas and Liu (2005) to assess understanding of statistical variability, the learner compared two different histograms to determine which had a greater standard deviation (see Appendix F). Following the completion of the pre-test (Section 1), learners were given feedback on how many items out of the 15 they got right, so that they would be engaged with all sections of the tutorial, including the post-test. They were also given a motivational prompt either to improve their score (if their score was 10 or less), enhance their understanding (if their score was 14 or 15), or both (if their score was between 11 and 13 inclusive) by completing the tutorial.

Section 2 introduced the standard deviation, *SD*, and its calculation based upon the sum of squared deviations from the mean for each observation (squared deviations are computed by squaring the distance of each observation from the sample mean, and these values are totaled to form a "sum of squares," or *SS*). The *SD* is then calculated by dividing the *SS* by the sample size, *N*, minus 1, and then square rooting the result:

$$SD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{SS}{N-1}}$$

These squared deviation calculations were represented visually in an example histogram, from which the learner had to identify relevant values and calculate the *SS* and S*D* (see Figure 2.1). In this same histogram, before calculating these values, the learners had to identify how many observations had a value of 2 and how many had a value of 3.

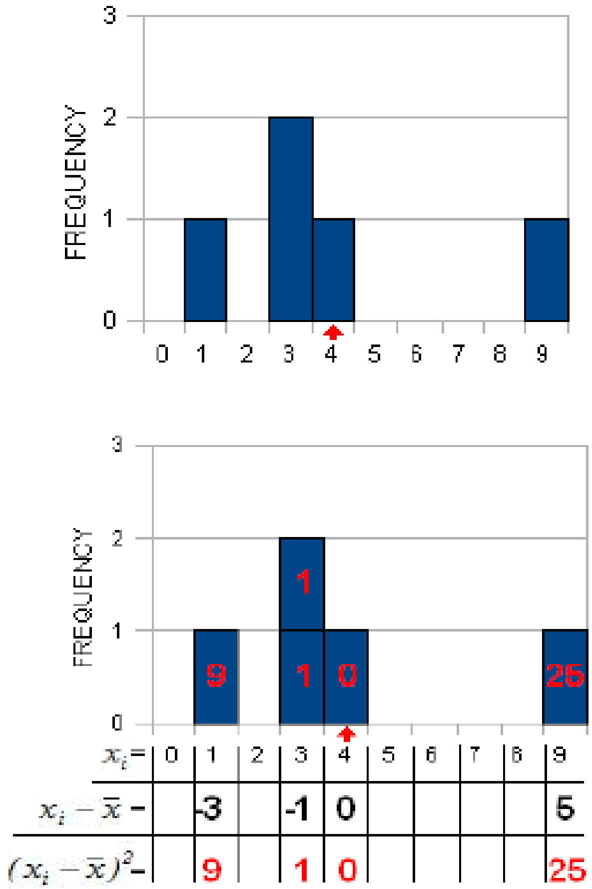


Figure 2.1. Introductory histogram in the Standard Deviation Tutorial. The top figure is the original figure and the bottom figure is presented in an optional pop-up window to illustrate how the sum of squared deviations is calculated. Here *SS* = 36 and *SD* = 3.

Following this overview, the tutorial featured a series of interactive conceptual self-assessment activities. These interactive activities were implemented in accordance with the principle that self-testing can improve learning (Roediger & Karpicke, 2006). Throughout most of the tutorial, students were asked to compare the standard deviations between pairs of histograms (see Figure 2.2 for an example) and choose a multiple-choice answer that reflected both the best answer and justification for the answer. On each of the histograms, the sample mean was marked by a red arrow. Multiple-choice distractors were constructed based upon specific common justifications—both correct and incorrect—used by students to produce the largest or smallest standard deviations in the delMas and Liu (2005) study. Justifications were often related to

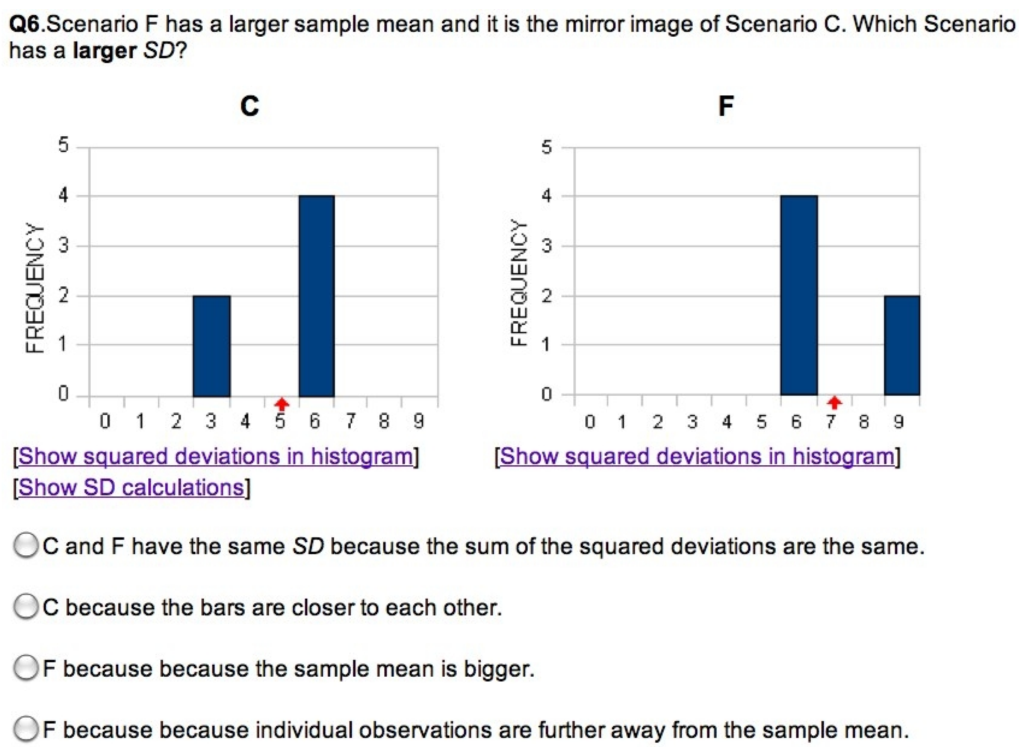


Figure 2.2. Example of a histogram-pair in Section 2. Participants compared the standard deviations of the histograms and chose a multiple-choice response to reflect best answer and justification.

comparing the shape, range, and how observations were distributed relative to the sample mean. To move forward in the tutorial, participants had to correctly answer each multiple-choice question.

For each of their responses, learners were told whether they were correct or not, and given a chance to view pop-up windows with explanatory histograms depicting the squared deviations and *SD* calculations for each of the original histograms (see Figure 2.3). Participants in the PC (Program Control) condition were automatically given a text explanation of why their response was correct or incorrect; in contrast, LC (Learner Control) participants were allowed to go on to the next questions without viewing explanative text feedback on why they answered correctly or incorrectly. To view explanative text feedback on each of their responses, learners in the LC condition needed to click on the "see why" link, which caused the explanation to appear in a pop-up window.

At any time, participants in both conditions could also click on a link to view only the squared deviations of observations in a histogram or the more detailed *SD* calculations. Depictions of squared deviations in histograms and *SD* calculations appeared in separate pop-up windows and could be viewed one at a time whenever the learners desired. Access to these scaffolds was allowed in both conditions so that comparisons of their usage across instructional conditions could be made.

Pairs of histograms were designed to highlight how shape and distribution affected the standard deviation, and the comparisons between pairs increased in difficulty progressing from 2-bar to 3-bar examples in Sections 2 and 3. Section 3 was more difficult conceptually than Section 2. For instance, in Section 2, histogram pairs illustrated the idea that a mirror image of a histogram has the same standard deviation (see Figures 2.2 and 2.3). In Section 3, range and shape needed to be integrated in comparing histograms (see Figure 2.4). Section 2 contained eight questions total: four questions on interpreting a histogram and calculating sums of squares and standard deviations, and four questions on comparing the standard deviation in pairs of histograms. Section 3 contained six questions on comparing pairs of histograms. The histogram pairs used in the tutorial involved only whole-number squared deviation scores and were less complicated than the ones presented in the pre-test and post-test, which depicted 4 to 8 bars of observations (see Appendix F).

At the ends of Sections 2 and 3, participants were advised that "The next three True-or-False questions are designed to review and integrate the principles that you worked on" in that section. Individuals in the PC condition were required to complete this set of questions before rating that section and then moving on to the next section. Regarding the three review questions, participants in the LC condition were advised, "You may either review them or skip them" and then given a choice to do them or not. As with the histogram pairs, these questions were easier on Section 2 than on Section 3. For instance, at the end of Section 2, the learner had to evaluate whether the following statement was true or false: "*SD* is the same when bars in the histogram are flipped to form a mirror

The general formula for the standard deviation of a sample is:

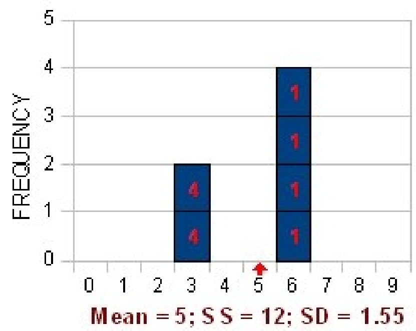
$$SD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{SS}{N-1}}$$

Scenario C
$\bar{x}$ = 5, N = 6:

$$SD = \sqrt{\frac{(3-5)^2+(3-5)^2+(6-5)^2+(6-5)^2+(6-5)^2+(6-5)^2}{6-1}}$$

$$SD = \sqrt{\frac{(-2)^2+(-2)^2+(1)^2+(1)^2+(1)^2+(1)^2}{5}}$$

$$SD = \sqrt{\frac{4+4+1+1+1+1}{5}} = \sqrt{\frac{12}{5}} = \sqrt{2.4} = 1.55$$

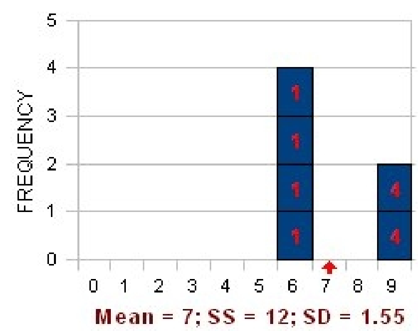


Scenario F
$\bar{x}$ = 7, N = 6:

$$SD = \sqrt{\frac{(6-7)^2+(6-7)^2+(6-7)^2+(6-7)^2+(9-7)^2+(9-7)^2}{6-1}}$$

$$SD = \sqrt{\frac{(-1)^2+(-1)^2+(-1)^2+(-1)^2+(2)^2+(2)^2}{5}}$$

$$SD = \sqrt{\frac{1+1+1+1+4+4}{5}} = \sqrt{\frac{12}{5}} = \sqrt{2.4} = 1.55$$



Although the sample mean in Scenario F is larger than that in Scenario C, both the sample size (N = 6) and the *sum of the squared deviations* (SS = 12) are the same, resulting in the same SD of 1.55.

Figure 2.3. Example of a histogram-pair compared and illustrated. Squared deviations for each observation were depicted visually and standard deviation calculations were given. This information appeared in an optional pop-up window.

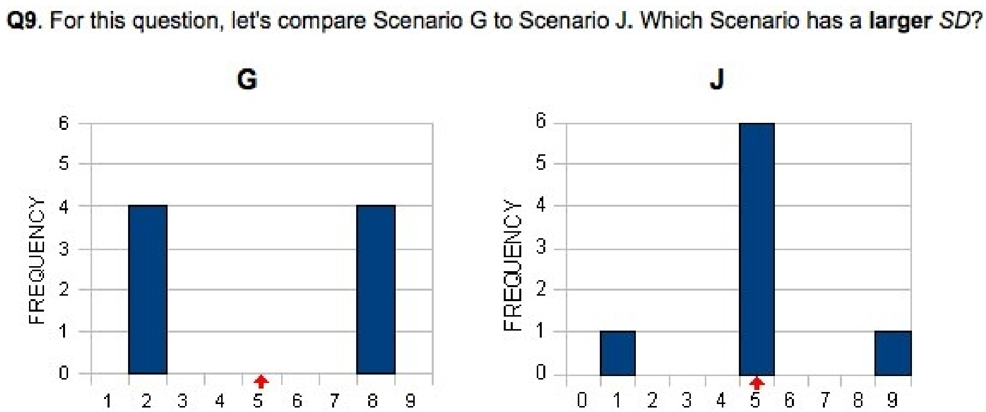


Figure 2.4. Example of a histogram-pair in Section 3. Participants compared the standard deviations of the histograms and chose a multiple-choice response to reflect best answer and justification.

image." Section 3 included more difficult statements to evaluate (e.g., "When the range is the same, a bell-shaped distribution always has a smaller *SD* than a U-shaped distribution.").

After completing each of the tutorial sections, Sections 2 and 3, learners in both conditions rated their mental load and performance in each section by completing four questions related to: (1) their effort exerted, (2) difficulty of the section, (3) how frustrating it was, and (4) how successful they believed they were on that section (see Appendix G). The measures of Effort, Difficulty, and Frustration were designed to provide an index of cognitive load. At the end of the tutorial, on Section 4, participants completed a post-test consisting of the same 15 items that were presented on the pre-test. Participants were encouraged to earn a higher score on Section 4 than they did on Section 1. After completing Section 4, participants were told how many items they got correct on that section. Participants were expected to complete the tutorial, pre-test, post-test, and all ratings in about 45 minutes. Time spent on the tutorial and each of its individual sections, as well as number of scaffolds used (optional *SD*/histogram pop-up windows and links to explanative feedback), were recorded for each participant.

Chapter 3: Results

3.1 *Demographic Information*

A total of 210 students completed the tutorial, 104 in the program-control (PC) condition and 106 in the learner-control (LC) condition. However, to ensure that the sample included only learners who engaged the tutorial and processed the presented questions, nine students who spent less than five minutes total on Sections 2 and 3 were eliminated from the sample. A total of 201 participants remained in the sample, 100 in the PC condition and 101 in the LC condition. Overall there were 169 women and 32 men. In the PC condition, there were 88 women and 12 men; in the LC condition, there were 81 women and 20 men. Most of the students majored in psychology (*n* = 126), followed by humanities (*n* = 28), biological sciences (*n* = 20), business or organizational sciences (*n* = 15), and other or undeclared (*n* = 12).

Concerning their statistical experience, the majority reported having learned about the standard deviation before completing the tutorial (*n* = 173). Regarding statistics courses, 116 participants reported having taken one or more (these participants will be referred to as "experts"), while 85 participants reported that they had not taken a statistics course or were taking their first course (these participants will be referred to as "novices"). There were 142 participants who were undergraduates or had completed their undergraduate studies but not continued on to graduate school, and 59 participants who were graduate students or had completed graduate school. Most graduate students had taken one or more statistics courses, whereas fewer than half of the undergraduate students had (see Table 3.1 for a breakdown of statistics courses by instructional control and education level).

Table 3.1

Frequencies of Participants by Statistics Courses, Instructional Control, and Education

			Statistics courses completed		
		Education Level	Fewer than one course	One or more courses	Total
Instructional Control	Program-control	Undergraduate	46	30	
		Graduate	2	22	
		Total	**48**	**52**	**100**
	Learner-control	Undergraduate	35	31	
		Graduate	2	33	
		Total	**37**	**64**	**101**
	Total		85	116	201

The majority of undergraduate participants reported their age to be 18 to 22 years old (132 of 142); the majority of graduate students reported themselves to be 23 to 29 (47 of 59; see Table 3.2 for a breakdown of age categories by instructional control). Thus, it can be concluded that the majority of undergraduate and graduate students were of traditional age for their educational status.

Table 3.2

Frequencies of Participants by Age Categories, Instructional Control, and Education

			Age (yrs)			
		Education Level	**18-22**	**23-29**	**30+**	Total
Instructional Control	Program-control	Undergraduate	69	7	0	
		Graduate	1	16	7	
		Total	**70**	**23**	**7**	**100**
	Learner-control	Undergraduate	63	3	0	
		Graduate	3	21	11	
		Total	**66**	**24**	**11**	**101**
	Total		136	47	18	201

3.2 *Reliability of Self-Reported and Section Ratings*

Following the demographic questions were 12 self-reported measures (see Appendix D) designed to capture self-efficacy, self-regulation of learning, and task value. These items showed reliability of varying amounts. Self-efficacy (SE) was measured by three items; Cronbach's alpha for the composite based on this set of items was .852, indicating high reliability. The seven self-regulation of learning (SRL) measures also demonstrated high reliability (Cronbach's alpha = .844). Thus, the three items for SE and the seven items for SRL were averaged, respectively, to form composite measures of each of these two learner characteristics. The two items designed to measure task value (TV) had unacceptable reliability (Cronbach's alpha = .152). Thus, only one of the items was retained to be used in subsequent analysis: "Learning about standard deviation is important to me." Deleted from subsequent analyses was the more global item that dealt with the importance of doing things well in general.

Reliability was moderately high for each cognitive measure as originally designed (see Appendix G). For the composite of three cognitive load measures on Section 2, Cronbach's alpha was .685. However, when the Effort rating was eliminated, Cronbach's alpha increased to .816. Indeed, Frustration and Difficulty ratings were more positively correlated to each other, $r = .70$, than Effort was to either, $r = .19$ and $r = .41$, respectively. Similarly, for the three cognitive load measures on Section 3, Cronbach's alpha was .710, but increased to .830 when the Effort rating was eliminated. Again, Frustration and Difficulty were more correlated to each other, $r = .71$, than either was to Effort, $r = .24$ and $r = .41$, respectively. Therefore, for both Sections 2 and 3, the Difficulty and Frustration ratings were summed as a measure of extraneous cognitive load, but the Effort and Success ratings were retained as single-item measures of germane cognitive load and performance self-evaluation, respectively.

Table 3.3 presents the correlations among Self-Regulated Learning (SRL), Self-Efficacy (SE), and Task Value (TV) with ratings of Effort, Difficulty, Frustration, and Success in each Section. SRL ratings were positively related to both Effort and Success ratings, but unrelated to Difficulty and Frustration ratings on Sections 2 and 3. Both SE and TV ratings were negatively related to Difficulty and Frustration ratings, but positively related to Success ratings and unrelated to Effort ratings on both Sections. These patterns further suggest that Difficulty and

Frustration ratings are distinct from Effort ratings, and that Effort is more related to behavioral measures such as SRL. Additionally, learners who rated themselves highly on SRL, SE, and TV beforehand also tended to rate themselves as more Successful during the tutorial.

Table 3.3

*Correlations between Self- and Section Ratings (*N *= 201)*

	Learner Characteristics		
Ratings	SRL	SE	TV
Self-Regulated Learning (SRL)	-		
Self-Efficacy (SE)	.44***	-	
Task Value (TV)	.34***	.34***	-
Section 2:			
Effort (E)	.21**	-.05	.08
Difficulty (D)	-.09	-.34***	-.25***
Frustration (F)	-.09	-.27***	-.26***
Success (S)	.23**	.37***	.35***
Section 3:			
Effort (E)	.15*	-.09	.08
Difficulty (D)	-.13	-.23***	-.20***
Frustration (F)	-.10	-.24***	-.25***
Success (S)	.29***	.38***	.27***

Note. ***p* <. 01, ****p* <. 001. SRL and SE are composites of seven and three items averaged, respectively, and TV was measured by one item. Significant negative correlations are highlighted. These correlations are only between the cognitive load measures and the learner motivational variables of SE and TV.

3.3 *Main Analyses Regarding Learning Outcomes*

The overall average pre-test score on knowledge of standard deviations and histograms was $M = 8.89$, $SD = 3.15$, and the overall average post-test score was $M = 10.18$, $SD = 3.06$. This increase on the post-test was significant, $t(200) = 6.74$, $p < .001$, with a Cohen's $d = .42$. A d of .50 is considered to reflect a "medium" effect (Cohen, 1988) and a d of .25 is considered to be small but practically significant in educational settings (Slavin, 1990). For participants in the PC condition, average pre-test scores ($M = 9.25$, $SD = 3.00$) increased on the post-test ($M = 10.53$, $SD = 2.92$), $t(99) = 4.57$, $p < .001$, $d = .43$. Participants in the LC condition showed comparable increases from the pre-test ($M = 8.53$, $SD = 3.26$) to the post-test ($M = 9.84$, $SD = 3.17$), $t(100) = 4.95$, $p < .001$, $d = 41$. Thus, the tutorial was effective in helping participants in both conditions learn about the standard deviation.

A hierarchical regression analysis on post-test scores was conducted, with the predictors entered in blocks in the following order: (1) pre-test scores, minutes spent on tutorial, and statistical expertise (i.e., whether or not the participant had completed one or more statistics courses); (2) task value, self-efficacy, and self-regulated learning; (3) instructional control; (4) cognitive load; and (5) instructional control x statistical expertise, and instructional control x self-regulated learning. The continuous predictors were centered prior to entry into the regression model and computation of interaction terms. The overall model was significant, $F(10, 190) = 17.46$, $p < .001$; $R^2 = .48$; adjusted $R^2 = .45$ (see

Table 3.4). About 45% of the variance in post-test scores can be explained by this set of predictors. The following predictions were tested:

H1. Greater task value and self-efficacy will be associated with greater learning.

Both task value and self-efficacy were significantly positively related to learning when ignoring all other predictors (r = .19, p < .01; and r = .24, p < .001 respectively; see Table 3.4). However, when controlling for other predictors in the model (pre-test knowledge, expertise, minutes on the tutorial, self-regulated learning, and cognitive load), neither task

Table 3.4

*Moderation Effects of Statistical Expertise on Learner Control in Predicting Learning Outcomes (*N *= 201).*

Step	Variable	*r*	R^2 Change	*B*	SE_B	*Beta*
1	Pre-test Knowledge	.616***	.403***	.494***	.060	.508**
	Statistical Expertise	.309***		.822*	.373	.133*
	Minutes on Tutorial	-.074		.005	.015	.019
2	Task Value	.192**	.012	-.082	.109	-.047
	Self-Efficacy	.241***		.189	.150	.081
	Self-Regulation	.120*		-.004	.202	-.001
3	Instructional Control (IC)	-.113[a]	.003	-.370	.351	-.061
4	Cognitive Load	-.416***	.019*	-.092*	.036	-.163*
5	IC x Statistical Expertise		.042**	2.559***	.681	.390***
	IC x Self-Regulation			-.551	.334	-.121
(Constant)				9.912	.313	

Note. *p <. 05, **p <. 01, ***p <. 001. Cumulative R^2 = .479; Adjusted R^2 = . 451. Continuous predictors were centered by subtracting their means.
[a]p < .10.

value nor self-efficacy had a unique association with learning (*beta* = -.05, $p = .45$; and *beta* = .08, $p = .21$, respectively).

H2. Self-regulated learning will have an even larger positive effect than task value and self-efficacy on learning.

When ignoring other predictors, self-regulated learning was positively and significantly related to learning ($r = .12$, $p < .05$) but not as strongly as task value and self-efficacy. Controlling for pre-test knowledge, expertise, minutes on the tutorial, self-efficacy, task value, and cognitive load, self-regulated learning was not statistically significant and had an even smaller beta weight (*beta* = -.00, $p = .98$) than did task value and self-efficacy. In fact, as a set of predictors, these three self-reported learner-characteristic variables did not add predictive value to the model beyond pre-test knowledge, expertise, and time spent on the tutorial (R^2 Change = .012, $p = .27$).

H3. Learning will be greater for the PC condition than the LC condition.

When controlling for pre-test knowledge, expertise, minutes on the tutorial, self-regulated learning, self-efficacy, task value, and cognitive load, instructional control did not reliably predict post-test scores (*beta* = -.06, $p = .29$).

H4. Higher levels of cognitive load will be related to less learning.

Both when ignoring other predictors ($r = -.42$, $p < .001$) and controlling for all other predictors in the model (*beta* = -.16, $p < .05$), cognitive load was negatively and significantly associated with learning outcomes. Moreover, the model was significantly improved with the inclusion of cognitive load (R^2 Change = .019, $p < .05$).

H5. Instructional control was predicted to interact with both statistical expertise and self-regulated learning strategy.

The inclusion of both of these interactions significantly improved the model beyond all the main effects (R^2 Change = .042, $p < .01$).

H5a. The benefits of PC instruction compared to LC instruction will be greater for novice learners than for expert learners.

The significant interaction between instructional control and statistical expertise indicates that novices and experts are affected differently by instructional control (*beta* = .39, $p < .001$). Novices demonstrated better learning in the PC condition (M = 10.44, SD = 3.00) than in the LC condition (M = 7.32, SD = 3.01). In contrast, experts did equally well regardless of instructional condition (M = 10.62, SD = 2.87; M = 11.30, SD = 2.22, respectively). Follow-up analyses (see Section 3.4) were conducted to examine the nature of this interaction.

H5b. The benefits of PC instruction compared to LC instruction will be greater for low self-regulating learners than high self-regulating learners.

The non-significant interaction between learner control and self-regulation of learning suggests there is no substantial difference between low and high self-regulators with regards to the effect of instructional control (*beta* = -.12, p = .13).

3.4 *Follow-Up Analyses Regarding Instructional Control and Statistical Expertise*

To examine the significant interaction between instructional control and statistical expertise, an ANCOVA was conducted to examine the effects of instructional control and statistical expertise on post-test scores, controlling for pre-test scores. The main effect of instructional control approached statistical significance, $F(1, 196) = 3.84$, $p = .051$, such that participants in the PC condition had higher adjusted post-test scores ($M = 10.31$, $SD = 2.30$) than did LC learners ($M = 9.69$, $SD = 2.41$). Statistical expertise was significant, $F(1, 196) = 11.04$, $p < .01$, such that more experienced learners scored higher on the post-test, controlling for pre-test scores, ($M = 10.60$, $SD = 2.01$) than did novice learners ($M = 9.45$, $SD = 2.78$). However, the main effects must be qualified by a highly significant interaction between instructional control and statistical expertise, $F(1, 196) = 15.31$, $p < .001$. Instructional control made a bigger difference for the novice learners than it did for the experts (see Figure 3.1). The novices in the LC condition had significantly worse post-test scores than novices in the PC condition even adjusting for pre-test scores, $F(1, 82) = 9.49$, $p < .01$. In contrast, among the expert learners, the difference between the PC and LC conditions was in the opposite direction and not reliable, $F(1, 113) = 2.87$, $p = .09$.

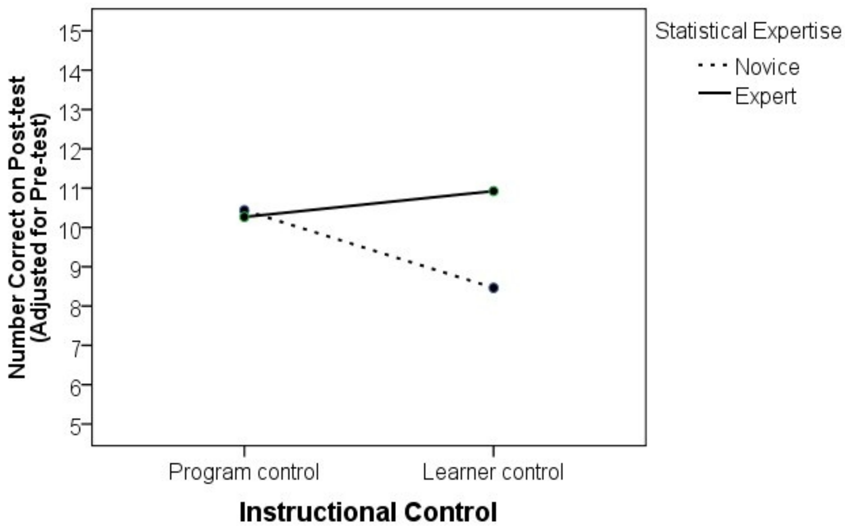


Figure 3.1. Number of correct answers on the 15-item post-test, adjusted for pre-test scores, as a function of instructional control and statistical expertise.

To follow-up on the statistical expertise by instructional interaction effect on post-test scores, separate ANOVAs were conducted on the three self-reported measures of self-efficacy, self-regulated learning, and task value, with statistical expertise and instructional control as independent variables. Statistical expertise was statistically significant for self-regulated learning ($F(1, 197) = 4.27$, $p < .05$, $d = .25$) and task value ($F(1, 197) = 10.88$, $p < .01$, $d = .47$), and nearly statistically significant for self-efficacy ($F(1, 197) = 3.56$, $p = .06$, $d = .28$). The more statistically expert participants in both instructional control conditions rated themselves higher on self-regulated learning and task value (see Table 3.5). The main effect of instructional control was not significant for any of the three self-ratings, p's > .17, nor were interactions between instructional control and statistical expertise statistically significant, p's > .74.

Table 3.5

Average Self-Ratings, SD*'s, and* F*'s by Instructional Control and Statistical Expertise*

		Statistical Expertise	Mean Self-Ratings (*SD*) Self-Regulation	Self-Efficacy	Task Value
Instructional Control	Program Control	Novice	5.14 (0.81)	4.74 (1.33)	4.04 (1.70)
		Expert	5.43 (1.00)	5.03 (1.44)	4.94 (1.70)
		PC Average	**5.29 (0.92)**	**4.89 (1.39)**	**4.51 (1.75)**
	Learner Control	Novice	5.02 (0.83)	4.42 (1.26)	4.00 (1.89)
		Expert	5.28 (1.02)	4.83 (1.17)	4.73 (1.66)
		LC Average	**5.19 (0.96)**	**4.68 (1.22)**	**4.47 (1.78)**
Statistical Expertise Averages		**Novice Average**	5.09 (.82)	4.60 (1.30)	4.02 (1.77)
		Expert Average	5.35 (1.01)	4.92 (1.30)	4.83 (1.68)
Overall Average			5.24 (0.94)	4.79 (1.31)	4.49 (1.76)

	F (1, 197) Self-Regulation	Self-Efficacy	Task Value
Instructional Control (IC)	1.00	1.86	.25
Statistical Expertise	4.27*	3.56[a]	10.88***
IC x Statistical Expertise	0.01	.09	.11

Note. *$p < .05$, ***$p < .001$.
[a]$p < .07$

The ratings of Effort, Frustration, Difficulty, and Success after Sections 2 and 3 were tested for relationships with instructional control and expertise. On Section 2, for Effort ratings, there were no significant main effects or interactions (see Table 3.6). In contrast, for Frustration ratings, both main effects and the interaction were significant. The LC tutorial was rated significantly more frustrating than the PC tutorial, $F(1, 197) = 4.97$, $p < .01$, and experts reported being less frustrated than novices, $F(1, 197) = 31.88$, $p < .001$. These main effects, however, are qualified by a significant interaction, $F(1, 197) = 6.92$, $p < .01$. Novices reported being reliably more frustrated in the LC condition than in the PC condition, $t(83) = 3.12$, $p < .01$, while experts were equally frustrated in the both instructional conditions, $t(114) = .31$, $p = .75$ (see Figure 3.2).

On Section 2, for both Difficulty and Success ratings, there was a significant main effect of statistical expertise and a significant interaction between statistical expertise and instructional control (see Table 3.6). Compared to experts, novices reported having more difficulty, $F(1,197) = 19.84$, $p < .001$, and being less successful, $F(1,197) = 10.41$, $p < .01$. Novices and experts did not significantly differ on Difficulty in the PC condition, $t(98) = 1.55$, $p = .12$, but did differ in the LC condition, $t(99) = 5.08$, $p < .001$. Similarly, novices and experts did not significantly differ on Success in the PC condition, $t(98) = .43$, $p = .67$, but did differ in the LC condition, $t(99) = 4.21$, $p < .001$. Thus, these differences between novices and experts were greater in the LC condition than the PC condition (see Figures 3.3 and 3.4).

Table 3.6

Average Section 2 Ratings, SD*'s, and* F*'s by Instructional Control and Statistical Expertise*

			Mean Section-Ratings (*SD*)			
		Statistical Expertise	**Effort**	**Frustration**	**Difficulty**	**Success**
Instructional Control	Program Control	Novice	4.40 (1.50)	3.73 (1.71)	3.94 (1.51)	5.15 (1.60)
		Expert	4.06 (1.72)	3.06 (1.69)	3.46 (1.55)	5.29 (1.73)
		PC Average	**4.22 (1.62)**	**3.38 (1.72)**	**3.69 (1.54)**	**5.22 (1.66)**
	Learner Control	Novice	4.43 (1.59)	4.81 (1.41)	4.27 (1.31)	4.27 (2.01)
		Expert	4.33 (1.63)	2.97 (1.37)	2.95 (1.23)	5.62 (1.23)
		LC Average	**4.37 (1.61)**	**3.64 (1.64)**	**3.44 (1.40)**	**5.13 (1.68)**
Statistical Expertise Averages		**Novice Average**	4.41 (1.53)	4.20 (1.67)	4.08 (1.42)	4.76 (1.83)
		Expert Average	4.21 (1.67)	3.01 (1.51)	3.18 (1.40)	5.47 (1.47)
Overall Average			4.29 (1.61)	3.51 (1.68)	3.56 (1.48)	5.17 (1.67)
			F (1, 197)			
			Effort	**Frustration**	**Difficulty**	**Success**
Instructional Control (IC)			0.44	4.97*	.19	1.35
Statistical Expertise			0.90	31.88***	19.84***	10.41**
IC x Statistical Expertise			0.25	6.92**	4.37*	6.82*

Note. *$p < .05$, **$p < .01$, ***$p < .001$.

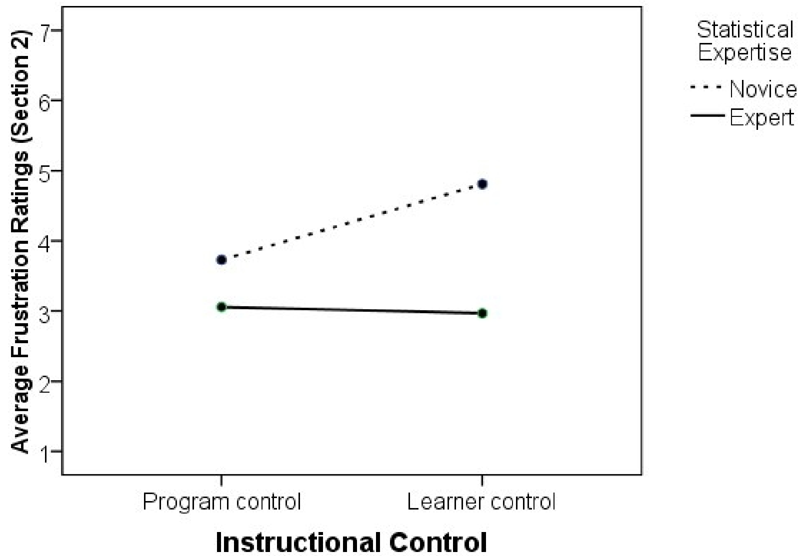


Figure 3.2. Average Section 2 Frustration ratings as a function of instructional control and statistical expertise.

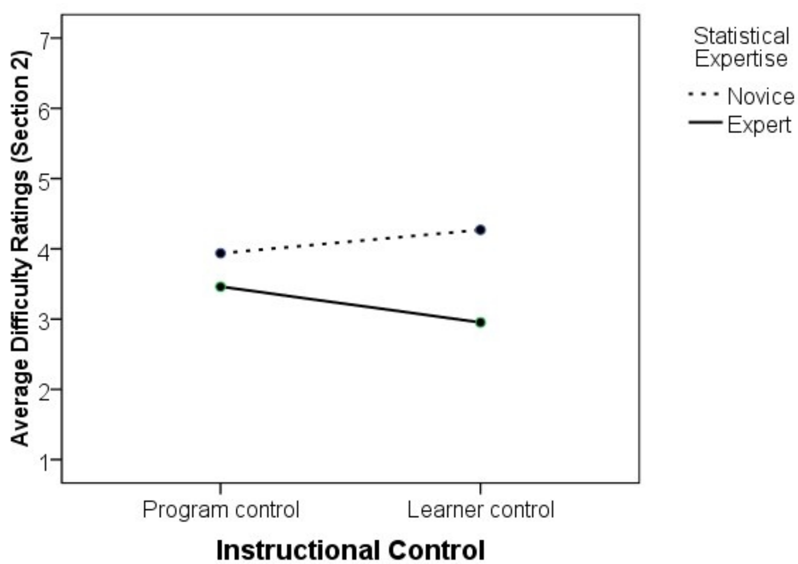


Figure 3.3. Average Section 2 Difficulty ratings as a function of instructional control and statistical expertise.

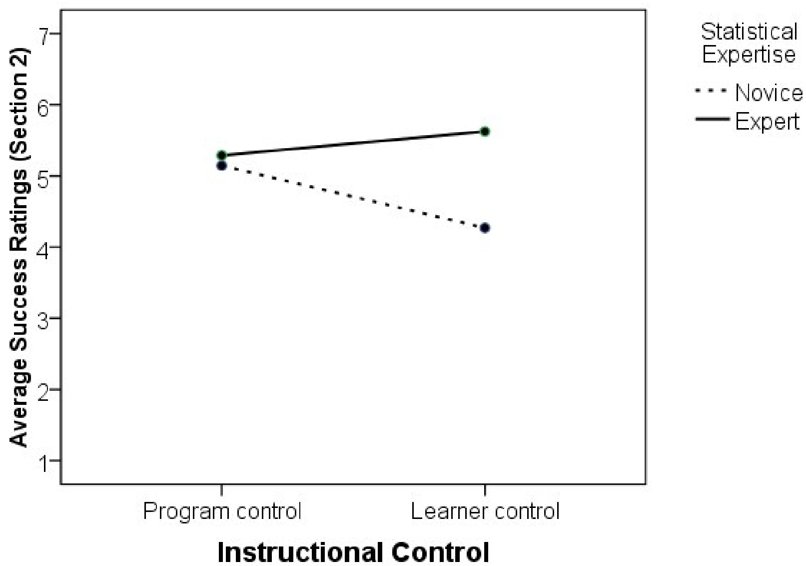


Figure 3.4. Average Section 2 Success ratings as a function of instructional control and statistical expertise.

Section 3 ratings showed similar patterns as did Section 2 ratings, although three effects that were significant in Section 2 did not attain statistical significance in Section 3: there was no significant main effect of instructional control on Frustration ratings, the interaction was marginally significant for Difficulty ratings, and no significant main effect of statistical expertise on Success ratings (see Table 3.7). For both Frustration and Difficulty ratings, novices had higher ratings than experts overall, $F(1, 197) = 14.46$, $p < .001$; and $F(1, 197) = 9.74$, $p < .01$, respectively. The interaction between statistical expertise and instructional control was significant for Frustration, $F(1, 197) = 6.44$, $p < .05$; and marginally significant for Difficulty, $F(1, 197) = 3.40$, $p = .06$. Novices and

Table 3.7

Average Section 3 Ratings, SD's*, and* F's *by Instructional Control and Statistical Expertise*

			Mean Section-Ratings (*SD*)			
		Statistical Expertise	**Effort**	**Frustration**	**Difficulty**	**Success**
Instructional Control	Program Control	Novice	4.40 (1.54)	3.67 (1.55)	3.94 (1.64)	5.10 (1.17)
		Expert	4.00 (1.53)	3.38 (1.62)	3.65 (1.61)	4.90 (1.60)
		PC Average	**4.19 (1.54)**	**3.52 (1.59)**	**3.79 (1.62)**	**5.00 (1.41)**
	Learner Control	Novice	4.24 (1.61)	4.46 (1.54)	4.35 (1.44)	4.32 (1.70)
		Expert	4.19 (1.79)	3.05 (1.50)	3.25 (1.47)	5.09 (1.43)
		LC Average	**4.21 (1.72)**	**3.56 (1.65)**	**3.65 (1.55)**	**4.81 (1.57)**
Statistical Expertise Averages		**Novice Average**	4.33 (1.56)	4.01 (1.59)	4.12 (1.56)	4.76 (1.47)
		Expert Average	4.10 (1.68)	3.20 (1.56)	3.43 (1.54)	5.01 (1.51)
Overall Average			4.20 (1.63)	3.54 (1.62)	3.72 (1.58)	4.91 (1.49)

	F (1, 197)			
	Effort	**Frustration**	**Difficulty**	**Success**
Instructional Control (IC)	.01	1.04	.00	1.93
Statistical Expertise	.92	14.46***	9.74**	1.80
IC x Statistical Expertise	.52	6.44*	3.40[a]	5.23*

Note. *$p < .05$, **$p < .01$, ***$p < .001$.
[a]$p < .07$

experts did not differ significantly on Frustration in the PC condition, $t(98) = .89$, $p = .38$, but did differ in the LC condition, $t(99) = 4.53$, $p < .001$. Following a similar pattern, novices and experts did not significantly differ on Difficulty in the PC condition, $t(98) = .87$, $p = .39$, but did differ in the LC condition, $t(99) = 3.66$, $p < .001$. Again, the differences between novices and experts were greater in the LC condition than the PC condition (see Figures 3.5 and 3.6). Neither main effect was significant for Success ratings, but the interaction was significant, $F(1, 197) = 5.23$, $p < .05$. Novices and experts had comparable Success ratings in the PC condition, $t(98) = .71$, $p = .48$, but novices had lower ratings in the LC condition, $t(99) = 2.43$, $p < .05$ (see Figure 3.7).

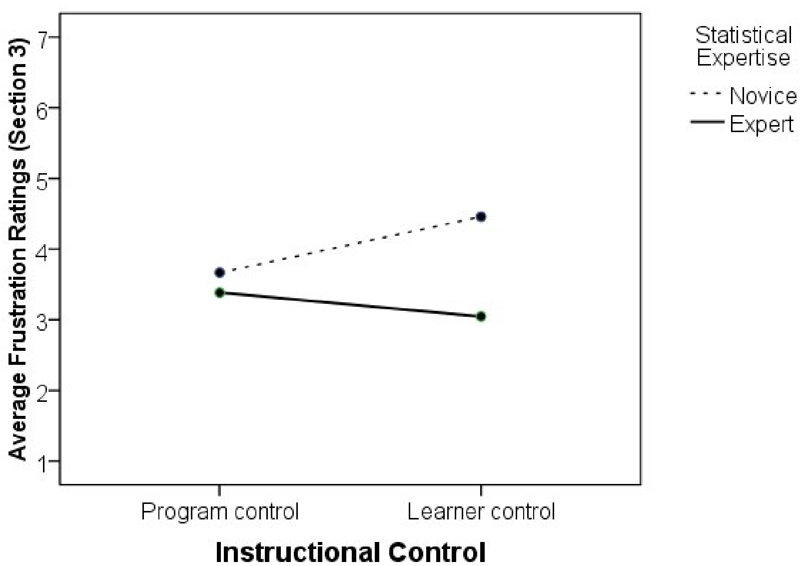


Figure 3.5. Average Section 3 Frustration ratings as a function of instructional control and statistical expertise.

Figure 3.6. Average Section 3 Difficulty ratings as a function of ratings as a function instructional control and statistical expertise.

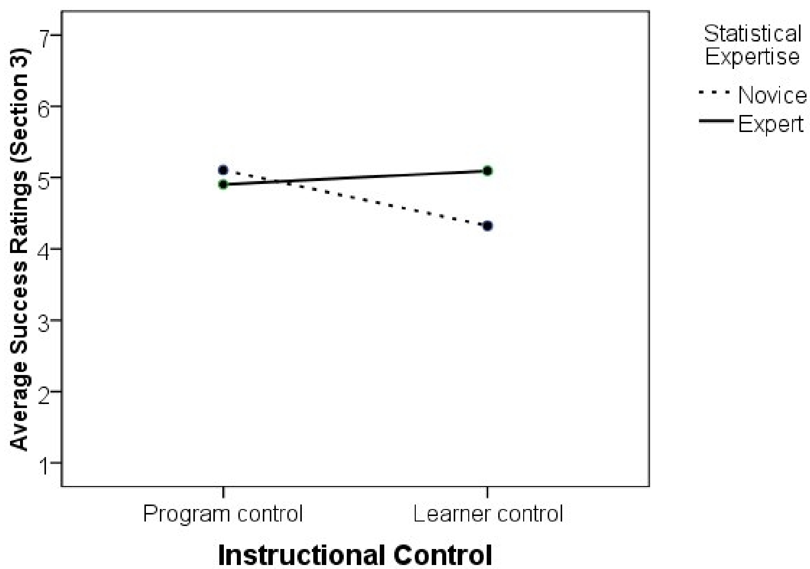


Figure 3.7. Average Section 3 Success ratings as a function of instructional control and statistical expertise.

Time spent on the tutorial was examined in an ANOVA for the different groups based upon instructional control and statistical expertise. Overall participants spent an average of 17.5 min (SD = 12.5, median = 13.1) on the tutorial. The instructional control groups differed significantly on how much time they spent on the tutorial, $F(1, 197) = 13.76$, $p < .001$. PC participants spent an average of 20.9 min (SD = 14.8, median 17.1); LC participants spent an average of 14.1 min (SD = 8.6, median = 11.3). In addition, novices spent more minutes on the tutorial (M = 20.8, SD = 15.5, median = 15.3) than did experts (M = 15.1, SD = 9.1, median = 11.7), $F(1, 197) = 8.35$, $p < .01$. The interaction between instructional control and statistical was not significant, $F(1, 197) = .13$, $p = .72$ (see Figure 3.8).

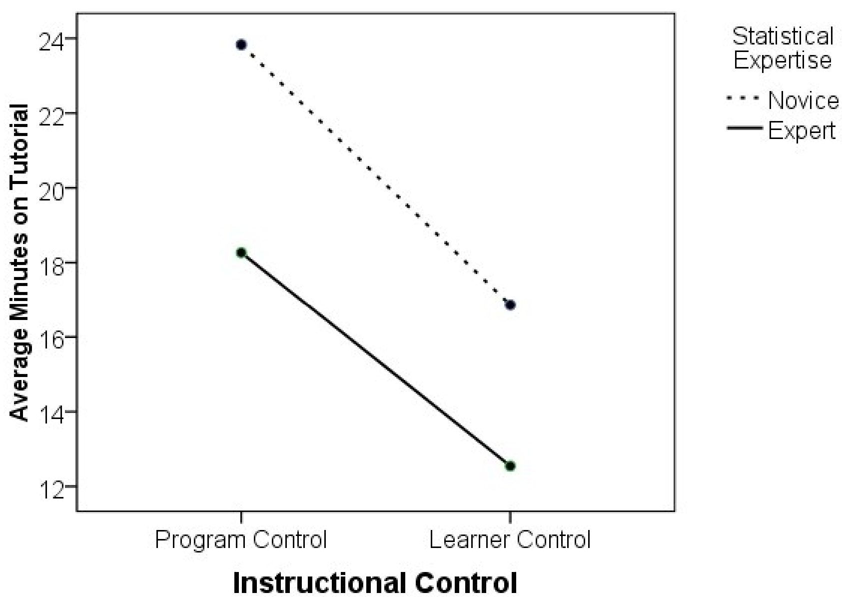


Figure 3.8. Average number of minutes spent on the tutorial as a function of instructional control and statistical expertise.

The number of *SD* and squared deviation histogram pop-up windows used was also compared for the different groups, with instructional control and statistical expertise as between-subjects factors in an ANOVA. Overall the average number of pop-ups selected was $M = 8.80$, $SD = 8.33$. To reduce the skew of this variable, it was log-transformed to meet the assumptions of ANOVA better. Neither main effect of instructional control or statistical expertise was significant, $F(1, 197) = 1.25$, $p = .27$; and $F(1, 197) = .53$, $p = .47$, respectively. Experts tended to use pop-ups equally across instructional conditions, while novices in the PC used more pop-ups ($M = 10.88$, $SD = 9.94$) than novices in the LC condition ($M = 7.14$, $SD = 6.33$), $t(83) = 2.00$, $p < .05$. (see Figure 3.9). However, the interaction between these factors failed to attain statistical significance, $F(1, 197) = 2.45$, $p = .12$.

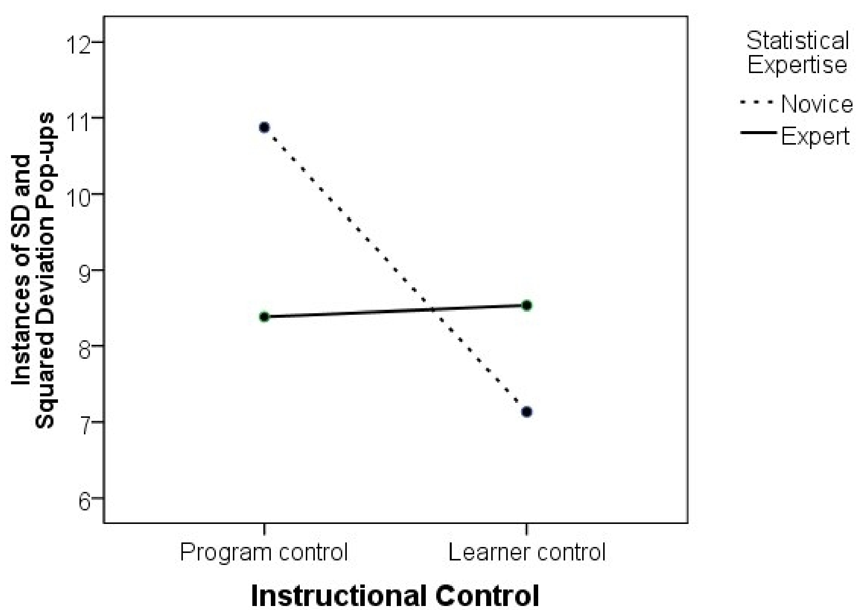


Figure 3.9. Average number of *SD* and squared deviation pop-ups as a function of instructional control and statistical expertise.

3.5 *Actual Performance and Accuracy of Self-Assessments*

To assess how accurate the different groups were in self-evaluating their performance during the tutorial, proportion of correct initial responses to questions on Sections 2 and 3 were compared to Success ratings. Actual performance on Sections 2 and 3 and differences between sections were first examined in a repeated-measures ANOVA with statistical expertise and instructional control as between-subjects factors, and section as the repeated-measure. As expected, participants did worse on the more conceptually difficult Section 3 (M = .69, SD = .24) than they did on Section 2 (M = .85, SD = .16), $F(1, 197) = 132.10$, $p < .001$. The section x instructional control interaction was significant, $F(1, 197) = 5.66$, $p < .05$. Differences between the sections in the PC condition were marginally significant, $t(199) = 1.97$, $p = .05$, but were significant in the LC condition, $t(199) = 2.21$, $p < .01$ (see Figure 3.10). Overall both main

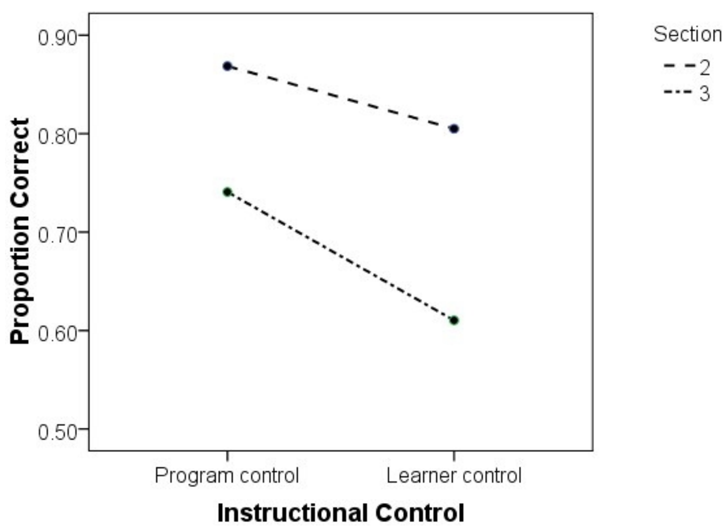


Figure 3.10. Proportion of initial responses on tutorial overall that were correct as a function of section and instructional control.

effects of instructional control and statistical expertise, and their interaction were significant across sections: $F(1, 197) = 16.87$, $p < .001$, $F(1, 197) = 11.57$, $p < .01$, and $F(1, 197) = 13.36$, $p < .001$, respectively.

Separate follow-up ANOVAs were conducted on proportion correct on each section. On Section 2, both the main effects of instructional control and statistical expertise were significant, as well as their interaction: $F(1, 197) = 8.16$, $p < .01$, $F(1, 197) = 14.17$, $p < .001$, and $F(1, 197) = 6.96$, $p < .01$, respectively. Both of these main effects were also significant on Section 3: $F(1, 197) = 16.78$, $p < .001$, $F(1, 197) = 5.83$, $p < .05$, and $F(1, 197) = 12.80$, $p < .001$, respectively. On Sections 2 and 3, learners in the PC condition did better (M =.87, SD = .14; and M = .74, SD = .19, respectively) than the learners in the LC condition (M =.82, SD = .18; and M = .64, SD = .26, respectively). In addition, experts outperformed novices on each section (M =.88, SD = .15 vs. M = .80, SD =.18, and M =.71, SD = .21 vs. M = .65, SD = .26, respectively).

However, these main effects are qualified by the significant interactions. There was no significant difference between novices and experts in the PC condition on Section 2 performance, $t(98) = .91$, $p = .36$, but there was a significant expertise difference in the LC condition, $t(99) = 4.05$, $p < .001$. Likewise on Section 3, there was no significant difference in the PC condition, $t(98) = .98$, $p = .33$, but there was in the LC condition, $t(99) = 3.71$, $p < .001$. Overall novices did worse than experts on each section, but these differences were mainly in the LC condition (see Figures 3.11 and 3.12).

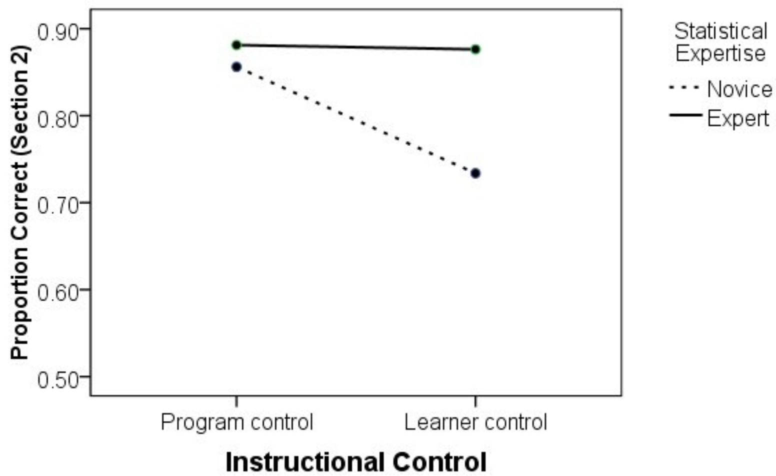


Figure 3.11. Proportion of initial responses on Section 2 that were correct as a function of instructional control and statistical expertise.

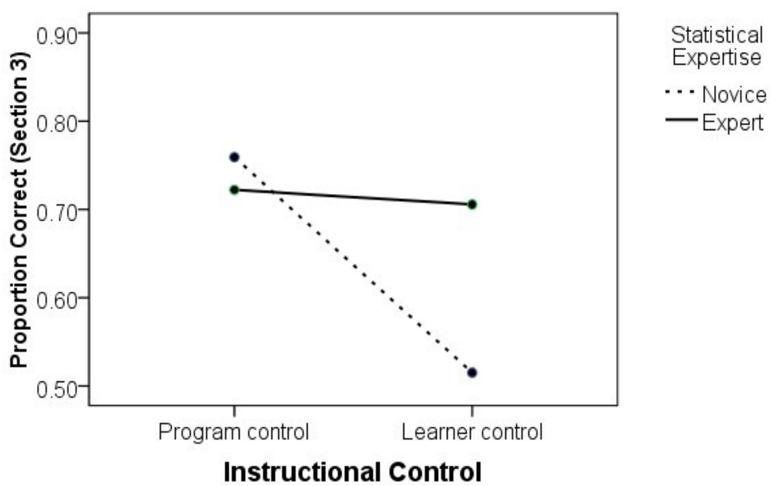


Figure 3.12. Proportion of initial responses on Section 3 that were correct as a function of instructional control and statistical expertise.

To evaluate if different learners were more accurate in their self-assessments on Sections 2 and 3, absolute deviation scores were calculated for each participant and on each section. Absolute deviation scores were computed in the following manner: (1) to convert the Success ratings to a scale comparable to proportion correct (0 to 1), these 1-7 ratings were subtracted by one and then divided by six; (2) deviation scores were calculated by subtracting proportion correct from the converted Success ratings; and (3) the absolute value of these deviation scores was computed.

These absolute deviation scores on each section were examined in a repeated-measures ANOVA with statistical expertise and instructional control as between-subjects factors, and section as the repeated-measure. There was an marginally significant three-way interaction between the variables, $F(1, 197) = 3.72$, $p = .055$; and significant effect of section, $F(1, 197) = 8.47$, $p < .01$. Overall participants were less accurate at self-assessing on Section 2 ($M = .21$, $SD = .19$) than on Section 3 ($M = .17$, $SD = .16$). Section did not interact significantly with either statistical expertise or instructional control, $F(1, 197) = .10$, $p = .75$; and $F(1, 197) = .00$, $p = .97$, respectively. Overall across sections, the main effect of statistical expertise was significant, $F(1, 197) = 6.83$, $p < .05$, $d = .36$. In self-assessing their performance on both tutorial sections overall, novices were less accurate ($M = .22$, $SD = .15$) than experts ($M = .17$, $SD = .13$). Neither the main effect of instructional control and their interaction were significant, $F(1, 197) = .01$, $p < .94$, and $F(1, 197) = 1.07$, $p =.30$, respectively.

Two separate ANOVAs were conducted on the absolute deviation scores from the two different sections, using statistical expertise and instructional control as between-subjects factors. On Section 2, there was a significant effect of statistical expertise, $F(1, 197) = 4.73$, $p < .05$, a marginally significant interaction between statistical expertise and instructional control, $F(1, 197) = 3.33$, $p = .07$, and no significant effect of instructional condition, $F(1, 197) = .01$, $p = .94$. Novices were less accurate than experts on Section 2 ($M = .24$, $SD = .21$ vs. $M = .18$, $SD = .17$). On Section 3, there was also a significant effect of statistical expertise, $F(1, 197) = 4.50$, $p < .05$, but no significant effect of instructional condition, $F(1, 197) = .00$, $p = .97$, nor significant interaction, $F(1, 197) = .08$, $p =$ 78. Novices were also less accurate than experts on Section 3 ($M = .20$, $SD = .18$ vs. $M = .15$, $SD = .14$). Thus, experts tended to be more accurate (i.e., have lower absolute deviation scores) than novices on both Sections 2 and 3 (see Figures 3.13 and 3.14).

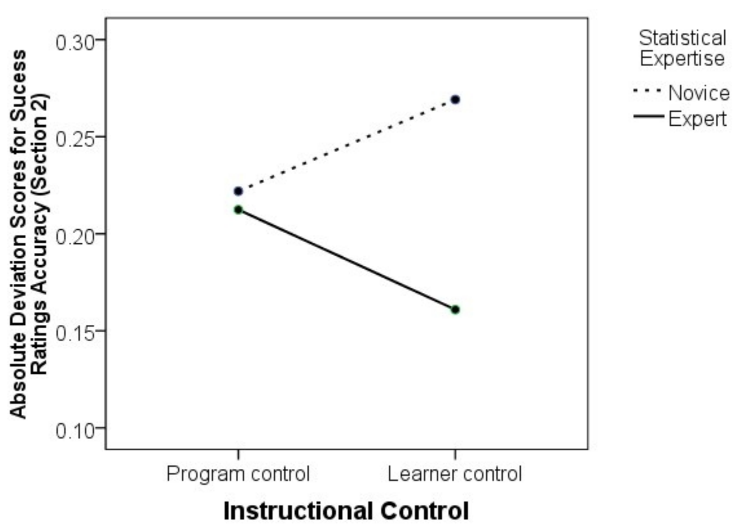


Figure 3.13. Absolute deviation scores on Section 2, as a function of instructional control and statistical expertise. These scores reflect accuracy of Success Ratings.

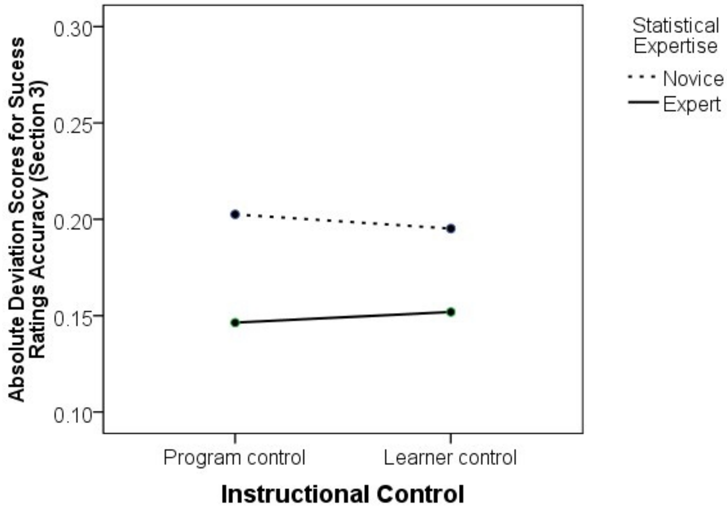


Figure 3.14. Absolute deviation scores on Section 3, as a function of instructional control and statistical expertise. These scores reflect accuracy of Success Ratings.

3.6 *What Did Participants Learn and Not Learn?*

To examine which concepts participants learned most from using the tutorial and which concepts remained difficult to understand, performance on individual items on the pre-test and post-test was examined. Table 3.8 presents the overall percent correct for these items and related concepts, on the pre-test and post-test for the five questions involving comparing histograms in sets (H1-H5, see Appendix E) and the ten involving comparison of pairs (P1-P10, see Appendix F). The five items that showed the most improvement were Items H5, P2, P3, P7, and P9. These items illustrate the effects on the *SD* when changing either the shape or range of a distribution while keeping the other dimension constant, except on H5 which involved comparing multiple histograms. On Items H5, P2,

Table 3.8

Percent Correct and Percent Changes from Pre-test to Post-Test by Item and Concept

Concept and Item	Pre-test	Post-test	Change
Interpreting histograms			
Item H1	70.6	73.1	+2.5
Item H2	57.7	61.2	+3.5
Item H3	64.7	72.1	+7.4
Identify smallest/largest *SD* of 5 histograms			
Item H4	54.2	68.7	+14.5
Item H5	**50.7**	**68.7**	**+18.0**
Location shift preserves *SD:* Item P1	81.6	83.6	+2.0
More U-shaped distribution has larger *SD* when keeping range constant			
Item P2	**62.2**	**78.1**	**+15.9**
Item P3	**63.7**	**84.6**	**+20.9**
More normal-shaped distribution has lower *SD* when keeping range constant: Item P4	41.6	43.6	+5.0
Larger range but same shape has larger *SD*: Item P5	78.6	87.6	+9.0
Mirror image preserves *SD*: Item P6	67.7	79.1	+11.4
Same shape and range, but with more scores farther from the sample mean has larger *SD*			
Item P7	**59.7**	**74.6**	**+14.9**
Item P9	**51.7**	**76.6**	**+24.9**
Normal distribution may have larger *SD* due to larger range			
Item P8	*45.3*	*28.9*	*-16.4*
Item P10	*39.3*	*35.3*	*-4.0*

Note. Top five positive percent-change items are bolded and negative percent-change items are italicized. Problems involving sets of histograms are preceded by “H” (see Appendix E) and those involving pairs are preceded by “P” (see Appendix F).

SD based upon its U-shape distribution. On Items P7 and P9, participants had to recognize that when the range is the same and shape is similar across pairs of histograms, the larger *SD* is present in the histogram with more scores farther from the sample mean.

As delMas and Liu (2005) also found, Items P8 and P10 were the most difficult for participants. On these items, participants were to integrate information from both the range and shape to make decisions about the *SD*; although normal distributions tend to have smaller *SD*'s than less normal distributions when the range is kept constant, the normal distributions in Items P8 and P10 have larger *SD*'s due to their larger ranges. The fact that participants were already sensitive to the effects of range on *SD* can be demonstrated by their relatively high pre-test and post-test performance on Item P5, where the shapes of the distributions are the same but one has been stretched out to occupy a larger range. Thus, on Items P8 and P10, the range is overlooked and the shape becomes the dominating factor in deciding which histogram has a larger *SD*.

3.7 *Supplemental Analyses Regarding Learner-Control Participants*

In the LC condition, participants could choose to view more detailed, explanative feedback on why they answered a question correctly or incorrectly. Novices tended to view more instances of this feedback (M = 2.30, SD = 3.33) than did experts (M = 1.50, SD = 2.12), but this difference was not significant, $t(99) = 1.47$, $p = .15$. The LC participants could also choose to skip or complete the optional review questions at the end of Sections 2 and 3. Tables 3.9 and 3.10

present the frequencies of the LC participants who skipped or completed these questions. On Sections 2 and 3, 51% and 49% of the novice completed the review questions, respectively. In contrast, 61% and 48% of experts did so, respectively. However, the chi-square tests of independence for both Sections 2 and 3 were not significant, χ^2 (1) =.88 , p = .35; χ^2 (1) = .00, p = .98, respectively. That is, statistical expertise is not reliably related to the frequency of doing the review questions. Thus, there is no evidence that novices sought out more information by doing these review questions than did experts.

Table 3.9

Frequencies of Learner-Control Participants Who Did Section 2 Review Questions as a Function of Statistical Expertise

	Statistical Expertise		
	Novice	Expert	Total
Skipped	18	25	43
Completed	19	39	58
Total	37	64	101

Table 3.10

Frequencies of Learner-Control Participants Who Did Section 3 Review Questions as a Function of Statistical Expertise

	Statistical Expertise		
	Novice	Expert	Total
Skipped	19	33	52
Completed	18	31	49
Total	37	64	101

Chapter 4: Discussion

4.1 *Summary of Findings and Implications*

This study provides evidence supporting the use of computer based tutorials as effective tools for teaching statistical concepts, specifically the concept of standard deviation. Furthermore, the findings from this study contribute to the body of research on learning theories and the use of technology in learning on ways to support conceptual learning for students with different levels of expertise. The computer-based tutorial was effective overall in teaching standard deviations, with an effect size of $d = .42$, reflecting a “medium” effect (Cohen, 1988) and surpassing Slavin’s (1990) threshold of .25 for an effect to be considered practical in educational settings. An effect size this large is impressive given the tutorial’s short duration, on average of 17.5 min, and the fact that the post-test items were more complex than those presented within the tutorial.

Although participants who completed the tutorial showed increased knowledge of standard deviation on the post-test, these learning gains were moderated by the learners’ statistical expertise and how much control they had over viewing feedback and doing integrative review questions on different sections. Indeed, statistical experts, who had completed one or more statistics courses, demonstrated different experiences and learning outcomes using the tutorial than did the novices, who were mostly completing their first statistics course. These experts had already learned about standard deviation in an introductory statistics course and were no doubt exposed to the idea of sum of squared deviations in some of the more advanced statistics courses. They also rated themselves higher on several learner characteristics (self-regulation of learning and task value) before beginning the tutorial.

The following section describes findings regarding the main hypotheses of the current study and discusses implications of the results:

H1. Greater task value and self-efficacy will be associated with greater learning.

When ignoring other factors, both task value (placing importance on learning about standard deviation) and self-efficacy (belief that the learner will be successful in learning about standard deviation) were significantly and positively related to learning. However, when controlling for other learner characteristic factors, including statistical expertise, neither task value nor self-efficacy was uniquely associated with learning.

It is noteworthy that self-regulation of learning is significantly and positively correlated with self-efficacy and task value, and that, compared to novices, experts rated themselves higher on all three of these learner characteristics. This difference between novices and experts may help explain why self-regulation of learning has only a small unique contribution in predicting learning, while statistical expertise may be a better predictor of learning.

H2. Self-regulated learning will have an even larger positive effect than task value and self-efficacy on learning.

Ignoring other factors, self-regulated learning was positively and significantly related to learning, but not as strongly as the motivational factors, task value and self-efficacy. Controlling for other factors, self-regulated learning was not significantly related to learning and had an even smaller impact on learning than did task value and self-efficacy. In fact, as a set of predictors, these three self-reported learner-characteristics did not add value to predicting learning outcomes. It may not be surprising, given their relationship with statistical expertise, that when statistical expertise is included in predicting learning, the effects of these learner characteristics on learning are greatly diminished. Perhaps the failure of self-regulated learning to emerge as a stronger predictor than self-efficacy and task value is because self-regulated learning was assessed as a more global and multi-faceted concept (which learners may find difficult to self-assess), while task value and self-efficacy were easier to self-assess and more directly related to the online tutorial at hand.

H3. Learning will be greater for the PC condition than the LC condition.

Learners demonstrated more learning in the PC condition than in the LC condition, but when controlling for other factors, including statistical expertise and cognitive load experienced during learning, the effect of instructional control on learning was eliminated. This finding indicates that the impact of learner control on improving knowledge about standard deviation may be affected by learner characteristics and experiences while using the tutorial. Perhaps these sources of variability can account for the inconsistent effects of learner control on learning expressed by other researchers (e.g., Lunts, 2002; Niemiec et al., 1996).

H4. Higher levels of cognitive load will be related to less learning.

As predicted, perceived cognitive load was negatively and significantly associated with learning, whether other factors were ignored or controlled. Even for learners of similar profiles (expertise and learner characteristics), cognitive load scores that are higher are associated with a lower score on learning. Adding cognitive load as a factor reliably improves predicting learning outcomes beyond all the other factors, including learner characteristics and instructional control.

H5a. The benefits of PC instruction compared to LC instruction will be greater for novice learners than for expert learners.

The significant interaction between instructional control and statistical expertise suggests that instructional control differentially affects novice and more expert learners. Follow-up analyses examining this interaction illustrate that, when controlling for pre-test scores, adjusted post-test scores for the novices were better in the PC instructional condition than in the LC instructional condition. On the other hand, experts demonstrated comparable learning using either LC instruction or PC instruction. Therefore, when compared to LC instruction, PC instruction enhanced learning more so for novices than for experts. Self-reported ratings on difficulty and frustration suggest that novices in the LC condition may have experienced cognitive overload that negatively influenced their learning.

H5b. The benefits of PC instruction compared to LC instruction will be greater for low self-regulating learners than for high self-regulating learners.

In contrast to the findings regarding statistical expertise, the non-significant interaction between instructional control and self-regulated learning suggests that the effect of instructional control on learning does not differ substantially between those classified as low versus high self-regulators based upon individuals' self-ratings of their self-regulation of learning behaviors.

This finding may not necessarily lead to the conclusion that self-regulation of learning abilities has no bearing on how the learner interacts with the learning environment to influence knowledge acquisition processes. Rather, it may reflect weakness in the measure of self-regulation and the relative difficulty of making accurate judgments about self-regulatory behaviors over time. In contrast, it is easier for learners to report accurately a more objective, specific number regarding statistical courses they have taken, which reflects statistical expertise. In addition, although both statistical expertise and self-regulation of learning measures are global, statistical expertise is more directly related to the tutorial in terms of content and knowledge of concepts related to standard deviations. The measures of self-regulation of learning were designed to capture relevant aspects of self-regulated learning that might take place on an interactive online tutorial that allowed for self-assessment opportunities and seeking additional information. Yet individuals may not have had enough experiences with online learning to make accurate subjective self-judgments of behaviors applicable to such environments (Joo et al., 2000; McManus, 2000), and aspects of self-regulated learning more influential to online learning may not have been measured by the items used.

Thus, expert learners who had completed one or more statistical courses did not suffer from having more learner control over their learning process (i.e., viewing feedback and selecting to do review questions or not) on the computer-based statistics tutorial, whereas novice learners *did* suffer and instead demonstrated better learning with PC instruction than with LC instruction. Learners who experienced higher cognitive load, as expressed by how difficult and frustrating a tutorial section was to the learner, demonstrated impaired learning. Novice learners in the LC condition had the highest perceived cognitive load ratings on both sections. However, certain concepts remained elusive for participants even after completing the tutorial. Even though the "expert" learners did better than the novice learners throughout the tutorial, they still did not fully master the items on the post-test, especially those items that required integrating information about both the shape and range of the distribution.

Still, statistical expertise moderated the effects of learner control on overall knowledge acquisition. Participants with more statistical expertise seemed to be more motivated in some regards; they reported that they valued learning about standard deviation more than participants who had less statistical expertise. Experts also reported demonstrating more self-regulation of learning behaviors. Self-efficacy, task value, and self-regulation of learning were all positively and significantly correlated with learning. Yet after controlling for each other, statistical expertise alone significantly predicted better learning outcomes (for the more expert learner) and moderated the effects of instructional control on cognitive load and learning.

Interactions between statistical expertise and instructional control were found throughout the tutorial in terms of performance and self-reported ratings regarding cognitive and motivational processes. There was only a hint of an expertise reversal effect. Compared to experts in the PC condition, experts in the LC version tended to do slightly better on the post-test, reported the tutorial to be slightly less frustrating and difficult, and reported feeling slightly more successful on each section. Although these differences did not attain statistical significance, the data on all of these measures was in the expected direction. These trends suggest that experts in the PC instruction may have experienced some cognitive constraints or reduced motivation by being exposed to unnecessary or unwanted feedback, review problems, or both.

Performance on the post-test revealed that even after completing the tutorial, the majority of learners still had difficulty integrating range and shape in making comparisons of variability, although the learners could more easily judge how changing just one of these dimensions affects standard deviation. This deficit in statistical understanding is also reflected by the relatively worse performance on the conceptually more difficult Section 3 than on Section 2. Both novices and experts did consistently worse on Section 3 than they did on Section 2. Section 2 dealt with easier principles such as the fact that histograms have the same standard deviation if they are mirror images. In contrast, Section 3 dealt with more difficult ideas, such as a histogram that had a smaller range could have the same or bigger standard deviation than a histogram with a larger range but a more normal shape. Thus, Section 3 presented histograms that represented more complex distributions that required integrating information about both the shape and range of distributions to make judgments about standard deviation.

In general, participants demonstrated over-reliance on shape in making standard deviation judgments, even on the post-test. Perhaps the stated goals in the overview of the tutorial should have also mentioned how “range” and not only “shape” affects the standard deviation. In addition, rather than just presenting a series of histogram pairs, perhaps more dynamic and interactive representations of histograms/distributions differing in standard deviations, such as by delMas and Liu (2005; 2007), would be useful in helping learners integrate information about different dimensions in making comparisons of variability.

Differences between novice and expert learners tended to diminish on Section 3 compared to Section 2, possibly reflecting experience with learning about standard deviation. Especially in the LC condition, experts tended to be more accurate in self-assessing their performance on a tutorial section than novices; however, this difference was reduced on Section 3. Presumably, with more experience with learning about statistics and a particular topic, novices can improve self-assessment of their performance. On Section 2, compared to novices, experts were more accurate in self-evaluating their performance, especially in the LC condition than the PC condition (although the interaction between statistical expertise and instructional control was only marginally significant, $p = .07$). Superior accuracy of experts in the LC condition may reflect that having more control over their learning helps the experts to be more self-reflective of their learning in terms of choosing to view explanative feedback or not for each response they gave. More accurate self-assessment of their own performance by experts compared to novices in the LC condition may also be due to the fact that the experts did more of the optional review questions on Section 2 than did the novices.

Cognitive load was measured by how difficult and frustrating a tutorial section was to a learner. Cognitive load ratings were negatively related to learning, controlling for other factors including pre-test scores, instructional control, time spent on the tutorial, statistical expertise, and other learner characteristics. Supporting Cognitive Load Theory, the novice learners in the LC condition, reported the highest levels of cognitive load and demonstrated the least learning; in contrast, expert learners learned equally well in either instructional control condition and tended to report equal amounts of cognitive load. These self-ratings were originally conceptualized to be measures of underlying cognitive resources allocated for the learning tasks, yet they can also been seen as measuring motivational aspects. When a task is frustrating or too difficult, a learner may become disengaged and discouraged. In fact, cognitive load ratings, based on Difficulty and Frustration on both Sections 2 and 3, were negatively associated with the motivational learner characteristics of self-efficacy and task value that were self-reported before the beginning the tutorial. That is, the more a learner reported being confident in doing well on the tutorial and the more they valued learning about the standard deviation, the less frustration and difficulty they reported while completing the tutorial. At the same time, compared to novices, experts reported higher task value and less frustration and difficulty overall.

Yet there is reason to believe that these cognitive load measures are not purely motivational or affective, but also reflect underlying cognitive processes. Novices in the LC condition did not differ significantly from novices in the PC condition on task value or how important it was to learn about standard deviation when beginning the tutorial. Across instructional control conditions, novices and experts did not differ on their self-reported Effort ratings on both Sections 2 and 3. In terms of tutorial time, novices took longer than experts overall and in the LC version, undermining the notion that novice learners were less motivated than experts in the LC condition. Additionally, in the LC condition, novices viewed instances of optional explanative text feedback in a way that was comparable to experts'.

The differential use of scaffolds may help explain the differences in cognitive load experienced by learners. Compared to novices in the LC condition, novices in the PC condition accessed more histograms that visually depicted the squared deviations of individual observations and calculations of *SD*s. Thus, novices, in the LC condition, not using these visual scaffolds may have lacked the supports to integrate new knowledge with their existing schemata, contributing to extraneous cognitive load. In contrast, across instructional conditions, experts used more equal instances of this visual scaffolding, which may have reduced cognitive load when coupled with their superior knowledge base, leading to better learning. The histograms may also be more helpful than explanative text feedback. These superiority of histograms may be due to their visual nature and ability to guide learners toward the correct responses, not just provide feedback only after responses are made. The possibly less useful and more cognitively demanding explanative text feedback may have put the novices at a greater disadvantage for learning. In the PC condition, novices and experts were presented this text feedback automatically for each of their responses. However, novices in the PC condition used the most histogram scaffolds out of the four groups, possibly enhancing their learning more so than novices in the LC condition who may have been cognitively overloaded by having to decide when to view the explanative text feedback as well as to choose when to view these histograms.

To assess efficiency of learning, time spent on instruction must be considered along with knowledge acquired. Considering this criterion, expert learners in the LC condition were the most efficient group. They showed learning comparable to experts in the PC condition but spent only about two thirds as much time on average to complete the tutorial ($M = 12.5$, $SD = 6.5$ vs. $M = 18.3$, $SD = 10.9$).

4.2 *Limitations and Future Research*

Garfield (2002) argued that for students to fully understand sampling, they need a variety of discovery learning activities and explicit instruction, including text or verbal explanations, concrete activities involving sampling, and interactions with simulated populations and sampling distributions. She concluded that teaching specific training rules is not adequate. On the other hand, supporting the view that teaching formal rules about reasoning can enhance learning, Fong, Krantz, and Nisbett (1986) improved the frequency and quality of statistical reasoning in sample size judgments concerning the law of large numbers (i.e., bigger samples result in more accurate sample means) with two different interventions that lasted as short as a half hour.

The current study seems to be in the middle in terms of teaching formal rules versus having more experiential experiences in teaching statistical concepts, giving validity to both viewpoints. The standard deviation tutorial in the current study was focused in duration and scope, and significantly increased the understanding of variability. While not teaching rules explicitly, the tutorial used interactive questions and a constructivist approach with structured examples to teach underlying principles regarding what affects variability. Such an approach resulted in overall yet inconsistent improvements in the statistical knowledge, suggesting a need for more expansive instruction as suggested by Garfield (2002). Yet even with more instruction, statistical misconceptions may form (Hodgson & Burke, 2000). This underscores the need for assessment and for considering both prior knowledge and new knowledge as it develops to ensure that incomplete understanding and misconceptions are effectively addressed by instruction.

This study represents one step in advancing statistics education research by examining the effects of scaffolding, cognitive processes, motivation, and expertise on learning, and also by illuminating aspects of computer-based instruction that may enhance statistical understanding. At the same time, some issues, particularly measuring cognitive load and the role of self-regulation of learning, remain unresolved. Future research should examine in more detail the self-regulation not only of cognitive processes but also of motivational processes in learning (Pintrich, 2004); however, measuring these constructs remains challenging. For instance, self-reported measures of self-regulated learning do not necessarily provide an accurate picture of actual self-regulatory behaviors (Puustinen & Pulkkinen, 2001). On the other hand, there is evidence that students can indeed accurately judge and report their learning behaviors. Their self-reported use of self-regulation of learning behaviors was highly correlated with teachers' ratings of students' self-regulation of learning behaviors, $r = .70$ (Zimmerman & Martinez-Pons, 1988).

In the current study, all three learner characteristics measured before the tutorial (self-regulation of learning, self-efficacy, and task) were positively related to learner outcomes when ignoring other factors. When statistical expertise and cognitive load are added as factors, these three learner characteristic factors did not provide statistically significant unique contributions to predicting learning, which is not surprising given their positive relationship with expertise. Yet self-regulation of learning, which was expected to have the biggest impact on learning outcomes, especially may be difficult to assess due to its more global and multi-dimensional nature. Using measures of actual behaviors reflecting self-regulatory practices, rather than or in addition to self-reported measures might be more useful to designing effective instruction, as noted by McManus (2000):

> Finding a way of assessing the actual use of self-regulated learning strategies within an environment through pattern analysis would be more effective for automatically individualizing instruction than self-report measures such as the MSLQ. (p. 248)

Self-regulation of learning and self-directed learning, as described by Song and Hill (2007) and Garrison (2003), share many similarities. In both frameworks, learners who are autonomous and self-motivated are predicted to more effectively use resources to enhance their understanding. Both frameworks postulate a cyclical process involving planning, monitoring, and evaluating the learning process. Both are learner-centered and emphasize the contributions of motivational and cognitive processes in impacting learning outcomes. The research on self-regulated learning may benefit from research done on self-directed learning. However, as Boekaerts (1999) recognized, self-regulation of learning, a multi-faceted construct, still presents challenges:

> The problem with a complex construct such as self-regulated learning (SRL) is that it is positioned at the junction of many different research fields, each with its own history. This implies that researchers from widely different research traditions have conceptualized SRL in their own way, using different terms and labels for similar facets of the construct. (p. 447)

Similarly, the measurement of cognitive load remains a challenge. Cognitive Load Theory posits there are different types of cognitive load (Sweller, van Merriënboer, & Paas, 1998). Intrinsic load and germane load are, respectively, neutral and beneficial to learning for experts. In contrast, intrinsic load and germane load can become extraneous load for novices who lack the capacity to process materials as efficiently as experts, interfering with their learning (Kalyuga, 2007). This makes it challenging to measure these different types of cognitive load for experts and novices using similar items and to manipulate the amount of cognitive load experienced differentially by both groups. In the current study, the cognitive load measures, reflected by Frustration and Difficulty ratings, seemed to represent extraneous load as they were negatively related to learning. Less clear is whether Effort ratings, which did not differ either by expertise or by instructional control, represented germane or intrinsic load, or some combination of the two. Effort ratings were positively correlated with self-regulation of learning, but this construct was not necessarily associated with better learning, or thus, higher germane processing. Intrinsic load reflects the inherent difficulty of the material and should vary according to the learner's expertise, but experts and novices did not differ on Effort ratings. Previously studies have used self-reported Effort ratings to reflect either germane load (e.g., Gerjets et al., 2009) or intrinsic load (e.g., DeLeeuw & Mayer, 2008), without strong support for either classification.

Measurement of statistical understanding is yet another challenge. Assessment is an integral component of statistics instruction (Franklin & Garfield, 2006; Garfield & delMas, 2010; Garfield et al., 2011). Traditionally in both educational and research settings, assessments are administered at the end of an instructional phase and have no bearing on the next instructional phase. The utility of assessments, however, may be increased when the assessments are used to guide the types of explanations and feedback to be presented during future instruction. As with pretraining or advance organizers, they may be especially useful even before instruction begins or even during online instruction to help learners self-assess their performance and make decisions about subsequent learning activities.

Garfield and Ben-Zvi (2008) declared that understanding the concept of statistical variability is much more difficult and complex than previous literature would suggest. Whereas traditional assessments of understanding of variability focus on calculations and simple interpretations of standard deviation, inter-quartile range, and range, Garfield and Ben-Zvi suggested assessing deeper conceptual understanding of variability by having students perform tasks such as interpreting summary measures and drawing and comparing graphs. In addition, Garfield and Ben-Zvi (2005; 2008) differentiated between statistical literacy (knowledge of the basic language and tools of statistics), statistical reasoning (making sense of statistical information and making conceptual connections), and statistical thinking (a higher order of reasoning that mimics experts' reasoning, including understanding of theoretical underpinnings). These different aspects of statistical cognitive abilities capture a range of associated skills and underlying knowledge that is often not assessed by traditional methods. Garfield and Ben-Zvi (2007) also observed that although there is some evidence for the effectiveness of certain types of training, there is less support for long-term retention and transfer of knowledge due to these training interventions. In measuring an aspect of transfer of learning, the current study used assessment items that were intended to measure understanding of specific statistical concepts in more complex problems than presented during instruction. By examining performance on these assessment items, it was easier to discern that even after completing the tutorial, learners still had issues with integrating information about the range and shape of distributions in making judgments about variability.

Teachers in classrooms seldom assess their students' motivation, specify concrete learning goals, or teach learning strategies, all of which would enhance students' ability to self-regulate their own learning (Zimmerman, 2002). Computer-based instruction introduces both advantages and disadvantages over traditional instruction. Dynamic and interactive representations presented in computer-based instruction, not otherwise possible with traditional instruction, can enhance learning (Larreamendy-Joerns & Leinhardt, 2006). Yet without a human instructor continuously monitoring the learner, it is more challenging to assess motivation and knowledge in online settings, and creative ways of doing so are required to build effective computer-based instructional tools. Cognitive engagement in distance education courses can be a critical component of learning (Bernard et al., 2009). More generally, both cognitive and motivational processes, as well as self-regulation of learning strategies, play roles in computer-based learning. Future research should examine how best to promote these processes and strategies that are conducive to learning in both traditional and computer-based settings.

Even though learner characteristics, including self-efficacy and self-regulation, are important aspects of online learning research, the features of the learning interface matter as well (Swan, 2004). There will always be the need to scaffold online instruction and provide instructional support (Artino & Stephens, 2009b) as well as to give learners control over their learning, although how much control and what kind of control are debatable (Chung & Reigeluth, 1992). Future research should further elucidate what kinds of scaffolds and learner control should be implemented to optimize learning and to minimize extraneous cognitive load.

Other research has shown that with continuing and extensive instruction, individuals can develop their domain expertise to the extent that certain scaffolds that were once helpful may later hamper learners as the learners become more proficient, reflecting the expertise reversal effect (Kalyuga, 2007). The current study did not replicate the expertise reversal effect by demonstrating that experts are at a large learning disadvantage using PC instruction. The current study did, however, indicate a non-significant trend for experts using LC instruction to experience less cognitive load and more effective and efficient learning compared to experts using PC instruction.

In addition, experience with online learning could possibly improve the accuracy of self-assessment. The development of expertise and the relationship between expertise and different instructional features, such as scaffolds and learner control, should be further investigated. In addition, for researchers to clarify which learning processes, including self-regulation practices, contribute to better learning outcomes, Lajoie (2008) recommended investigating how experts in a given domain go about learning. Instructional design could then be used to encourage these self-regulated learning practices, especially important in online instructional settings (Artino, 2008). Lajoie's recommendation highlights the influence of domain in determining what constitutes beneficial and effective self-regulatory practices; some learning practices may be more helpful in some domains or contexts than others.

4.3 *Concluding Remarks*

From 2002 to 2008, the use of online instruction has grown substantially, far exceeding the growth of total enrollment at higher education institutions and mostly at the undergraduate level (Allen & Seaman, 2010). The current study has implications especially relevant for designing effective hypermedia, computer-based instruction that will be an essential part of online learning. It shows that the benefits of learner control depend upon learner characteristics, including prior experiences. Furthermore, it sheds light on the effects of expertise and cognitive load on learning, which points to important avenues for future research (Kostons et al., 2009). These findings have implications for the design and implementation of computer-based instruction in general education as well as statistics education; for instance, the findings highlight the need to be thoughtful about how much learner control should be given and the factors this decision depends upon, including the expertise of learners. The findings direct researchers and designers where to focus their resources, thereby optimizing benefits while reducing costs. Especially important is promoting self-regulation of learning and self-evaluation when learners are using computer-based instruction (Vovides et al., 2007). Computer-based instruction also needs to address cognitive demands placed on learners and provide different kinds of scaffolding to experts and novices (Lambert, Kalyuga, & Capan, 2009).

In *Statistics Education Research Journal*'s special issue on reasoning about statistical distributions, Pfannkuch and Reading (2006) identified four themes across the five articles in the issue: (1) educational research is becoming more cognitive based, (2) research is generating meaningful qualitative data, (3) qualitative data provides rich information, and (4) statistical variation is a key concept closely linked to understanding data distributions. These themes are consistent with the issues raised in this study. Research relying on constructivist approaches to learning is cognitive-based as it advocates accounting for students' existing knowledge and cognitive load, as well as considering the best external representations that may be presented. Examining how instructional practices differentially impact procedural and conceptual knowledge can help guide instructional development. An integrative approach is warranted in assessing students' statistical knowledge and, in particular, understanding of variability and distributions. Such an approach recommends a variety of assessment techniques not just for collecting pre-test and post-test data but also for delivering optimized instruction and helping learners self-assess their own understanding and choose appropriate follow-up learning tasks. The effectiveness of computer-based instruction can be enhanced with appropriate consideration of prior knowledge, expertise, and ongoing cognitive and motivational processes that occur during learning and the self-regulation of learning.

References

Aberson, C. L., Berger, D. E., Healy, M. R., Kyle, D. J., & Romero, V. J. (2000). Evaluation of an interactive tutorial for teaching the central limit theorem. *Teaching of Psychology*, *27*, 289–291.

Aiken, L. S, & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage Publications.

Allen, I. E., & Seaman, J. (2010). Learning on demand: Online education in the United States, 2009. Retrieved from http://sloanconsortium.org/publications/survey/pdf/learningondemand.pdf.

Amadieu, F., Tricot, A., & Mariné, C. (2009). Prior knowledge in learning from a non-linear electronic document: Disorientation and coherence of the reading sequences. *Computers in Human Behavior*, *25*, 381–388.

Aquilonius, B. C. (2005). *How do college students reason about hypothesis testing in introductory statistics courses?* (Doctoral dissertation, University of California, Santa Barbara). Retrieved from ProQuest Digital Dissertations (ATT 3163105).

Artino, A. R. (2008). Motivational beliefs and perceptions of instructional quality: predicting satisfaction with online training. *Journal of Computer Assisted Learning*, *24*, 260–270.

Artino, A. R. & McCoach, D. B. (2008). Development and initial validation of the online learning value and self-efficacy scale. *Journal of Educational Computing Research*, *38*(3), 279-303.

Artino, A. R., & Stephens, J. M. (2009a). Academic motivation and self-regulation: A comparative analysis of undergraduate and graduate students learning online. *Internet and Higher Education*, *12*, 146–151.

Artino, A. R., & Stephens, J. M. (2009b). Beyond grades in online learning: Adaptive profiles of academic self-regulation among naval academy undergraduates. *Journal of Advanced Mathematics*, *20*(4), 568-601.

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

Ausubel, D . P. (1960). The use of advance organizers in learning and retention of meaningful material. *Journal of Educational Psychology*, *51*, 267-272.

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

Ausubel, D. P. (1978). In defense of advance organizers: A reply to the critics. *Review of Educational Research*, *48*, 251-257.

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist*, *40*, 199–209.

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition – Implications for the design of computer-based scaffolds. *Instructional Science*, *33*, 367-379.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review, 84*, 191–215.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, *28*(*2*), 117-148.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Barnard, L., Lan, W. Y., To, Y. M. Paton, V. O., & Lai, S. (2009). Measuring self-regulation in online and blended learning environments. *Internet and Higher Education*, *12*, 1–6.

Barnes, B. R., & Clawson, E. U. (1975). Do advance organizers facilitate learning? Recommendations for further research based on an analysis of 32 studies. *Review of Educational Research*, *45*, 637-659.

Bernard, M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, *79*, 1243–1289.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction*, *7*(2), 161–186.

Boekaerts, M. (1999). Self-regulated learning: where we are today. *International Journal of Educational Research*, *31*, 445-457.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn.* Washington, DC: National Academy Press.

Burke, P. A., Etnier, J. L., & Sullivan H. J. (1998). Navigational aids and learner control in hypermedia instructional programs. *Journal of Educational Computing Research, 18*(2), 183-196.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245-281.

Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods*, *4*(1), 113-117.

Chance, B., delMas, R. & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi and J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-323). Dodrecht, the Netherlands: Kluwer.

Chen, C., & Rada, R. (1996). Interacting with hypertext: A meta-analysis of experimental studies. *Human-Computer Interaction*, *11*, 125–156.

Chen, S. Y., Fan, J. P., & Macredie, R. D. (2006). Navigation in hypermedia learning systems: Experts vs. novices. *Computers in Human Behavior*, *22*, 251–266.

Chung, J., & Reigeluth, C. M. (1992). Instructional prescriptions for learner control. *Educational Technology, 32*(10), 14-20.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, *100*, 223–234.

delMas, R. C., Garfield, J., & Chance, B. L. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education, 7*(3). Retrieved from http://www.amstat.org/publications/jse/secure/v7n3/delmas.cfm.

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6(2)*, 28-58. Retrieved from http://www.stat.auckland.ac.nz/serj.

delMas, R., & Liu, Y. (2003). Exploring students' understanding of statistical variation. In C. Lee (Ed.), *Proceedings of the Third International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-3*). [CD-ROM, with video segments] Mount Pleasant, Michigan: Central Michigan University.

delMas, R. & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal, 4(*1), 55-82. Retrieved from http://www.stat.auckland.ac.nz/serj.

delMas, R. C. & Liu, Y. (2007). Students' conceptual understanding of the standard deviation. In M.C. Lovett and P. Shah (Eds.), *Thinking with data* (pp. 87-116). Lawrence Erlbaum, NY.

Dillon, A., & Gabbard, R. (1998). Hypermedia as an educational technology: A review of the quantitative research literature on learner comprehension, control, and style. *Review of Educational Research*, *68*, 322–349.

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychological Review*, *20*, 391-409.

Duncan, T. G., & McKeachie, W. J. (2005). The making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, *40*, 117-128.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18*, 253-292.

Franklin, C. A., & Garfield, J. B. (2006). The GAISE project: Developing statistics education guidelines for grades pre-K-12 and college courses. In G. F. Burrill (Ed.), *Thinking and reasoning with data and chance: Sixty-eighth annual yearbook of the National Council of Teachers of Mathematics* (pp. 345-375). Reston, VA: National Council of Teachers of Mathematics.

Garfield (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, *10*(3). Retrieved from www.amstat.org/publications/jse/v10n3/garfield.html.

Garfield, J. & Ahlgren, A. (1988). Difficulties in learning basic concepts in statistics: Implications for research. *Journal for Research in Mathematics Education,* 19, 44-63.

Garfield, J. B., & Ben-Zvi, D. (2005). Research on statistical literacy, reasoning, and thinking: issues, challenges, and implications. In D. Ben-Zvi and J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 397-409). Netherlands: Springer.

Garfield, J. B., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, *75*(3), 372-396.

Garfield, J.B. & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Netherlands: Springer.

Garfield, J. & delMas, R. (2010). A web site that provides resources for assessing students' statistical literacy, reasoning and thinking. *Teaching Statistics*, *32*(1), 2-7.

Garfield, J., Zieffler, A., Kaplan, D., Cobb, G. W., Chance, B., & Holcomb, J. P. (2011). Rethinking assessment of student learning learning in statistics courses. *The American Statistician*, *65*(1), 1-10.

Garrison, D. R. (2003). Cognitive presence for effective asynchronous online learning: the role of reflective inquiry, self-direction and metacognition. In J. Bourne & J. C. Moore (Eds.), *Elements of quality online education:, practice and direction* (pp. 47-58). Needham, MA: Sloan Center for Online Education.

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked-out examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction*, *16*, 104–121.

Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., & Eysink, T. H. S. (2009). Learning with hypermedia: The influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior*, *25*, 360–370.

Hannafin, J. D., & Sullivan, H. J. (1995). Learner control in full and lean CAI programs. *Educational Technology Research and Development*, *43*, 19-30.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society 50th annual meeting* (pp. 904-908). Santa Monica, CA: Human Factors & Ergonomics Society.

Hart, S. G., & Staveland, L. E. (1988). Development of 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: Elsevier.

Hartley, J., & Davies, I. K. (1976). Preinstructional strategies: The role of pretests, behavioral objectives, overviews, and advance organizers. *Review of Educational Research*, *46*, 239-265.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112.

Hodgson, T., & Burke, M. (2000). On simulation and the teaching of statistics. *Teaching Statistics*, *22*, 91-96.

Hsu, Y. (2003). *The effectiveness of computer-assisted instruction in statistics education: A meta-analysis* (Doctoral dissertation, University of Arizona, Tucson). Retrieved from ProQuest Digital Dissertations (AAT 3089963).

Joo, Y., Bong, M., & Choi, H. (2000). Self-efficacy for self-regulated learning, academic self-efficacy, and Internet self-efficacy in Web-based instruction. *Educational Technology Research and Development, 48*(2), 5–17.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). Judgment under uncertainty: Heuristics and biases. Cambridge, UK: Cambridge University Press.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19,* 509–539.

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

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

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

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

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology*, *31*, 187–194.

Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory & Cognition*, *34*(5), 959-972.

Kostons, D., Van Gog, T., & Paas, F. (2009). How do I do? Investigating effects of expertise and performance process records on self-assessment. *Applied Cognitive Psychology*, *23*, 1256-1265.

Kostons, D., Van Gog, T., & Paas, F. (2010). Self-assessment and task selection in learner-controlled instruction: Differences between effective and ineffective learners. *Computers & Education*, *54*, 932-940.

Kulik, C. C., & Kulik, J. A. (1986). Effectiveness of computer-based education in colleges. *AEDS Journal, 19*(1-2), 81-108.

Kulik, C. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior, 7*, 75-94.

Kulik, C. C., Kulik, J. A., & Cohen, P. A. (1980). Instructional technology and college teaching. *Teaching of Psychology, 7*(4), 199-205.

Larreamendy-Joerns, J., & Leinhardt, G. (2006). Going the distance with online education. *Review of Educational Research*, *76*, 567–605.

Larreamendy-Joerns, J., Leinhardt, G., & Corredor, J. (2005). Six online statistics courses: Examination and review. *American Statistician*, *59*, 240–251.

Lajoie, S. P. (2005). Extending the scaffolding metaphor. *Instructional Science*, *33*, 541-557.

Lajoie, S. P. (2008). Metacognition, self-regulation, and self-regulated learning: A rose by any other name? *Educational Psychology Review*, *20*, 469–475.

Lambert, J., Kalyuga, S., & Capan, L. A. (2009). Student perceptions and cognitive load: What can they tell us about e-learning Web 2.0 course design? *E–Learning*, *6*. Retrieved from http://www.wwwords.co.uk/ELEA.

Lane, D. M., & Tang, Z. (2000). Effectiveness of simulation training on transfer of statistical concepts. *Journal of Educational Computing Research*, *22*(4), 383-396.

Lane-Getaz, S. J. (2007). *Development and validation of a research-based assessment: Reasoning about p-values and statistical significance* (Doctoral dissertation, University of Minnesota, Minneapolis). Retrieved from ProQuest Digital Dissertations (ATT 3268997).

Lipson, K., Kokonis, S., & Francis, G. (2003). Investigation of students' experiences with a web-based computer simulation. *Proceedings of the 2003 IASE Satellite Conference on Statistics Education and the Internet.* Berlin. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/6/Lipson.pdf.

Lovett, M. C. & Greenhouse, J. B. (2000). Applying cognitive theory to statistics instruction. *American Statistician*, *54*(3), 196-206.

Luiten, J., Ames, W., & Ackerson, G. (1980). A meta-analysis of the effects of advance organizers on learning and retention. *American Educational Research Journal*, *17*, 211-218.

Lunts, E. (2002). What does the literature say about the effectiveness of learner control in computer-assisted instruction? *Electronic Journal for the Integration of Technology in Education*, 1(2). Retrieved from http://ejite.isu.edu.

Lynch, R. & Dempo, M. (2004). The relationship between self-regulation and online learning in a blended learning context. *International Review of Research in Open and Distance Learning*, *5*(2). Retrieved from http://www.irrodl.org.

Mayer, R. E. (1979a). Can advance organizers influence meaningful learning? *Review of Educational Research*, *49*, 371-383.

Mayer, R. E. (1979b). Twenty years of research on advance organizers: Assimilation theory is still the best predictor of results. *Instructional Science*, *8*, 133-167.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, 59, 14-19.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*, 43–52.

McManus, T. F. (2000). Individualizing instruction in a web-based hypermedia learning environment: nonlinearity, advance organizers, and self-regulated learners. *Journal of Interactive Learning Research*, *11*, 219-251.

Milheim, W. D., & Martin, B. L. (1991). Theoretical bases for the use of learner control: Three different perspectives. *Journal of Computer-Based Instruction*, *18*, 99-105.

Mills, J. D. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education, 10*(1). Retrieved from http://www.amstat.org/publications/jse/v10n1/mills.html.

Moos, D. C., & Azevedo, R. (2008a). Monitoring, planning, and self-efficacy during learning with hypermedia: The impact of conceptual scaffolds. *Computers in Human Behavior*, *24*(*4*), 1686–1706.

Moos, D. C., & Azevedo, R. (2008b). Self-regulated learning with hypermedia: the role of prior domain knowledge. *Contemporary Educational Psychology*, *33*(*2*), 270–298.

Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*, 309–326.

Niemiec, R. P., Sikorski, C., & Walberg, H. J. (1996). Learner-control effects: A review of reviews and a meta-analysis. *Journal of Educational Computing Research*, *15*, 157-174.

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

Pfannkuch, M., & Reading, C. (2006). Reasoning about distribution: A complex process. *Statistics Education Research Journal*, *5*(2), 4-9. Retrieved from http://www.stat.auckland.ac.nz/serj.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, *31*, 459-470.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(4), 385-407.

Pintrich, P. R., & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*, 33-40.

Puustinen, M. & Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research*, *45*(3), 269-286.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181-210.

Romero, V. L., Berger, D. E., Healy, M. R., & Aberson, C. L. (2000). Using cognitive learning theory to design effective on-line statistics tutorials. *Behavior Research Methods, Instruments, & Computers, 32*(2), 246-249.

Saldanha, L., & Thompson, P. (2003). Concepts of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, *51*, 257-270.

Saw, A.T., Berger, D.E., Mary, J.C., & Sosa, G. (2009, April). *Misconceptions of hypothesis testing and p-values*. Paper presented at the meeting of the Western Psychological Association, Portland, Oregon.

Scheiter, K. & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review*, *19*, 285–307.

Schenker, J. (2007). *The effectiveness of technology use in statistics instruction in higher education: A meta-analysis using hierarchical linear modeling* (Doctoral dissertation, Kent State University, Kent). Retrieved from ProQuest Digital Dissertations (AAT 3286857).

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207-217.

Schunk, D. H. (1981). Modeling and attributional feedback effects on children's achievement: A self-efficacy analysis. *Journal of Educational Psychology, 74***,** 93–105.

Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist, 26*, 207-231.

Shapiro, A., & Niederhauser, D. (2004). Learning from hypertext: Research issues and findings. In Jonassen, David H. (Ed.), *Handbook of research on educational communications and technology* (2nd ed., pp. 605-620). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Shin, E., Schallert, D., & Savenye, C. (1994). Effects of learner control, advisement, and prior knowledge on young students' learning in a hypermedia environment. *Educational Technology Research and Development*, *42*(*1*), 33–46.

Slavin, R. E. (1990). IBM's writing to read: Is it right for reading? *Phi Delta Kappan*, *72*, 214–216.

Song, L., & Hill, J. R. (2007). A conceptual model for understanding self-directed learning in online environments. *Journal of Interactive Online Learning*, *6*, 27-42.

Sosa, G., Berger, D. E., Saw, A. T., & Mary, J. C. (2011). Effectiveness of computer-based instruction in statistics: A meta-analysis. *Review of Educational Research*, *81*(1), 97–128.

Sosa, G., Berger, D. E., Saw, A. T., & Mary, J. C. (2009, April). *Understanding the misconceptions of p-values among graduate students: A qualitative analysis.* Paper presented at the meeting of the Western Psychological Association, Portland, Oregon.

Stone, C. L. (1983). A meta-analysis of advance organizer studies. *The Journal of Experimental Education*, *51*, 194-199.

Swan, K. (2004). Learning online: current research on issues of interface, teaching presence and learner characteristics. In J. Bourne & J. C. Moore (Eds.), *Elements of quality online education, into the mainstream* (pp. 63-79). Needham, MA: Sloan Center for Online Education.

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

Tripp, S., & Roby, W. (1990). Orientation and disorientation in a hypertext lexicon. *Journal of Computer-Based Instruction*, *17*(4), 120-124.

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

Vovides, Y., Sanchez-Alonso, S., Mitropoulou, V. & Nickmans, G. (2007). The use of e-learning course management systems to support learning strategies and to improve self-regulated learning. *Educational Research Review*, *2*, 64–74.

Vygotsky, L. S. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.

Well, A. D., Pollatsek, A., & Boyce, S. J. (1990). Understanding the effects of sample size on the variability of the mean. *Organizational Behavior and Human Decision Processes, 47*, 289-312.

Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational Psychologist*, *30*(4), 173-187.

Winne, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences*, *8*(4), 327-353.

Winters, F. I., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychological Review*, *20*, 429-444.

Young, J. D. (1996). The effect of self-regulated learning strategies on performance in learner controlled computer-based instruction. *Educational Technology Research and Development*, *44*, 17-27.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*(1), 3-17.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*, 82–91.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, *41*(2),64-70.

Zimmerman, B. J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal, 31*, 845–862.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, *29*, 663–676.

Zimmerman, B. J., & Martinez-Pons, M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology*, *80,* 284-290.

Zumbach, J. (2006). Cognitive overhead in hypertext learning reexamined: Overcoming the myths. *Journal of Educational Multimedia and Hypermedia*, *15*, 411-432.

Zumbach, J., & Mohraz, M. (2008). Cognitive load in hypermedia reading comprehension: Influence of text type and linearity. *Computers in Human Behavior*, *24*, 875–887.

Appendix A: Informed Consent Form

You are being asked to participate in a dissertation research project conducted by Amanda Saw, a graduate student in the School of Behavioral and Organizational Sciences at Claremont Graduate University.

PURPOSE: The purpose of this study is to examine how features of an online statistics tutorial influence learning.

PARTICIPATION: You will be asked to complete an online tutorial on standard deviation, which will include questions designed to assess your knowledge and enhance your learning. We expect your participation to take about 30 to 50 minutes of your time.

RISKS & BENEFITS: No risks are anticipated, beyond those associated with completing an online tutorial. If at any time you feel uncomfortable about giving a response, you may discontinue your participation without penalty. We expect the project to benefit you by enhancing your knowledge of fundamental statistical topics.

COMPENSATION: No reimbursement or payment is offered. However, if you are completing this study as part of a class assignment, your instructor may give you course credit or a comparable assignment for credit.

VOLUNTARY PARTICIPATION: Please understand that participation is completely voluntary. Your decision whether or not to participate will in no way affect your current or future relationship with Claremont Graduate University or its faculty, students, or staff. You have the right to withdraw from the research or refuse to answer any questions at any time without penalty.

CONFIDENTIALITY: Your individual privacy will be maintained in all publications or presentations resulting from this study. Your name and all individual responses will be kept confidential by the researcher (you will be given a number for identification purposes).

If you have any questions or would like additional information about this research, please contact Amanda Saw via email at: amanda.saw@cgu.edu. You can also contact my research collaborator/advisor Dr. Dale Berger at Dept. of Psychology, Claremont Graduate University, 123 East Eighth St., Claremont CA 91711, or via email at: dale.berger@cgu.edu. The CGU Institutional Review Board, which is administered through the Office of Research and Sponsored Programs (ORSP), has reviewed this project. You may also contact ORSP at (909) 607-9406 with any questions.

You may print this form before proceeding onto the tutorial.

By checking the box, I indicate the following:

1) I understand the above information and have had all of my questions about participation on this research project answered.
2) I voluntarily consent to participate in this research and may be receiving course credit.
3) I am at least 18 years of age.

To continue, please indicate your consent by checking the box above and then clicking on the "Continue" button below.

Appendix B: Demographic Questions

Before beginning the tutorial, we would like to ask you a few questions about yourself.

1. Have you learned about standard deviations before?
   __ Yes   __ No

2. What is your experience with statistics (check all that apply):
   __ None
   __ Student
   __ Instructor (I teach or have taught statistics)
   __ Professional (I use statistics in my profession)
   __ Other: ________

3. How many statistics courses have you completed?
   __ None
   __ Currently taking first course
   __ Completed one course only
   __ Completed multiple courses

4. What is your highest level of education completed?
   __ Less than high school
   __ High school
   __ Currently in college
   __ College (B.A.)
   __ Currently pursuing Masters
   __ Masters
   __ Currently pursuing PhD
   __ PhD

5. If you are a current student, what field are you in? ______________
6. What is your gender? __ Female   __ Male
7. What is your age (in years)?
   __18-22
   __ 23-29
   __ 30+

8a. Which institution/college are you affiliated with? ______________
8b. If your instructor gave you a code, please enter it (or your name) here:
____

Appendix C: Self-Efficacy and Self-Regulation Subscales of MSLQ (Motivated Strategies for Learning Questionnaire, Pintrich & DeGroot, 1990)

Self-efficacy for Learning and Performance **(α = .93, 8 items)**
5. I believe I will receive an excellent grade in this class.
6. I'm certain I can understand the most difficult material presented in the readings for this course.
12. I'm confident I can learn the basic concepts taught in this course.
15. I'm confident I can understand the most complex material presented by the instructor in this course.
20. I'm confident I can do an excellent job on the assignments and tests in this course.
21. I expect to do well in this class.
29. I'm certain I can master the skills being taught in this class.
31. Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this class.

Metacognitive Self-Regulation **(α = .79, 12 items)**
33. During class time I often miss important points because I'm thinking of other things. (REVERSED)
36. When reading for this course, I make up questions to help focus my reading.
41. When I become confused about something I'm reading for this class, I go back and try to figure it out.
44. If course readings are difficult to understand, I change the way I read the material.
54. Before I study new course material thoroughly, I often skim it to see how it is organized.
55. I ask myself questions to make sure I understand the material I have been studying in this class.
56. I try to change the way I study in order to fit the course requirements and the instructor's teaching style.
57. I often find that I have been reading for this class but don't know what it was all about. (REVERSED)
61. I try to think through a topic and decide what I am supposed to learn from it rather than just reading it over when studying for this course.
76. When studying for this course I try to determine which concepts I don't understand well.
78. When I study for this class, I set goals for myself in order to direct my activities in each study period.
79. If I get confused taking notes in class, I make sure I sort it out afterwards.

Appendix D: Self-Ratings
(adapted from Pintrich & DeGroot, 1990)

For each of these twelve items, the learner rated how true these statements were of themselves on a 1-7 scale, from “Not true at all” to “Very true of me.”

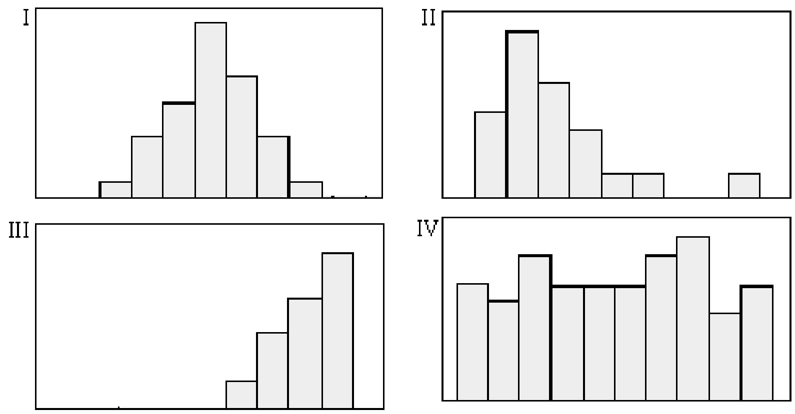
Self-efficacy (3 items)
2. I’m certain I can understand the most difficult material presented in this tutorial.
6. I expect to do well on this tutorial.
11. I’m confident I can learn the basic concepts taught in this tutorial.

Self-regulation of learning (7 items)
3. I ask myself questions to make sure I understand the material I have been reading.
4. I summarize my learning to examine my understanding of what I have learned.
5. When I don’t understand something I’m reading, I go back and try to figure it out.
7. I try to change the way I study in order to fit the material and learning goals.
8. When I’m told I’m wrong, I look for more information.
9. I try to think through a topic and decide what I am supposed to learn.
10. When reading I try to connect new information with what I already know.

Intrinsic value (2 items)
1. Learning about standard deviation is important to me.
12. It is important to me to do well in everything I do.

Appendix E: Items from CAOS Test Assessing Knowledge of Distributions and Variability (from delMas et al., 2007)

Four histograms are displayed below. For each item, match the description to the appropriate histogram.



1. A distribution for a set of quiz scores where the quiz was very easy is represented by:

   A. Histogram I. B. Histogram II. C. Histogram III. D. Histogram IV.

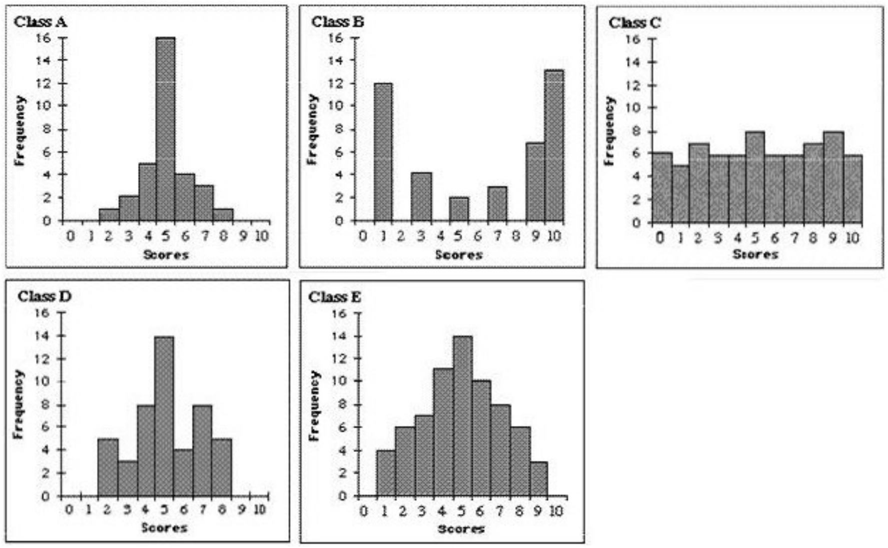
2. A distribution for a set of wrist circumferences (measured in centimeters) taken from the right wrist of a random sample of newborn female infants is represented by:

   A. Histogram I. B. Histogram II. C. Histogram III. D. Histogram IV.

3. A distribution for the last digit of phone numbers sampled from a phone book (i.e., for the phone number 968-9667, the last digit, 7, would be selected) is represented by:

   A. Histogram I. B. Histogram II. C. Histogram III. D. Histogram IV.

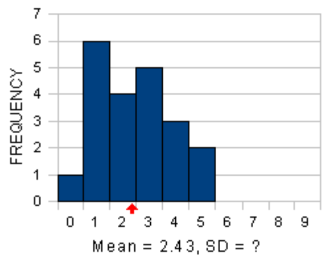
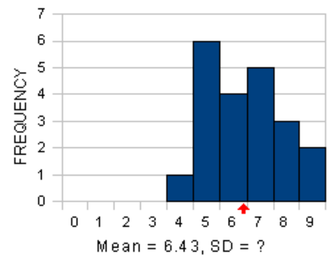
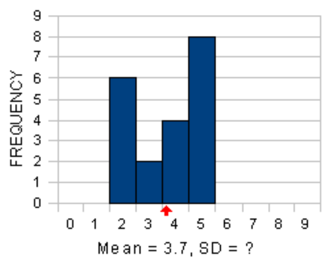
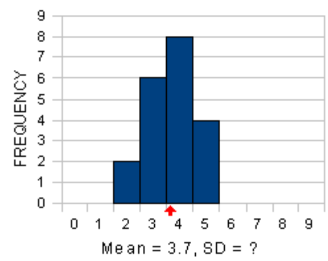
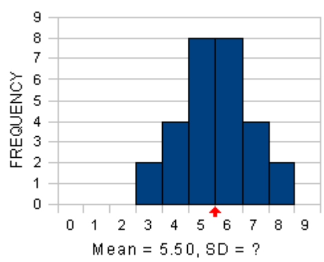
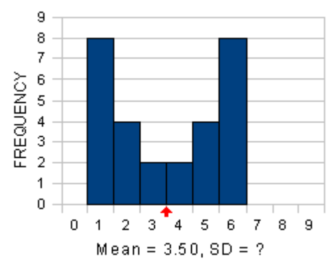
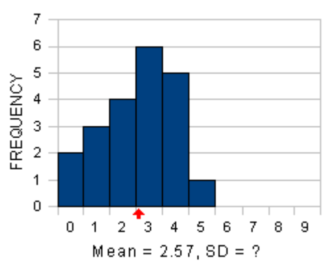
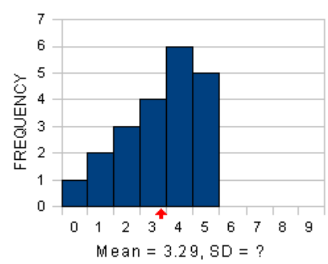
Five histograms are presented below. Each histogram displays test scores on a scale of 0 to 10 for one of five different statistics classes.

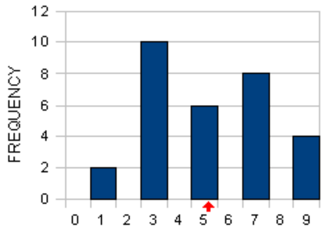
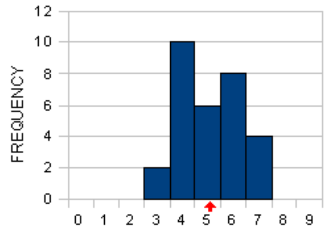
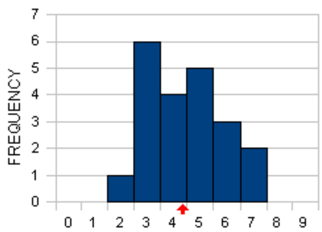
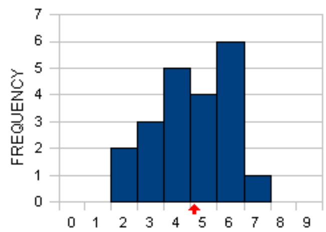
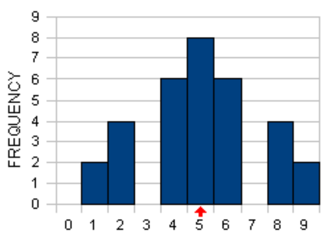
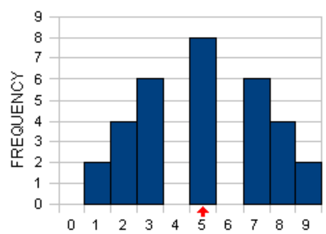
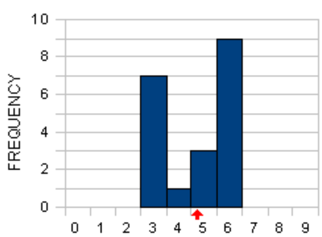
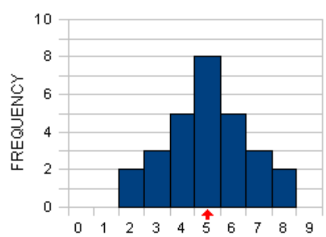
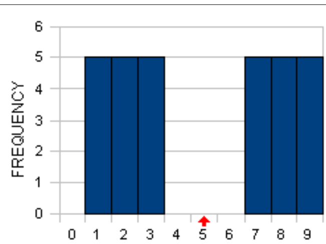
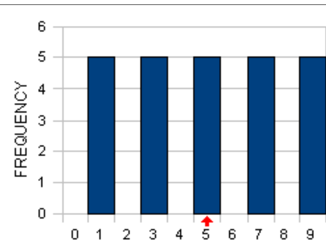


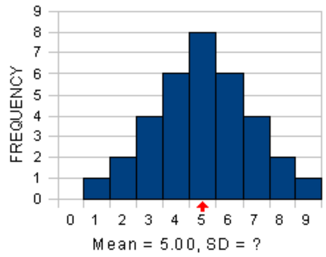
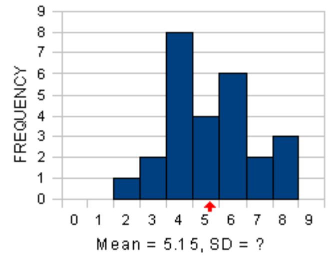
4. Which of the classes would you expect to have the smallest standard deviation, and why?
    - Class A, because it has the most values close to the mean.
    - Class B, because it has the smallest number of distinct scores.
    - Class C, because there is no change in scores.
    - Class A and Class D, because they both have the smallest range.
    - Class E, because it looks the most normal.

5. Which of the classes would you expect to have the greatest standard deviation, and why?
    - Class A, because it has the largest difference between the heights of the bars.
    - Class B, because more of its scores are far from the mean.
    - Class C, because it has the largest number of different scores.
    - Class D, because the distribution is very bumpy and irregular.
    - Class E, because it has a large range and looks normal.

Appendix F: Test Items Assessing Knowledge of Standard Deviation
(adapted from delMas & Liu, 2005)

Item	A	B
P1		
P2		
P3		
P4		

P5	 Mean = 5.13, SD = ?	 Mean = 5.07, SD = ?
P6	 Mean = 4.43, SD = ?	 Mean = 4.57, SD = ?
P7	 Mean = 5.00, SD = ?	 Mean = 5.00, SD = ?
P8	 Mean = 4.70, SD = ?	 Mean = 5.00, SD = ?
P9	 Mean = 5.00, SD = ?	 Mean = 5.00, SD = ?

P10		

Appendix G: Tutorial Section Ratings

You have completed the second [third] section. Now it's your turn to rate this section.

1. Rate your mental effort on the previous part of the tutorial.

O	O	O	O	O	O	O
Low						High

2. How difficult was this part of the tutorial?

O	O	O	O	O	O	O
Easy						Difficult

3. How frustrating was this part of the tutorial?

O	O	O	O	O	O	O
Relaxing						Frustrating

4. How successful do you think you were on this part of the tutorial?

O	O	O	O	O	O	O
Not very successful						Very Successful