INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Exploring the development of graduate learners' monitoring proficiencies and task understandings in a complex writing task

Vivek Venkatesh

A Thesis

in

The Department

of

Education

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Arts at

Concordia University

Montreal, Quebec, Canada

December 2002

ABSTRACT

Exploring the development of graduate learners' monitoring proficiencies and task understandings in the context of a complex writing task

Vivek Venkatesh

This thesis study presents an exploration of the development of graduate learners' monitoring proficiencies and task understandings in the context of a complex writing task. Participants were seventeen students enrolled in a graduate Learning Theories course, engaged in writing six, weekly learning logs; logs were based on the content being covered in the course. Data was collected primarily using a self-assessment tool, the Task Analyzer and Performance Evaluator, which accompanied each learning log. Measures analysed included instructors' performance assessments, students' predicted performances, students' prediction confidence scores, independently scored performances on the logs, criteria justification scores, and monitoring proficiencies, which included bias and discrimination. Results suggest that learners showed improved monitoring proficiencies as instruction progressed. Findings also reveal that while learners exhibited signs of a general monitoring ability across the six logs, prediction confidence, bias and discrimination abilities were mostly unrelated to one another. While learners showed improved criteria justification scores as instruction progressed, there were no relationships found between this measure of task understanding and monitoring abilities. The findings are discussed in relation to recent research on the domain-generality of monitoring and instructional features that promote self-regulation. Due to the exploratory nature of this study and the small sample size, findings cannot be generalized and must be treated with caution. Further research is needed to better characterize monitoring abilities in adult learners engaged in complex, writing tasks.

ACKNOWLEDGEMENTS

I would like to acknowledge the following individuals, in no particular order, for their role in making this thesis one of the more interesting journeys I have undertaken in this lifetime.

I thank my supervisor, Dr. Allyson Hadwin, for her support, encouragement and exhaustive feedback on drafts of the document, as well as for her help in conceptualizing the phenomenon of 'task understanding' over the course of my masters program.

I am indebted to my committee member, Dr. Robert Bernard, for his belief in my abilities to undertake the proposed analysis of the data, and for his close supervision of the repeated measures, pairwise comparison and correlation procedures that ultimately formed the backbone of my results chapter. Dr. Bernard's unparalleled ability to lucidly explain the theoretical bases of the statistical analyses that I undertook gave me the confidence to plow through the analyses at much faster a rate than I had expected.

I tip my hat to my committee member, Dr. Steven Shaw, who, over endless cups of tea (sometimes coffee) and sympathy, patiently endured my early ministrations on the correlational and repeated measures procedures that I proposed to perform on the data I had collected. Dr. Shaw's encouragement, support and creativity in looking at various theoretical and analytical aspects of investigating 'performance monitoring' allowed me to paint a holistic portrait of a complex phenomenon, and provided me enough fodder to carry with me through to the doctoral program.

I thank my friend and colleague, Lori Wozney, for her patience in helping me conceptualize the study, and for her constant encouragement and belief in my abilities. I thank Jennifer Sclater and Gretchen Lowerison, my friends and colleagues, both of who helped me pull my hair out when I was frustrated, and who patted me on my back when I needed it. I thank Dr. Philip Abrami, Peter Wallet, Gretchen and Lori for their comments on early drafts of the methods and results chapters, which we discussed during our statistics classes. I thank Jennifer for reading and commenting on my discussion chapter.

I thank Anne Brown for keeping an eye out for me, and making sure that I met all the deadlines with respect to my defense, thesis submission and program requirements.

I thank my mother-in-law, Francine Detière, for her heart of gold, her love and unquestioning support throughout my masters program. I thank my parents, Krishnaswamy Venkatesh and Vijayalakshmi Venkatesh, for their belief in my abilities.

I dedicate this thesis to my wife, Laurence Detière, for her selfless support of my aspirations, for her love and care, for her ability to always keep a humorous and light outlook towards life through all periods of time, and for her ability to listen when I needed it the most. Perhaps, most importantly, I thank Laurence for helping me figure out how to calculate the monitoring proficiency measure of 'bias'; late one autumn evening as we were waiting for our fondue dinner to simmer, and we feverishly scribbled on umpteen paper napkins, Laurence was the first to reach a solution, probably so that we could get on with our meal!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Preface And Overview

The idea for this thesis arose with my interest in the developmental and instructional aspects of how individuals evaluate their own performance at learning tasks. This thesis study explores the development of graduate learners' monitoring proficiencies and aspects of task understanding in the context of a complex writing task. Monitoring, or learners' abilities to evaluate their own performance on an academic task, has been widely researched under the umbrella term of 'metacognition', mainly in the context of reading comprehension, and recently, in multiple-choice test-taking situations. Researchers have characterized graduate learners' monitoring proficiencies as being a phenomenon that varies from one individual to the next, with variations accorded to factors such as task difficulty, affective states, feedback structures as well as the task environment. Research has also shown that monitoring displays both domain-specific and domain-general characteristics. Following from these current research issues, this thesis study will investigate (a) how well graduate learners' monitor their performance on a complex writing activity, as well as (b) whether these monitoring abilities are specific to each instance of the writing task or are generalizable across an entire period of instruction.

Task understanding refers to learners' perceptions of both the nature of an academic task and the requirements for completing the task. Recent research has posited that these perceptions of the task interact with a host of individual motivational and affective variables. Research has also revealed that there is a lack of empirical evidence of instructional features that promote learners' task understanding in naturalistic, learning

environments. This thesis study will see the implementation of empirically supported instructional features to attune students' perceptions of the requirements of a complex writing task, and match these perceptions with those of the instructor.

In chapter 1, I review the research on monitoring and task understanding, using a theoretical framework provided by prior foundational research on self-regulation and metacognition. This literature review will guide the development of the research purpose and ultimately, the research questions for the thesis study. In chapter 2, I outline the research methodology that I used to conduct my thesis study. This chapter includes detailed descriptions of the research context, the instruments used and the scoring procedures employed. Chapter 3 presents the results of the analyses conducted on the measures of performance, monitoring and task understanding; these results will paint a picture of how monitoring proficiencies and task understanding developed in the context of the research. Chapter 4 then discusses the findings in light of the theoretical framework that was introduced in the literature review. Finally, chapter 5 provides concluding words to the thesis and possible directions for future research.

Chapter 1: Literature Review

Self-Regulation

What is Self-Regulation?

Academic self-regulation involves the strategic application and adaptation of learners' cognitive and metacognitive thought processes in influencing their own behaviors while tackling academic tasks (Pintrich, 2000; Zimmerman, 1990, 1994, 2000; Winne & Hadwin, 1998), taking into account their emotions (McCombs & Marzano, 1990) as well as motivational states (Pintrich, 2000; Pintrich & De Groot, 1991; Winne & Hadwin, 1998) within a specific learning context or environment (Winne & Hadwin, 1998; Zimmerman, 2000). Models of self-regulated learning (SRL) have adopted various perspectives, ranging from socio-cognitive (e.g., Schunk & Zimmerman, 1997; Zimmerman, 2000), affective (e.g., McCombs & Marzano, 1990), motivational (e.g., Pintrich & De Groot, 1991; Rheinberg, Vollmeyer & Rollett, 2000), and context-specific discussions of SRL constructs (e.g., goal setting, Latham & Locke, 1991). Other models of SRL (e.g., Pintrich, 2000; Winne & Hadwin, 1998) acknowledge the need for regulating all the five elements of cognition, affect, motivation, behavior and context in explaining individual self-regulating processes. Most models of SRL promote goal-setting, strategic planning and execution of plans, reflection, self-monitoring, self-efficacy, and self-evaluation as essential skills to be developed by learners who engage with complex tasks requiring resource management skills, individual and group analyses of problem situations, as well as strategic use of feedback and contextually available resources (Butler & Winne, 1995; Ertmer, Newby, & MacDougall, 1996; Paris &

Newman, 1990; Zimmerman, 1990, 1994, 2000). Of specific interest, in this thesis study, is the exploration of learners' *monitoring proficiencies*, or learner's abilities to evaluate their own performance, in the context of complex writing tasks.

Monitoring of Learning, Performance and Comprehension

Metacognition and Monitoring

Monitoring falls under the general umbrella term of metacognition; metacognition, in turn, has been discussed within the theory of self-regulation. For example, while self-regulated learning is conceived as including the processes of monitoring, controlling and regulating cognition as well as motivation, affect, volition and effort (Schunk & Zimmerman, 1994; Zimmerman, 1986, 2000; Zimmerman & Schunk, 1989), these processes of monitoring, controlling and regulating are related to and dependent on the *metacognitive* knowledge about self and cognition (Garcia & Pintrich, 1994; Pintrich, Wolters, & Baxter, 2000).

Metacognition, put simply, is the ability of a learner to be an agent of one's own thoughts. Metacognition has been defined as "knowledge of cognition and monitoring and control of cognitive activities" (Hacker, 1998, p. 2). Models of metacognition take into account the interactions between constructs that include metacognitive knowledge, metacognitive experiences, awareness, control, goals, strategies and regulation of strategies (e.g., Brown, Bransford, Ferrara & Campione, 1983; Flavell, 1979; Hacker, 1998; Nelson & Narens, 1990; Paris & Winograd, 1990). Researchers such as Pintrich et al. (2000), recognize and distinguish between three aspects of metacognition, including (a) metacognitive knowledge, (b) metacognitive judgments and monitoring, as well as (c) self-regulation and control of cognition. Metacognitive judgment and monitoring are

associated with the process of reflecting on one's metacognitive awareness and other metacognitive activities, as one is engaged with a learning task (Pintrich et al., 2000; Nelson & Narens, 1990). These metacognitive activities include thinking about and acting upon (a) judgments of task difficulty, (b) reactions to learning and comprehension monitoring, (c) feelings of knowing, and (d) confidence judgments (Pintrich et al., 2000; Winne & Hadwin, 1998). In this thesis study, I will be focusing on the processes associated with *learning, performance* and *comprehension monitoring* in graduate learners in the context of complex writing tasks.

Definitions and Measurement of Monitoring

The generic definition of monitoring centers on the ability of learners to evaluate their performance at a given point in time. Comprehension monitoring, long viewed under the umbrella of metacognitive skills, empowers learners to not only evaluate but also alter, and hopefully improve, their performance (Butler & Winne, 1995; Pressley & Ghatala, 1990). Self-monitoring of metacognitive processes has been long considered as a prerequisite for learners to assume and take control of their learning, as well as bridge the gap between what learners know about their learning and performance and what they do not know (Brown, 1980; Flavell, 1979, Pintrich, Wolters, & Baxter, 2000; Schraw & Impara, 2000). However, the measurement of high-level processes in metacognition, specifically, that of monitoring, is considered to be especially laborious, difficult, and context-specific (Pintrich et al., 2000; Tobias & Everson, 2000). Some of the relevant conclusions reached by Pintrich et al. (2000) in discussing the issue of assessing metacognition within a framework of SRL are that (a) metacognition is measured in a variety of ways, from think-aloud protocols to self-report surveys to observations; (b)

different measures of components of metacognition assess the same components in different ways; (c) there is a lack of theoretical links between metacognition and SRL; (d) the issue of domain-generality and domain-specificity of metacognition needs to be further explored; and (e) performance assessments may help in measuring constructs related to metacognition across and within domains.

In my thesis study, I will be focusing on learner's abilities to calibrate their performance (Glenberg, Sanocki, Epstein & Morris, 1987; Schraw, Dunkle, Bendixen & Roedel, 1995; Schraw & Nietfeld, 1998; Schraw & Roedel, 1994). Calibration specifically refers to learners' abilities to evaluate performance upon immediate completion of a task or test item. Following Schraw et al.'s (1995), lead, I use the general term *monitoring*, throughout this thesis, as it is a term that is more familiar than the term *calibration*, with readers. In the next section, I briefly review the literature on monitoring in test-taking situations. Note that I use the term "test-taking" to signify an academic task that undergoes an instructor-based evaluation and is assigned a grade or measure of performance of some form.

Factors Influencing Monitoring in Test-Taking Contexts

Reviews of learners' monitoring capabilities while taking tests have revealed that generally, individuals are better able to evaluate their performance during or after a test, than before it (see Baker, 1989, Pressley & Ghatala, 1990 and Schraw & Moshman, 1995, for recent reviews). Effective monitoring is dependent on constraints such as the nature of the test, individual characteristics of the test taker as well as the test environment. In discussing the nature of the test, research has focused on the difficulty and format of the test. Prior research has demonstrated that difficult tests lead to poorer

monitoring because of a failure to adjust to performance expectations (Schraw & Roedel, 1994). Recognition tests lead to poorer monitoring than recall tests because the recognition test-takers mistakenly accord themselves a higher level of mastery than those taking recall tests (Ghatala, Levin, Foorman & Pressley, 1989). Monitoring proficiency has been seen to improve when learners are tested on detailed information rather than main ideas (Pressley, Ghatala, Woloshyn & Pirie, 1990).

Test-taking individuals possess characteristics that influence monitoring capabilities, including familiarity with the domain, intellect, and dispositions. Research investigating familiarity with domains has a mixed set of findings. While Glenberg & Epstein (1987) found a negative relationship between expertise and monitoring, Morris' (1990) research demonstrates that domain knowledge was unrelated to monitoring proficiency even though it was related to the ability to answer questions effectively in that domain. Schraw and Roedel (1994) reported that college students monitored their test performance with equal accuracy in three domains once test difficulty was controlled. Maki and Serra (1992) interestingly found that monitoring improved as individuals acquired more information from the learning material that was being used during the instruction.

A number of studies by Pressley and colleagues, cited in Pressley & Ghatala's (1990) review, reveal that learning ability does not necessary lead to high-skill levels of monitoring. On the other side of this spectrum, Walczyk & Hall (1989a) discussed how children's ability to monitor was seriously affected by cognitive impulsivity. Slife and Weaver (1992) found that depressed individuals monitored their comprehension less

effectively than non-depressed individuals and also showed less control of metacognitive skills.

The environment in which the test is taken also affects monitoring skills. When given incentives to monitor accurately, Schraw, Potenza and Nebelsick-Gullet (1993) found that test-takers monitored more accurately than a control group who were not given incentives. Moreover, test takers who were given a reward for normatively accurate monitoring outperformed the control group. Elsewhere, Pressley, Snyder, Levin, Murray, & Ghatala (1987) showed how perceived readiness for testing improved when additional questions were included during study. Similarly, students who were provided with feedback during testing showed improved monitoring skills (Glenberg, Sanocki, Epstein & Morris, 1987; Walczyk & Hall, 1989b). In further support of the use of consequential, engaging activities that promote processing during test-taking situations, Maki, Foley, Kajer, Thompson, & Willert (1990) found that students who generated missing information for text provided in a test, monitored more accurately than those who did not.

Schraw et al. (1995) propose four general characteristics of monitoring proficiencies. First, monitoring proficiency is dependent on the timing of the confidence judgments made during test-taking situations. Second, a high degree of domain knowledge does not automatically qualify a learner to possessing superior monitoring proficiencies. Third, monitoring proficiency is dependent on the nature of the test and the instructions that accompany the test in aiding the learner to successfully complete the test. Finally, monitoring proficiency seems to be unrelated to intellectual ability or processing speed, but it might be affected by dispositional factors, such as mood, impulsivity, and emotional states that a learner might possess.

As Schraw and his colleagues (Schraw et al., 1995; Schraw & Neitfeld, 1998) observed, monitoring in test-taking situations is best characterized as an "idiosyncratic phenomenon" (Schraw et al., 1995, p. 434), influenced by individual learner characteristics and the nature of the test, as opposed to the general skill that the term metacognition suggests. Thus, while monitoring skills might inherently exist or be learned in a test-taker, there are likely to be a range of utilizations from person to person due to the inherent eccentricities in the nature and measurement of monitoring. In fact, investigations on college and adult learners' monitoring of academic performance (e.g., see Schraw, 1994, 1997, 1998; Schraw et al., 1993), suggest that most adult populations possess metacognitive knowledge about their learning even though a large proportion do not use this knowledge to improve their on-line regulation of performance.

It should be noted, though, that most research on learning, performance and comprehension monitoring has been primarily focused within the domain of reading comprehension in a school-based population. My thesis will explore the development of monitoring proficiencies in a different domain, namely, complex writing tasks, with adult, graduate learners. Such research is necessary in further developing the notion of monitoring proficiencies in various academic contexts as well as in exploring whether monitoring is context-dependent or context-general. The issue of domain-specificity and domain-generality of monitoring is, hence, discussed in the next section.

Domain Specificity versus Domain Generality of Monitoring

The literature seems to be divided in its description of the nature of metacognition; metacognition has been described, on one hand, as a higher-order type of knowledge that regulates comprehension and performance within a single domain, while

on the other as a higher-order type of knowledge that regulates performance and understanding across all domains (Pintrich et al., 2000; Schraw et al., 1995; Schraw & Nietfeld, 1998). The two opposing views on metacognition lead to two competing hypotheses on the nature of monitoring. The *domain-specific hypothesis* of monitoring supports the notion that monitoring in one domain is unrelated to monitoring in separate, distinct domains. Expert problem-solving subscribes mostly to the domain-specific hypothesis (e.g., Bereiter & Scardamalia, 1993; Glaser & Chi, 1988; Voss, Green, Post & Brenner, 1983; Voss, Wolfe, Lawrence, & Engle, 1991), in its propositions that experts assess and rationalize problems, select strategies to tackle these problems, and evaluate the validity of their solutions with greater accuracy than novices would. Central to the domain-specific notion of monitoring is the assumption that monitoring proficiency is dependent on the level of domain-related knowledge (Schraw et al., 1995). According to the domain-specific view, high levels of monitoring can only be seen if domain-related knowledge and domain-specific regulatory skills are simultaneously present and interact.

On the other hand, the *domain-general hypothesis* subscribes to the notion that monitoring in any one domain is dependent both on general metacognitive skills as well as domain-specific knowledge and regulatory skills. Empirical findings of domain knowledge bearing no relation to monitoring proficiencies, including Glenberg & Epstein (1987), Morris (1990) and Schraw and Roedel (1994), lend support to the domain-general hypothesis. Moreover, theories proposed by Borkowski, Chan and Muthukrishna (2000) and Pressley, Borkowski and Schneider (1990), which look at metacognitive theories as explaining the triumphs and the failings of strategy generalizations, lend support to the domain-general monitoring theory. In the domain-general view, as Schraw et al. (1995)

explain, monitoring proficiency is determined more by domain-general metacognitive awareness than domain-specific awareness; examples include evaluating the sufficiency of domain-related knowledge, selecting and applying appropriate strategies in a given situation, and assigning appropriate levels of cognitive and metacognitive resources based on task demands. Schraw and his colleagues proposed that, given a set of performance and monitoring scores across a variety of domains, the domain-general view would be most strongly supported by uncorrelated performance scores and correlated monitoring proficiency scores across all domains; this would suggest that a general monitoring skill is present even when a performance skill is not. The domain-specific view, however, would be best represented by a strong performance correlations and unrelated monitoring scores across all domains, thereby suggesting that measures of monitoring are unrelated even in the face of related performances.

Schraw et al. (1995) conducted two experiments to test the domain-specific and domain-general assumptions by assessing students' performance and confidence in correctly answering eight sets of multiple-choice test. Each of the multiple-choice tests reflected a different domain of knowledge, and required students to mainly recall semantics from their long-term memories on factual information (e.g., U.S. presidents, geography, etc.). The measures of performance and confidence yielded two measures of monitoring proficiencies. The first is termed as *discrimination*, and refers to the ability of students to assign an appropriate level of confidence to their performance on a test item. Discrimination was calculated as the difference between confidence for correct items and incorrect items (Lundeberg, Fox, & Puncochar, 1994). The second measure calculated was *bias* (Keren, 1990; Yates, 1991), which measured the extent to which students were

over or under-confident for each of the eight tests. Bias was calculated by taking the difference between the average confidence and average performance for each of the eight test items. In experiment 1, Schraw and his colleagues found that performance and discrimination accuracy were not correlated across the eight domains, lending support to the domain-specific hypothesis because it suggested that feelings of confidence and derived measures of monitoring proficiency were unrelated. However, in experiment 1, confidence and bias were correlated, lending support to the domain-general hypothesis, because this suggested that a general monitoring skill existed even when a general performance skill did not exist. In experiment 2, after variability due to difference in domains was eliminated on the eight tests, performance and confidence measures were collected, and correlations were computed among performance, confidence and the two measures of monitoring proficiency, discrimination and bias. Results from experiment 2 showed all four measures to be correlated across all or most domains; in addition, confidence was correlated even after the effect of performance was removed. The results of Schraw et al.'s (1995) two experiments show some support for the domain-general argument on monitoring.

In a follow-up experiment, Schraw & Nietfeld (1998) tested adults' performances and confidences in drawing novel inferences on eight different measures (domains) of fluid and crystallized ability as opposed to the simpler tests on retrieval of declarative and factual knowledge seen in Schraw et al. (1995). Fluid ability measures the processes underlying mental activity, whereas crystallized ability measures the sum of acquired knowledge experience in learners (for more detailed descriptions see Schraw & Nietfeld, 1998, p. 237). In this study, monitoring proficiency was represented by discrimination (as

described in Schraw et al., 1995) and accuracy, which represented the absolute value of the difference between average confidence and average performance for each test. Accuracy provided a measure of how far learners' predictions of their performances were from their actual performances, regardless of whether they overestimated or underestimated their performance. Findings from Schraw & Nietfeld's (1998) study supported two main conclusions, the first being that monitoring scores were correlated across multiple domains, and the second, that individuals may possess separate general monitoring skills for fluid and crystallized tasks. Further, the data from Schraw & Nietfeld's (1998) study were best explained by domain-general theories of monitoring proficiencies, as opposed to information-encapsulation theory (e.g., domain-specific views on performance and monitoring accuracy) or a modular perspective (e.g., the belief that biological structures support cognitive functions).

Monitoring Proficiencies in Complex Writing Tasks

In my thesis study, I will extend Schraw and his colleagues' exploration of the nature of monitoring proficiencies in the context of a complex writing task for graduate learners. I believe it is necessary for monitoring proficiencies to be investigated in manner similar to Schraw et al (1995) and Schraw & Nietfeld (1998), but in a context aside of multiple-choice test-taking situations. Complex writing tasks related to a specific curriculum in a graduate learning environment provide a more representative, adult learning context in which learners engage in consequential activities designed to promote deep processing activities and higher-order thinking. My thesis study will explore whether monitoring proficiencies for learners engaged in complex writing tasks develop in a manner that is similar to those involved in multiple-choice test taking situations. In

addition, I will explore the domain-generality and domain-specificity of learners' monitoring skills in the context of complex writing tasks.

Self-Regulation and Instructional Design

A secondary purpose of this thesis study is to explore the notion of learners' task understandings in the context of complex writing tasks. Specifically, through the use of empirically supported instructional design (ID) principles, my thesis explores the possibilities of attuning learners' perceptions of the assessment criteria for a complex writing task with the criteria laid out by the instructor.

Despite the widespread research on SRL-based instruction, there is a paucity of experimental evidence of instructional methods that promote the various aspects of learners' academic self-regulation. Although Ley and Young (2001) have suggested principles of instruction for self-regulation in classrooms, these principles are not supported by empirical findings. This lack of research led to our conducting a review of ID features that promote self-regulation (Venkatesh & Hadwin, 2002). In this review of the literature on SRL-based instructional strategies, various strategies that emerged from the literature were classified as one of three types. The first was coined as *instructional processes* and referred to strategies that focused on the manner in which teachers interacted with students while delivering instruction (e.g., modeling, scaffolding, teacher questioning, etc.). The second was termed as *classroom culture*; these were strategies aimed at influencing the environment in which learners applied themselves, (e.g., promotion of a supportive social environment, fostering positive attitudes towards learning). Third, *task structuring* included strategies aimed at explaining how the task had been designed (i.e., individual activity, collaborative, case-study, problem-based), what

tools the instructor provided for completing the task (e. g., recording criteria-based progress, recording performance, using planning sheets, aiding comprehension of the task) as well as what type of feedback structure was being employed (e.g., teacher feedback on performance, peer feedback, self-evaluations). While the three types of strategies outlined relied heavily on the cognitive, metacognitive and behavioral aspects of SRL, we acknowledged the role each design feature plays in shaping the *motivational* and *affective* reactions of the learner. Of special interest, in the present investigation, is the issue of how to explore and instructionally promote learners' task understanding, that is, how we can better instructionally promote *task structuring* in the context of complex, writing tasks.

Task Understanding

Critical Components of Task Understanding

Task understanding draws on two distinct, but interacting elements; these include individuals' perceptions of the *academic task*, as well as of *themselves* as a *learner* within a particular academic context (c.f., Winne & Hadwin, 1998). Learners' perceptions of the academic task include both the *nature* of the task, and the *assessment criteria* associated with the task. Learners reflect on their perceptions of the *nature* of the task, including (a) the rationale for performing the task; (b) the procedures that need to be undertaken to perform the task and the required outputs; (c) the materials that are available to perform the task; as well as (d) the contextual conditions under which the task has to be performed. Learners also need to grapple with the *assessment criteria* that the instructor will be using in judging their performance on the task. It is therefore clear that task

understanding involves a close interaction between learners' perceptions and the instructor's perceptions of the academic task.

In addition to the task-associated elements, task understanding is influenced by the learner's *knowledge of "self-as-learner"*. Such knowledge includes preferred learning styles and learning needs, prior content and task-specific knowledge, current motivational and emotional levels of anxiety and efficacy, as well as motivational and emotional levels associated with a specific type of task environment (Lin, 2001; Randi & Corno, 2000; Winne & Hadwin, 1998). Similar to the distinctions I have made above in the components of task understanding, Winne and Hadwin (1998) distinguish between the task and cognitive conditions that influence students' comprehension of an academic task. Whereas *task conditions* refer to the nature and assessment criteria of a task, *cognitive conditions* are the content-related strategies, prior knowledge and experiences, affective states, beliefs and motivational attributes that affect the extent to which learners develop accurate perceptions of the academic task.

Task Understanding as a Phase of Self-Regulation

Models of SRL have conceived of task understanding as either a phase or a key element of a phase of self-regulation. Different researchers, though, have varying conceptions of terms related to task understanding. For example, a triadic, socio-cognitive model of self-regulation, which takes into account the personal, behavioral and environmental effects on self-regulation (see Zimmerman, 2000), describes the three cyclical phases of forethought, performance or volition control and self-reflection. The first phase of forethought includes a component Zimmerman terms as task analysis. However, task analysis is explained in terms of learner's abilities to set goals and

strategically adopting a plan of action in achieving these goals. No mention of comprehension of task requirements is made in the model. Zimmerman's socio-cognitive model, however, does acknowledge the important roles of self-motivation beliefs, interest, value placed by learner on the task, as well as goal orientation.

Elsewhere, Pintrich (2000) outlines four phases of self-regulation similar to Zimmerman's (2000) model. In Pintrich's (2000) model, the four phases of goal-setting, monitoring, control and regulation processes each apply to the four areas of regulation of cognition, motivation or affect, behavior and context. The first phase, goal-setting, according to Pintrich, regulates (a) *cognition*, by activating prior knowledge; (b) *motivation* or *affect*, by considering efficacy judgments and goal orientations; (c) *behavior*, by accounting for time and effort management strategies; and finally (d) *context*, by acknowledging that students develop perceptions of the context and the task itself. In comparison to Zimmerman's (2000) model, Pintrich's first phase emphasizes the importance of framing task understanding as an internal cognitive and affective activity, as well as a regulation of external contextual factors.

Finally, Winne and Hadwin (1998) explicitly introduce task understanding as the first phase of self-regulation; the other three phases being 'goal setting and planning', 'enactment of strategies' and 'evaluating and updating'. Task understanding in this model, as explained earlier, is influenced by the task and cognitive conditions in a specific academic context. Winne and Hadwin propose that learners cycle through the four phases of self-regulation throughout their engagement with an academic activity, but do not necessarily follow a specific order through the four phases. For example, a student's engagement with a strategy could result in a failure to achieve a goal. The

student might then cycle back to rethink the goals set for the task, which in turn, could affect a component of task understanding, including, for example, perceptions of task difficulty level or motivation and anxiety levels. Therefore, task understanding might be developed across the phases of self-regulation, as the learner interacts with the task in a contextualized environment (Hadwin, 2000; Winne & Hadwin, 1998).

Development of Task Understanding Across All Phases of SRL

While I acknowledge the importance of task understanding as a critical, first phase of SRL, I agree with Hadwin (2000) that it is not very often that students develop a complete perception of the academic task at the very beginning of their engagement with the task. Just as the three models of SRL described above subscribe to a cyclical development of self-regulation, I agree with both Hadwin (2000) and Winne and Hadwin (1998) that task understanding continuously develops as students cycle through the various phases of self-regulation. For example, in the context of a complex academic task often encountered in a graduate classroom setting, information contributing to students' task understanding might include (a) the rationale for performing a task, (b) the instructor's assessment criteria, (c) the resources available in the given environment, and (d) the prior knowledge and knowledge of "self-as-learner" that the student brings to the task. The extent to which these elements interact to form an initial representation of the task varies from student to student at the beginning of their engagement with the task. Exploration of the task by performing a few preliminary activities, setting a few proximal goals and trying to attain them, followed by feedback from the instructor on initial progress on the task might help in building the students' individual task understanding. Moreover, in building an impression of oneself as a learner *while* engaging with the task,

the student's knowledge of "self-as-learner" is continuously developing to reflect changes in task understanding, and in turn, influences the strategic engagement of the student with the task (see Randi & Corno, 2000 for an innovative example of building learners knowledge of "self-as-learner" through the development of metacognitive knowledge and beliefs). Knowledge of "self-as-learner" interacts with, and is continually influenced by the task conditions including the nature of the task, assessment criteria and rationale, as well as the cognitive conditions imposed by the learner including prior knowledge, metacognitive knowledge and awareness, beliefs, values and presuppositions.

Task understanding, therefore, does not necessarily develop as a first phase of self-regulation. Rather, the cyclical nature of SRL demands that students revisit and redefine the task as their knowledge of both the task and self are influenced and grow over the time spent engaging with the task.

Instructional Principles in Promoting Task Understanding

Our review (Venkatesh & Hadwin, 2002) has revealed a lack of research on how to improve learners' task understandings. Therefore, while the concept of *task structuring* was exemplified mainly through explicit activities for students to set goals as well as plan and execute strategies in achieving goals, very few researchers developed instruction specifically to improve learners' understanding of a specific academic task. In fact, our review revealed that research studies that proposed SRL-based instruction very often required students to jump into goal-setting and planning situations, without providing students with an idea about what the academic task entailed, and without providing support for developing the critical, metacognitive knowledge of "self-as-learner".

Our review (Venkatesh & Hadwin, 2002), however, pointed to studies that explicitly provided instruction to support learner's task understanding; these included, for example, Butler (1998), Englert, Raphael, Anderson, Anthony, and Stevens (1991), Ertmer et al. (1996), Perry (1998), Perry and VandeKamp (2000), and Perry, VandeKamp, Mercer, and Nordby (2002).

Task understanding was addressed in Butler's (1998) evaluation of the Strategic Content Learning (SCL) approach to developing self-regulation in undergraduate students with learning disabilities. Butler used one-on-one tutoring sessions, where students were taught strategies to better comprehend the requirements of an academic task in terms of existing knowledge and beliefs, set attainable and individualized goals based on their unique needs, and implement strategies towards the attainment of these goals. The tutors in Butler's study helped students (a) choose learning areas that were problematic, (b) set their own learning goals, (c) explicitly state and set assessment criteria to judge their progress, and (d) choose strategies to achieve their goals. Students were also taught to monitor their progress towards their goals, and adjust their approaches based on perceptions of their progress.

Perry's (1998) work in second and third grade classrooms using portfolio activities provides a different exemplar of developing instruction to promote task understanding and self-regulation. In her research with second and third-grade classroom-based portfolios, Perry classified those classrooms as "high self-regulated" ones, where students were provided with (a) choices in their writing activities (i.e., choice of what, where, when to write and who to write about); (b) control over the amount of challenge they experienced in the class; (c) opportunities for self-evaluation; and (d) instrumental

peer and teacher support. Perry found that classroom contexts affected student beliefs, values, expectations and actions in the classroom, thereby highlighting the importance of developing knowledge of "self-as-learner" throughout the phases of self-regulation.

Englert et al. (1991) developed an intervention called Cognitive Strategy Instruction in Writing (CSIW) to improve the expository writing abilities of fourth and fifth-graders. In their efforts to develop students' perceptions of the rationale of performing the writing task, teachers in Englert et al.'s study used scaffolding, modeling, questioning and peer discussions as key instructional processes. During the writing activities, students were given worksheets with queries directing students to plan their writing (who am I writing for?; why am I writing this?; what do I [already] know?; how can I group my ideas?; how will I organize my ideas?); similar worksheets were also used to revise and edit students' essays within groups of peers.

Finally, task understanding was addressed by Perry et al. (2002) and Perry and VandeKamp (2000) in the context of complex reading and writing activities with students from kindergarten to grade 3. Perry and her colleagues point to the use of instrumental support from both instructor as well as peers in better developing an understanding of (a) the nature of the reading or writing task and (b) the assessment criteria for the reading or writing task, that students were engaged in. This instrumental support includes regular feedback on learners' progress in completing a task, clarifying the meaning and rationale behind reading and writing assignments, discussing the assessment criteria with learners and encouraging peer discussion of assignments.

Exploring Task Understanding and Monitoring in Complex Writing Tasks

In this thesis study, I use a self-assessment tool, the Task Analyzer and Performance Evaluator (TAPE), to develop a more accurate understanding of a complex writing activity in graduate students. Specifically, the TAPE tool will help attune students' comprehensions of the writing task's assessment criteria to match those of the instructor.

The TAPE tool is designed keeping in mind the instructional principles that emerged from our review (Venkatesh & Hadwin, 2002), including providing instrumental, instructional support to help make students' understandings of the task criteria explicit and to provide feedback on students' perceptions of the task criteria. I acknowledge that I have not included any questions in the TAPE tool that elicit the motivational, affective or emotional states of students while they performed their self-assessments. At the outset, therefore, I recognize the limited view of task understanding that the TAPE tool provides; the learner's perceptions of the assessment criteria for the writing task are the only measures I will be able to collect from students' responses to items related to task understanding. A second function of the TAPE tool is to assess the development of monitoring proficiencies of learners as they tackled the learning log task; to this end, the TAPE tool will be used to collect measures of performance prediction and prediction confidence for the writing task that learners will engage in over the course of instruction.

Purpose

Extending Schraw and his colleagues' work on the development of monitoring abilities, the main purpose of this study will be to explore the development of adult,

graduate learners' monitoring proficiencies in the context of a complex writing task. Given the current debate on the domain-generality and domain-specificity of monitoring proficiencies (Pintrich et al., 2000; Schraw et al., 1995; Schraw & Nietfeld, 1998), a second purpose of this study will be to explore the generality and specificity of learners' monitoring abilities in complex writing tasks, and compare these trends with those that Schraw and his colleagues have uncovered in the context of multiple-choice test taking situations. Third, following from our review of SRL-based ID (Venkatesh & Hadwin, 2002), and extending Perry and her colleagues' work in addressing task understanding within a framework of self-regulation, this study will use empirically supported instructional features to investigate the development and promotion of learners' task understanding in a complex writing task. Specifically, this study will attempt to shed light on whether, and how learners' perceptions of the assessment criteria may contribute to the cyclical and recursive development of task understanding. Fourth, this study will investigate the relationship between task understanding and monitoring proficiencies for graduate learners tackling a complex writing task. The specific research questions I will be addressing are:

1. How do graduate learners' monitoring proficiencies develop over the course of instruction as they tackle a complex writing task?

1. Are learners' abilities to meet the assessment criteria for a complex writing task reflected in the instructor's assessment of their performance?

1. Are learners' understandings and perceptions of the assessment criteria for a complex writing task related to their monitoring proficiencies over the period of instruction?

Chapter 2: Method

Research Context

Participants

Participants were 17 volunteers registered in a graduate Learning Theories course conducted in the summer session of 2002 (May-June) at a large Eastern Canadian University. The course ran for a period for a period of seven weeks, with two 2-hour classes scheduled per week, except for the last week where only one class was held, thereby yielding a total of 13 classes. Of the 17 participants, five were male. All but three of the students were required to compulsorily complete the Learning Theories course as part of the requirement for their graduate degree.

Research Context

Students enrolled in the Learning Theories course were expected to read a set of readings for every week and subsequently prepare question-and-answer sets based on the contents of the readings. These question-and-answer sets are termed as "learning logs". Three types of question-and-answer sets were outlined for this course: (a) multi-structural, wherein students drew explicit connections between two or more related concepts; (b) relational, wherein students drew implicit links between two or more constructs; and (c) extended abstract, wherein students created and supported hypotheses based on learning theories, or demonstrated and supported real-life applications of learning theories. The three levels of question-and-answer sets are based on Biggs' SOLO taxonomy (Biggs, 1991, 1996). Students were provided with a set of criteria for the evaluation of a question-and-answer set for each of the three sets outlined above.

Please see Appendix A for an extract from the course outline describing each level and the assessment criteria for the question-and-answer sets in more detail. Students were also informed, by the instructor, that grades for the learning logs would reflect the accuracy of the content written for each learning log and students' growth in demonstrating skills in tackling the learning log. Logs were submitted by students for grading on a weekly basis; the instructor tried to provide feedback on every log before the subsequent log was due. Due to the tight schedule within which the course was conducted, feedback was sometimes delayed. Specifically, feedback for most or all students on log 2, for half the students on log 3 and for two students on log 4, was delayed. Students also received an extension on completing and submitting their sixth and final learning log until after the final examination. The instructional aspects of the course, and the structure of the learning logs were designed entirely by the instructor.

Learning logs. There were a total of six learning logs. Each of the first five logs consisted of one multi-structural question-and-answer set and either a relational or extended abstract question-and-answer set. For the sixth and final log, students were required to produce one multistructural, one relational and one extended abstract question-and-answer set. For log 1, students were provided with a multistructural and relational question; students provided the answers to these two questions. For logs 2 and 4, students produced one multistructural and one relational question-and-answer set. For logs 3 and 5, students produced one multistructural and one extended abstract question-and-answer set. Each learning log contributed to six percent of the final grade, yielding a total of 36%.

Instructional aids. Students were exposed to two types of instructional aids to help them create their learning logs. The first was termed as "worked examples" of question-and-answer sets, which consisted of completed learning logs from previous years' students of the same Learning Theories course, accompanied by the instructor's comments. The second was termed "question builders"; it consisted of a two-columned table, one of which provided a set of terms and the other with the types of queries that might help students with framing a question at the appropriate level (i.e., multistructural, relational or extended abstract). Based on the results of a test of prior knowledge, students underwent a matched random assignment into two groups to ensure group equivalence in prior knowledge; one group had nine students while other had eight. Students from either group did not receive any instructional tips for logs 1 and 6 (i.e., the first and last week of the course). For the next two weeks of the course (i.e., logs 2 and 3), students in one of the groups received "worked examples" as an instructional aid for their learning logs, whereas students in the other group received "question builders". For the latter two weeks of the course (i.e., logs 4 and 5), the instructional aids were switched so as to allow all students in the course to benefit from both the instructional aids.

TAPE description. For the purposes of my thesis study, I designed a self-assessment tool, the Task Analyzer and Performance Evaluator (TAPE), to accompany each learning log (the TAPE tool is described in the "Instruments" section of this chapter and a sample can be seen in Appendix B). At the end of the seven-week course, students produced six learning logs with six completed TAPES, one for each learning log. Students wrote the answers to the questions posed in the TAPE only after they completed their learning log for that week. All learning log question-and-answer sets as well as their

accompanying TAPEs were required to be submitted online to a First Class® conference space allocated for the course. Each TAPE contributed to one percent of the final grade. It was expected that the instructor's feedback on both the contents of the learning log, as well as the students' responses to the TAPE would enable students to (a) develop a more *accurate understanding* of the task in terms of the task requirements and criteria for assessment; and (b) improve the *accuracy* of their *monitoring* of *academic performance* over the course of seven weeks of instruction.

Researcher Roles

Research was being conducted under the umbrella of the course instructor's research grant. As such, data were collected for the instructors' research agenda as well as for my thesis. While my thesis questions focused on the development of students' task understanding and monitoring proficiencies, the instructor's questions concerned the effectiveness of the two instructional aids, "worked examples" and "question builders". My responsibility was to design the TAPE tool, and coordinate the research aspects of the course; I was not involved in any of the instructional or evaluation activities. The instructor and teaching assistant, facilitated all online and classroom activities of the students in the course. All data collection procedures complied with guidelines set out by the American Psychological Association. My thesis research was approved by the Departmental Ethics Committee, while the instructor's research agenda was approved by the University Ethics Committee before any data were collected.

Research Procedures

Method of Recruitment

During the first class on the first week, I briefed all students about the nature of the research being undertaken and how it related to the course activities. It was likely that some of the students might have encountered the instructor, teaching assistant or myself if they had already taken courses in the department where the course was being offered. Students were then informed of the research purposes as outlined in the consent form (see Appendix C). Students were informed that the research involved no additional commitment beyond required course activities apart from optional interviews with me. I explained that the three interviews of approximately 30 minutes each were oriented toward understanding students' reasons for choosing and writing a particular question-and-answer learning log, the criteria the student used to self-assess their own performance, and whether they found the online instructional scaffolds to be useful. The interviews were NOT a required aspect of the course, and I emphasised the voluntary nature of the interviews to the students. The instructor was on hand to answer any queries related to the course and the research. Further, the instructor reiterated that participation in the research would remain confidential until the final grade for the course was submitted. Consenting participants chose to (a) allow access to their course work and performance measures, and/or (b) be interviewed.

Students were informed that they would be allowed to withdraw their consent to either aspect of the research by meeting with the Director of Graduate Programs at the University. Our contact information was listed on the consent form. Should students have

chosen to withdraw, their identities would not be disclosed to either the instructor or the teaching assistant. All this information was clearly stated in the consent form.

Students were given two copies of the consent form. One was to be retained by students for their own records. The other was filled out and signed to indicate willingness to participate in the one of the following aspects of the research project: (a) consenting to release their learning logs, self-evaluations, instructor assessment of the learning logs and examination grades for research purposes, or (b) consenting to being interviewed as well as releasing items listed in (a) for research purposes.

Research Design

A quasi-experimental, counter-balanced research design was employed for between-subjects comparisons; the between-subjects factor was the type of instructional aid received (i.e., "worked examples" or "question builders"). Measures of performance, confidence, monitoring proficiencies (i.e., bias and discrimination), independent performance scores, and justification scores were collected over the course of seven weeks and were analysed using a repeated measures procedure. These measures will be discussed in the subsequent sections on "Instruments" and "Scoring Procedures". Please see Table 1 for a summary of the research design used in the study.

Table 1
Research Design

Timeline	Event	Breakdown of Activities	Type of Instructional Aid*		Measures Collected
Week 1	Test of Prior Knowledge	27 Multiple Choice Items on Learning Theory content	None		Performance on Test of Prior Knowledge
Week 1	Log 1	Multistructural and Relational questions provided, students required to answer	None		Performance Assessments, Performance Predictions, Prediction Confidence, Prediction Accuracy, Bias, Discrimination, Independent Multistructural, Relational and Overall Performance, Multistructural and Relational Answer Justification Scores
Week 2	Log 2	Multistructural and Relational Questions and Answers	Group A Group B	WE QB	Performance Assessments, Performance Predictions, Prediction Confidence, Prediction Accuracy, Bias, Discrimination, Independent Multistructural, Relational and Overall Performance, Relational Question and Answer Justification Scores
Week 3	Log 3	Multistructural and Extended Abstract Questions and Answers	Group A Group B	WE QB	Performance Assessments, Performance Predictions, Prediction Confidence, Prediction Accuracy, Bias, Discrimination, Independent Multistructural, Extended Abstract and Overall Performance, Extended Abstract Question and Answer Justification Scores
Week 5 (note: no logs administered during Week 4)	Log 4	Multistructural and Relational Questions and Answers	Group A Group B	QB WE	Performance Assessments, Performance Predictions, Prediction Confidence, Prediction Accuracy, Bias, Discrimination, Independent Multistructural, Relational and Overall Performance, Relational Question and Answer Justification Scores
Week 6	Log 5	Multistructural and Extended Abstract Questions and Answers	Group A Group B	QB WE	Performance Assessments, Performance Predictions, Prediction Confidence, Prediction Accuracy, Bias, Discrimination, Independent Multistructural, Extended Abstract and Overall Performance, Extended Abstract Question and Answer Justification Scores
Week 7	Log 6	Multistructural, Relational and Extended Abstract Questions and Answers	None		Performance Assessments, Performance Predictions, Prediction Confidence, Prediction Accuracy, Bias, Discrimination, Independent Multistructural, Relational, Extended Abstract and Overall Performance, Multistructural, Relational and Extended Abstract Question and Answer Justification Scores

*WE = Worked Examples; QB = Question Builder

Instruments

Test of Prior Knowledge

Students' prior knowledge on learning theories was measured via a 27-item multiple choice test (see Appendix D). Results of the test of prior knowledge were used to divide the students into two groups based on a procedure of matched random assignment.

Task Analyzer and Performance Evaluator (TAPE)

As mentioned earlier in this chapter, for the purposes of my thesis study, I designed a self-assessment TAPE tool, to accompany each learning log. Recall also that the items in TAPE tool were designed, based on empirically supported instructional features to promote task understanding and extract measures of monitoring proficiencies. The TAPE tool (Appendix B) included six questions in total, querying students as to (a) why a particular set was a relational or extended abstract question-and-answer (2 items); (b) what grade they expect to score for their question-and-answer set (1 item), and how well students thought they had performed in relation to their peers (1 item); (c) how confident they were that the mark they have awarded to themselves is, in fact, what the instructor would award them (1 item); and (d) how useful they thought the instructional approach adopted (i.e., "worked examples" or "question builders") for the learning log was (1 item). Not all the TAPE items were used in the analyses for this study, hence I outline those that are relevant to the forthcoming analyses in the subsequent section called "Scoring Procedures".

Scoring Procedures

This section outlines the scoring procedures used for each of the measures collected and used in subsequent analyses. I also provide variable names or labels for each of the measures, so as to help the reader keep track of the variables being used in this study. These variable names are summarized in table 9 at the end of this section.

Instructor's Assessment of Performance on Learning Logs

The instructor assigned a grade to reflect students' performance on each learning log; these grades are termed as *performance assessments* for the purposes of this study. The instructor made it clear that a single grade was assigned for each learning log, and that each question-and-answer set contributed an equal weight to the final grade for a learning log. Therefore, in log 1, each of the two answers contributed equally to the grade assigned (the questions were already provided); in logs 2, 3, 4 and 5, each of the two question-and-answer sets contributed equally to the respective grades on the logs; in log 6, each of the three question-and-answer sets contributed a third of the final grade. Grades assigned by the instructor ranged from A+ to C. For analyses purposes, predicted grades were coded into the following numerical values: 7(A+), 6(A), 5(A-), 4(B+), 3(B), 2(B-), 1(C). Since there was no assigned grade of C+, I assumed an equal-interval scale for the scoring of all grades in this study. Grades were assigned based on how well students had met the criteria for the levels of question-and-answer sets generated for a particular log, the accuracy of the content in the learning logs, as well as how much growth the students had demonstrated in writing their learning logs.

Scoring of TAPE-Related Measures

Justifications of meeting assessment criteria. Items 1 and 2 in the TAPE tool from logs 1 to 5, and items 1 to 6 in the TAPE tool for log 6, were intended to draw out students' perceptions of the task requirements and assessment criteria. Students justified why they thought their question and answer could be described as multistructural, relational or extended abstract. Upon discussion with the instructor, I decided to limit these items to a total of two per learning log for the first five learning logs, so as not to make the self-assessment too tedious for the students. As a result, for logs 2 and 4, the two items were related to the relational question and answer respectively, while in logs 3 and 5, the two items were related to the extended abstract question and answer respectively. In week 1, since two questions, the first multistructural and the second relational, had already been provided for the students, item 1 asked the students to justify why their first answer was multistructural, while item 2 asked students to justify why their second answer was relational. In log 6, students were required to produce three question-and-answer sets, one from each level. Hence, for this sixth and final log, I asked students to respond to why they thought each question and each answer was at the desired level (i.e., multistructural, relational, or extended abstract) yielding a total of six items. In total, students responded to 16 items, over the course of six learning logs, justifying how well they had met the criteria for their questions and answers at a particular level. In answering these items, students justified meeting the criteria that the instructor had set out in the course outline and reinforced in class for writing multistructural, relational and extended abstract questions and answers. The instructor's feedback to these responses

provided a measure of how accurate student's perceptions of the task requirements in relation to the instructor's own, at a particular stage of the instruction.

To facilitate a quantitative analysis of each participant's 16 responses to items that required them to justify meeting the criteria for a particular level of question and answer, I independently scored students' answers to the first two items in logs 1,2,3, 4 and 5 and the first six items in log 6. A score of 0 was assigned to responses that did not mention the proper criteria for the required level of question or answer (e.g., the student did not succeed at all in justifying why the question written was relational); a score of 1 was assigned to responses that partially succeeded in responding to the item (e.g., the student gave an incomplete explanation for why a particular answer was multistructural); a score of 2 represented a complete and correct response to the query posed in the item. Table 2, provides a summary of the scoring for TAPE items related to students' justifications of meeting the criteria for writing a question and answer at the required level.

Table 2

Scoring of TAPE Items 1 and 2 (Justifying why question/answer meets criteria)

Scale	Score	Criteria
Justification	0	Inappropriate criteria and/or incorrect explanation provided
of Meeting	1	Partially correct explanation provided
Criteria	2	Complete and correct explanation provided

Prediction of performance. Item 3 in the TAPE tool asked students to assign or predict a grade, ranging from A+ to C, for their completed learning log; this measure is

labeled as *performance predictions*. For analyses purposes, these predicted grades were coded into the following numerical values: 7(A+), 6(A), 5(A-), 4(B+), 3(B), 2(B-), 1(C). The range of grades is reflective of the range that the instructor assigned in the assessment of students' performance in the learning logs.

Accuracy in prediction. Students' accuracy in predicting their grades for the learning logs is the absolute value of the signed difference between the predicted grade on a particular learning log and the instructor's assessment of performance on the same learning log. This measure is labeled as *prediction accuracy*. Integer values of prediction accuracy ranged from 6 to 0. A value of zero represents a perfectly accurate prediction. The larger the value of accuracy, the more inaccurate a student's prediction. Note that this calculation of accuracy does is *not* equivalent to the calculation of accuracy performed in Schraw & Nietfeld's (1998) study, who characterized accuracy as a monitoring proficiency. Recall also that the only prediction participants made in Schraw & Nietfeld's (1998) study was for confidence in performing correctly on the multiple-choice tests of fluid and crystallized abilities; in my thesis study, however, the difference between predicted and actual performance represents accuracy in prediction, and does not factor the value of confidence.

Confidence in self-assessment of academic performance. Item 4 asked students to rate how confident they were about receiving the same grade they have awarded themselves from the instructor; this measure is labeled as *prediction confidence*. Students responded on a 4-point Likert-type scale, including, 4("Very Confident"), 3("Somewhat Confident"), 2("Not So Confident") and 1("Not Confident At All").

Monitoring Proficiencies: Discrimination. The three measures of performance predictions, performance assessments, and prediction confidence were used to calculate two measures of monitoring proficiency. The first measure of monitoring proficiency calculated was *discrimination,* which measured the degree to which learners could assign an appropriate level of confidence to their predictions of the grade for each learning log. Discrimination was calculated by taking the signed difference between the average prediction confidence scores for accurate predictions and the average prediction confidence scores for inaccurate predictions. Discrimination was calculated for each of the logs, yielding a total of six discriminations scores over the course of instruction. The value of discrimination ranged from −1 to +1. A negative value represents confidence for inaccurate predictions, while positive values represent confidence for accurate predictions. A discrimination value close to zero suggests that the learner was incapable of discriminating between accurate and inaccurate predictions. This means that students with a large, positive value of discrimination (i.e., close to +1) are very proficient in monitoring as it suggests that they can assign a high value of confidence when accurately predicting their grades on the learning log assignment. The closer the value of discrimination to 1, the more accurate was a student's monitoring.

To calculate discrimination, prediction confidence scores were first converted to proportions; each score was divided by 4, which is the maximum possible value of confidence. For log 1, if the students' performance prediction was equal to the performance assessment, then the converted prediction confidence score was assigned as the discrimination score; if the prediction was inaccurate, the negative value of the converted prediction confidence score was assigned as the discrimination score. For

subsequent logs, discrimination was calculated by taking the average of the signed, converted prediction confidence score (using the same procedures as described for log 1) and the previous log's discrimination score. This means that the score of discrimination for log 2 represents the student's ability to discriminate, based on predictions from both logs 1 and 2. Discrimination scores for log 6 provide a measure of the students' abilities to discriminate, based on predictions from all six learning logs. Note that this procedure for calculating discrimination does *not* mirror the procedure used in Schraw and his colleagues' studies (Schraw et al., 1995; Schraw & Nietfeld, 1998), since they did not factor students' predictions of performance in their procedures.

Monitoring Proficiencies: Bias. The second measure of monitoring proficiency calculated was *bias*. Bias measured the degree to which individuals were over or under-confident for each TAPE self-evaluation made. Bias was calculated by taking the signed difference between the prediction confidence score and prediction capability. Like the discrimination score, bias ranged in value from –1 to +1. A negative value of bias indicated under-confidence, whereas positive values indicated overconfidence in predicting grades; the larger the negative value of bias, the more under-confident the learner, the larger the positive value, the more overconfident the learner in predicting grades. This would suggest that students with a score of bias close to 0 have good monitoring proficiency, as they assign an appropriate level of confidence to their predictions. Bias was calculated for each log, giving each student six bias scores. To calculate bias, the confidence scores were converted to proportions by dividing each confidence score by 4. Prediction capability was calculated by taking the ratio of the smaller of the two values of performance prediction and performance assessment with the

larger of the two. Prediction capability hence measured how well the student had predicted a grade for a particular learning log. For example, if a student predicted a B+ and received a B+ from the instructor, the value of prediction capability would be calculated as the ratio of 4 to 4 (4 is the score for a grade of B+), yielding a score of 1, suggesting 100% prediction capability. If the student overestimates performance and predicts an A for the learning log, but, in fact receives a B, then prediction capability is calculated as the ratio of 3 to 6 (3 and 6 are the scores for B and A respectively), yielding a score of 0.5. This suggests that the student was able to predict only 50% of the actual grade received, thereby yielding a score of prediction capability of 0.5. If the student underestimates performance by predicting, for example, a B, but receiving an A- from the instructor, the prediction capability was calculated as the ratio of 3 to 5 (3 and 5 being the scores for B and A- respectively), which gives a score for prediction capability as 0.6. This suggests that the student was able to predict 60% of the final grade received. This procedure for calculating bias is different from the one used by Schraw et al. (1995) and Schraw & Nietfeld (1998); Schraw and his colleagues' procedures did not factor predictions of performance into the calculation of bias.

It is important to acknowledge that the value of bias is influenced more heavily by the converted prediction confidence scores than by the calculated prediction ability scores. This is because the converted prediction confidence scores can only take one of four values, with a fixed interval of 0.25. The values possible for converted prediction confidence are 0.25, 0.50, 0.75 or 1(i.e., the ratio of the integers 1, 2, 3 and 4 with 4). However, the calculated prediction ability can take the value of any ratio of two integers between 1 and 7, the smaller of the two integers being the numerator of the ratio except if

the two integers are equal (which yields a ratio of 1). Obviously, there are opportunities for intervals smaller than 0.25 to exist from one calculated score of prediction ability to another. For bias to be equally influenced by prediction confidence and prediction ability scores, I would have had to measure prediction confidence and performance assessments on the same scale of measurement. Given that the grading scheme for the logs was decided by the instructor, this would have meant trying to create a 7-point scale for measuring students' prediction confidence. However, such an exercise would have provided too fine-grained a scale for measuring prediction confidence, and might have proven to be tedious enough for the students that it would have resulted in them not responding, or inaccurately responding, to the item in the TAPE. Therefore, my calculation of bias is affected by the difference in the scales of measuring performance and confidence; in order to get students to respond in as honest a manner as possible to the items on prediction confidence, I was forced to measure it on a 4-point scale.

Independently Scored Student Performances on Learning Logs

While the instructor's performance assessments on individual students' learning logs provided one measure of academic performance, these grades were not always the sole indicator of students' abilities to meet the criteria for writing each level of learning log question-and-answer sets. Students were informed by the instructor that they would be graded for an element of growth in their ability to write the learning logs; the grade assigned to each log was not just a reflection of the student being able meet the assessment criteria, but also of the student being able to demonstrate growth in developing skills the instructor perceived was necessary in writing learning logs at a graduate level of studies. Further, the grade gave a combined score of students' responses

to multiple question-and-answer sets; for example, it was not possible to discern, from the final grade for a log, how well a student had performed in the multistructural level as in comparison with the relational or extended abstract level. Therefore, in order to represent students' academic performances on each question-and-answer set of each learning log, independent of growth across learning logs, I scored each student's learning logs, at each of the levels (i.e., multistructural, relational and extended abstract), based on the instructor's criteria for evaluating the question-and-answer sets. The scoring scheme was developed in close discussion with the instructor and teaching assistant of the course to ensure that I was, in fact, scoring for criteria that were integral to the assignment. This scoring procedure helped me to develop a parallel set of *independent multistructural performance scores, independent relational performance scores, independent extended abstract performance scores* and *independent overall performance scores.*

Questions. All three levels of questions (i.e., multistructural, relational and extended abstract) were scored on a scale of 0 to 2 for appropriateness in questioning style (e.g., differences, compare-contrast, cause-effect, providing example, hypothesis testing, evaluation, etc.) and appropriate choice of concept(s) for the desired level of the question. A score of 0 represented inappropriate choice of concepts *and* inappropriate questioning style (e.g., choosing a concept that is too simple for relational level, too complex for a multistructural level or asking a complex hypothesis testing question at the multistructural level). A score of 1 represented choosing inappropriate concepts for the desired level (e.g., comparing one simple and one complex idea in a relational question, or choosing a simple concept which is not explicitly detailed in the text for a multistructural question) *or* inappropriate questioning style for the desired level. Finally,

a score of 2 represented appropriate choice of concepts for the desired level of the question *and* appropriate questioning style used for the desired level. Table 3 summarizes the scoring scheme for the learning log questions.

Table 3

Scoring of Multistructural, Relational and Extended Abstract Learning Log Questions

Scale	Score	Criteria
Appropriate Choice of Concept and Questioning Style	0	Inappropriate concept(s) chosen AND inappropriate questioning style for level of question
	1	Inappropriate concept(s) chosen for the type of question OR inappropriate questioning style for level of question
	2	Appropriate concepts chosen AND appropriate questioning style for level of question

Multistructural Answers. Students' answers to their multistructural learning log questions were scored for quality of explaining and linking concepts, measured on a scale from 0 to 2. A score of 0 represented incorrect explanations of the concepts and lack of linkage between the concepts, if and when a linkage is necessary (e.g., incorrect definitions or explanations of concepts and no linkage between the concepts). A score of 1 represented either incorrect/unclear explanations of concepts or lack of linkage between the concepts, if and when a linkage is necessary. Finally, a score of 2 represented a cohesive and correct explanation of concepts as well as a linkage between concepts, when necessary. Table 4 summarizes the scoring scheme for multistructural answers.

Table 4

Scoring of Multistructural Answers

Scale	Score	Criteria
Quality of Explanation and Linkage of Concepts	0	Incorrect explanation AND lack of linkage between concepts (when necessary)
	1	Incorrect/unclear explanation OR lack of linkage between concepts (when necessary)
	2	Cohesive and clear explanations AND linkage between concepts (when necessary)

Relational Answers. Students' answers to their relational learning log questions were scored for quality of explanation, measured on a scale from 0 to 2. A score of 0 represented incorrect explanations of the concepts (e.g., incorrect definitions or explanations of concepts). A score of 1 represented an attempt at explaining the concepts which falls short of being cohesive and clear. Finally, a score of 2 represented a cohesive and correct explanation of concepts.

Relational answers were also scored for level of complexity employed in explaining the link between concepts, measured on a scale of 0 to 2. A score of 0 represented an oversimplification of the concepts leading to weak linkages between the concepts; or a lack of links drawn between concepts, sometimes offering disjointed explanations of key terms and concepts raised in the question. A score of 1 represented a higher level of complexity in links than a score of 0, however these links were explicitly

drawn from the text, or there was evidence of some implicitness in the links, but it was not sufficiently explored. Finally, a score of 2 represented links that students had implicitly generated, or had expanded substantially on material from the text. Table 5 summarizes the scoring scheme for relational answers.

Table 5

Scoring of Relational Answers

Scales	Score	Criteria
Quality of Explanation	0	Incorrect explanation of concepts
	1	Partially clear or incomplete explanations of concepts
	2	Cohesive and clear explanations of concepts
Level of Complexity of Links	0	Weak linkages due to oversimplification of concepts or no links drawn between concepts, only disjointed explanations
	1	Explicit linkages between concepts drawn directly from text or beginnings of implicit linkages seen but not explored
	2	Student generates own implicit linkages or expands substantially on material from text

Extended Abstract Answers. Students' extended abstract answers were scored on two scales. The first, quality of explanation, was the same as those used for relational answers. The second scale, applicable only to extended abstract answers, measured, on a scale from 0 to 2, the level of connection made in tying components of the extended abstract answer to the theory being exemplified. A score of 0 represented an absence of

or a minimal connection to the theory referred to in the extended abstract question (e.g., naming theoretical procedures or concepts in relation to the application). A score of 1 represented the presence of a partial but insufficient connection to the theory; for example, if an answer lacked an explanation of how the "nuts and bolts" of the theory fit with the practical application or hypothesis. Finally, a score of 2 represented an acceptable and relatively complete connection made to the theory, with an effort made at explaining the "nuts and bolts" of the theory that supports the application or hypothesis. Table 6 summarizes the scoring scheme for extended abstract answers.

Table 6

Scoring of Extended Abstract Answers

Scales	Score	Criteria
Quality of Explanation	0	Incorrect explanation of concepts
	1	Partially clear or incomplete explanations of concepts
	2	Cohesive and clear explanations of concepts
Level of Connection Made to Theory	0	Absence of or minimal connection to theory
	1	Partial connection made to theory
	2	Relatively complete connection made to theory

Match Between Question and Answer. For each level of question and answer generated by students (i.e., multistructural, relational, and extended abstract), the match between the question and answer was scored on a scale of 0 to 1. A score of 0 represented

a complete mismatch between the levels of question and answer (e.g., if the question-and-answer set was supposed to be multistructural: the question was multistructural in nature, but the answer was more relational in its complexity). A score of 1 represented a match between the levels of the question and answer (e.g., for an extended abstract log, the author made an attempt to keep both question and answer at an extended abstract level). Table 7 summarizes the scoring scheme for the match between levels of question and answer.

Table 7

Scoring of Match Between Question and Answer

Scale	Score	Criteria
Match between Levels of Question and Answer	0	Mismatch between levels of question and answer
	1	Match between level of question and answer

Calculating Total Scores. For the multistructural level, the maximum total score was 5 (2 from the questions, 2 from the answers and 1 for the match); the relational level had a maximum total score of 7 (2 from the question, 4 from the answer and 1 from the match); the extended abstract level had a maximum total score of 7 (2 from the question, 4 from the answer and 1 from the match). For each of the levels of questions, total scores

can be easily compared within and across students. Table 6 summarizes the calculation of maximum total scores discussed above.

Table 8

Calculation of Maximum Total Scores for Each Level

		Level		
		Multistructural	Relational	Extended Abstract
Max. Score	Question	2	2	2
	Answer	2	4	4
	Match	1	1	1
Total Max. Score		5	7	7

All scoring underwent inter-rater reliability checks with the teaching assistant of the course; 95% reliability was achieved. All disagreements were resolved through discussion.

Facilitating comparison of total scores across logs. Using the scoring scheme outlined in table 6, total scores were computed for each student on each of the 11 question-and-answer sets generated from learning log 2 through learning log 6. For each student, therefore, five *independent multistructural performance scores* (one each from logs 2, 3, 4, 5 and 6), three *independent relational performance scores* (one each from logs 2, 4 and 6), and three *independent extended abstract performance scores* (one each from logs 3, 5 and 6) were calculated. In addition, there was one multistructural answer and one relational answer from log 1 (the questions were provided by the instructor). In

order to make the scores from log 1 comparable with those from the other logs, all scores were converted to a percentage.

An *independent overall performance score* was also calculated for each student on each learning log; this overall performance score served as the counterpart to the grade assigned by the instructor for each learning log (i.e., performance assessments). Independent overall performance scores for learning log 1 attached an equal weight to the independent multistructural performance score and independent relational performance score for log 1; this meant taking the average of the independent multistructural performance score and independent relational performance score for log 1. For learning logs 2 and 4, the independent overall performance score was the average of the independent multistructural performance score and independent relational performance score for the respective logs. For learning logs 3 and 5, the independent overall performance score was the average of the independent multistructural performance score and independent extended abstract performance score for the respective logs. Finally, for log 6, the calculated total score was the average of the independent multistructural, relational and extended abstract performance scores from that log. Due to the nature of the scoring scheme, I acknowledge a discrepancy. Note that the independent multistructural performance scores contain an equal weighting for the questions and the answers, with maximum possible scores of 2 each. However, the independent relational and extended abstract performance scores have unequal weighting for the questions and answers, with maximum possible scores of 2 and 4 for questions and answers respectively. Hence, independent relational and extended abstract performance scores carry more weight for the answers than the questions, as opposed to the multistructural

level, where both question and answer carried equal weight. I discussed this issue with the teaching assistant of the course, and we came to an agreement that such a balance in weighting made sense, because the relational and extended abstract answers were expected to be more difficult to write than multistructural answers. The implicit linkages that needed to be drawn in relational answers and the connections with theories that had to be made for an answer to be considered at an extended abstract level were expected to be more difficult to achieve than the direct and explicit comparisons that made an answer multistructural.

Summary. Table 9 provides a summary of the measures described in this "Scoring Procedures" section.

Table 9

Details of variable names/labels for measures collected or derived

Measure	Variable Name/Label	No. of Scores
Instructor's assessments of performances	Performance Assessments	6
Students' predicted performances	Performance Predictions	6
Students' confidences in predictions	Prediction Confidence	6
Students' accuracies in predictions	Prediction Accuracy	6
Students' biases in predictions	Bias	6
Students' discriminations in predictions	Discrimination	6
Independently calculated scores for overall performances	Independent Overall Performance Score	6

Table 9 (continued)

Measure	Variable Name/Label	No. of Scores
Independently calculated scores for performances on relational level	Independent Relational Performance Score	4
Independently calculated scores for performances on extended abstract level	Independent Extended Abstract Performance Score	3
Independently calculated scores for justifications of meeting criteria for creating multistructural-level question	Multistructural Question Justification Score	1
Independently calculated scores for justifications of meeting criteria for creating multistructural-level answer	Multistructural Answer Justification Score	2
Independently calculated scores for justifications of meeting criteria for creating relational-level question	Relational Question Justification Score	3
Independently calculated scores for justifications of meeting criteria for creating relational-level answer	Relational Answer Justification Score	4
Independently calculated scores for justifications of meeting criteria for creating extended abstract-level question	Extended Abstract Question Justification Score	3

Table 9 (continued)

Measure	Variable Name/Label	No. of Scores
Independently calculated scores for justifications of meeting criteria for creating extended abstract-level answer	Extended Abstract Answer Justification Score	3

Chapter 3: A Description of Results and Findings

Results

Missing Values

Data were entered into a statistical analysis package, namely, the Statistical Package for Social Sciences (SPSS), and were verified twice for errors. There were four instances of missing TAPE data from four different students. Two of these were scores of peer evaluation, one in log 1 and one in log 3, each with unique sources. The other missing value was for a score for usefulness of instructional aid received in log 3 by a student different from those who did not respond to the peer evaluation items. The fourth case of missing data came from an individual who did not submit learning log 6 for grading. This student, therefore, did not have any performance scores (both independently calculated and the instructor's assessments) for log 6. Missing values for any variable resulted in the case in question being excluded from any analyses involving the variable. Due to the small sample size and the small number of missing values, I decided not to replace any missing data for fear of misrepresenting students' responses to the TAPE items and performance.

Establishing Group, Gender Equivalences

I confirmed both group and gender equivalence on the test of prior knowledge before students received any instruction. Recall that grouping was based on order of instructional aid received. One group of students received "worked examples" for logs 2 and 3 followed by "question builders" for logs 4 and 5 (M score for test of prior knowledge=8.00, SD=2.00), while the other received "question builders" for logs 2 and 3

and "worked examples" for logs 4 and 5 (M=8.67, SD=2.00). Since participants underwent a random matched assignment to groups, no differences were expected in each group's mean performance on the test of prior knowledge. Independent-samples t-test procedures indicated that there was no statistically detectable difference between the groups on their mean scores on the test of prior knowledge, $t(15)$=.686, p=.50. I used a similar procedure to establish gender equivalence; there was no significant difference between male (M=8.00, SD=3.39) and female (M=8.50, SD=1.17) performance on the test of prior knowledge, $t(4.40)$=.32, p=.76.

Collapsing Groups

In order to conduct analyses on a *single* sample of the 17 consenting students, I needed to provide sufficient evidence to collapse the two experimental groups, which were based on order of instructional aid received, into one single group.

Independent sample t-tests were conducted between the two experimental groups, on the list of dependent measures seen in table 9 (from the previous chapter).

Experimental group differences. The two experimental groups differed significantly only on three scores from those described in table 9. Table 10 details the scores wherein the group differences were found.

Table 10

Observed differences between experimental group conditions for all measures

Measure	Group[a]	*M*	*SD*	*t*	df	*p*	95% confidence interval for mean difference (*MD*)	Effect Size [b]
Relational Answer Justification Score for Log 2	Group A (*n*=8)	1.88	.35	2.47	15	.026	.089<*MD*<1.22	1.21
	Group B (*n*=9)	1.22	.67					
Independent Relational Performance for Log 4	Group A (*n*=8)	55.00	27.77	-2.20	15	.044	-53.63<*MD*<-.81	-1.07
	Group B (*n*=9)	82.22	23.33					
Prediction Confidence Score for Log 3	Group A (*n*=8)	2.88	.35	-2.20	14.34	.045	-.90<*MD*<-.013	-1.03
	Group B (*n*=9)	3.33	.50					

[a] Group A refers to between-groups condition in which students received "worked examples" for logs 2 and 3, and "question builders" for logs 4 and 5 as instructional aids. Group B refers to between-groups condition in which the order of instructional aids was reversed.

[b] Effect size was calculated by taking the signed difference between the mean of the measure for group A and the mean of the measure for Group B, and dividing the resultant difference by the pooled standard deviation of the two samples.

Although the effect sizes for the differences are rather large, this can be explained by the small sample size; it would take only one outlier to increase the mean of any given group, and hence increase the effect size. Furthermore, the fact that three out of 71 scores (4.2%) showed significant differences between experimental groups can be attributed to being an artifact of the experimental process; these differences can be attributed to chance and not as a result of the order of instructional aids received.

Interaction effects. To test the significance of the interaction effect between the order of instructional aids received and the various measures outlined in table 9, I conducted a set of between-groups, repeated measures, analyses of variance. No interaction effect was found between the order of instructional aid received and the following set of repeated measures: performance assessments, performance predictions, prediction accuracy, prediction confidence, bias and discrimination, $F(30, 345)=.786$, $p=.785$, $\eta^2=.06$. An interaction effect, was however found, between the order of instructional aid received and the following repeated measures: independent overall performance score and independent multistructural performance score, $F(10, 140)=2.58$, $p=.007$, $\eta^2=.16$. No interaction effects were found between the order of instructional aid received and the repeated measures of independent relational performance score and relational answer justification score, $F(6, 84)=1.97$, $p=.08$, $\eta^2=.12$. Finally, no interaction effects were found between the order of instructional aid received and the independently calculated repeated measures of independent extended abstract performance score, relational question justification score, extended abstract question justification score, extended abstract answer justification score, $F(8, 52)=1.09$, $p=.38$, $\eta^2=.14$.

The only interaction effect that was found to be significant can be accounted for by one of the chance statistically significant differences noted in table 10: the order of instructional aid received affected the independently calculated score for multistructural level sets for log 4. Based on the lack of interactions in the other analyses conducted, I decided to collapse the data for the two groups, and conduct all subsequent analyses on one sample of 17 students.

Organization of Results

To answer the first research question, namely, how graduate learners' monitoring proficiencies develop over the course of instruction, I conducted analyses on the measures of performance assessments, performance predictions, prediction confidence, prediction accuracy, bias and discrimination. For the second research question, namely, whether learners' abilities to meet the assessment criteria for the writing task were reflected in the instructor's performance assessments, I performed analyses on the independent scores of performance on the multistructural, relational, extended abstract levels as well as overall performance, and compared these scores to the performance assessments analysed for the first research question. For the third and final question, which explores the relationship between students' understandings of assessment criteria and the monitoring proficiencies, I analysed the justification scores for relational, extended abstract and multistructural questions and answers and explored their relation with the monitoring proficiencies analysed for the first research question.

Each of the measures, identified above by the three research questions, underwent most of the following analytical procedures: (a) within-group, repeated measures analyses of variance, (b) trend analyses, (c) unrestricted, corrected pairwise comparisons

analyses, and (d) correlational analyses. Given the complexity of the analyses involved, I will be following Salovey's (2000) suggestions about embedding discussion issues in my results, as well as justifying the selection of statistical tests and procedures in the results section. This way, the reader will get a clear picture of the rationale behind the use of certain statistical analyses, and the answers to the research questions will emerge, piece by piece, from the findings of each analysis. To help the reader navigate the results, I include sections that provide a summary of the findings from the repeated measures, trend and pairwise comparisons analyses for a particular set of measures (e.g., independent performance scores, justification scores). These summarized findings, combined with the findings of the subsequent correlational analyses will provide the basis for the contents of the discussion chapter.

The results section begins with descriptive statistics for the various measures described in table 9. Next, the first category of measures, namely, the performance assessments, performance predictions, prediction confidence, prediction accuracy, as well as bias and discrimination undergo repeated measures, trend and pairwise comparisons analysis. The findings related to these performance and monitoring proficiencies are then summarized and briefly discussed. The second category of measures, namely, the independent scores of performance on multistructural, relational, and extended abstract levels, as well as those for overall performance undergo repeated measures, trend and pairwise comparisons analysis. The findings related to the independent scores of performance are then summarized. Finally, the justification scores undergo repeated measures, trend and pairwise comparison analyses, which are then summarized. As discussed earlier, through these three summaries, I hope to provide the reader with a

general idea of how the analyses of the (a) performance and monitoring-related measures, (b) independent scores of performance, and (c) justification scores, contribute to partially answering of each of the research questions. Finally, I conduct a set of correlational analyses for each of the three categories of measures, and provide a description of the findings from each of these analyses. Together, the repeated measures, trend, pairwise comparisons, and correlational analyses provide answers to the three research questions posed.

Descriptive Statistics

Table 11 provides descriptive statistics (means and standard deviations) for performance assessment, performance prediction, prediction confidence, prediction accuracy, bias and discrimination, across the six logs. Table 12 provides descriptive statistics for the independent scores of overall performance, as well as performance on the multistructural, relational and extended abstract levels, in addition to the justification scores for multistructural, relational, and extended abstract questions and answers.

Table 11

Descriptive statistics for Performance Assessments and TAPE-related Measures

Log	Performance Assessments (maximum score: 7)		Performance Predictions (maximum score: 7)		Prediction Confidence (maximum score: 4)		Bias (range: -1 to 1)		Discrimination (range: -1 to 1)		Prediction Accuracy (range: 0 to 6)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	4.06	.66	3.82	1.38	3.00	.50	-.025	.22	-.31	.72	1.18	1.13
2	3.88	.86	3.94	1.09	2.88	.49	-.016	.26	-.088	.50	.53	.72
3	4.53	.62	4.18	.95	3.12	.49	-.025	.23	-.037	.56	.59	.71
4	4.82	.64	4.47	1.01	3.35	.61	-.055	.25	-.011	.54	.71	.92
5	5.24	.75	4.76	.83	3.41	.51	-.065	.18	.11	.50	.47	.72
6	5.06	1.29	4.63	.89	3.44	.51	.073	.16	-.047	.45	.94	.93

Note. n for all logs was 17, except for log 6, where n was reduced to 16

Table 12 *Descriptive Statistics for Independently Calculated Scores for Justification and Performance*

Log	Independently Calculated Scores																	
	Multistructural Question Justification (maximum score: 2)		Multistructural Answer Justification (maximum score: 2)		Relational Question Justification (maximum score: 2)		Relational Answer Justification (maximum score: 2)		Extended Abstract Question Justification (maximum score: 2)		Extended Abstract Answer Justification (maximum score: 2)		Independent Multistructural Performance Score (maximum score: 100)		Independent Relational Performance Score (maximum score: 100)		Independent Extended Abstract Performance Score (maximum score: 100)		Independent Overall Performance Score (maximum score: 100)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	-		1.29	.77	-		1.47	.80	-		-		76.47	22.87	47.06	31.58	-		61.76	23.92
2	-		-		.82	.95	1.53	.62	-		-		72.94	24.43	76.47	23.66	-		74.71	16.50
3	-		-		-		-		1.41	.94	1.29	.69	78.82	23.95	-		73.95	20.35	76.39	16.32
4	-		-		1.47	.72	1.35	.79	-		-		69.41	28.39	88.24	14.49	-		78.82	16.38
5	-		-		-		-		1.59	.62	1.59	.71	87.06	19.93	-		83.19	18.37	85.13	14.83
6	.87	.96	.75	.93	.94	.68	1.00	.82	1.69	.70	1.56	.63	81.25	23.63	82.14	23.04	73.21	18.72	78.87	12.87

Note: *n* for all the logs was 17, except for log 6, where the *n* was reduced to 16.

Repeated Measures Analysis for Performance, Confidence and Monitoring Measures

Performance assessment, performance predictions, and prediction confidence showed statistically significant differences across the six logs, as evidenced by the main effects reported in table 13. Bias, discrimination and prediction accuracy, on the other hand, fluctuated according to chance across the six logs, as evidenced in the statistically insignificant main effects revealed in table 13.

Table 13

Repeated measures, within-group one-way analyses of variances on measures of performance assessments, performance predictions, prediction confidence, bias, discrimination, and prediction accuracy.

Measure	*F*	df	*MSE*	*p*	η^2	Power	Effect Size [a]
Performance Assessments	11.33	5	.38	<.001	.43	1.00	.87
Performance Predictions	5.70	3.73[b]	1.22	.001	.28	.96	.62
Prediction Confidence	3.60	5	.22	.006	.19	.91	.48
Bias	1.29	5	.03	.28	.08	.43	.29
Discrimination	1.61	5	.18	.17	.10	.53	.33
Prediction Accuracy	2.21	5	.56	.06	.13	.69	.39

[a] Effect size is calculated using Cohen's (1998) procedure for calculation of *f*, which is equivalent to $\sqrt{\frac{\eta^2}{1-\eta^2}}$, where η^2 is interpreted as the proportion of the total variability in the dependent variable that is accounted for by variation in the independent variable; η^2 is the ratio of the between groups sum of squares to the total sum of squares.

[b] Degrees of freedom was reduced using the Huynh-Feldt correction for violation of Mauchly's test of sphericity.

Trends. Performance assessments, performance predictions, and prediction confidence, each revealed a statistically significant, increasing linear trend (see table 14 for results of trend analyses and figures 1 to 3 for a visual representation). Prediction accuracy, which showed no main effect for differences across the six logs, revealed, however, a statistically significant quadratic trend, $F(1, 15)=8.55$, $p=.01$, $\eta^2=.36$ (see figure 6 for a visual representation). It is also interesting to note that the fluctuations in performance assessment across the six logs revealed less, yet statistically significant, quadratic, $F(1, 15)=5.87$, $p=.03$, $\eta^2=.28$; and cubic trends, $F(1, 15)=15.37$, $p=.001$, $\eta^2=.51$, as compared to the more significant (p values less than .001) linear trend shown in table 14.

Table 14

Main effects for trend analyses (linear) for performance assessments, performance predictions, and prediction confidence scores

Measure	*F*	df	*MSE*	*p*	η^2	Power	Effect Size [a]
Performance Assessments	25.40	1	12.64	<.001	.63	1.00	1.30
Performance Predictions	14.33	1	20.36	.002	.49	.94	.98
Prediction Confidence	9.20	1	3.43	.008	.38	.81	.78

[a] Effect size is calculated using Cohen's (1998) procedure for calculation of *f*, which is equivalent to $\sqrt{\frac{\eta^2}{1-\eta^2}}$, where η^2 is interpreted as the proportion of the total variability in the dependent variable that is accounted for by variation in the independent variable; η^2 is the ratio of the between groups sum of squares to the total sum of squares.

Pairwise Comparisons For Performance, Confidence and Monitoring Measures

Choice of tests: Fisher and Tukey. In order to determine where the differences lay for each of the measures of performance assessments, performance predictions, prediction confidence, prediction accuracy, bias and discrimination, I next conducted unrestricted pairwise comparisons using two correction procedures, namely, the Fisher test (also known as the protected Least Significant Differences [LSD] test) and the Tukey test (for detailed explanations of how to conduct these two tests, see Keppel, 1982). While the Tukey test maintains the family-wise alpha rate at the value of .05 for the

entire set of pairwise comparisons, the Fisher test exerts its control by conditionalizing the decision to conduct pairwise comparisons on the significance of the omnibus F test. The Tukey test is, therefore, more conservative than the Fisher test; if statistically significant differences are found between a pairwise comparison using the Tukey test, the same difference will be statistically significant using the Fisher test.

Another critical difference that between the two correction procedures worth exploring is that the Fisher test for pairwise comparisons can only be conducted when the omnibus F test is found to be statistically significant. The Fisher test, therefore, does not involve any special correction once the omnibus *F* is found to be statistically significant. The Tukey test however, employs a distinct correction procedure, and can be performed on any of the aforementioned measures, regardless of whether the omnibus *F* test yields a statistically significant value. With respect to my thesis study, in the case of the repeated measures analyses explored so far, the Fisher test would be applicable only to the instructor's assessments of performance, students' predictions of performance, and students' confidences at prediction.

I decided to use two, separate correction procedures, following Keppel's (1983) call for a procedure to allow for suspension of judgment (p. 163). Under such a procedure, if the calculated difference found for a pairwise comparison exceeded the uncorrected critical difference (i.e., the critical difference in the Fisher test) but does not exceed the corrected critical difference (i.e., the critical difference found in the Tukey test), then I would decide to suspend judgment on whether a statistically significant difference existed for the comparison. Only if the critical difference exceeded the corrected critical difference from the Tukey test, then I would be able to declare the

difference as statistically significant at a p level of .05. By deciding to take no formal action of rejection or acceptance of the null hypothesis, I would, therefore, avoid committing either a Type I or Type II error.

The formula for calculating the critical difference for the Tukey test (d_T) is $q_T \times \sqrt{\frac{MSE}{n}}$; where q_T refers to an entry in a table of the studentized range statistic under the number of means being compared and the degrees of freedom associated with the error term, *MSE* refers to the mean square error term from the omnibus F test, and n refers to the number of participants contributing to each instance of the measure. The formula for calculating the critical difference for the Fisher test (d_F), once the omnibus F test has been shown to be significant, is $t \times \sqrt{\frac{2 \times MSE}{n}}$; where t is the value of the t-distribution under the chosen value of alpha and the degrees of freedom associated with the error term, *MSE* refers to the mean square error term from the omnibus F test, and n refers to the number of participants contributing to each instance of the measure. Note that the assumption I am making in the calculation of both d_T and d_F is that the value of n is equal in each instance of the measure being collected. This assumption is not violated in the calculation; the value of n is 16 from a maximum possible value of 17, since all data for the participant who did not complete log 6 and the accompanying TAPE have been excluded from any repeated measures analyses.

Comparisons for performance assessments and performance predictions. Performance assessments showed statistically significant differences only after students had written their first three logs, while performance predictions showed statistically

significant differences only across logs that were at least four periods apart (e.g., between logs 1 and 5; see table 15).

Table 15

Pairwise comparisons for performance assessments (lower triangle) and performance predictions (upper triangle)

	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	-	-.19	-.50	-.75*	-1.06**	-1.31**
Log 2	-.19	-	-.31	-.56	-.88*	-1.13**
Log 3	.44	.63*	-	-.25	-.56	-.81*
Log 4	.75**	.94**	.31	-	-.31	-.56
Log 5	1.25**	1.44**	.81**	.50*	-	-.25
Log 6	.56*	.75**	.13	-.19	-.69**	-

Absolute value of d_T for performance assessments = .64

Absolute value of d_F for performance assessments = .44

Absolute value of d_T for performance predictions = 1.02

Absolute value of d_F for performance predictions= .69

Each entry in the table represents the mean difference calculated by taking the signed difference between the measure for the row and the measure for the column. Non-asterisked entries represent non-significant calculated mean differences.

* Suspend judgment on whether calculated mean difference is statistically significant, because calculated mean difference was found to be greater than or equal to the uncorrected critical difference found by the Fisher test, but less than the corrected critical difference found from the Tukey test.

** Statistically significant differences found between calculated mean differences, which is greater than or equal to the corrected critical difference found from the Tukey test.

Comparisons for prediction confidence and prediction accuracy measures.

Prediction confidence showed statistically significant differences between logs 1 and 5, and logs 2 and 6, while prediction accuracy showed no significant differences across the six logs (see table 16). The six measures of accuracy in predictions did not yield a significant omnibus *F* value from the repeated measures analysis of variance, however, a glance at pairwise comparisons between their means revealed some statistically significant differences, without any correction for maintaining a family-wise alpha rate of .05. Therefore, I decided to conduct a corrected, pairwise comparison for the six accuracy measures, using only the corrected critical difference derived from the Tukey test, since the Fisher test could not be used.

Table 16

Pairwise comparisons for prediction confidence (lower triangle) and prediction accuracy (upper triangle)

	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	-	.63	.69	.50	.69	.25
Log 2	-.13	-	.06	-.13	.06	-.38
Log 3	.13	.25	-	-.19	0	-.44
Log 4	.31	.44*	.19	-	.19	-.25
Log 5	.38*	.50**	.25	.06	-	-.44
Log 6	.44*	.56**	.31	.13	.06	-

Absolute value of d_T for prediction confidence = .49

Absolute value of d_F for prediction confidence = .33

Absolute value of d_T for prediction accuracy = .78

Each entry in the table represents the mean difference calculated by taking the signed difference between the measure for the row and the measure for the column. Non-asterisked entries represent non-significant calculated mean differences.

* Suspend judgment on whether calculated mean difference is statistically significant, because calculated mean difference was found to be greater than or equal to the uncorrected critical difference found by the Fisher test, but less than the corrected critical difference found from the Tukey test.

** Statistically significant differences found between calculated mean differences, which is greater than or equal to the corrected critical difference found from the Tukey test.

Comparisons between monitoring proficiency measures. Recall that no main effects were found for either of bias and discrimination measures (see table 13 for results of analyses, and figures 4 and 5 for visual representation). Further, unlike prediction

accuracy, neither bias nor discrimination revealed any statistically significant, uncorrected pairwise differences. Hence, I did not conduct any corrected test of pairwise differences on bias and discrimination measures.

Plots Of Means Of Measures of Performance Assessments, Performance Predictions, Prediction Confidence, Prediction Accuracy and Monitoring Proficiencies

In order to provide a graphical view of trends, both significant and non-significant, in the repeated measures discussed thus far, Figures 1 to 6 show the estimated marginal means for performance assessments, performance predictions, prediction confidence, bias, discrimination and prediction accuracy, respectively.

Figure 1

Means for Instructor's Performance Assessments Across Six Logs

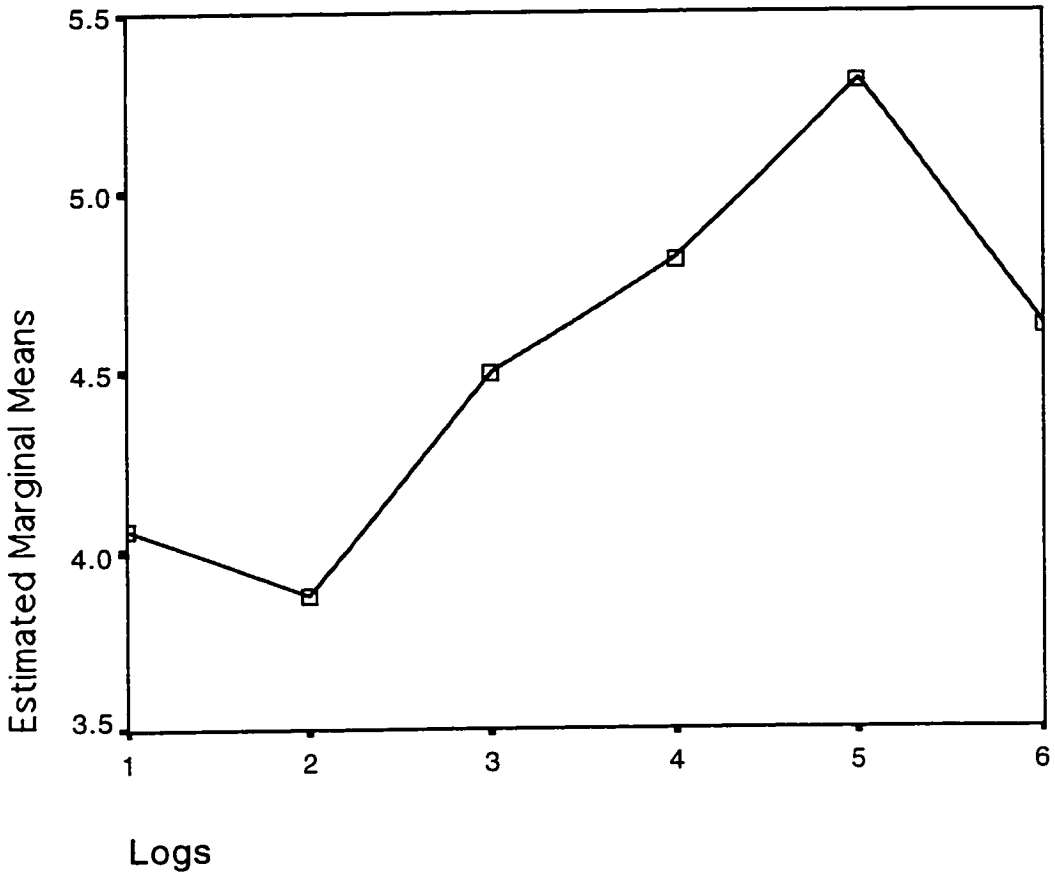


Figure 2

Means for Students' Performance Predictions Across Six Logs

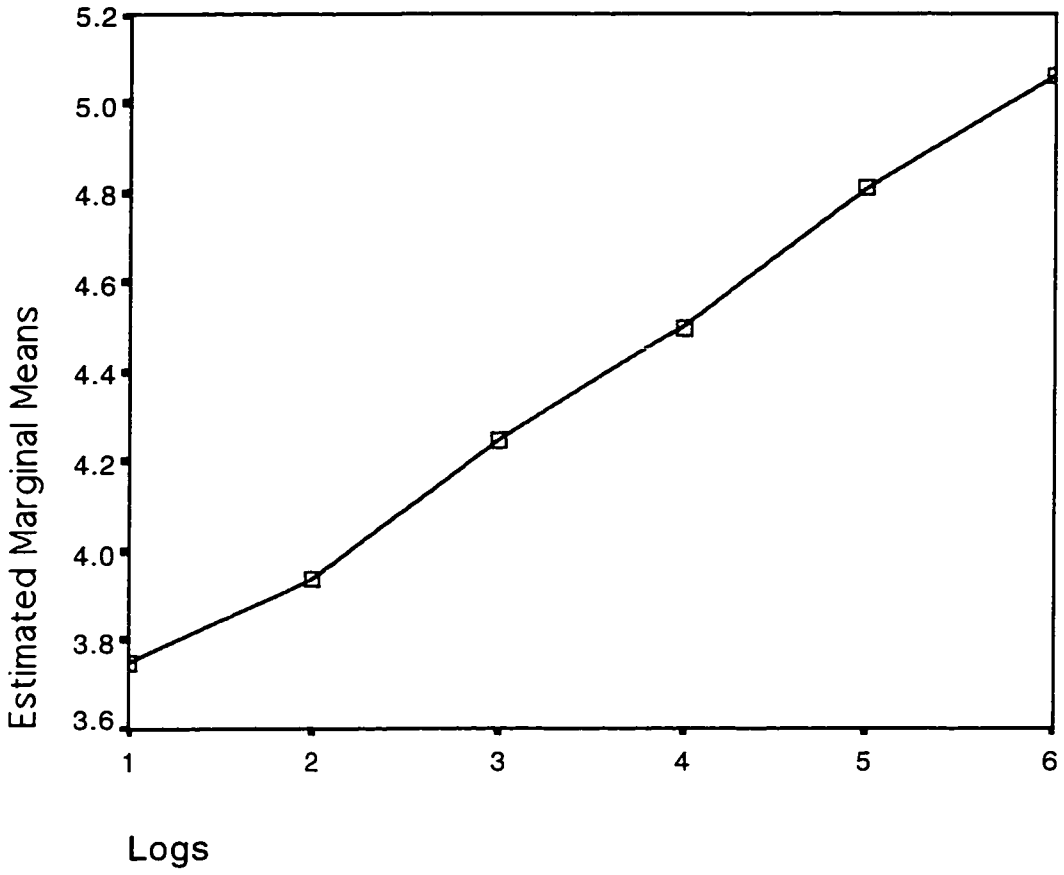


Figure 3

Means for Students' Prediction Confidence Across Six Logs

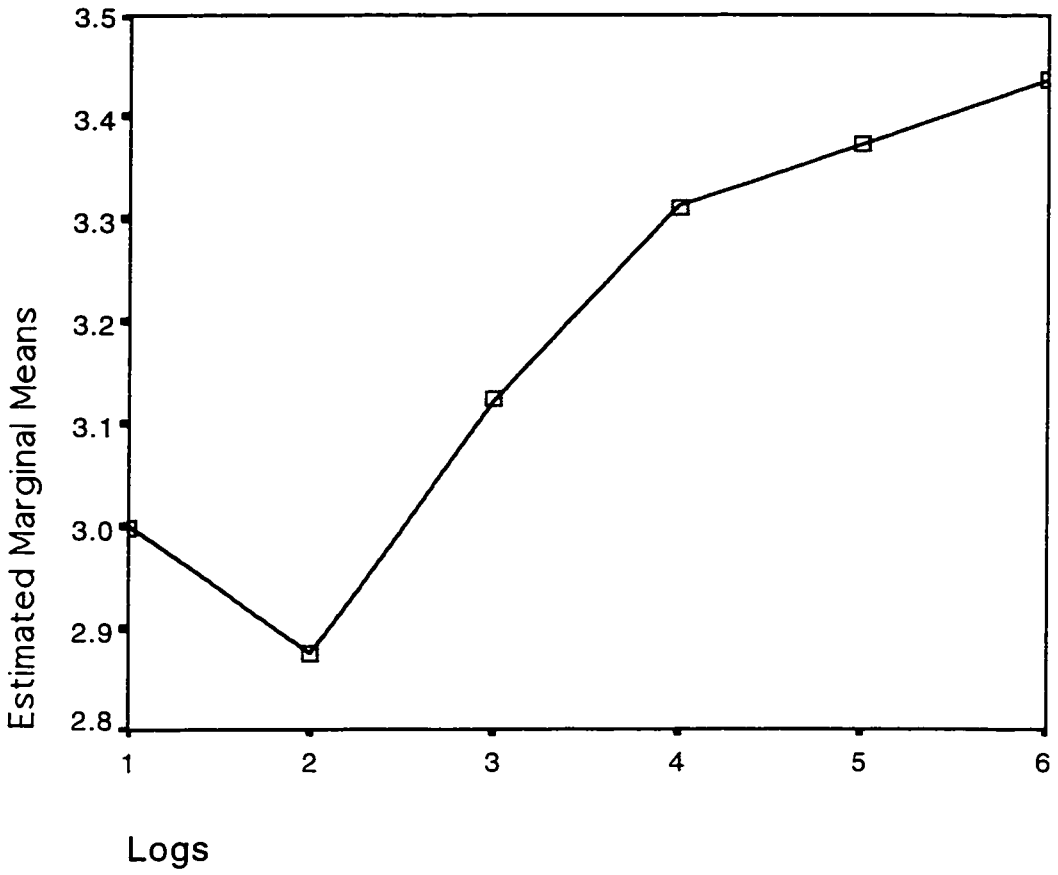


Figure 4

Means for Students' Bias Across Six Logs

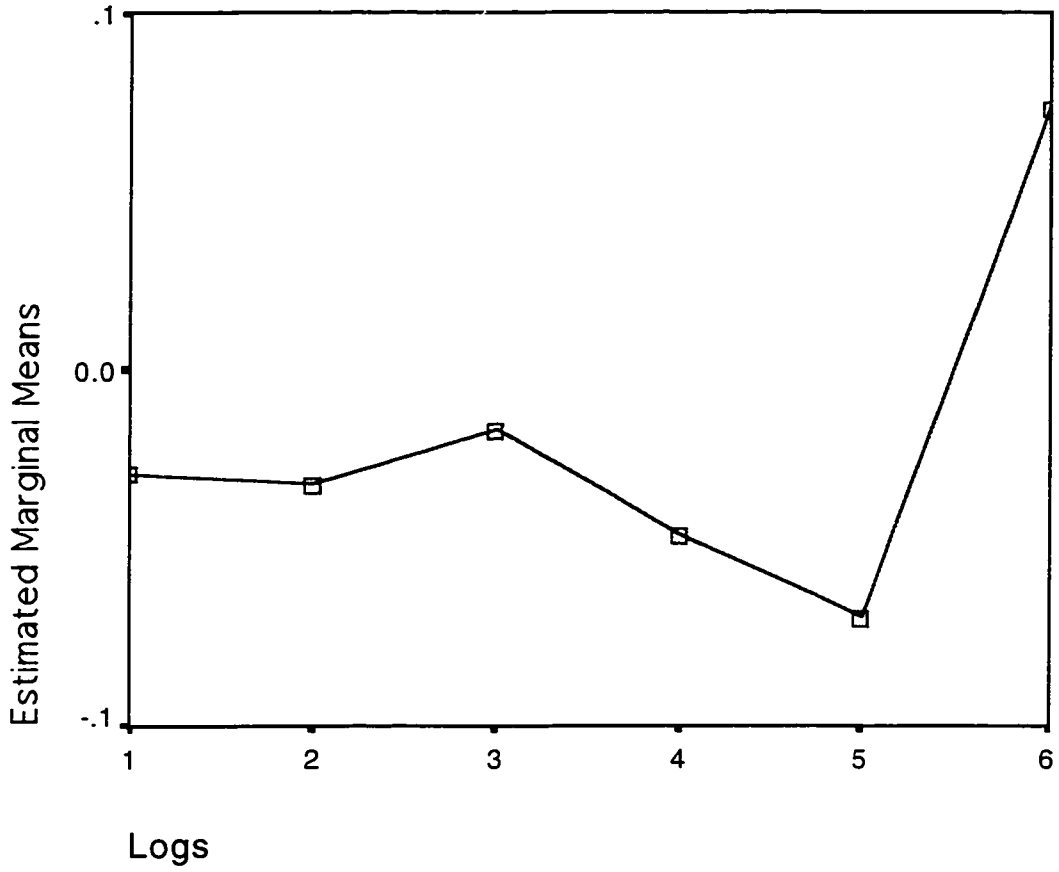


Figure 5

Means for Students' Discrimination Across Six Logs

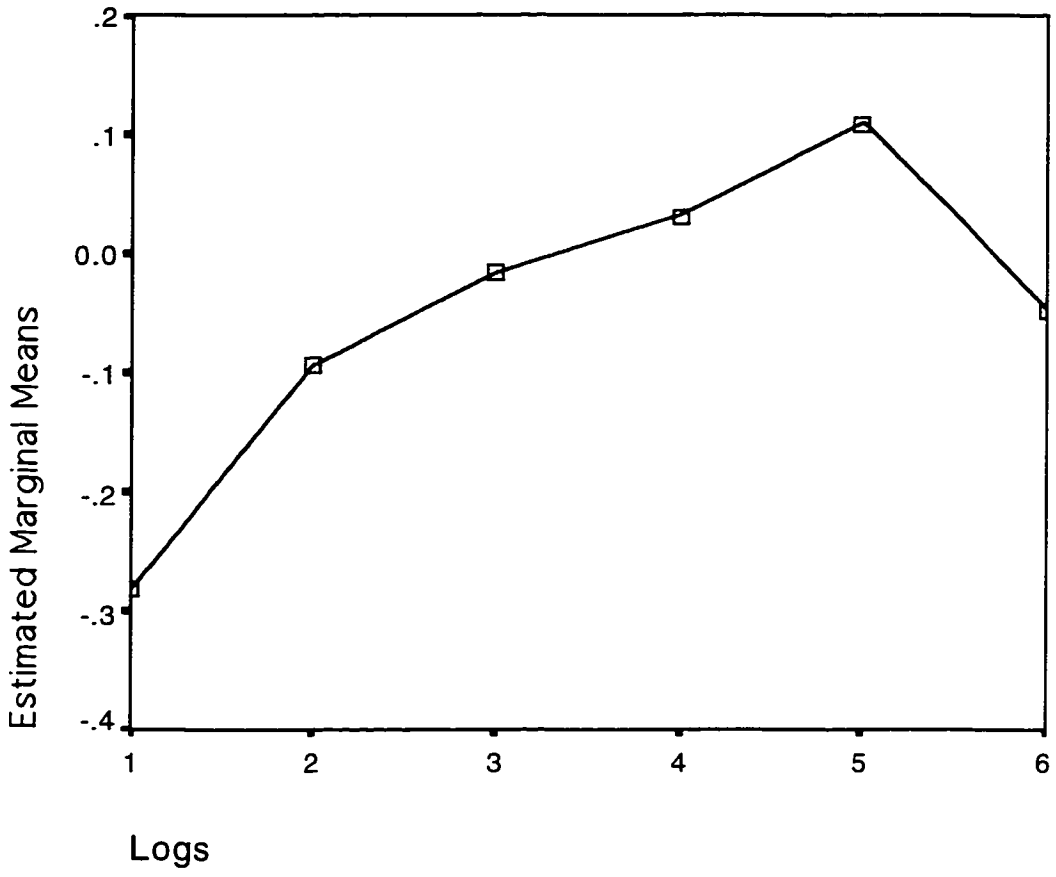
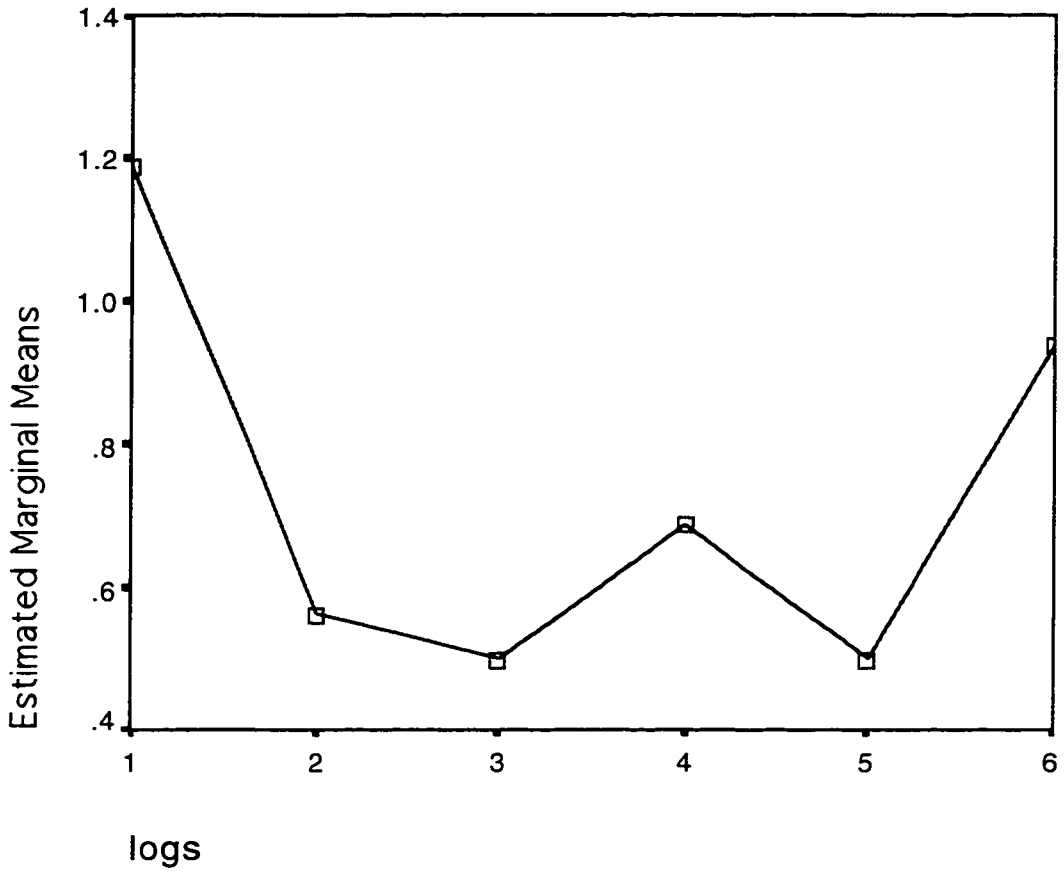


Figure 6

Means for Students' Prediction Accuracy Across Six Logs



Findings Related to Performance, Predictions, Confidence and Monitoring

Following from the results reported in the preceding sections on descriptive statistics, repeated measures analyses, pairwise comparisons and plots, I begin, in this section, to paint a portrait of how students' performance and monitoring proficiencies varied across the six logs.

Performance assessments. Performance assessments produced a main effect, $F(5, 75)=11.33$, $p<.001$, $\eta^2=.43$, and showed an increasing, linear trend, $F(1, 15)=25.40$, $p<.001$, $\eta^2=.63$ (see tables 13, 14 and Figure 1). Apart from the performance dips in week 2 (as compared to week 1) and week 6 (as compared to both logs 5 and 4), from the instructor's viewpoint, students demonstrated improved performance when progressing from the first to the sixth log. Despite the fact that statistically significant quadratic and cubic trends are revealed in the performance assessments, it makes more sense to interpret the data in light of the most significant, increasing linear trend. The quadratic results can be explained by the dip in performance in log 6, as compared to logs 5 and 4, while the cubic trend can be explained by the additional dip in performance in week 2, as compared to week 1 (see Figure 1). Results from the pairwise comparisons for performance assessments, also show that performance on logs 4 and 5 are each significantly greater than each of logs 1 and 2, and that there is a significant increase in performance from logs 3 to 5, and from logs 2 to 6 (see table 15). Note also that some of the differences between adjacent logs almost reach statistical significance (i.e., p values of .05 or less); in fact, I have suspended judgment on the differences seen between logs 2 and 3, and logs 4 and 5. Statistically significant differences in performance are not present between logs 1, 2 and 3. In fact, there is a decrease in performance between logs 1 and 2, but this change is most likely a function of chance, as evidenced by the non-significant difference reported in table 15. Subsequent improvements in performance as students progressed from log 3 to 4, and from log 4 to 5, yield statistically significant increases in performance when logs are compared across two periods (e.g., logs 2 and 4, logs 3 and 5), three periods (e.g., logs 1 and 4, logs 2 and 5), and four periods (e.g., logs 2

and 6; see table 15); note that I use the term period to denote the passage of time from one log to the next. Interestingly, students seem to have shown statistically significant improvements in performance when tackling logs of similar structure; for example, improvements in performance were recorded from log 2 to 4, both of which contained multistructural and relational question-and-answer sets, and from logs 3 to 5, both of which contained multistructural and extended abstract question-and-answer sets. The significant decrease in performance between logs 5 and 6 can be attributed, in part, perhaps, due to the fact that the posttest (final examination) for the course was conducted after log 5 was submitted, and that an extension was granted for the submission of log 6. Students' performances in log 6, might have, therefore, been affected by their disengagement from the course after having completed the posttest. Other factors include the structure of log 6, which contained all three multistructural, relational and extended abstract sets; this might have made log 6 more difficult to perform in than the other 5 logs, all of which contained only two of the three possible sets. Overall, students showed a general improvement in performance, as judged by the instructor, over the first five logs, but showed a statistically significant deterioration in their performance from the fifth to the sixth, and final, log.

A possible explanation for statistically significant changes in performance assessments could be the instructor's criteria for evaluating the logs. While the assessment criteria for each level were clearly stated in the course outline, recall that the instructor had informed students that they needed to demonstrate growth across their logs. This element of growth is subjectively judged by the instructor, and an understanding for what criteria are being used for assessing growth, while not clear form

the course outline, might have been implicitly co-constructed between individual students and instructor. In learning situations that involve complex writing activities, like that of the learning log, such implicit development of criteria for evaluation are, not uncommon, and need to be factored into the analyses. Subsequent repeated measures analyses on the independent scores of performance, and correlational analyses between performance assessments and the independently scores of performance will shed more light on this issue.

Performance predictions. Performance predictions produced a statistically significant main effect, $F(3.73, 56)=5.70$, $p=.001$, $\eta^2=.28$, and followed a statistically significant, increasing linear trend, $F(1, 15)=14.33$, $p=.002$, $\eta^2=.49$ (see tables 13, 14 and Figure 2). For any given log, students' consistently expected to score a better grade than any of their prior predictions. Statistically significant differences between predictions across logs are only evident, though, between logs that are at least four periods apart (e.g., logs 1 and 5, logs 2 and 6, logs 1 and 6). From table 15, notice also that I suspend judgment on the significance of the observed differences in predictions between logs 1 and 4, logs 2 and 5, as well as logs 3 and 6. It is therefore, interesting to note that despite the steadily increasing predictions, differences in expected grades between adjacent logs (e.g., between logs 1 and 2, logs 2 and 3, etc.), and across logs that are two periods apart (e.g., between logs 1 and 3, logs 2 and 4, logs 3 and 5, and logs 4 and 6) are likely a function of chance. Predictions showed statistically significant increases only after a sufficiently lengthy period of time of engagement with the logs, namely, at least four to five weeks.

Prediction accuracy. One way of investigating accuracy in students' predictions across the six logs is to determine if there were any statistically significant differences between each of the six pairs of scores of performance predictions and performance assessments. Paired-sample t-tests between the six pairs yielded only one statistically significant difference; performance predictions on log 5 (M=4.76, SD=.83) were found to be significantly smaller than the performance assessments on the same log (M=5.24, SD=.75), $t(16)$=-2.704, p=.016. Overall, students seemed to demonstrate accurate prediction abilities; while students generally underestimated their performances on logs 1, 3, 4 and 5, and overestimated their performances on logs 2 and 6, average differences between predicted and instructor-assessed performances for the six logs ranged from -.47 on the lower end of the underestimation, to .059 to the upper end of overestimation. However, such a measurement of accuracy in prediction can be misleading. The range of values for the difference between predicted and instructor-assessed performance seem to suggest that students' accuracies in predictions were within a single grade of their performance. Since the differences can take both positive and negative values, one students' overestimation cancels out another's underestimation of the same magnitude, thereby providing a misrepresentative view of accuracy in prediction. The absolute difference between performance prediction and performance assessment, which I have labeled *prediction accuracy* in this thesis study, measured the degree to which the students' predictions matched their performance, regardless of whether students over or underestimated their performances. The plot in Figure 6 illustrates that students' average predictions (M=1.19, SD=.29) were inaccurate by more than one grade for log 1. Between logs 2 and 5, however, the average accuracies were lowered and remained within the

range of .50 to .69 (*SD*s ranged from .18 to .24); this discrepancy can be explained by chance fluctuations. The value for accuracy in prediction rose to a comparatively inaccurate level of .94 with an *SD* of .23 for log 6, most likely, because students were not highly engaged while completing the sixth and final log, since it was due for submission after the final examination.

Prediction confidence. Prediction confidence revealed a statistically significant main effect, $F(5, 75)=3.60$, $p=.006$, $\eta^2=.19$ across the six learning logs, and produced a statistically significant, linear trend, $F(1, 15)=9.20$, $p=.008$, $\eta^2=.38$ (see tables 13, 14 and Figure 3). Regarding pairwise comparisons, the only two statistically significant differences in confidence lie between log 2 and 5, and log 2 and 6 (see table 16). Notice also, from table 16, that I have suspended judgment on the differences observed between log 1 and 5, log 1 and 6, as well as log 2 and 4. All other differences in confidence are likely a function of chance. It is interesting to note that the only two statistically significant differences involve the average value for confidence in prediction for log 2, which happens to be the lowest average confidence recorded throughout the period of instruction ($M=2.88$, $SD=.125$), as well as the only time when confidence in predictions dipped (see Figure 3). A likely reason for the dip in confidence could be conjectured from the students' relatively inaccurate predictions for log 1. Subsequently increasing values of confidence, until log 6, can be attributed to the relatively accurate predictions made from logs 2 to 5. Despite the steadily increasing values of confidence from logs 2 to 6, note that statistically significant levels of differences between scores of prediction confidence were reached after engagement with the logs of at least three periods of time.

Monitoring proficiencies. With regards to the calculated monitoring proficiencies of discrimination and bias, the repeated measures analyses of variance did not reveal any main effects (see table 13), suggesting that, across the six logs, students' discrimination and bias fluctuated in a manner that was most likely due to chance. The largely negative values of average bias across the six logs show that, in general, students were underconfident of their abilities (see table 11). Notice, from Figure 4, that the value for bias becomes increasingly negative from log 3 to log 5, reflecting increasing levels of underconfidence. The positive value of bias for log 6 reflects the overconfidence students demonstrated in their prediction of a grade for the final log. Average discrimination measures progress from being negative in log 1 (M=-.28, SD=.18), to becoming increasingly more positive until log 5 (M=.11, SD=.13), before dipping back to a negative value for log 6 (M=-.05, SD=.11; see Figure 5). The increasingly positive values of average discrimination from log 1 to 5 represent students' progressive abilities to assign a higher level of confidence for accurate prediction (see Figure 5). Students' average discrimination plummeted in log 6, thereby bucking the increasing trend, perhaps due to their inability to accurately predict their grade for the sixth, and final log.

Summary. In summary, an initial analysis of the measures of instructor-assessed performance, predicted grades, confidence, and monitoring proficiencies yielded the following findings. First, performance assessments, as gauged by the instructor, generally improved by adopting a significant linear trend, from log 1 to log 6, statistically significant improvements were seen only after students had engaged in writing their first three logs. Second, performance predictions, like their performance, also demonstrated an increasing, significant, linear trend from log 1 to log 6; statistically significant increases

were observed only across logs that were at least four periods apart. Third, prediction confidence, showed an increasing, statistically significantly linear trend; improvements in confidence were statistically significant only across logs that were at least three periods apart. Fourth, monitoring proficiencies, which included discrimination and bias in predictions, fluctuated as a function of chance. Fifth, performance assessments dipped on the final log, most likely due to the effect of writing of the posttest between the submissions of logs 5 and 6; monitoring proficiencies fluctuated accordingly for log 6, with students demonstrating overconfidence (relatively large, positive average bias) and assigning a relatively higher level of confidence for inaccurate predictions (relatively large, positive average discrimination).

Repeated Measures Analysis for Independent Multistructural, Relational, Extended Abstract and Overall Performance Scores

Recall that the independently calculated scores of performance reflected the extent to which students had met the assessment criteria for each level of question-and-answer set in their learning logs. The independent relational performance scores and overall performance scores revealed a main effect across the six logs, while the independent multistructural and extended abstract performance scores did not reveal a main effect (see table 17 for results of analyses).

Table 17

Repeated measures one-way analyses of variances on independent scores of performance

Measure	No. of Scores	*F*	df	*MSE*	*p*	η^2	Power	Effect Size [a]
Independent Multistructural Performance Score	6	1.56	5	489.51	.18	.09	.52	.31
Independent Relational Performance Score	4	13.19	3	449.96	<.001	.47	1.00	.94
Independent Extended Abstract Performance Score	3	2.66	2	292.80	.09	.15	.49	.42
Independent Overall Performance Score	6	5.20	5	216.25	<.001	.26	.98	.59

[a] Effect size is calculated using Cohen's (1998) procedure for calculation of *f*, which is equivalent to $\sqrt{\frac{\eta^2}{1-\eta^2}}$, where η^2 is interpreted as the proportion of the total variability in the dependent variable that is

accounted for by variation in the independent variable; η^2 is the ratio of the between groups sum of squares to the total sum of squares.

Trends. The independent overall performance scores on the logs revealed a significant linear trend, $F(1, 15)=9.90$, $p=.007$, $\eta^2=.40$, as well as a less, yet statistically significant, quadratic trend, $F(1, 15)=7.47$, $p=.02$, $\eta^2=.33$. Similarly, the independent relational performance scores also revealed a significant linear trend, $F(1, 15)=17.41$, $p=.001$, $\eta^2=.54$, as well as a less, yet statistically significant, quadratic trend, $F(1, 15)=13.26$, $p=.002$, $\eta^2=.47$.

Facilitating comparison between different levels of question-and-answer sets. Recall that the instructor's performance assessment consisted of a single grade for each learning log, equally distributed between the multistructural and relational or extended abstract level, depending on the levels required for any given log. Such a measure of performance does not allow for comparisons to be made within and between different levels of question-and-answer sets. However, the independent scoring of performance in meeting assessment criteria for each level of the logs allow for such a comparison. For example, table 17 reveals the results of tests of main effects of students' scores within each of the multistructural, relational and extended abstract levels. To facilitate a comparison between the relational and extended abstract levels of questions and answers, I conducted a repeated measures, within-group, one-way analyses of variance on the eight independently calculated scores of performance on the relational (four scores) and extended abstract (three scores) levels of question-and-answer sets. A statistically significant main effect was found for the seven mean scores of performance on relational and extended abstract levels, $F(6, 90)=9.20$, $p<.001$, $\eta^2=.38$. In addition, a statistically

significant quadratic trend was revealed for the mean scores of performance on relational and extended abstract levels, $F(1, 15)=42.11$, $p<.001$, $\eta^2=.74$, along with a less significant linear trend, $F(1, 15)=7.96$, $p=.013$, $\eta^2=.38$.

Pairwise Comparisons for Independent Scores of Performance

Similar to the procedures employed earlier in the section on pairwise comparisons of the measures related to instructor's assessments of performance and the TAPE-related monitoring proficiency measures, I conducted unrestricted pairwise comparisons using the Fisher and Tukey tests, on all independent scores of performance which produced a main effect in their omnibus F test. I calculated critical differences for both the Fisher (d_F) and the Tukey test (d_T), for each measure undergoing the pairwise procedure, and compared the calculated differences with the values of d_F and d_T. If the calculated difference was lesser than the value of d_F, I rejected the null hypothesis that there were statistically significant differences between the pair of measures being compared; if the calculated value was greater than or equal to the value of d_F, but lesser than the value of d_T, I suspended judgment on making a decision as to whether the observed difference was statistically significant or not; if the calculated difference was greater than or equal to the value of d_T, I rejected the null hypothesis and concluded that a statistically significant difference did exist between the pair of measures being compared.

Independent overall performance score. Students' independent scores of overall performance showed statistically significant differences between logs 1 and 4, logs 1 and 5, and logs 1 and 6 (see table 18 for results). Between logs 2 and 6, the independent overall performance scores fluctuated with chance.

Table 18

Pairwise Comparisons for Independent Overall Performance Scores

	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	-					
Log 2	13.15*	-				
Log 3	14.76*	1.61	-			
Log 4	16.90**	3.75	2.14	-		
Log 5	25.57**	12.41*	10.80*	8.66	-	
Log 6	17.83**	4.67	3.07	.92	-7.73	-

Absolute value of d_T for independently overall performance score = 15.24

Absolute value of d_F for independently overall performance score= 10.37

Each entry in the table represents the mean difference calculated by taking the signed difference between the measure for the row and the measure for the column. Non-asterisked entries represent non-significant calculated mean differences.

* Suspend judgment on whether calculated mean difference is statistically significant, because calculated mean difference was found to be greater than or equal to the uncorrected critical difference found by the Fisher test, but less than the corrected critical difference found from the Tukey test.

** Statistically significant differences found between calculated mean differences, which is greater than or equal to the corrected critical difference found from the Tukey test.

Independent relational and extended abstract performance score. The mean independent relational performance score for log 1 was significantly lower than the mean independent relational performance scores for each of the logs 2, 4 and 6 (see table 19),

as well as lower than the mean independent extended abstract performance scores for logs 3, 5 and 6 (see table 20).

Table 19

Pairwise comparisons for the Independent Relational Performance Scores for logs 1, 2, 4 and 6

	Log 1	Log 2	Log 4	Log 6
Log 1	-			
Log 2	30.89**	-		
Log 4	43.39**	12.50	-	
Log 6	37.14**	6.25	-6.25	-

Absolute value of d_T for independent relational performance score = 20.02

Absolute value of d_F for independent relational performance score = 15.15

Each entry in the table represents the mean difference calculated by taking the signed difference between the measure for the row and the measure for the column. Non-asterisked entries represent non-significant calculated mean differences.

** Statistically significant differences found between calculated mean differences, which is greater than or equal to the corrected critical difference found from the Tukey test.

Table 20

Pairwise comparisons for the means of the independent relational and extended abstract performance scores across the six logs

Mean Differences of Independent Relational and Extended Abstract Performance

	Log 1: Relational	Log 2: Relational	Log 3: Extended Abstract	Log 4: Relational	Log 5: Extended Abstract	Log 6: Relational	Log 6: Extended Abstract
Log 1: Relational	-						
Log 2: Relational	30.89**	-					
Log 3: Extended Abstract	29.11**	-1.79	-				
Log 4: Relational	43.39**	12.50	14.29*	-			
Log 5: Extended Abstract	40.71**	9.82	11.61	-2.68	-		
Log 6: Relational	37.14**	6.25	8.04	-6.25	-3.57	-	
Log 6: Extended Abstract	28.21**	-2.68	-.89	-15.18*	-12.50	-8.93	-

Absolute value of d_T for independent relational and extended abstract performance scores = 20.35

Absolute value of d_F for independent relational and extended abstract performance scores = 13.39

Each entry in the table represents the mean difference calculated by taking the signed difference between the measure for the row and the measure for the column. Non-asterisked entries represent non-significant calculated mean differences.

* Suspend judgment on whether calculated mean difference is statistically significant, because calculated mean difference was found to be greater than or equal to the uncorrected critical difference found by the Fisher test, but less than the corrected critical difference found from the Tukey test.

** Statistically significant differences found between calculated mean differences, which is greater than or equal to the corrected critical difference found from the Tukey test.

Plots for Independently Calculated Scores of Performance

Figures 7 to 11 provide a graphical view of the estimated marginal means for the following measures, described in prior analyses: independent multistructural performance scores across the six logs (figure 7), independent relational performance scores across logs 1, 2, 4 and 6 (figure 8), independent extended abstract performance scores across logs 3, 5 and 6 (figure 9), independent overall performance scores (figure 10), and independent relational and extended abstract performance scores across the six logs (figure 11).

Figure 7

Means of Independent Multistructural Performance Scores across six logs

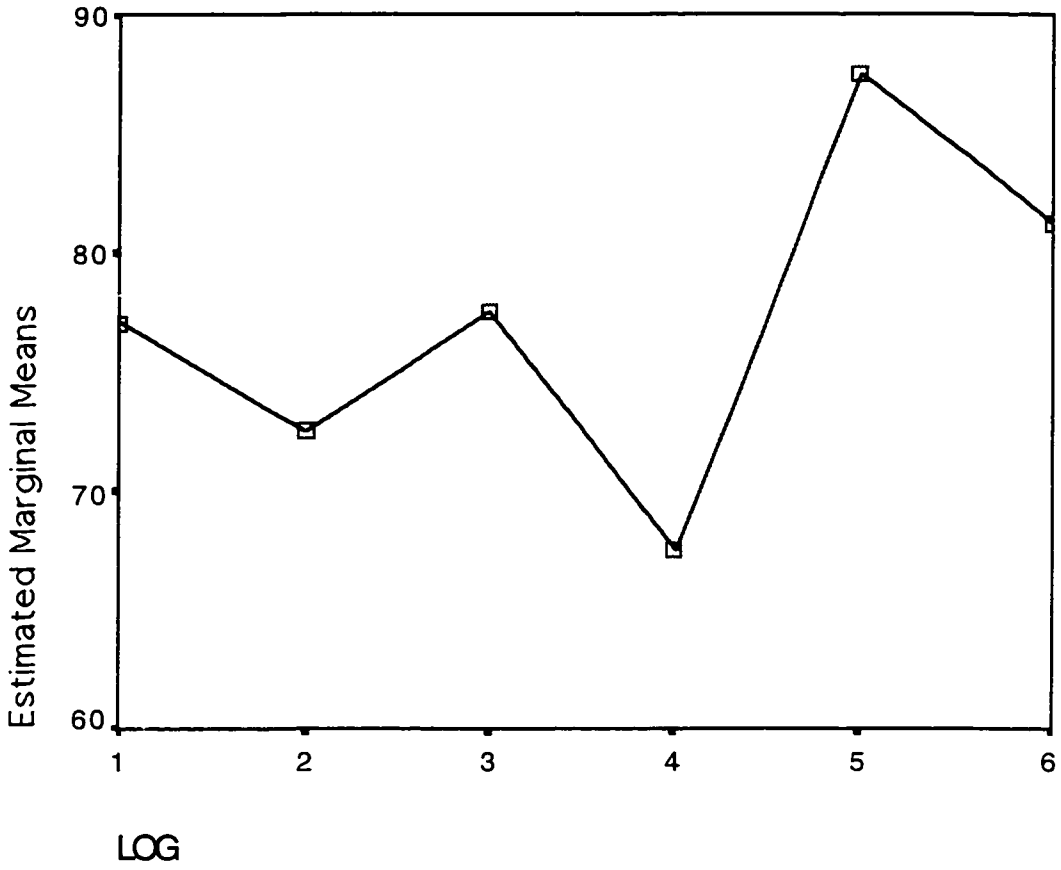


Figure 8

Means of Independent Relational Performance Scores across logs 1, 2, 4 and 6

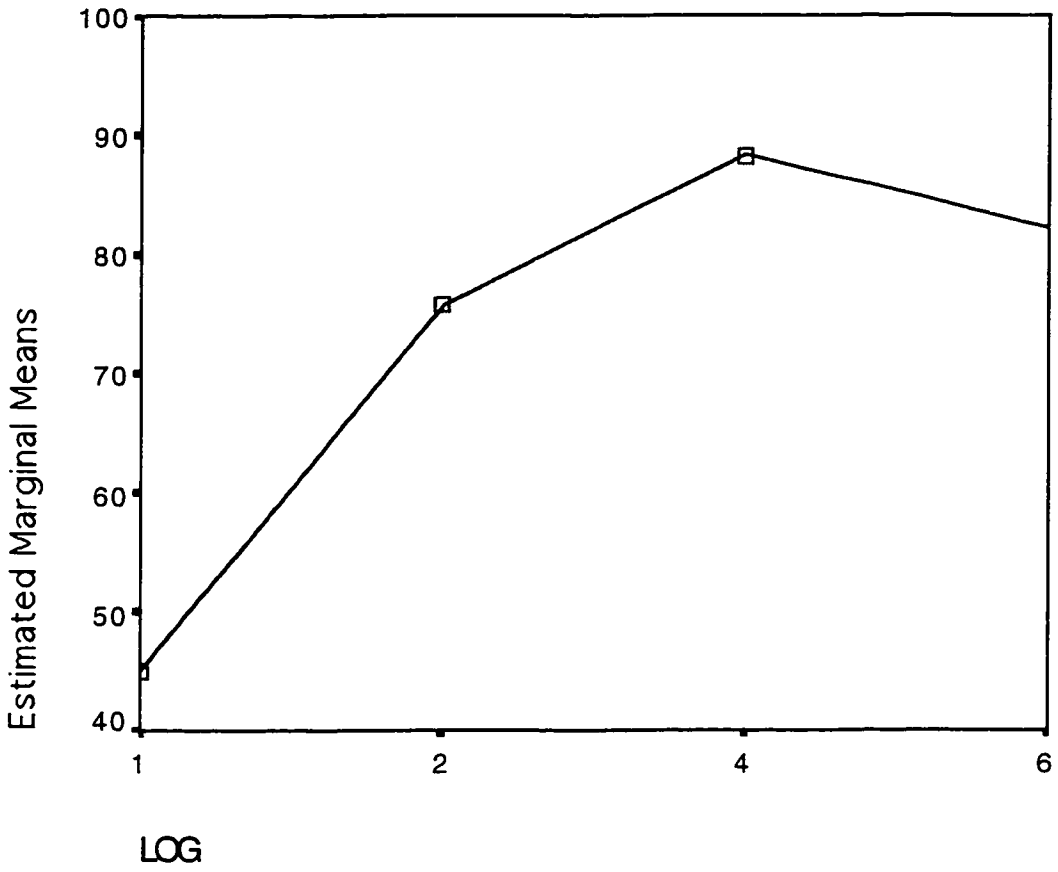


Figure 9

Means of Independent Extended Abstract Performance Scores across logs 3, 5 and 6.

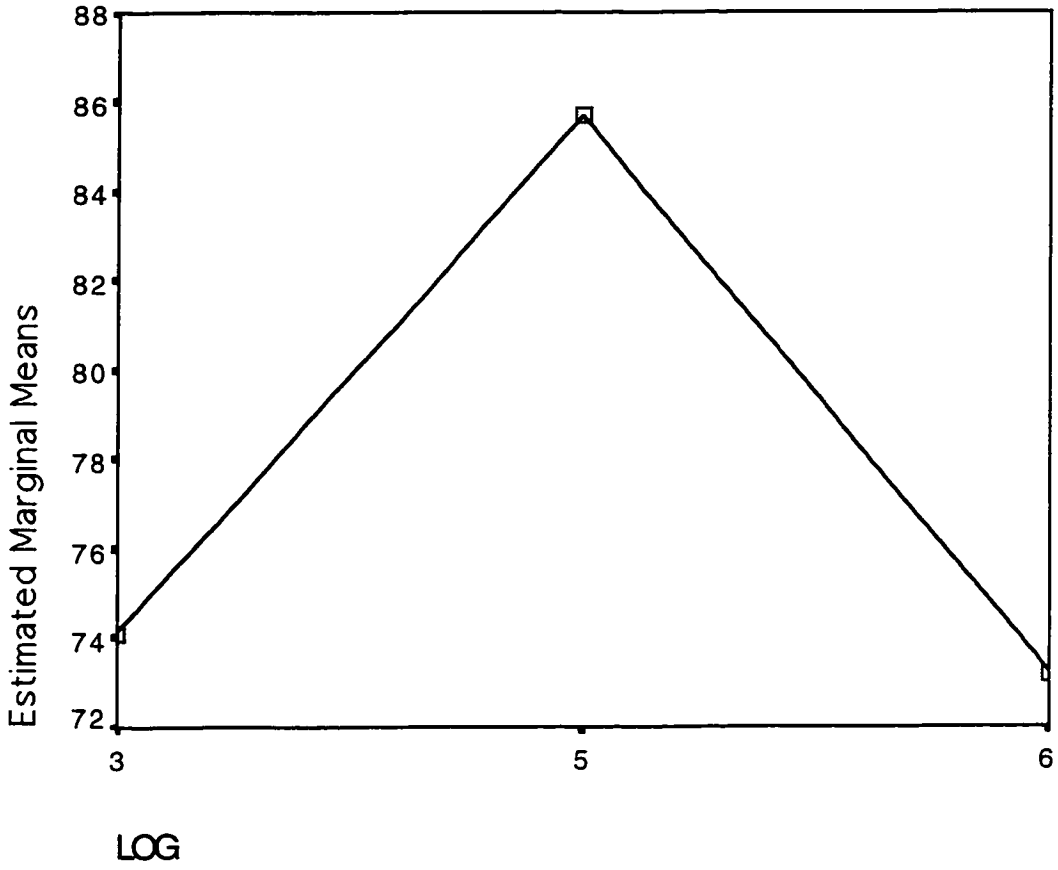


Figure 10

Means of Independent Overall Performance Scores across six logs.

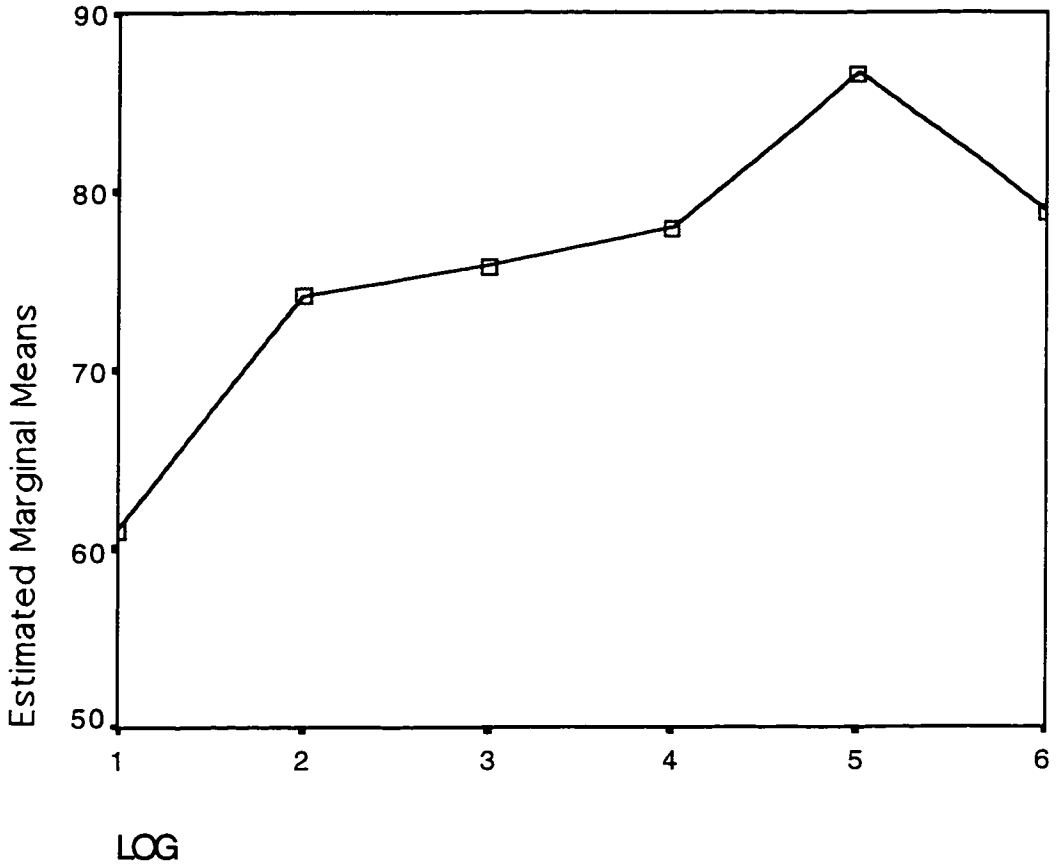
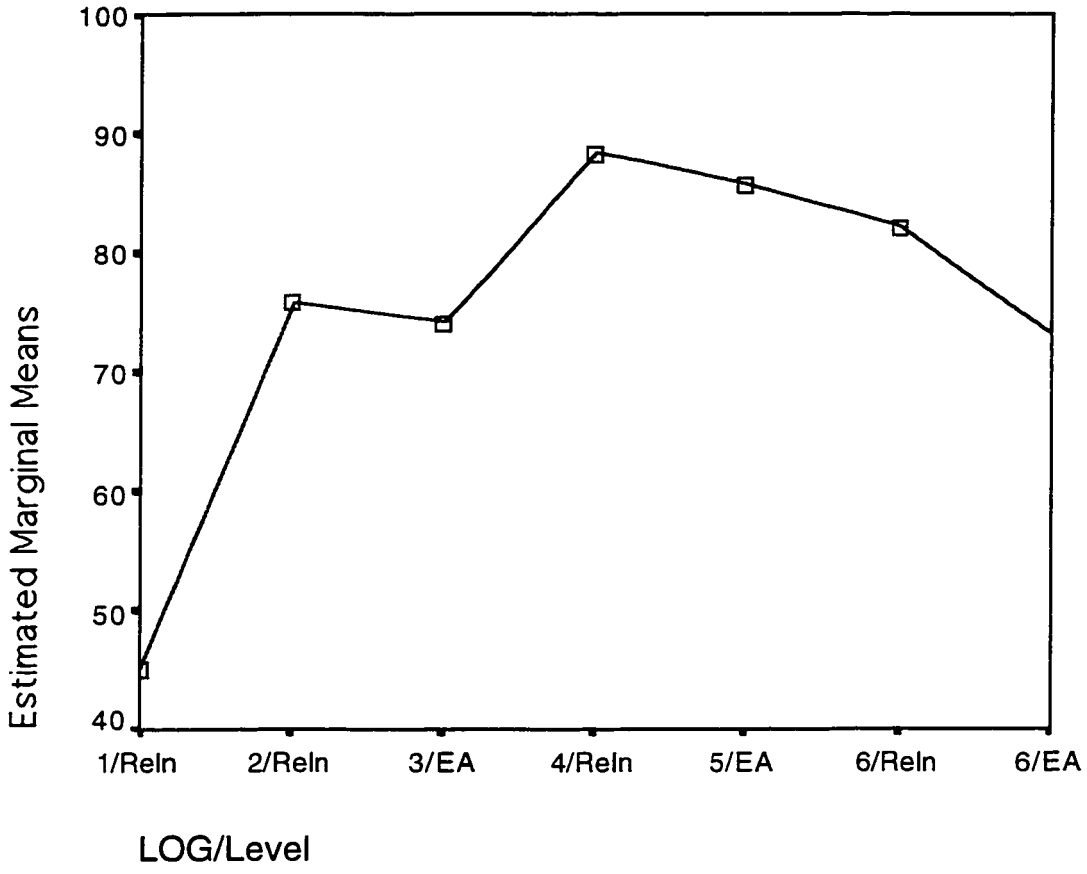


Figure 11

Means of Independent Relational (Reln) and Extended Abstract (EA) Performance Scores across six logs.



Findings Related to Independent Multistructural, Relational, Extended Abstract and Overall Performance Scores

Based on the repeated measures analyses, pairwise comparisons and the plots, I now paint a picture of how the various independent performance scores reveal how well students met the assessment criteria that were laid out in the course outline.

Independent multistructural and extended abstract performance scores. Independent multistructural and extended abstract performance scores fluctuated largely

as a function of chance, as is seen from the non-significant main effects reported in table 17. This suggests that students met the assessment criteria, for both multistructural and extended abstract levels, to the same extent across the six logs, despite the differences in course content that served as sources of information for each log. Moreover, the difference in structure of log 1 (question provided by instructor, answer generated by student) as compared to the rest of the logs (question and answer generated by student), did not affect students' abilities to meet the assessment criteria for the multistructural level to the same extent throughout all six logs. Even though students improved in their ability to meet the assessment criteria for extended abstract levels from log 3 (M=74.11, SD=21.01) to log 5 (M=85.71, SD=15.65), and showed reduction in this ability between log 5 and log 6 (M=73.21, SD=18.72; see plot in Figure 9), these differences are accorded to chance fluctuations. Note also that students, for the most part, were able to successfully meet the assessment criteria for the multistructural level, as is evidence by the high scores across the six logs (see values of Ms and SDs in table 12 and plot in figure 7, range of Ms: 67.50 to 87.50, range of SDs: 20.49 to 28.17).

Independent relational performance scores. A statistically significant main effect was found for the independent relational performance scores for logs, 1, 2, 4 and 6, $F(3, 45)=13.19$, $p<.001$, $\eta^2=.47$. The reported statistically significantly, increasing linear trend, $F(1, 15)=17.41$, $p=.001$, $\eta^2=.54$ (see Figure 8) prompted an exploration of where exactly the differences in performance lay. Table 19 reveals that the average ability of a student in meeting the criteria for relational levels for log 1 (M=45.00, SD=31.41) is statistically, significantly lower than the average scores for each of the logs 2 (M=75.89, SD=24.31), 4 (M=88.39, SD=14.95) and 6 (M=82.14, SD=23.04). There are, however, no

other statistically significant differences among these four means. The pairwise comparisons suggest that students' abilities to meet the assessment criteria for the relational level question-and-answer sets improved from log 1 to log 2, but subsequently, their improved abilities to meet the criteria fluctuated with chance between logs 2, 4 and 6. Recall also that for log 1, students were given a relational level question, and were required to answer the question, whereas for the other three logs, students were required to generate their own questions and answers. This difference in the structure of the learning log might account for the observed differences in students' abilities to meet the assessment criteria for the relational level.

Independent overall performance scores. Independent overall performance scores revealed a main effect across the six logs, $F(5, 75)=5.20$, $p<.001$, $\eta^2=.26$ (see table 17); the reported significant, increasing linear trend (see plot in Figure 10) resulted in further explorations through pairwise comparisons seen in table 18. It is interesting to note that despite the steadily increasing trend in the independently calculated students' scores of performance, mean differences reveal statistical significance only between logs 4 and 1, logs 5 and 1, as well as logs 6 and 1. I suspend judgment on the significance of observed differences between sets of adjacent logs in the first half of the course (e.g., logs 2 and 1, logs 3 and 2), as well as between logs at least two periods apart, for example, logs 5 and 2, and logs 5 and 3. The results of the pairwise comparisons suggest that students' abilities to meet the assessment criteria for the required levels of question-and-answer sets improved significantly only after students had written their first four learning logs (as evidenced by the significant differences observed between logs 4 and 1, logs 5 and 1, and logs 6 and 1). However, notice that from log 2 through to log 6, students' abilities to meet

the assessment criteria for each of the logs were met, for the most part, to similar extents, with any fluctuations accorded to chance. Therefore, no significant differences were found for performances from log 2 through to log 6, with the exception of the two cases of suspension of judgment, despite the differences in (a) the structure of the learning logs, for example, logs 2 and 4 required relational levels, while logs 3 and 5 required extended abstract levels of question-and-answer sets, and (b) the content of the course covered from log 2 through to log 6.

Comparison of independent performance scores on relational and extended abstract levels. As reported earlier, a statistically significant main effect was found for the seven independently calculated scores of performance on relational and extended abstract levels of question-and-answer sets across the six logs. In addition, a statistically significant quadratic trend was also observed (see plot in Figure 11). Pairwise comparisons, shown in table 20, reveal that all statistically significant differences involve students' performance in answering the relational level question for log 1. In fact, students' abilities to meet the assessment criteria for answering the relational level question for log 1 (M=45.00, SD=31.41) is significantly lower than *all* the other measures of abilities to meet assessment criteria for writing relational or extended abstract levels of question-and-answer sets. Two other differences call for a suspension of judgment in relation to their statistical significance; these include the improved ability to meet the criteria for writing a relational level question-and-answer set for log 4 as compared to writing, both at an extended abstract level for log 3 (mean difference of 14.29, standard error of 4.88) and at an extended abstract level for log 6 (mean difference of 15.18, standard error of 5.75). Apart from these two suspensions of judgment, all other

differences did not reach statistical significance. The results suggest that students showed sustained, improved performance on their relational and extended abstract level question-and-answer sets when compared to their performance in answering a relational level question for log 1. This improved performance fluctuated largely due to chance across the relational and extended abstract levels of question-and-answer sets between logs 2 and 6. Students' performance in meeting assessment criteria peaked in their writing of a relational level question-and-answer set for log 4. Subsequent dips in performance on the relational level for log 6, and the extended abstract levels for logs 5 and 6 can be accorded to chance, but help in explaining the statistical significant quadratic trend, as observed in the plot in figure 11.

Summary. The repeated measures analyses, pairwise comparisons and plots of the independent performance scores revealed the following findings. First, students met the assessment criteria for the multistructural levels to the same extent across all six logs. Second, students met the assessment criteria for the extended abstract level to the same extent across logs 3, 5 and 6. Third, students showed an improved performance in meeting the assessment criteria for relational levels from log 1 to log 2, however, across logs 2, 4 and 6, students met the assessment criteria for relational levels to the same extent. Fourth, students met the assessment criteria for relational and extended abstract questions to the same extent from logs 2 to 6. Fifth, students' overall performance in meeting the assessment criteria for the logs improved significantly only after students had written their first four logs; students' overall performance at meeting assessment criteria for the logs showed largely chance fluctuations from logs 2 to 6.

Repeated Measures Analyses on Justification Scores

The justification scores for multistructural, relational, extended abstract questions and answers (see table 12 for descriptive statistics, table 9 for labels) are based on students' responses to the first two items of each TAPE from log 1 to log 5, and the first six items for the TAPE for the sixth and final log, asked students to explain why their questions and answers were at a multistructural, relational or extended abstract level, as the case may have been (see Methods section for complete description). Relational question justification scores revealed a statistically significant main effect, $F(1.58, 23.67)=4.34$, $p=.03$, $\eta^2=.23$; no other justification score revealed a main effect (see table 21). In addition, justifications for meeting relational question criteria displayed a significant, quadratic trend, $F(1,15)=20.59$, $p<.001$, $\eta^2=.58$.

Table 21

Repeated Measures One-way Analyses of Variance on Justification Scores for Relational and Extended Abstract Questions and Answers

Repeated Measure	No. of Scores	*F*	df	*MSE*	*p*	η^2	Power	Effect Size [a]
Relational Question Justification	3	4.34	1.58[b]	.47	.03	.23	.63	.56
Relational Answer Justification	4	1.58	3	.50	.21	.10	.39	.32
Extended Abstract Question Justification	3	.354	2	.41	.71	.02	.10	.14
Extended Abstract Answer Justification	3	.188	2	.41	.64	.03	.12	.18

[a] Effect size is calculated using Cohen's (1998) procedure for calculation of *f*, which is equivalent to $\sqrt{\frac{\eta^2}{1-\eta^2}}$, where η^2 is interpreted as the proportion of the total variability in the dependent variable that is accounted for by variation in the independent variable; η^2 is the ratio of the between groups sum of squares to the total sum of squares.

[b] Degrees of freedom was reduced using the Huynh-Feldt correction for violation of Mauchly's test of sphericity.

Multistructural question and answer justification scores. There were only two instances of justifications of meeting multistructural answers (i.e., TAPE items on logs 1 and 6, see table 12); their respective average, independently calculated scores were compared using a paired-samples t-test procedure. The average score for the 16 valid responses for justifications for meeting multistructural answer criteria for log 1 (M=1.31, SD=.79) was significantly better than the score for log 6 (M=.75, SD=.93), $t(15)$=2.18, p=.05.

Justifications of meeting question and answer assessment criteria. In order to determine if there was a main effect of students' scores for justification items on the TAPE related to *questions* and *answers*, at all three levels (i.e., multistructural, relational and extended abstract), I conducted two more within-group, repeated measures one-way analyses of variance. Recall that over the course of six logs, students responded to seven TAPE items asking for justifications of meeting criteria for relational, multistructural, or extended abstract *questions* (one item each from the TAPEs for logs 2 to 5, and three items from the TAPE for log 6); students also responded to a total of nine items related to justifications for meeting criteria for multistructural, relational and extended abstract *answers* (two items from the TAPE for log 1, one each from the TAPEs for logs 2 to 5, and three items from the TAPE for log 6). A statistically significant main effect was revealed for scores on question-related justification items, $F(6,90)=5.37$, $p<.001$, $\eta^2=.26$; similarly, a main effect was also present for scores on answer-related justification items, $F(8,120)=2.46$, $p=.02$, $\eta^2=.14$.

Pairwise Comparisons for Justification Scores

Relational question justification scores. Students' relational question justification scores for log 4 were significantly higher than their scores on log 6, as seen from the results of pairwise comparisons in table 22.

Table 22

Pairwise Comparisons for Relational Question Justification Scores for logs 2, 4 and 6.

	Log 2	Log 4	Log 6
Log 2	-		
Log 4	.69**	-	
Log 6	.19	-.50	-

Absolute value of d_T for relational question justification scores = .68

Absolute value of d_F for relational question justification scores = .56

Each entry in the table represents the mean difference calculated by taking the signed difference between the measure for the row and the measure for the column. Non-asterisked entries represent non-significant calculated mean differences.

** Statistically significant differences found between calculated mean differences, which is greater than or equal to the corrected critical difference found from the Tukey test.

Justifications of meeting question and answer assessment criteria. Table 23, below, reveals the results of the unrestricted pairwise comparisons of the averages of the seven justification scores for *questions* at all three levels, across all six logs. Following this, table 24, displays the results of the pairwise comparisons made on the means of the nine justification scores for *answers* at all three levels, across the six logs.

Table 23

Pairwise Comparisons of Justification Scores for Questions

Mean Differences of Justification Scores for Questions							
	Log 2: Relational	Log 3: Extended Abstract	Log 4: Relational	Log 5: Extended Abstract	Log 6: Multi-structural	Log 6: Relational	Log 6: Extended Abstract
Log 2: Relational	-						
Log 3: Extended Abstract	.75**	-					
Log 4: Relational	.69*	-.06	-				
Log 5: Extended Abstract	.88**	.13	.19	-			
Log 6: Multi-structural	.13	-.63*	-.56*	-.75**	-		
Log 6: Relational	.19	-.56*	-.50*	-.69*	.06	-	
Log 6: Extended Abstract	.94**	.19	.25	.06	.81**	.75**	-

Absolute value of d_T for justification scores for questions at all three levels= .72

Absolute value of d_F for justification scores for questions at all three levels = .48

Each entry in the table represents the mean difference calculated by taking the signed difference between the measure for the row and the measure for the column. Non-asterisked entries represent non-significant calculated mean differences.

* Suspend judgment on whether calculated mean difference is statistically significant, because calculated mean difference was found to be greater than or equal to the uncorrected critical difference found by the Fisher test, but less than the corrected critical difference found from the Tukey test.

** Statistically significant differences found between calculated mean differences, which is greater than or equal to the corrected critical difference found from the Tukey test.

Table 24

Pairwise Comparisons of Justification Scores for Answers

Mean Differences of Justification Scores for Answers									
	Log 1: Multi-structural	Log 1: Relational	Log 2: Relational	Log 3: Extended Abstract	Log 4: Relational	Log 5: Extended Abstract	Log 6: Multi-structural	Log 6: Relational	Log 6: Extended Abstract
Log 1: Multi-structural	-								
Log 1: Relational	.13	-							
Log 2: Relational	.19	.06	-						
Log 3: Extended Abstract	.06	-.06	-.13	-					
Log 4: Relational	0	-.13	-.19	-.06	-				
Log 5: Extended Abstract	.25	.13	.06	.19	.25	-			
Log 6: Multi-structural	-.56*	-.69*	-.75*	-.63*	-.56*	-.81**	-		
Log 6: Relational	-.31	-.44	-.50*	-.38	-.31	-.56*	.25	-	
Log 6: Extended Abstract	.25	.13	.06	.19	.25	0	.81**	.56*	-

Absolute value of d_T for justification scores for answers at all three levels = .78

Absolute value of d_F for justification scores for answers at all three levels = .49

Each entry in the table represents the mean difference calculated by taking the signed difference between the measure for the row and the measure for the column. Non-asterisked entries represent non-significant calculated mean differences.

* Suspend judgment on whether calculated mean difference is statistically significant, because calculated mean difference was found to be greater than or equal to the uncorrected critical difference found by the Fisher test, but less than the corrected critical difference found from the Tukey test.

** Statistically significant differences found between calculated mean differences, which is greater than or equal to the corrected critical difference found from the Tukey test.

Plots For Justification Scores

Figures 12 to 17 provide graphical views of the estimated marginal means of the justification scores for: relational questions across logs 2, 4 and 6 (figure 12), relational answers across logs 1, 2, 4 and 6 (figure 13), extended abstract questions across logs 3, 5 and 6 (figure 14), extended abstract answers across logs 3, 5 and 6(figure 15), questions from all three levels across the six logs (figure 16), answers from all three levels across the six logs (figure 17).

Figure 12

Means of Relational Question Justification Scores across logs 2, 4 and 6

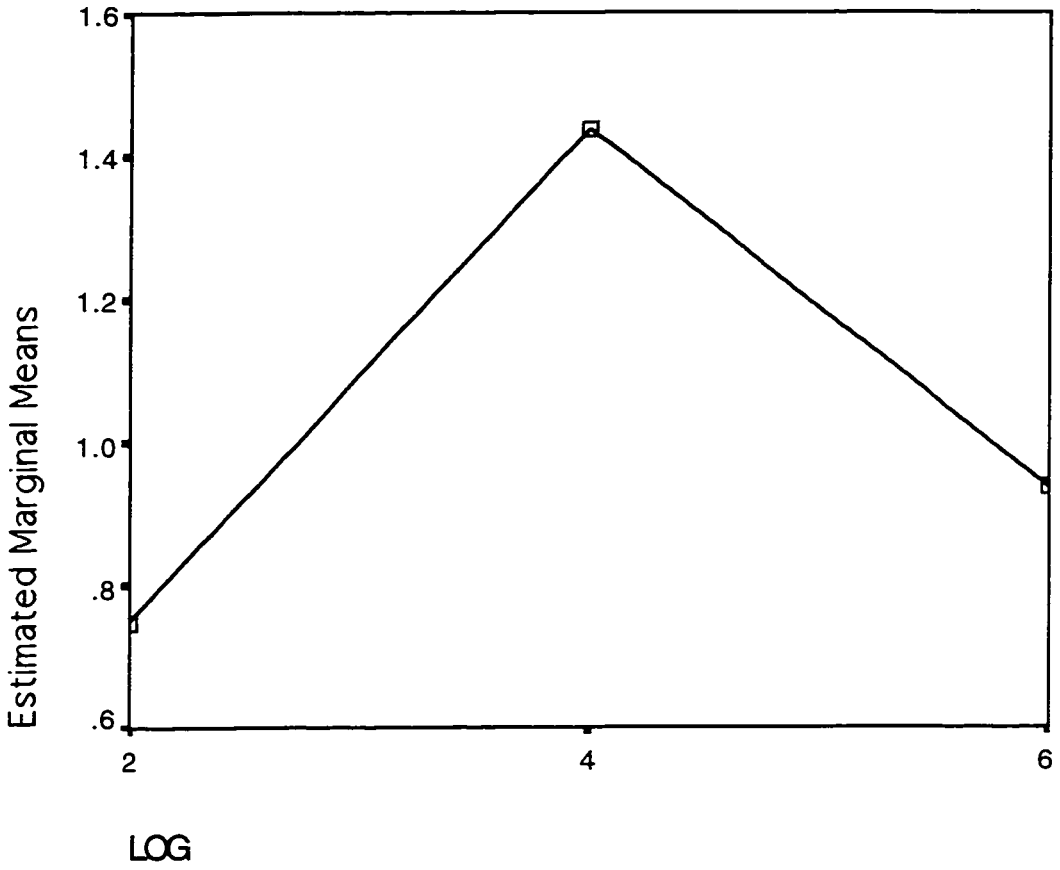


Figure 13

Means of Extended Abstract Question Justification Scores across logs 3, 5 and 6

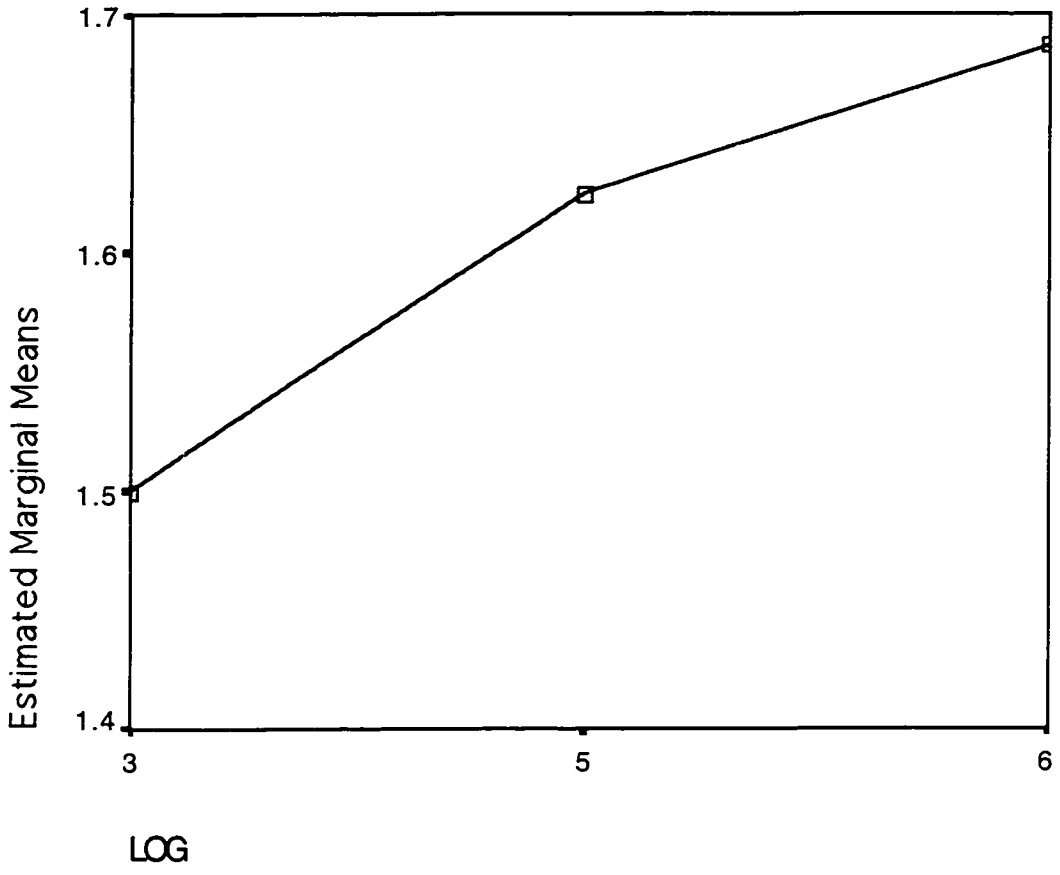


Figure 14

Means of Relational Answer Justification Scores across logs 1, 2, 4 and 6

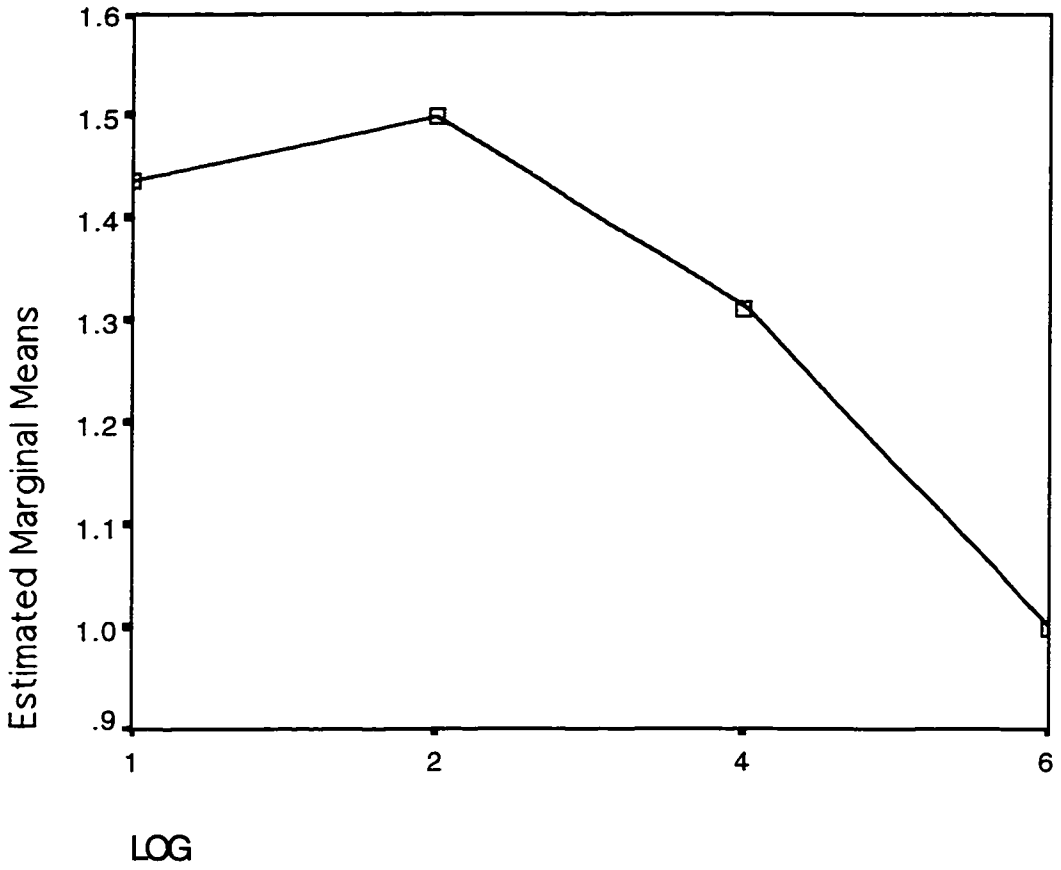


Figure 15

Means of Extended Abstract Answer Justification Scores across logs 3, 5 and 6

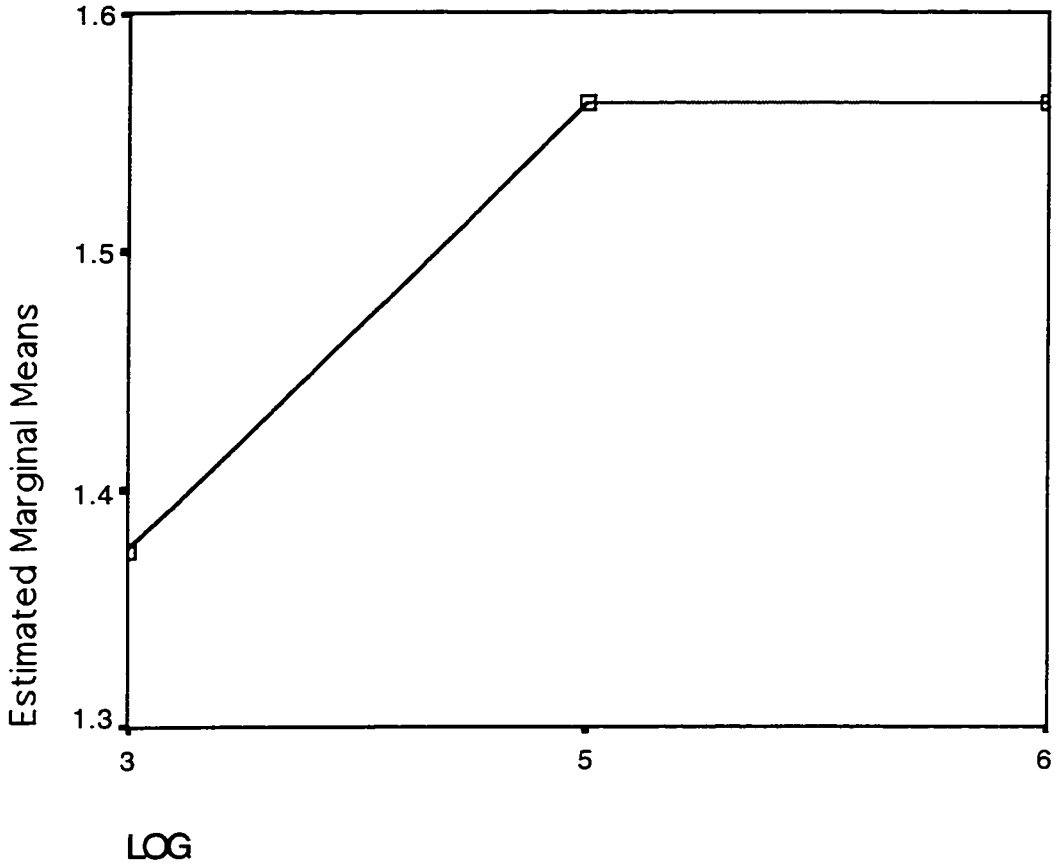
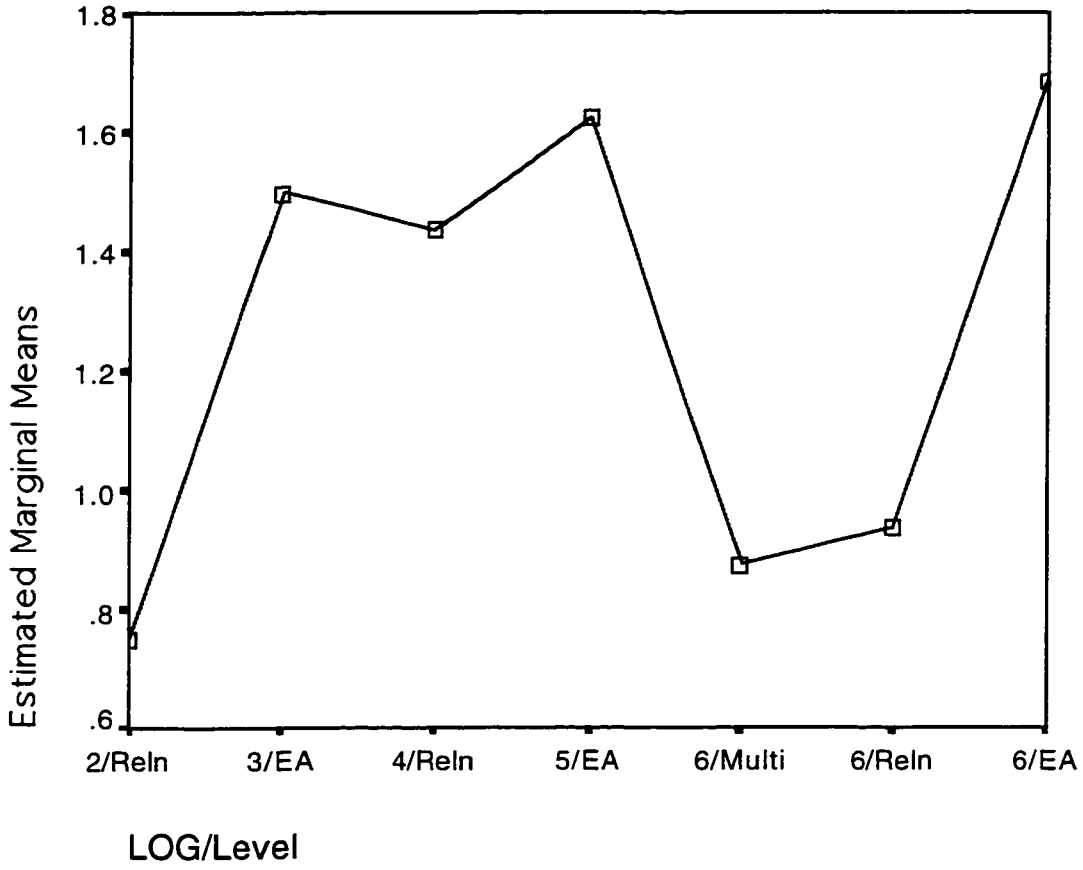


Figure 16

Means of Justification Scores for Questions from all three levels across six logs



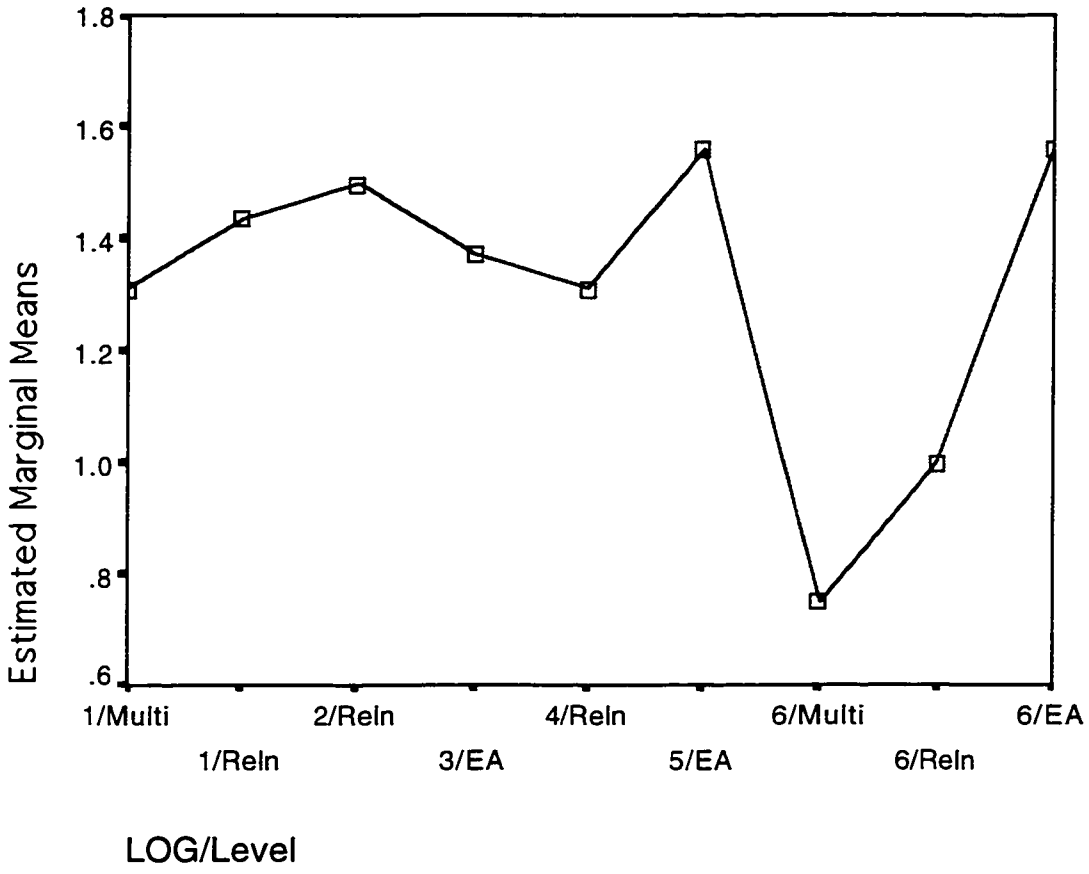
Multi =Multistructural level

Reln =Relational Level

EA = Extended Abstract Level

Figure 17

Means of Justification Scores for Answers from all three levels across six logs



Multi =Multistructural level

Reln =Relational Level

EA = Extended Abstract Level

Findings Related To Justification Scores

Multistructural question and answer justification scores. Students' multistructural answer justification scores showed statistical significance in revealing lower scores for log 6 (M=.75, SD=.93) as compared to log 1 (M=1.31,SD=.79). While the result can be

interpreted as students being less able in providing reasons for why their answers were multistructural in log 6 as compared to their ability to do the same in log 1, there is a possibility that a lack of practice in answering justifications related to multistructural levels might have contributed to the reduced performance.

Relational question justification scores. As reported earlier, relational question justification scores revealed both a main effect (see table 21) and a statistically significant, quadratic trend (see plot in figure 12). Pairwise comparisons, seen in table 22, show that students showed statistical significance in better relational question justification scores for log 4 (M=1.44, SD=.73) as compared to log 2 (M=.75, SD=.93). In addition, the reduced scores seen from log 4 to 6 (M=.94, SD=.68), explain the existence of statistically significant quadratic trend.

Relational and extended abstract answer justification scores. Students' relational answer justification scores in logs 1, 2, 4 and 6 show a non-significant, decreasing trend, (see plot in figure 12). However, the decrease in ability can only be explained as chance fluctuations, as evidenced by the non-significance of the omnibus F-test in table 21. With regards to the extended abstract level, students' abilities to justify meeting the criteria for creating extended abstract questions and answers shows an increasing trend (see plots in figures 13 and 14); the differences can be accorded to chance, as is seen from the non-significance of the omnibus F-tests in table 21.

Justification scores for questions from log 2 to log 5. Following from the main effect found in the seven justification scores for questions at all the three levels, pairwise comparisons between the seven scores are reported in table 23. Students' relational question justification scores in log 2 (M=.75, SD=.93) showed statistical significance in

being smaller than the extended abstract question scores for log 3 (M=1.50, SD=.89) and log 5 (M=1.62, SD=.62), and relational question scores for log 4 (M=1.44, SD=.73) and log 6 (M=.94, SD=.68; see table 23). This suggests that, from log 2 through to log 5, students' abilities to justify meeting the criteria for creating questions showed a statistically significant improvement between logs 2 and 3; however, any differences in justification scores for questions between logs 3, 4, and 5 can only be attributed to chance fluctuations. The relatively higher values of scores of abilities for log 3 (M=1.50, SD=.89), log 4 (M=1.44, SD=.73) and log 5 (M=1.62, SD=.62) as compared to those for log 2 (M=.75, SD=.93) also suggest that students were able to *more correctly* state what made their question relational or extended abstract between logs 3 and 5 than they did in log 2.

Justification scores for questions in log 6. Students' multistructural question justification log 6 was statistically, of a significantly lower ability, the extended abstract question justification for log 5 (see table 23). This result is not surprising, given that students had never been required to justify why their questions were of a multistructural nature until log 6. Students' relational question justification scores for log 6 (M=.94, SD=.68)were lower than each of the justification scores for questions in logs 3, 4 and 5; in fact, I suspend judgment on the significance of each of the following mean differences for justification scores related to questions: log 6, relational and log 3, extended abstract (M=1.50, SD=.89); log 6, relational and log 4, relational (M=1.44, SD=.73); log 6, relational and log 5, extended abstract (M=1.62, SD=.62; see table 23). In fact, students' ability to justify meeting the criteria for creating relational questions in log 6 was comparable to that of relational questions in log 2 (M=.75, SD=.93). Note also that

students' extended abstract question justification scores in log 6 (M=1.69, SD=.70) were significantly higher than those for the multistructural question in log 6 and relational question in log 6 (see table 23); in fact, the mean extended abstract question justification score for log 6 (M=1.69) is the highest mean of all seven justification scores related to questions. The low mean score for relational question justifications in log 6 is surprising, given that students had ample practice in conducting such a justification and receiving feedback from the instructor in logs 2 and 4, and that their justifications of meeting the criteria for creating relational questions had, statistically, improved significantly from log 2 to log 4. This low score may be attributed to the fact that log 6 was written after the posttest, and hence, students might not have engaged in writing log 6 and its accompanying TAPE to the same degree as they would for the previous logs. However, note that the hypothesized lack of engagement in completing log 6 did not affect students' justification scores for extended abstract questions. Perhaps students found it easier to understand how to justify criteria for extended abstract questions than relational questions while completing their TAPEs for log 6, because the criteria for writing an extended abstract question are more familiar to students than those for relational levels. Whereas relational level questions try to integrate ideas across readings or chapters and make implicit connections, extended abstract questions basically looked for "real-life" applications of theories.

Justification scores for answers. The main effect revealed for the nine justification scores for answers at all three levels (multistructural, relational and extended abstract) across the six logs, prompted a procedure of pairwise comparisons, the results of which have been detailed in table 24. For the most part, students' justification scores

between log 1 and log 5 showed fluctuations that can be explained as chance occurrences (range of *M*s: 1.31 to 1.56, range of *SD*s: .62 to .81). Moreover, the high values of the means (the maximum possible score was 2) indicated that students were, for the most part, *correctly* stating their justifications in meeting the criteria for creating multistructural, relational and extended abstract answers. Students' multistructural answer justification scores for log 6 (*M*=.75, *SD*=.23) was statistically found to be significantly lower than both the extended abstract answer justification scores in log 5 (*M*=1.56, *SD*=.18), and the extended abstract answer justifications in log 6 (*M*=1.56, *SD*=.16; see table 24). The relatively low multistructural answer justification scores in log 6 is not surprising, given that students only justified meeting criteria for creating multistructural answers *twice* over the course of the learning logs, once in log 1 and once in log 6; the lack of experience in writing such justifications explains the reduced score. Another set of relatively low scores to note is the relational answer justification for log 6 (*M*=1.00, *SD*=.20). Note from table 24 that I have suspended judgment on the difference between relational answer justification scores in log 6 and those for relational answers in log 2 (*M*=1.50, *SD*=.16), extended abstract answers in log 5 (*M*=1.56, *SD*=.18) and extended abstract answers in log 6 (*M*=1.56, *SD*=.16); although these differences did not reach statistical significance, the decision to suspend judgment indicates that the justifications scores of meeting assessment criteria for relational answers in log 6 were relatively low. A possible explanation for the dip in relational answer justification scores in log 6 could be the disengagement of students in writing their sixth and final log, as it was due for submission after the posttest was completed. However, similar to the discussion on justification scores for questions in log 6, the relatively high scores for

extended abstract answer justifications in log 6 (M=1.56, SD=.16) is not explained by the hypothesized disengagement of students. Again, a possible explanation for this high score is a better understanding and familiarity with the criteria for extended abstract answers, which looked for applications of theories, as oppose to the more implicit connections that were required in a relational level answer.

Summary. The repeated measures analyses, pairwise comparison procedures, and plots revealed the following about students' justification scores for multistructural, relational, and extended abstract questions and answers. First, students' multistructural answer justification scores showed a statistically significant depressed performance when measured twice (logs 1 and 6), possibly due to a lack of exposure to this type of justification. Second, students showed non-significant improvement in their extended abstract question and answer justification scores across logs 3, 5 and 6. Third, students showed a non-significant depression in their relational answer justification scores across logs 1, 2, 4 and 6. Fourth, students' relational question justification scores for log 4 was, statistically, significantly better than their ability to do so in both logs 2 and 6. Fifth, students' justification scores for questions, from log 2 to 5, showed a statistically significant improvement from log 2 to log 3, and chance fluctuations between logs 3, 4 and 5. Sixth, students' abilities in justifying meeting the criteria for creating answers revealed chance fluctuations from log 1 to 5. Seventh, students' extended abstract justification scores in log 6 were, for the most part, statistically found to be significantly higher than the both the multistructural and relational justification scores for log 6. This seventh finding suggests that students were, perhaps, more familiar with and better

understood the criteria for the extended abstract level than the relational or multistructural level.

Relationships Between Performance Assessments and Prediction Confidence

As a starting point to investigating the nature of the monitoring proficiencies, relationships between performance assessments and prediction confidences were explored. Table 25 shows the intercorrelations between performance assessments (upper triangle) and prediction confidence (lower triangle), across the six logs.

Table 25

Intercorrelations between Performance Assessments (upper triangle) and Prediction Confidence (lower triangle)

	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	-	.46	.38	.18	.22	.26
Log 2	.52*	-	.47	.42	.24	.28
Log 3	-.26	.06	-	.09	.25	.36
Log 4	.21	.15	.49*	-	.35	.56*
Log 5	0	.21	.05	.31	-	.09
Log 6	0	-.03	.03	.39	.62*	-

*Computed correlation coefficients are significant at or below the .05 level

While the performance assessment scores were largely uncorrelated, except for one statistically significant correlation between the grades assigned for logs 4 and 6 (r=.56, p=.02), confidence scores for three pairs of successive logs showed statistically

significant correlations (see table 25). The relatively low and insignificant intercorrelations for performance assessments seem to suggest that the topics covered within the learning theory curriculum were, most likely, diverse in content and varied in their relative difficulty. The uniformly insignificant intercorrelations between performance assessments across the six logs might give reason to dispel the notion of a log-general performance ability.

Investigation of intercorrelations between prediction confidence scores. Prediction confidence scores showed statistically significant correlations between the following three pairs of logs: logs 1 and 2, r=.52, p=.03; logs 3 and 4, , r=.49, p=.05; logs 5 and 6, , r=.62, p=.01 (see table 25). No other pairs of intercorrelations were found to be statistically significant. A possible explanation for the statistically significant intercorrelations could be the *structure of the logs* and the *timing of instructor feedback on the quality of the log*. Since logs 1 and 2 had similar types of question-and-answer sets (multistructural and relational), it might be of no surprise that confidence scores correlated between the two. Further, students had an opportunity to internalize feedback on the quality of the answers received for log 1 while tackling their assignment on log 2; students' confidence in prediction for log 2 might have been influenced by their performance feedback on log 1.

The lack of a statistically significant correlation between prediction confidence scores for logs 2 and 3 might be explained by the difference in the structure of the logs; log 3 consisted of one multistructural and one extended abstract question-and-answer set, while log 2 had one multistructural and one extended abstract question-and-answer set. In addition, some students had not received feedback from the instructor on log 2while

writing log 3, and probably had to contend with using the feedback on log 1 to predict their grades for log 3. However, feedback on the relational level question-and-answer sets in log 1 would, perhaps, not be very useful to the students in writing their question-and-answer sets for log 3, because log 3 was of a different structure than logs 1 and 2. Moreover, log 3 was the first occasion for students to tackle writing a question-and-answer set at the extended abstract level. Therefore, a more likely explanation for the unrelated confidence scores between logs 2 and 3 could be the fact that students were tackling a new structure of learning logs for the first time when they were completing log 3.

Despite the difference in structure between logs 3 and 4, students' prediction confidence scores for these two logs show a statistically significant correlation (r=.49, p=.05). This suggests that prediction confidence scores for logs 3 and 4 are independent of the types of question-and-answer sets that students were required to answer in these learning logs. In the case of log 4, students may not have relied on their confidence ratings for a similarly structured learning log (e.g., log 2), but instead on the confidence ratings for the most recently graded learning log (i.e., log 3). This can be seen from the lack of a statistically significant correlation between the prediction confidence scores on logs 2 and 4, both of which had the same structure, but instead a statistically significant correlation between prediction confidence scores for logs 3 and 4, as noted above.

Interestingly, the prediction confidence scores between logs 4 and 5, two logs which are differently structured, do not show a statistically significant correlation. A plausible explanation for this could be the fact that some students did not receive their feedback for learning log 4 until after they had submitted log 5. This lack of feedback, as

seen in the case of the prediction confidence scores between logs 2 and 3, might have resulted in insignificant correlations between prediction confidence scores for logs 4 and 5 (r=.31, p=.23) as well. The low, insignificant correlation between the prediction confidence scores for log 3 and 5 (r=.05, p=.86) indicates that similarity in structure of logs did not influence the prediction confidence scores. The prediction confidence scores between logs 5 and 6 (r=.62, p=.01) show a statistically significant correlation, though; this points again to the lack of dependence between log structure and prediction confidence scores. Another possible avenue for explaining the strong correlation in prediction confidence scores between logs 5 and 6 is the fact that, since log 6 was written after the examination and students had already received their feedback on log 5, students might have used their most recently received feedback to make their predictions and confidence ratings for log 6.

Partial intercorrelations between prediction confidence scores. To better investigate the within-log relationship between performance assessments and prediction confidence, I computed partial correlations for the prediction confidence scores across the logs, by controlling for any variance attributable to performance assessments. These partial correlations yielded two statistically significant coefficients. The first was prediction confidence scores between logs 3 and 4, partial r=.76, p=.01; the second was prediction confidence scores between logs 5 and 6, partial r=.65, p=.04. Across the six logs, all but one of the partial correlation coefficients represented an increase in magnitude as compared to their counterpart simple, bivariate correlations reported earlier; the only exception was the coefficient for prediction confidence scores between logs 1 and 2, which showed a reduced, insignificant, partial correlation of .31 at a p level of .38.

It remained now to explain why the partial correlation coefficient for prediction confidence scores between logs 1 and 2 reduced in its significance, while those for all other correlations increased. One possible explanation would be that the prediction confidence scores for log 1 might show significant correlations with various values of the performance assessments scores across the six logs. I therefore computed intercorrelations between the six performance assessment scores and six prediction confidence scores (see table 26 below) and discovered statistically significant coefficients between prediction confidence scores in log 1 scores with performance assessments on logs 2, 4 and 5. These strong relationships help explain that some performance assessment scores did in fact contribute to the variation in students' prediction confidence scores for log 1, thereby causing the reduction in partial correlation coefficients involving prediction confidence scores for log 1.

Domain-generality of confidence judgments. The general improvement in significance of the partial correlation coefficients for prediction confidence scores when the variance accounted for by performance assessments is removed suggest that, for the most part, the prediction confidence scores on one log were related to prediction confidence scores on another log over and above the instructor's assessments of performance. However, the lack of a uniformly statistically significant set of correlations between confidence in prediction scores makes it difficult to support the domain-general hypothesis; if the domain-general hypothesis were true, confidence judgments should show statistically significant associations regardless of performance assessments. Given the small sample size being investigated in this study, I would conjecture that the improved partial correlation coefficients between prediction confidence scores when

performance assessments are controlled for, is an encouraging finding. Moreover, the explanations offered for the patterns of statistically significant intercorrelations and improved partial correlations between confidence judgments, show that for each of the pairs of logs 1 and 2, logs 3 and 4, as well as logs 5 and 6, confidence judgments were strongly related regardless of performance on the tests. Therefore, three distinct confidence abilities or patterns might have been operating, one each across the three pairs of logs 1 and 2, 3 and 4, and 5 and 6. While this does not support the existence of a general confidence ability, which would in turn, support the domain- general monitoring hypothesis, the three sets of confidence abilities reveal that confidence was not a log-specific ability. While confidence judgments might have been related more significantly and uniformly in a more controlled environment, the results from these analyses show how confidence judgments unravel in a realistic learning environment for a complex, writing task.

Relationship between instructor-assessed performances and confidence in predictions. The different patterns of correlations observed in the intercorrelations for performance assessments and prediction confidence scores lead me to contend that within-log performance assessment and prediction confidence are not strongly correlated. The within-log correlations between instructor assessed performance and confidence in predictions (table 26) were as follows: log 1, .38; log 2, .12; log 3, .19; log 4, .01; log 5, -.27; log 6, .53. Only the coefficient calculated for log 6 was statistically significant at a p level of .034. Therefore, within each log, except for log 6, prediction confidence scores were unrelated to performance assessments. Students' prediction confidence scores for log 1 showed statistically significant relationships with the performance assessments for

logs 2, 4 and 5, however, there does not seem to be any meaningful explanation for these relationships. Results from table 26 also reveal that students' prediction confidence scores for any one log were, for the most part, unrelated to their performance assessments on the previous log; this suggests that an increased performance on one log did not necessarily translate to an improved prediction confidence on the next log.

Table 26

Correlations between Performance Assessments and Prediction Confidence scores across six logs

Performance Assessment Scores	Prediction Confidence Scores					
	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	.38	.41	-.02	.57*	.11	.11
Log 2	.58*	.12	-.12	.33	-.31	-.02
Log 3	.20	.01	.19	.47	-.14	.10
Log 4	.59*	.13	-.54*	.01	.24	.26
Log 5	.50*	.25	-.25	-.06	-.27	.15
Log 6	0	-.26	-.34	.24	.34	.53*

*Computed correlation coefficients are significant at or below the .05 level

Summary. This exploratory correlation procedure and its accompanying discussion leads me to contend that, first, performance assessments across logs 1 to 6 were largely unrelated possibly due to varying levels of difficulty in the content of the course, as well as possibly due to the inherent differences in the domains of the content

covered on a weekly basis. Second, prediction confidence scores showed statistically significant relations between two successive logs, possibly *only* when feedback on performance on the earlier of the two logs was received by the student before the later of the two logs was submitted. Third, students' prediction confidence scores between each of the three pairs of logs 1 and 2, logs 3 and 4, as well as log 5 and 6 were statistically found to be significantly related regardless of the performance assessment, providing support for the existence of three distinct confidence abilities or patterns, each acting across a distinct pair of logs. Fourth, students' prediction confidence scores on any one log acted, for the most part, without any significant relation to the performance assessment on both the same log and the previous log. Fifth, students' prediction confidence scores were largely unaffected by the changing structure of the learning log assignment.

Relations Between Performance Predictions, Performance Assessments and Prediction Confidence

Intercorrelations between performance predictions. I next ran procedures to compute intercorrelations among the performance prediction scores for all six logs; calculated correlation coefficients are shown in the lower triangle of table 27. Students' predicted performance on their learning logs showed several statistically significant intercorrelations, specifically, between logs 1 and 4, logs 2 and 4, logs 3 and 4, logs 3 and 5, logs 3 and 6, logs 4 and 5, as well as logs 5 and 6. This suggests that, possibly, as instruction progressed, students' predictions on a log were related to their predictions on the two or three most recently completed logs, most notably, regardless of the difference in structure of the logs. In support of this contention, note how predictions for log 4 are

related to those on logs 1,2 and 3; predictions on log 5 are related to those on logs 3 and 4, while predictions on log 6 are related to those on logs 3, 4 and 5. It is interesting to note that predictions on logs 1, 2 and 3 are unrelated to one another, suggesting that, during the first half of the course (i.e., logs 1 to 3), students predictions might have been influenced by factors other than their prior predictions. The largely uniform, significant intercorrelations between logs 3, 4, 5 and 6 suggest the appearance of a general prediction ability after the first two logs.

Table 27

Intercorrelations among Performance Predictions (lower triangle) and Partial Intercorrelations among Performance Predictions, controlling for Prediction Confidence (upper triangle)

	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	-	.43	.52	.72*	.50	.19
Log 2	.45	-	.30	.35	.13	-.36
Log 3	.41	.31	-	.50	.61	.46
Log 4	.69*	.48*	.56*	-	.54	.36
Log 5	.34	.33	.61*	.51*	-	.80*
Log 6	.34	-.09	.54*	.33	.76*	-

*Computed correlation coefficients are significant at or below the .05 level

Relation between within-log performance predictions and prediction confidence scores. To explore relations between prediction confidence scores and performance predictions, partial intercorrelations were computed among performance predictions

across the six logs, controlling for variation due to prediction confidence (see upper triangle in table 27). There were no clear patterns of increased or decreased partial correlations in comparison with the simple, bivariate intercorrelations shown in the lower triangle. This suggests that within-log prediction confidence scores and performance predictions displayed a complex, yet mostly insignificant relationship, as confirmed by the six non-significant computed intercorrelations between prediction confidence scores and prediction performances (see table 28): log 1, .36; log 2, .22; log 3, -.04; log 4, .12; log 5, -.20; log 6, .36. The non-uniformity of the signs of the correlations shows that students' prediction confidence scores did not always increase with higher performance predictions.

Relation between performance assessments and performance predictions. I also explored relations between within-log performance predictions and performance assessments (see table 29); the computed within-log correlation coefficients were as follows: log 1, -.19; log 2, .60; log 3, .46; log 4, .14; log 5, .59; log 6, .37. Within-log correlations for logs 2 and 5 were statistically significant at the .05 level. From table 29, the other statistically significant correlations are between (a) performance assessments on log 2 and performance predictions on log 3 (r=.49, p=.05), (b) performance assessments on log 2 and performance predictions on log 4 (r=.58, p=.02), as well as, (c) performance assessments on log 6 and performance predictions on log 3 (r=.61, p=.01). The statistically significant correlations between performance assessments and performance predictions for logs 2 and 5 suggest that more students were accurate in their predictions for these two logs than in the other logs. In fact, for log 2, ten out of the sixteen students perfectly predicted their grades, while for log 5, eleven students showed perfect

predictions. Students' predicted grades on log 3 being related to their performance on log 2 is surprising, because half the students did not receive feedback on their performance in log 2 until after they had submitted log 3, and were preparing for log 4. Further, log 3 was of a different structure than log 2 and students were tackling the extended abstract level for the first time in log 3. The performance predictions on log 4 being related to the performance assessments on log 2 suggests that the delayed feedback on log 2 might have affected the students' performance predictions while they were writing log 4; there is also a possibility that the common structure between logs 2 and 4 might explain the strong correlation. To explore the effect of performance assessments on students' performance predictions, I computed partial intercorrelations between predicted performances, controlling for any variation in the performance assessments. No uniform pattern of increased or decreased partial intercorrelations for performance predictions were observed, suggesting that, for this sample, performance predictions and performance assessments influenced each other in an indiscernible manner.

Overall, performance predictions and performance assessments seem to share a complex relationship. Performance assessments on log 2 show statistically significant relations to students' performance predictions on the two subsequent logs, despite the change in structure of logs and the difference in content being covered across the three logs. Statistically significant within-log relationships between performance assessments and performance predictions on logs 2 and 5 suggest increased number of students making accurate predictions.

Effect of prediction confidence on relationship between performance predictions and performance assessments. To better understand how performance predictions and

performance assessments were related, I decided to view their relationship controlling for the variability in prediction confidence scores. Within-log, partial correlation coefficients between performance predictions and performance assessments, removing the variation due to prediction confidence were as follows, log 1, .002; log 2, .62; log 3, .63; log 4, -.23; log 5, .51; log 6, .47. While none of the partial correlation coefficients were statistically significant at the .05 level, there is an increase in their values for logs 1, 2, 3 and 6 as compared to their counterpart simple, bivariate correlation coefficient (see table 28 for the intercorrelations between prediction confidence and performance predictions). This result suggests that the within-log relationship between predicted performance and performance assessments for logs 1, 2, 3 and 6 existed over and above students' prediction confidence, to the point where the relationship was strengthened when confidence was kept constant. While the partial intercorrelations between performance predictions and performance assessments for logs 4 and 5 is lower than the simple, bivariate correlations, note how the change in intercorrelations for log 4, from an *r* of .14 to a partial *r* of -.23, is more drastic than that for log 5, from an *r* of .59 to a partial *r* of .51. A glance at the intercorrelations between performance assessments and prediction confidence scores (table 25) and intercorrelations between performance predictions and prediction confidence (table 28) reveals possible explanations for the reduction in partial correlation between performance predictions and performance assessments for log 4. Performance predictions on log 4 shows statistically significant correlations with prediction confidence scores on log 1 with a coefficient of .50, at a *p* level of .04. Performance assessments on log 4 display a statistically significant, positive correlation with confidence in prediction on log 1 (r=.59) and a statistically significant, negative

correlation with confidence in prediction on log 3 (r=-.54), both at a p level of .05. These statistically significant correlations reveal the source for the anomalous behavior of the partial within-log coefficient between predicted performance and instructor's assessment of performance for log 4, but, unfortunately, do not provide a likely explanation for the phenomenon. Apart from this anomaly, the within-log relation between students' performance predictions and the performance assessments was strengthened, for the most part, when variability due to prediction confidences was controlled.

Table 28

Intercorrelations between Prediction Confidence Scores and Performance Predictions across logs.

Performance	Prediction Confidence					
Predictions	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	.36	-.13	.13	.08	.11	-.12
Log 2	.46	.22	.01	.13	.05	.17
Log 3	.13	-.36	-.05	-.01	-.16	.18
Log 4	.50*	-.01	.01	.12	-.16	-.06
Log 5	.30	-.07	-.24	-.32	-.20	.21
Log 6	0	-.30	-.01	-.11	.17	.36

*Computed correlation coefficients are significant at or below the .05 level

Table 29

Intercorrelations between Performance Assessments and Performance Predictions across logs

Performance	Predicted Performances					
Assessments	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	-.19	.27	-.02	.05	-.32	-.39
Log 2	.35	.60*	.49*	.58*	.31	.01
Log 3	.19	.23	.46	.28	.14	.20
Log 4	.03	.26	.26	.14	.39	.33
Log 5	-.14	.32	.29	.34	.59*	.20
Log 6	-.14	-.03	.61*	.07	.17	.37

*Computed correlation coefficients are significant at or below the .05 level

Summary. The exploration of the relationship between students' performance predictions, instructor's performance assessments, and students' prediction confidences leads me to the following conjectures. First, student' performance predictions are revealed as a general ability only after the first two logs. Second, students' performance predictions share a complex relationship with the prediction confidence scores, with higher confidences not always resulting in higher predictions. Third, the within-log relationship between students' performance predictions and instructor's performance assessments is strengthened when variation due to students' prediction confidence scores are controlled.

Intercorrelations Among Prediction Accuracy Scores

The lower triangle of table 30 reveals the intercorrelations between the scores of prediction accuracy across the six logs. Recall that prediction accuracy was calculated for each student on each log by taking the absolute value of the difference between the performance assessment and performance prediction for that log. Prediction accuracy scores for log 3 showed statistically significant correlations with those for log 1 (r=.77, p=.001), and for log 2 (r=.51, p=.05). In addition, prediction in accuracy for log 5 showed statistically significant correlations with those for log 2 (r=.57, p=.02) and for log 3 (r=.58, p=.02). No other intercorrelations were found to be significant. The pattern of intercorrelations seen in the prediction accuracy is neither similar to those seen for performance assessment nor performance predictions. The results suggest that for the first three logs, students' prediction accuracy followed a similar pattern. The difference in the structure of logs 1, 2 and 3 did not seem to affect the pattern with which students were developing their prediction accuracy. Note, however, that subsequently, for log 4, prediction accuracy scores were not related to any of the other accuracy scores on other logs. Prediction accuracy scores for log 5 revealed strong relations with those for logs 2 and 3, suggesting that students reverted to predicting their performance with the same accuracy as they were doing earlier in the course. Accuracies for log 6 are not related to any of the other accuracies, perhaps due to the disengagement of students in completing the final log after their posttest.

Relation between prediction confidence and prediction accuracy. To explore whether students' prediction accuracy scores were affected by their prediction confidence scores, I calculated partial intercorrelation coefficients for prediction accuracy scores

across the six logs, after removing variance accounted for by students' prediction confidence scores (see upper triangle of table 30). While there are no clear patterns of increased or decreased partial correlations as compared to the simple, bivariate correlations in the lower triangle, notice how the partial correlations for prediction accuracy scores between logs 1 and 3 (partial r=.80, p=.006), logs 2 and 3 (partial r=.51, p=.14), as well as logs 5 and 2 (partial r=.62, p=.05) show an increase from the simple, bivariate, statistically significant correlations discussed earlier. Only one of the statistically significant intercorrelations between prediction accuracy scores, namely between logs 5 and 3, showed a decreased partial correlation coefficient when prediction confidence was kept constant.

Summary. The results suggest that, first, students' prediction accuracy showed patterns of a general ability only in logs 1, 2, 3 and 5. Second, students' prediction accuracy was related between logs 1 and 3, logs 2 and 3, and between logs 2 and 5 regardless of students' confidence in predictions.

Table 30

Intercorrelations among Prediction Accuracy scores (lower triangle) and Partial Intercorrelations among Prediction Accuracy scores after removing the effect of Prediction Confidence (upper triangle)

	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	-	.28	.80*	-.20	.24	.68*
Log 2	.34	-	.51	.17	.62*	.38
Log 3	.77*	.51*	-	-.35	.32	.45
Log 4	.30	.27	.06	-	-.46	.20
Log 5	.43	.57*	.58*	-.05	-	.20
Log 6	.20	.15	.17	.13	-.05	-

*Computed correlation coefficients are significant at or below the .05 level

Relationships Involving Monitoring Proficiencies of Discrimination and Bias

Table 31 provides the intercorrelations for the two derived monitoring proficiencies; correlations across the six discrimination scores is in the lower triangle, while those for bias are in the upper triangle. Recall that discrimination measured students' ability to assign an appropriate level of confidence for their predicted performances, while bias provided a measure of over and under-confidence in students' judgments of prediction.

Table 31

Intercorrelations among Discrimination scores (lower triangle) and Bias scores (upper triangle).

Discrimination	Bias					
	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	-	.28	.62*	.53*	.25	.33
Log 2	.67*	-	.52*	.32	.58*	.17
Log 3	.67*	.80*	-	.58*	.45	.16
Log 4	.18	.55*	.59*	-	.30	.15
Log 5	.31	.48	.56*	.52*	-	.28
Log 6	.14	.35	.26	.27	.36	-

*Computed correlation coefficients are significant at or below the .05 level

Exploring relations between discrimination scores. The correlations revealed in the lower triangle of table 31 suggest the existence of two discrimination abilities or patterns across two sets of logs, the first being logs 1, 2 and 3 and the second being logs 2, 3, 4 and 5. Discrimination scores for log 6 were not correlated to any of the other scores, most likely, due to the disengagement of students in writing their sixth and final log, after having completed their posttest. Only one of the intercorrelations of discrimination scores between logs 2, 3, 4 and 5 is anomalous; the intercorrelation between logs 2 and 5 almost reaches significance (r=.48, p=.054). Apart from this anomalous correlation value, the uniform set of statistically significant correlations between logs 1, 2 and 3, as well as between logs 3, 4 and 5 reveal two distinct sets of

discrimination abilities. This would suggest that students' ability to discriminate correct and incorrect items, by assigning an appropriate level of confidence to each of their logs' predictions, showed evidence of possessing two distinct patterns, one across logs 1 to 3, and another across logs 3 to 5. This finding is further supported by the increase in the values of all statistically significant correlations when partial intercorrelation coefficients were computed for discrimination scores, after removing the variation due to the performance assessments. This suggested that the observed patterns of discrimination in predictions acted over and above the performance assessments.

Relations between discrimination scores, performance assessments, and prediction confidence scores. To further corroborate the existence of the two discrimination abilities across the six logs, I calculated correlations between (a) discrimination scores and performance assessments (table 32), (b) discrimination scores and prediction confidence scores (table 33). Results of the correlational procedures reveal few statistically significant correlations; those that were statistically significant between discrimination and prediction confidence lay between the values of -.50 to -.63 (see table 33), while the only significant correlation between discrimination and performance assessment was of the value -.71. Moreover, a glance at tables 32 and 33 also reveal non-uniformity in the valence (sign) of the correlations. This suggests that both the relationships between discrimination and performance assessment, as well as between discrimination and prediction confidence are idiosyncratic, varying from log to log with no discernible patterns in both magnitude and sign.

Table 32

Correlations between Discrimination Scores and Performance Assessments.

Performance	Discrimination					
Assessments	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	-.32	-.18	-.26	-.21	-.71*	-.05
Log 2	-.17	.21	.33	.11	.11	.34
Log 3	-.38	-.04	-.14	-.06	-.15	.22
Log 4	.29	.30	.34	.19	.15	.27
Log 5	-.12	.02	.09	.45	-.13	-.02
Log 6	.14	.01	.14	.04	.03	.47

*Computed correlation coefficients are significant at or below the .05 level

Table 33

Correlations between Discrimination and Prediction Confidence scores

Prediction	Discrimination					
Confidence	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	-.18	.25	.14	.33	-.01	0
Log 2	.07	.15	-.10	-.23	-.52*	-.39
Log 3	-.38	-.31	-.31	-.37	-.30	-.07
Log 4	-.63*	-.38	-.42	-.37	-.50*	-.03
Log 5	.24	.15	-.05	-.12	-.15	.05
Log 6	-.05	-.18	-.07	.05	-.02	.05

*Computed correlation coefficients are significant at or below the .05 level

Relationships between bias scores. The values of correlation coefficients in the upper triangle of table 31 reveal intercorrelations between students' bias scores across the six logs. The uniformly positive signs for the coefficients suggest that students' tended to be uniformly over or under-confident; from the descriptive statistics for the bias in prediction scores (table 11), it is clear the students were, for the most part, under-confident of their predictions. The statistically significant correlations between logs 1 and 3 (r=.62, p=.01), logs 1 and 4 (r=.53, p=.03), as well as logs 3 and 4 (r=.58, p=.02) suggest that a bias ability or pattern exists between these three logs. However, across the other sets of logs, there are only two other statistically significant correlations, namely, between bias in predictions for logs 2 and 3 (r=.52, p=.03), and for logs 2 and 5 (r=.58, p=.02). It is tempting to suggest that a bias pattern exists between logs 2, 3 and 5, considering, in addition to the two significant correlations between bias in prediction scores for logs 2 and 5 and logs 2 and 3, that the correlation between bias in predictions for logs 3 and 5 almost reaches significance (r=.45, p=.07). Overall, the results suggest two distinct bias patterns, one across logs 1, 3 and 4, and another, less evident one, across logs 2, 3 and 5. This finding is further supported by the increase in the values of all statistically significant correlations when partial intercorrelation coefficients were computed for bias scores, after removing the variation due to the performance assessments. This suggested that the observed patterns of bias in predictions acted over and above the performance assessments.

Relationship between bias scores, prediction confidence scores and performance assessments. To better understand the nature of the bias patterns that emerged from the results of correlational analyses seen in table 31, I computed intercorrelations between

bias scores and performance assessments (table 34), as well as between bias in prediction scores and prediction confidence scores (table 35). The lack of statistically significant correlations between bias and performance assessments, as well as the non-uniformity in valence (sign) of the correlations (see table 34) show bias and performance assessments to be related in an eccentric fashion, with unpredictable variation from one log to the next, in both magnitude and sign. A glance at table 35 demonstrates that similar suggestions can also be made about the relationship between bias and prediction confidence scores due to the lack of significant correlations (only two of 36 possible) and the non-uniformity in the sign of the correlations.

Table 34

Correlations between Bias scores and Performance Assessments

Performance	Bias					
Assessments	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	-.16	.16	.10	-.07	.26	-.44
Log 2	0	.22	.31	.16	.43	-.24
Log 3	-.40	.15	0	.27	-.17	-.40
Log 4	-.22	.39	.08	0	.52*	.14
Log 5	-.34	-.05	-.09	.07	.14	.09
Log 6	-.42	.55*	-.06	-.15	.18	-.18

*Computed correlation coefficients are significant at or below the .05 level

Table 35

Correlations between Bias and Prediction Confidence scores

Prediction	Bias					
Confidence	Log 1	Log 2	Log 3	Log 4	Log 5	Log 6
Log 1	.15	.20	.39	.25	.51*	.03
Log 2	.27	.07	.21	.20	.39	.08
Log 3	.26	-.11	.31	.28	-.23	.11
Log 4	-.10	.34	.22	.17	.18	-.19
Log 5	.28	.56*	.09	.16	.15	.08
Log 6	-.08	.44	0	-.02	.22	-.05

*Computed correlation coefficients are significant at or below the .05 level

Summary. The results of the correlational analyses on discrimination and bias in predictions, reveals the following. First, students' discrimination and bias in predictions were each related in complex, idiosyncratic, but insignificant fashion with instructor's assessment of performances and confidence in predictions, suggesting that both these measures of monitoring proficiency were affected, but not to a large extent, by performance assessment and prediction confidence scores. Second, discrimination scores revealed two distinct patterns, one across logs 1 to 3, and the other across logs 3 to 5. Third, bias also revealed two distinct patterns, the first was across logs 1, 3 and 4, while the second was across logs 2, 5 and 6.

Exploring Relationships Between Performance Assessments And Independent Multistructural, Relational, Extended Abstract and Overall Performance Scores.

In order to determine whether the performance assessments across the six logs took into account how well students' had met the assessment criteria for creating their learning logs, I computed within-log correlations between the instructor's assessments of performance and the independent overall performance scores; these correlations were as follows: log 1, .56; log 2, .67; log 3, .33; log 4, .73; log 5, .57; log 6, .85. Apart from the coefficient for log 3, all other correlations between instructor's assessment of performances and independently calculated scores of performance were significant at a *p* value of .02 or less. It remained now to see whether there were any relationships between the independent multistructural, relational, or extended abstract performance scores and the performance assessments.

Relations between independently calculated scores for multistructural level sets and instructor's assessments of performance. Within-log correlation coefficients between the independent multistructural performance scores and performance assessments across the six logs were as follows: log 1, .37; log 2, .20; log 3, .13; log 4, .51; log 5, .22; log 6, .15. Apart from the coefficient for log 4, which was significant at a *p* value of .04, all other correlations were insignificant.

Relations between independent relational and extended abstract performance scores and performance assessments. Recall that students created relational level question-and-answer sets only for logs 1, 2, 4 and 6, while extended abstract level question-and-answer sets were created for log 3, 5 and 6. I computed within-log correlation coefficients between the independent relational performance scores with the performance assessments for logs 1, 2, 4 and 6; the correlation coefficients, all statistically significant at a *p* value of .02 or less, were as follows: log 1, .58; log 2, .63;

log 4, .63; log 6, .63. Finally, within-log correlation coefficients between independent extended abstract performance scores and performance assessments for logs 3, 5 and 6 were as follows: log 3, .38; log 5, .69; log 6, .79; apart from the coefficient for log 3, the other two correlations were significant at a p level of .01 or less.

Summary. The correlational procedures employed to determine the nature of the relationship between the independent performance scores and the performance assessments across the six logs revealed the following. First, performance assessments showed strong and statistically significant within-log correlations with the independent overall performance scores. Second, independent multistructural performance scores showed no significant relation with the performance assessment, however, the independent relational and extended abstract performance scores showed significant correlations with the instructor' assessment of performance, except for log 3. This second finding suggests that the instructor's performance assessment on the log seemed to be based solely on the criteria that were set for the relational and extended abstract level question-and-answer sets, and not on the criteria set for multistructural level question-and-answer sets.

Relation Between Justification of Meeting Criteria for Creating Learning Log Questions/Answers And Monitoring Proficiencies

The final set of analyses pertains to the issue of whether, across the six logs, students' understanding of the assessment criteria had any relation with the monitoring proficiencies of discrimination and bias. The TAPEs from each of the logs attempted to measure students' understandings of assessment criteria by asking them to justify how they had met the criteria in creating, for the most part, their relational and extended

abstract learning log questions and answers. To investigate whether students' justification scores for questions and answers at the relational and extended abstract levels was related to their monitoring proficiencies, I conducted a set of correlational analyses, which are detailed in this section. I do not include students' justifications of meeting criteria for multistructural level questions and answers, because students had limited opportunity to write such justifications; in fact, the justification for meeting criteria for creating multistructural question was required only in log 6, while those for multistructural answers were conducted twice, namely, log 1 and 6. On the other hand, from logs 1 to 5, students were asked to justify meeting the criteria for creating their relational or extended abstract level set, and in the case of log 6, students were asked to justify meeting the criteria for creating both relational and extended abstract sets. It therefore, makes more sense to determine the associations between students' justifications for the relational and extended abstract level of questions and answers with their monitoring proficiencies; table 36 details these relevant correlations. Note also that bias and discrimination were both measured on a scale of 1 to −1, while the justification items were measured on a scale of 0 to 2.

Relation between justifications of meeting criteria for questions and bias scores. Justification scores related to questions reveal a pattern in their within-log correlations with bias across the six logs (table 36). Correlations begin with a small, insignificant positive value of .06 (relational question for log 2) and steadily increase till a value of .33 (relational question for log 4) before dropping back to finally reach a small, negative correlation of -.02 (relational question for log 6) and small, positive correlation of .10 (extended abstract question for log 6). Note, from table 11, that as students moved from

logs 2 to 5, their scores for bias become increasingly negative, suggesting that students' became progressively more under-confident in their predictions. From table 12, notice that there is a steady increase in scores from logs 2 to 5, in students' abilities to justify meeting the criteria for question-related TAPE items. Bias scores for log 6 show students to be relatively overconfident, while their justification scores for question-related items vary significantly between the relational question for log 6 (M=.94, SD=.68) and the extended abstract question for log 6 (M=1.69, SD=.70), as seen from the results of pairwise comparisons conducted earlier.

The descriptive statistics from tables 11 and 12 and correlations reported between the question-related justification items and the scores for bias in table 36 reveal that, first, while students' bias scores were at their most proficient during log 2, as understandings of the assessment criteria for relational and extended abstract levels improved, students' under-confidences increased. Second, while the correlations between the measure of understanding the assessment criteria for questions and bias improved from logs 2 to 4, they still remained insignificant, suggesting that bias and understanding of criteria for questions are not strongly related. Third, the decrease in the significance of correlations from logs 4 to 6 reveal that while students showed more variation in their understanding of assessment criteria for questions (SDs for justification of criteria for questions from logs 4 to 6 ranged from .68 to .72) , their scores of bias were more clustered (SDs for bias in predictions from logs 4 to 6 ranged from .16 to .25). Fourth, for the most part, the relation between bias in predictions and understandings of assessment criteria did not seem to be affected by the changing structure of the logs; as noted from table 36, the

within-log correlations showed an increase, peaking at log 4, subsequently decreasing regardless of log structure.

Relation between justifications of meeting criteria for questions and discrimination scores. Recall from table 11 that between logs 1 and 5, students showed improved abilities to discriminate, that is, as logs progressed from log 1 to 5, students were better able to assign an appropriate level of confidence to their predictions. A distinct pattern emerges for correlations between students' understandings of criteria for questions and discrimination scores. Table 36 shows that, except for log 4, within-log correlations between justifications of criteria for questions and discrimination are negative and insignificant. These negative correlations range from -.19 to -.03, while the only positive correlation is a small, insignificant value of .11. The results indicate that discrimination and understanding of assessment criteria for questions are for the most part, not related to one another across the six logs; however, the largely negative values of the correlations suggests that increased understandings of criteria for questions do not necessarily translate into improved abilities of assigning appropriate confidence levels to predictions made.

Relation between justifications of meeting criteria for answers and bias scores. Results from table 12 show that from logs 1 to 5, students' justification scores for answers showed slight fluctuations (range of *M*s: 1.29 to 1.59, range of *SD*s: .62 to .80); improvements were seen from log 1 to 2, after which, a slight depression in performance occurred through logs 3 and 4, followed by an improvement in performance in log 5. In log 6, results from the pairwise comparison conducted earlier show that there was a significant difference between students' abilities to justify meeting criteria for the

relational question (M=1.00, SD=.82) and the extended abstract question (M=1.56, SD=.63).

Correlations between justifications of meeting criteria for answers and bias did not show any clear pattern. There were moderate correlations between justifications for the relational answer in log 1 and bias for log 1 (r=.42, p=.09), and between justification for the extended abstract answer in log 3 and bias for log 3 (r=-.42, p=.08). Overall, however, the relationship between students' bias scores and their understanding of the assessment criteria for relational and extended abstract answers is log specific; the difference in the valences across the six logs, as well as the insignificance of all the correlations reveal the relationship to be, for the most part, idiosyncratic and weak.

Relation between justifications of meeting criteria for answers and discrimination scores. Results from table 36, show that within-log correlations form logs 1 to 6, between students' understandings of assessment criteria for answers and their discrimination scores are insignificant and fluctuate in their valence. This suggests that, while discrimination scores showed improvement from logs 1 to 5, and students' understandings of assessment criteria for relational and extended abstract answers showed an overall improvement between logs 1 and 5, the relationship between discrimination and justification of criteria for answers was, mostly, a log-specific phenomenon.

Summary. The results of the correlational procedures between the monitoring proficiencies and the justifications for meeting the question and answer criteria revealed the following. First, the relationship between students' bias scores and their relational and extended abstract question justification scores showed increasing associations from logs 2

to 4, and decreasing associations between logs 4 and 6, despite differences in the structure of the logs. Second, all other relationships between monitoring proficiencies and justifications related to questions or answers were insignificant and for the most, part log-specific phenomena.

Table 36 *Within-log Correlations Between Justifications of Meeting Assessment Criteria and Monitoring Proficiencies*

	Justifications of Meeting Assessment Criteria												
	Relational Answer Log 1	Relational Question Log 2	Relational Answer Log 2	Extended Abstract Question Log 3	Extended Abstract Answer Log 3	Relational Question Log 4	Relational Answer Log 4	Extended Abstract Question Log 5	Extended Abstract Answer Log 5	Relational Question Log 6	Relational Answer Log 6	Extended Abstract Question Log 6	Extended Abstract Answer Log 6
B1	.42	-	-	-	-	-	-	-	-	-	-	-	-
D1	-.03	-	-	-	-	-	-	-	-	-	-	-	-
B2	-	.06	0	-	-	-	-	-	-	-	-	-	-
D2	-	-.10	.26	-	-	-	-	-	-	-	-	-	-
B3	-	-	-	.31	-.44	-	-	-	-	-	-	-	-
D3	-	-	-	-.19	.03	-	-	-	-	-	-	-	-
B4	-	-	-	-	-	.33	-.14	-	-	-	-	-	-
D4	-	-	-	-	-	.11	-.26	-	-	-	-	-	-
B5	-	-	-	-	-	-	-	.22	.07	-	-	-	
D5	-	-	-	-	-	-	-	-.08	-.03	-	-	-	-
B6	-	-	-	-	-	-	-	-	-	-.02	.12	.10	-.05
D6	-	-	-	-	-	-	-	-	-	-.06	.43	-.13	.24

Note. *n* for all the logs was 17, except for log 6, where the *n* was reduced to 16. B denotes bias, D denotes Discrimination, the numerals 1 to 6 denote log numbers (e.g., B1 represents bias score for log 1). No correlations were found to be significant. Dashes denote no correlations were calculated between scores

Chapter 4: Discussion

Interpretation of Findings

Overview

One of the purposes of this study was to explore the development of monitoring proficiencies of graduate learners in the context of a complex writing task; the writing task, in this case, was a set of six weekly learning logs in the context of a graduate Learning Theories course. Monitoring proficiencies included (a) discrimination, or learners' abilities to assign an appropriate level of confidence to their predictions of grades on their learning logs, and (b) bias, or the degree to which learners were over or under-confident in their predictions of grades on their learning logs. A second purpose of this study was to explore the domain-generality and domain-specificity of graduate learners' monitoring abilities as they engaged with the learning log task. The domain-general hypothesis contends that learners use context-general as well as context-specific monitoring skills while completing learning tasks, while the domain-specific hypothesis subscribes to the notion that learners' monitoring skills are tied down to the context and vary from one learning task to the next. Third, this study aimed to explore whether learners' performance on the learning log task was an accurate reflection of the extent to which they were meeting the assessment criteria for the task. The fourth and final purpose of this study was to explore the relationship between learners' task understanding and their monitoring proficiencies. Specifically, this study investigated the relationship

between learners' perceptions of the assessment criteria for the learning log task and their monitoring abilities, as measured by discrimination and bias.

The discussion chapter of this thesis is organized in terms of the research questions; these questions are directly related to the purposes of the study, which have been restated above. The first of these questions explores how graduate learners' monitoring proficiencies developed over the course of the instruction in the context of a complex writing task. The second asks whether learners' abilities to meet the assessment criteria for a complex writing task are reflected in the instructor's assessments of their performances. The third and final question asks whether a relationship exists between learners' understandings of the assessment criteria for a complex writing task and their monitoring proficiencies.

How do graduate learners' monitoring proficiencies develop over the course of instruction as they tackle a complex writing task?

Support for domain-generality of monitoring hypothesis. The results of this study point to some interesting facets of graduate learners' monitoring proficiencies in the context of a complex writing task. While the performance assessments were, in large part, a log-specific phenomena, with performance on one log mostly unrelated to performance on another log, prediction confidence scores revealed three distinct abilities or patterns across three successive pairs of logs. The three confidence patterns observed were between logs 1 and 2, logs 3 and 4, and logs 5 and 6. In addition, the three pairs of confidence scores for logs 1 and 2, logs 3 and 4, and logs 5 and 6, were related to one another, over and above the performance assessments. This provides some support for the presence of a general confidence ability, which manifested itself in the form of three

unique patterns. While it is tempting to suggest that the results lend some support to the domain-general hypothesis of monitoring, the discovery of three patterns of confidence, cannot be taken to imply that there is a general confidence, and hence, a general monitoring ability across the logs. Further research is necessary in determining the nature of and the relationship that exists between distinct confidence abilities in the context of complex writing tasks.

The partial support for a general confidence ability in adult learners engaging in varying levels of difficulty with a complex writing task, mirrors, to a small extent, some of the results revealed in Glenberg et al. (1987), Schraw et al (1995), Schraw & Nietfeld (1998) and Weaver (1990). These researchers suggest that unrelated performance measures and uniformly related confidence measures point towards domain-generality of monitoring. Schraw and his colleagues' examined the notion of domain-general confidence and monitoring abilities in the context of tasks composed of multiple-choice questions. My study provides encouraging evidence of such general confidence abilities existing in a different context and points to the need to further explore how confidence and monitoring abilities develop in the context of a set of complex writing tasks which vary in their relative difficulty.

The results also suggested that learners' prediction confidence scores on any one log was not necessarily bound to their performance assessments on that log, which was consistent with the results seen in Glenberg et al. (1987), Schraw et al. (1995) and Schraw & Nietfeld (1998). Further analyses also revealed that prediction confidence on any one log was related neither to performance assessment on the previous log nor to performance assessment on logs of a similar structure. In other words, not only was there

some evidence of a general confidence ability, which acted over and above performance assessments, but also, prediction confidence scores and performance assessments were, for the most, part unrelated across the six logs, in any meaningful way.

My results also suggest that prediction confidence develops as a unique pattern across successive logs when feedback was available for the earlier log; this contention needs to be further explored in future research within a framework of the nature and type of feedback that promotes confidence and improved monitoring skills (see Butler & Winne, 1995 for a review of feedback in the context of self-regulation).

Factoring the notion of performance predictions. An important aspect of my thesis study is the investigation of the notion of performance predictions, and its relation to the instructor's performance assessments and students' prediction confidence scores. Neither of Schraw and his colleagues' recent investigations (Schraw et al., 1995; Schraw & Nietfeld, 1998) dealt with the notion of students' performance predictions and how these predictions might be related to their actual performance and confidence. Schraw and his colleagues investigated monitoring in the context of multiple-choice questions, and hence, students did not predict *how* correct their responses were, rather they stated their confidence that their answers were correct. In fact, in Schraw and his colleagues' studies, students' were implicitly predicted perfect performance. Further, in these studies, monitoring proficiencies were calculated using performance and confidence scores. In my study, the notion of performance predictions adds a new dimension to measuring monitoring proficiencies. Both the measures of monitoring proficiencies, namely, discrimination and bias, take into account performance predictions, performance assessments, and prediction confidence. I have shown that monitoring of performance in

the context of complex writing activities needs to take into account students' performance predictions. I propose that when performance is not gauged simply in terms of "right" and "wrong" answers, but is instead, mostly graded on a scale, then students' monitoring abilities need to account for any over or under-estimation of performance before considering the effect of their prediction confidence.

Performance prediction, performance assessment and prediction confidence: A complex relationship. Findings in my study indicate that as the logs progressed, students' consistently predicted higher grades and had greater confidence in their predictions. However, the relationship between prediction confidence and performance predictions was highly log-specific, with no discernable patterns across logs.

One reason why both the instructor's performance assessments and students' performance predictions did not seem to have an effect on the learners' prediction confidence could be the fact that the content covered for the course may have varied largely in its levels of difficulty. Recall that this difficulty factor also seems to have played a large role in the log-specificity of the instructor's assessment of performance. Despite the fact that the students were performing the same task (learning log writing), learners' prediction confidence in their grades may have been guided by a factor such as content difficulty or, even by a general monitoring ability, and not by their levels of performance prediction or performance assessment. Put simply, an increase in performance assessment or performance prediction did not necessarily prompt an increase in prediction confidence. Interestingly, students' performance predictions seemed to develop into a general ability or pattern after the first two logs. This result suggests the performance prediction behaved very differently from the instructor's

performance assessments. While performance assessments were log-specific, findings suggest that performance predictions developed, after some engagement with the first two learning logs, as a pattern, which acted across the remaining four logs. More research is needed to reveal how performance predictions and performance assessments vary in the context of complex writing tasks.

My findings also show that the within-log relationship between performance predictions and performance assessments, though complex, acted over and above the learners' prediction confidence scores. The relationship between predicted and instructor-assessed performance is for the most part, log-specific, variable in valence and insignificant. The increased association between performance predictions and performance assessments when variation due to confidence in predictions is removed, supports the notion that measures of students' performance need not necessarily, be affected or, in turn, affect, their measures of confidence as they engage in complex writing tasks.

Prediction accuracy pattern. Prediction accuracy, which was measured as the absolute difference between performance prediction and performance assessment on any particular log, showed signs of developing into a general ability or pattern across logs 1, 2, 3 and 5. While the mean accuracy for log 1 was relatively high, between logs 2 and 5, students', on the average, were able to predict their performance to within a grade of accuracy. These findings suggest that students' possessed an ability to accurately predict their grade across the four logs 1, 2, 3, and 5, regardless of the structure of the log and the content covered in the course. Moreover, the increased association between the majority of these four prediction accuracy scores when any variation due to confidence in

predictions was removed, further supports the notion of a general accuracy ability or pattern in graduate learners engaged in a complex writing task. This accuracy pattern mirrors the prediction pattern discussed earlier, because it stands in stark contrast to the log-specific performance ability.

Discrimination in predictions. Results suggest that students showed an increased ability to discrimination, that is, as the logs progressed students were better able to assign an appropriate level of confidence to their performance predictions. Findings from my thesis study reveal the possible existence of two distinct discrimination patterns in graduate learners engaged with complex writing tasks; the first pattern was found across logs 1, 2, and 3, while the second was found across logs 3, 4 and 5. These results oppose those that are found in Schraw et al.'s (1995) study, where discrimination was a domain-specific phenomenon. One possible reason for this discrepancy is that the manner in which discrimination was calculated for Schraw et al.'s study was very different from that used in this thesis study. While a unique discrimination was calculated for each individual multiple-choice test taken by learners in Schraw and his colleagues' work, discrimination for any one log, in the context of my thesis study, was related to discrimination abilities calculated for previous logs. Given the difference in the context which I was investigating, namely, complex writing tasks, I had decided to adapt the calculation procedures for the measure of discrimination to take into account (a) the progressive nature of the learning log task, (b) students' performance predictions, (c) instructor's performance assessments, and (d) students' prediction confidence scores. The existence of two patterns of discrimination in students engaged in a complex writing task, and the absence of a general discrimination ability in students engaged in semantic

memory recall-based, multiple-choice tests for different domains reveal that students' abilities to assign an appropriate level of confidence for their performance predictions might vary from one type of academic task to the next. Thus, while my results support the existence of two log-general patterns of discrimination, the differences observed in these results with those of Schraw et al. (1995), suggest that discrimination ability might be context-specific, and might vary with fluctuations in task difficulty. Due to the exploratory nature of my thesis study, further research is necessary in exploring the validity of comparing cross-contextual discrimination measures, some of which account for learners' predictions in performance and others, which do not.

Discrimination also revealed a complex relation with both prediction confidence scores and performance assessments in terms of magnitude and valence, however, these relations were mostly insignificant. Significantly correlated discrimination scores showed improved association, over and above the instructor's assessments of performance, lending weight to the proposition that a general discrimination, and hence a general monitoring ability was acting across the six logs. However, the lack of association between confidence and discrimination, despite findings that supported the existence of unique confidence and discrimination patterns, seem to diminish the support for the domain-general hypothesis. If a general monitoring skill was apparent across the logs, students' abilities to appropriately assign a confidence level to predictions (discrimination) should be associated with their prediction confidence. Similar insignificant associations between discrimination and confidence are reported in Schraw et al.'s (1995) study, however, they are unable to explain the reason behind this occurrence. In the context of this thesis study, the difference in patterns of

intercorrelations between confidence and discrimination offer a possible explanation for the lack of relationship between the two measures; while confidence revealed three distinct patterns between logs 1 and 2, logs 3 and 4, and logs 5 and 6, discrimination abilities were revealed across logs 1 to 3 and across logs 3 to 5. This finding suggests that discrimination and confidence develop independent of one another, at different stages, across the six logs.

Bias in predictions. Results of analyses on bias scores revealed that students, were, for the most part, under-confident of their performance. The results also suggest that a general bias ability may exist across the logs and may manifest itself as two unique patterns, one across logs 1, 3 and 4, and the other across logs 2, 5 and 6. This notion of a general bias ability was supported by the increased association between significantly correlated bias scores when variation due to the performance assessments was removed. The findings mirror, to a small extent, those observed in Schraw et al. (1995) and Schraw & Nietfeld (1998), where a general bias ability was found to be acting across different domains of multiple-choice tests. However, in contrast to Schraw and his colleagues' findings, bias and confidence did not show strong intercorrelations. The weak associations between confidence and bias measures suggest that confidence and bias patterns acted across a different set of logs. While confidence patterns were seen between logs 1 and 2, logs 3 and 4, and logs 5 and 6, bias patterns were evident across logs 1,3 and 4, and to a smaller extent, between logs 2, 5 and 6. Bias and confidence abilities, therefore, develop independent of one another, at different stages, across the six logs.

Investigating monitoring proficiencies in the context of complex writing tasks. The above discussion provides a picture of how monitoring proficiencies developed in

seventeen graduate learners across the six logs. The exploratory procedures employed in the analyses provide preliminary evidence that learners' monitoring proficiencies showed a propensity towards being a general phenomenon across the logs, as opposed to being specific to each log. While the measures of prediction confidence, performance prediction, prediction accuracy, discrimination and bias, each revealed one or more patterns that spanned across a set of logs, successful performance, as gauged by the instructor, was the only variable that retained an essence of being specific to each learning log. Support for the domain-general hypothesis would have been strongest if performance measures were uncorrelated, and confidence, discrimination and bias were uniformly correlated across the logs. Such a pattern of correlations would mean that student prediction confidence and monitoring abilities were related across the logs despite performance being a unique phenomenon to each log. The absence of such uniform correlations suggests that the generality of monitoring proficiencies needs to be further explored in the context of complex writing tasks.

The naturalistic context of this study makes it difficult to study whether any domain-specific features of monitoring were present across the logs because (a) performance feedback was not consistently timed with some students receiving delayed feedback on logs 2 and 4, and (b) the difficulty level of the content varied across the six logs. Further, the difference in the difficulty of the content being covered proves to be a stumbling block in investigating the role that co-varying performance plays in the variability seen in monitoring proficiencies like discrimination and bias. Future research should control for these above-mentioned factors in an effort to develop a clearer picture of monitoring abilities in the context of complex writing tasks.

Contribution to theory and practice. One of the purposes of this study was to explore the development of monitoring proficiencies in graduate learners in the context of a complex writing task. In achieving this purpose, I set out to determine whether the results provided support for a domain-general or domain-specific monitoring skill. Traditional modular theories have viewed cognitive skills as domain-specific (Fodor, 1983, Gardner, 1983, Glaser & Chi, 1988; Hirschfeld & Gelman, 1994), while information-processing theorists have proposed and found support for the existence of more domain-general skills (Borkowski & Muthukrishna, 1992; Paris & Byrnes, 1989; Brown, 1987; Pressley et al., 1987; Schneider & Pressley, 1989). Studies by Schraw and his colleagues (Schraw et al., 1995; Schraw & Nietfeld, 1998) have supported the existence of both domain-specific and domain-general types of monitoring skills; these studies have been conducted mostly in the context of tests involving multiple-choice questions that required recall of information from semantic memory or those that tested fluid and crystallized ability, in college learners. My thesis study explores monitoring proficiencies in the context of a more complex writing task with adult, graduate learners. While monitoring ability has been shown to be a complex phenomenon in this study, the results from analyses point towards the existence of a general monitoring ability that spans across the writing task, namely, the six learning logs. It is also interesting to note that confidence, discrimination and bias each reveal patterns across a different set of logs, suggesting that monitoring ability might manifest itself in different types of proficiencies depending on contextual factors such as difficulty of content being covered, nature and timing of feedback, or structure of the writing task. It must be noted, however, that the small sample size used in this study poses a stumbling block in making generalizations

about the nature of learners' monitoring proficiencies in complex writing tasks; future research must therefore investigate this phenomenon with a larger sample.

Metacognition and monitoring are generally understood to be domain-general phenomena (Brown, 1987; Pintrich et al., 2000; Schraw et al., 1995; Schraw & Impara, 2000; Schraw & Nietfeld, 1998; Tobias & Everson, 2000), however, it should be reiterated that domain-general monitoring skills, while independent of domain-specific monitoring skills and knowledge, generally complement them. Future research in the investigation of monitoring of learning and performance in complex writing tasks should, therefore, investigate which types of domain-specific monitoring abilities are, in fact, present and are utilised by learners in such contexts. Moreover, it would be interesting to test the existence of two types of general monitoring skills, one related to fluid and one related to crystallized ability (c.f., Schraw & Nietfeld, 1998) in learners engaged in complex writing tasks as opposed to tests of multiple-choice questions. Future research should also investigate the relationship between discrimination and bias, and whether these two proficiencies co-exist across similar types of tasks, or work independently of one another. An important reason for investigating the existence of domain-specific monitoring abilities is that effective self-regulation depends on proficient monitoring (Pintrich, 2000; Winne & Hadwin, 1998; Zimmerman, 2000); if evidence exists that monitoring proficiencies are linked with specific domains or contexts of learning, then educators need to cater their instruction to improving monitoring proficiencies within these domains in addition to encouraging the development of general monitoring abilities.

The results of this study also provide a strong platform for the investigation of the developmental aspects of general monitoring knowledge and skills, an area of research

that is currently being investigated by Schraw and his colleagues (see Schraw & Impara, 2000). Further research is needed to verify the possibility that monitoring in contextualized domains is progressively generalized until it becomes a metacognitive skill that spans cognitive domains. This developmental sequence has been well researched over the past decade and a half as the *good information-processing model*; after being initially proposed by Pressley et al. (1987), it has been elaborated by Schneider and Pressley (1989), as well as Borkowski and Muthukrishna (1992), and most recently applied to measurement issues in metacognition by Borkowski et al. (2000). In short, the good information processing model contends that learners with higher-order cognitive skills (a) initially attain strategy knowledge within a particular domain of learning, (b) use this strategy knowledge to develop conditional metacognitive knowledge of when and how to use specific strategies, and (c) build a repertoire of general metacognitive and metastrategy knowledge for application across domains. Further research with students engaged in complex writing activities should explore whether and how monitoring proficiencies become more domain-general or domain-specific in nature. For example, if the propositions from the good information processing model are held to be true, then learners who are engaged in writing learning logs across different graduate classroom settings might develop a general monitoring proficiency after sufficient exposure and engagement with that specific type of writing task, across different learning contexts, with each context varying in its level of difficulty.

The present study has provided some evidence of general monitoring abilities. However, more research is needed to better describe the nature of these proficiencies and their relation to one another. The small sample size for this study makes it hard to find

moderately high correlations (e.g., values of .20 to .40) that are significant. Perhaps a larger sample size might help in better pinpointing the behavior of the monitoring proficiencies characterized in this study. Moreover, the inherent difficulties in using different scales of measurement for performance, confidence, discrimination and bias also pose serious threats to finding significant associations between the mentioned measures. Since my thesis study's investigations were conducted in a naturalistic, graduate classroom environment, the difference in the scales of measurement is a reflection of my adapting the instruments used in the study to the instructional protocol developed by the instructor of the class. Further investigations of monitoring skills in learners tackling complex writing tasks across a variety of academic contexts needs to take into account this hurdle in the development of its scales of measurement.

Are learners' abilities to meet the assessment criteria for a complex writing task reflected in the instructor's assessment of their performance?

Learners' performances on their six learning logs were graded by the instructor, and they were also independently scored for the extent to which each of the question-and-answer sets had created for the learning logs matched the assessment criteria outlined by the instructor. The results revealed some interesting relationships and differences between the instructor's performance assessments and the independent overall performance scores across the six logs. Both these measures of performance showed statistically significant, increasing linear trends, indicating that overall, students' performance on the learning logs improved as instruction progressed, whether performance was assessed by the instructor or was independently scored. When performance was assessed by the instructor, students' performances showed significant

differences between successive logs after the completion of log 3. In addition, the instructor's performance assessments on log 4 were significantly greater than each of those on logs 1 and 2, while performance assessments on log 5 was significantly greater than each of those on logs 1, 2 and 3, indicating that students' performances rose sharply after log 3. In comparison, when overall performance was independently calculated, students' performance between logs 2 and 6 showed chance fluctuations. The only significant differences in independent overall performance scores that analyses revealed were between logs 1 and 4, logs 1 and 5, and logs 1 and 6. Therefore, while the independent overall performance scores on logs indicated that students were meeting the assessment criteria, for the most part, to the same extent between logs 2 and 6, the instructor's assessment suggested that students' performed significantly better once they had completed their third log. One reason for the difference in observed trends could lie with the fact that the instructor was assessing the students on *more* than just the explicit assessment criteria; the element of *growth* that instructor intended to assess across the learning logs was not clearly described in the course outline, and was perhaps, implicitly constructed between the instructor and individual student as the instruction progressed (see McCaslin & Good, 1996 for a related theoretical discussion on co-regulated learning). This analysis would best be triangulated with qualitative data in the form of interviews with the instructor and student; such a discussion is beyond the scope of this thesis, but remains an avenue to be explored in the future.

Another reason for the observed difference in trends between the performance measures can be found in the results of the correlational analyses between the instructor's performance assessments and the independent multistructural, relational and extended

abstract performance scores. Recall that each of the logs 1 to 5 consisted of two question-and-answer sets, one was multistructural, while the other was either relational or extended abstract (log 6 consisted of all sets from all three levels). The final grade for any given log intended to reflect the student's overall performance on log, with an equal weighting applied to each question-and-answer set. For the independently calculated scores, the overall performance on any one log for any one student was the average of the scores for the individual levels required for that log. Strong within-log correlations were mostly observed between the instructor's performance assessments and the independent overall performance. Interestingly, though, performance assessments were not correlated with the independent multistructural performance scores, but were strongly associated with the independent relational and extended abstract performance scores. This indicates that the instructor's performance assessments seemed to be based mostly on the criteria for relational and extended abstract levels and very little on those for the multistructural level. This discrepancy can account for the observed differences in trends between the two measures of performance, namely, instructor's assessments of performances and the independently calculated scores of overall performance on the logs. This finding can be best triangulated with qualitative interview data with the instructor as well as with the learners. Future research should explore how an understanding of the assessment criteria was developed among the learners.

The preceding discussion of whether the instructor's performance assessment takes into account learners' ability to meet the assessment criteria is related to the development of students' understanding of academic tasks. In as much as ability to meet assessment criteria is dependent on learner's understanding of assessment criteria, the

results suggest that students' understandings of the criteria for the learning log task played a role in their performance. In my thesis study, I measured students' understanding of the assessment criteria for creating questions and answers, mostly at the relational and extended abstract level. I explicitly asked students to respond to questions such as, "Why do you think your question/answer is relational/extended abstract?"; these questions were embedded in the self-assessment TAPE tool for every learning log. Students' showed, for the most part, a good understanding for the criteria for creating answers across logs 1 to 5, regardless of whether the answers were relational or extended abstract; this measure of understanding showed only chance fluctuations between logs 1 and 5. Similarly, students' understanding of the criteria used for creating relational and extended abstract questions was also good, and showed chance fluctuations between logs 3 and 5. However, these measures of understanding of assessment criteria are insufficient in explaining how students' performance assessments showed significant improvements, as revealed by the analyses, across logs 1 to 5.

As put forward by Hadwin (2000), as well as Winne and Hadwin (1998), I subscribe to the notion task understanding involves not just learners' perceptions of the assessment criteria for the task, but also their "knowledge of self-as-learner", which includes their prior content and task-related knowledge, motivational and affective beliefs as well as emotional states. Students' task understandings develop dynamically as instruction progresses, and hence, a static measure of students' understandings of assessment criteria, which does not take into account the "knowledge of self-as-learner", does not accurately reflect how students' understandings of the learning log task developed. The encouraging result, however, is that students' understanding of the

assessment criteria for both questions and answers was maintained at a relatively high level for most of the logs; this provides evidence for the use of an instructional feature that explicitly asks students for their perceptions of the assessment criteria for the academic task (Butler, 1998; Perry, 1998, Perry et al., 2002; Perry & VandeKamp, 2000; Venkatesh & Hadwin, 2002). It is important, however, to reiterate that a qualitative analysis of interview data with the students will help better paint a picture of students' task understanding; the interview data will help trace the various aspects of the "knowledge of self-as-learner" that develop alongside the perceptions of the assessment criteria for the task. Such a discussion, however, is beyond the scope of this thesis, but should be considered as ideas that can be implemented for future research.

Are learners' understandings and perceptions of the assessment criteria for a complex writing task related to their monitoring proficiencies over the period of instruction?

The results of analyses involving the relational and extended abstract justification scores and the monitoring proficiencies do not suggest more than the existence of a log-specific relationship between monitoring ability and understandings of assessment criteria for the learning log task. At the present time, to the best of my knowledge, research has not focused on how monitoring abilities are affected, or in turn, affect learners' changing understandings of an academic task. Further research is necessary to explore the relationship, if any, between task understanding and monitoring ability in the context of complex writing activities; perhaps, qualitative methodologies can shed light on their relationship.

Chapter 5: Conclusion

The findings from this study lend some support towards the existence of a general monitoring ability in the seventeen graduate learners across the six learning logs. Learners' performances on their logs were a log-specific phenomenon, most likely due to the differences in the difficulty level of the content being covered from one log to the next. Prediction confidence, predicted performance, discrimination, bias and prediction accuracy each displayed propensities towards a general ability and manifested itself as one or more unique patterns spanning across a set of logs. The results also pointed to a discrepancy between observed trends in the instructor's performance assessments and the independent overall performance scores, indicating that the instructor might have assessed learners' performances on the logs based on more than just the explicit assessment criteria outlined. Findings also revealed that learners' understandings of the assessment criteria were not related to their monitoring proficiencies.

Future research needs to explore the generality of the monitoring abilities investigated in the context of complex writing activities. Research should also address the extent to which learners use context-specific monitoring skills so as to better equip instructors in improving both domain-general and domain-specific monitoring skills. While the small sample size for this study hinders the generalizability of the findings, this study provides a platform for future studies investigating monitoring proficiencies and task understandings in naturalistic environments. Formal, systematic and longitudinal investigations of key self-regulatory processes is essential in (a) finding empirical support

for developmental issues in self-regulation and (b) implementing instructional features that promote self-regulation in learning environments.

References

Bereiter, C., & Scardamalia, M. (1993). *Surpassing ourselves: An inquiry into the nature and implications of expertise.* Chicago, IL: Open Court.

Biggs, J. B. (1991). Student learning in the context of school. In J. B. Biggs (Ed.), *Teaching for Learning: The View from Cognitive Psychology* (pp. 7-29). Hawthorn, VA, Australia: The Australian Council for Educational Research Ltd.

Biggs, J. B. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347-364.

Borkowski, J., & Muthukrishna, N. (1992). Moving metacognition into the classroom: "Working models" and effective strategy instruction. In M. Pressley, K. Harris, & J. Guthrie (Eds.), *Promoting academic competence and literacy in school* (pp. 477-501). New York: Academic Press.

Borkowski, J. G., Chan, L. K. S., & Muthukrishna, N. (2000). A process-oriented model of metacognition: Links between motivation and executive functioning. In G. Schraw & J. C. Impara, *Issues in the Measurement of Metacognition* (pp. 1-42). Lincoln, NE: Buros Institute of Mental Measurements

Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. B. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 453-481). Hillsdale, NJ: Lawrence Erlbaum.

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. Weinert & R. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-116). Hillsdale, NJ: Erlbaum.

Brown, A., Bransford, J., Ferrara, R., & Campione, J. (1983). Learning, remembering and understanding. In P. H. Mussen (Series Ed.) & J. Flavell, & E. Markam (Vol. Eds.), *Handbook of child psychology: Vol. 3. Cognitive development* (4th ed., pp. 77-166). New York, NY: Wiley.

Butler, D. L. (1998). The strategic content learning approach to promoting self-regulated learning: A report of three studies. *Journal of Educational Psychology, 90*, 682-697.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

Englert, C. S., Raphael, T. E., Anderson, L. M., Anthony, H. M., & Stevens, D. D. (1991). Making strategies and self-talk visible: Writing instruction in regular and special education classrooms. *American Educational Research Journal, 28*, 337-372.

Ertmer, P. A., Newby, T. J., & MacDougall, M. (1996). Students' responses and approaches case-based instruction: The role of reflective self-regulation. *American Educational Research Journal, 33*, 719-752.

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist, 34*, 906-911.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Garcia, T., & Pintrich, P. R. (1994). Regulating motivation and cognition in the classroom: The role of self-schemas and self-regulatory strategies. In D. H.

Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 127-153). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Ghatala, E. S., Levin, J. R., Foorman, B. R., & Pressley. M. (1989). Improving children's regulation of their reading PREP time. *Contemporary Educational Psychology, 14,* 49-66.

Glaser, R. & Chi, M. T. (1988). *The nature of expertise.* Hillsdale, NJ: Erlbaum.

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15,* 84-93.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of performance. *Journal of Experimental Psychology: General, 116,* 119-136.

Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.) *Metacognition in educational theory and practice* (pp. 1-23). Hillsdael, NJ: Lawrence Erlbaum Associates.

Hadwin, A. F. (2000). *Building a case for self-regulating as a socially constructed phenomenon.* Unpublished doctoral dissertation. Simon Fraser University, Burnaby, BC, Canada.

Hirschfeld, L. A., & Gelman, S. A. (1994). Toward a topography of mind: An introduction to domain-specificity. In L. A. Hirschfeld, & S. A. Gelman (Eds.), *Mapping the mind: Domain-specificity in cognition and culture* (pp. 3-35). Cambridge, UK: Cambridge University Press.

Keppel, G. (1982). *Design and analysis: A researcher's handbook* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall, Inc.

Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 77*, 217-273.

Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Making, 50*, 212-247.

Ley, K., & Young, D. B. (2001). Instructional principles for self-regulation. *Educational Technology Research & Development, 49*(2), 93-103.

Lin, X. (2001). Designing metacognitive activities. *Educational Technology Research & Development, 49*(2), 23-40.

Lindeberg, M. A., Fox, P. W., & Puncochar, J. (1994). Highly confident but wrong.: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology, 86*, 114-121.

Maki, R. H., Foley, M. J., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 609-616

Maki, R. H., & Serra, (1992). The basis of test predictions for text materials. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 116-126.

McCaslin, M., & Good, T. (1996). *Listening in classrooms*. New York, NY: Harper Collins.

McCombs, B. L., & Marzano, R. J. (1990). Putting the self in self-regulated learning: The self as an agent in integrating will and skill. *Educational Psychologist, 25*, 51-69.

Morris, C. C. (1990). Retrieval processes underlying confidence in comprehension

judgments. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16,* 223-232

Nelson, T., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-141). New York, NY: Academic Press.

Paris, S. G., & Byrnes, J. P. (1989). The constructivist approach to self-regulation and learning in the classroom. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theory, research and practice* (pp. 169-200). New York, NY: Springer-Verlag.

Paris, S. G., & Newman, R. S. (1990). Developmental aspects of self-regulated learning. *Educational Psychologist, 25,* 87-102.

Paris, S. G., & Winograd, P. (1990). How metacognition can promote academic learning and instruction. In B. F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 15-51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perry, N. E. (1998). Young children's self-regulated learning and contexts that support it. *Journal of Educational Psychology, 90,* 715-729.

Perry, N. E., & VandeKamp, K. J. O. (2000). Creating classroom contexts that support young children's development of self-regulated learning. *International Journal of Educational Research, 33,* 821-843.

Perry, N. E., VandeKamp, K. J. O., Mercer, L. K., & Nordby, C. J. (2002). Investigating teacher-student interactions that foster self-regulated learning. *Educational Psychologist, 37,* 5-15.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451-502). San Diego, CA: Academic Press.

Pintrich, P. R., & De Groot, E. V. (1991). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33-40.

Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw & J. C. Impara, *Issues in the Measurement of Metacognition* (pp. 43-98). Lincoln, NE: Buros Institute of Mental Measurements.

Pressley, M., Borkowski, J. G., & Schneider, W. (1987). Cognitive strategies: Good strategy users coordinate metacognition and knowledge. In R. Vasta & G. Whitehurst (Eds.), *Annals of child development* (Vol. 5, pp. 89-129). Greenwich, CT: JAI Press.

Pressley, M., Borkowski, J. G. , & Schneider, W. (1990). Good information processing: What it is and how education can promote it. *International Journal of Educational Research, 2*, 857-867.

Pressley, M., & Ghatala, E. S. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist, 25*, 19-33.

Pressley, M., Ghatala, E. S., Woloshyn, V., & Pirie, J. (1990). Sometimes adults miss the main ideas and do not realize it: Confidence in responses to short-answer and multiple-choice comprehension questions. *Reading Research Quarterly, 25*, 232-249.

Pressley, M., Snyder, B. L., Levin, J., Murray, H. G., & Ghatala, E. S. (1987). Perceived

readiness for examination performance (PREP) produced by initial reading of text and containing adjunct questions. *Reading Research Quarterly, 22*, 219-236.

Randi, J., & Corno, L. (2000). Teacher innovations in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 651-685). San Diego, CA: Academic Press.

Rheinberg, F., Vollmeyer, R. & Rollett, W. (2000). Motivation and action in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 503-529). San Diego, CA: Academic Press.

Salovey, P. (2000). Results that get results. In R. J. Sternberg (Ed.), *Guide to publishing in psychology journals*. New York, NY: Cambridge University Press.

Schneider, W., & Pressley, M. (1989). *Memory development between 2 and 20*. New York: Academic Press.

Schraw, G. (1994). The effect of metacognitive knowledge on local and global monitoring. *Contemporary Educational Psychology, 19*, 143-154.

Schraw, G. (1997). The effect of generalized metacognitive knowledge on test performance and confidence judgments. *The Journal of Experimental Education, 65*, 135-146.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*, 113-125.

Schraw, G., Dunkle, M. E., Bendixen, L. D., & Roedel, T. D. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology, 87*, 433-444.

Schraw, G., & Impara, J. C. *Issues in the measurement of metacognition*. Lincoln, NE: Buros Institute of Mental Measurement.

Schraw, G. & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*, 351-371.

Schraw, G. , & Nietfeld, J. (1998). A further test of the general monitoring skill. *Journal of Educational Psychology, 90*, 236-248.

Schraw, G., Potenza, M. T. & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. *Contemporary Educational Psychology, 18*, 455-463.

Schraw, G., & Roedel, T. D. (1994). Test difficulty and judgment bias. *Memory & Cognition, 22*, 63-69.

Schunk, D. H., & Zimmerman, B. J. (1994). *Self-regulation of learning and performance: Issues and educational applications*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schunk, D. H., & Zimmerman, B. J. (1997). Social origins of self-regulatory competence. *Educational Psychologist, 32*, 195-208.

Slife, B. D., & Weaver, C. A. III (1992). Depression, cognitive skill, and metacognitive skill in problem-solving. *Cognition and Emotion, 6*, 1-22.

Tobias, S. & Everson, H. (2000). Assessing metacognitive knowledge monitoring. In G. Schraw & J. C. Impara, *Issues in the Measurement of Metacognition* (pp. 147-222). Lincoln, NE: Buros Institute of Mental Measurements

Venkatesh, V., & Hadwin, A. F. (2002). *Instructional principles to promote self-regulation: A review of research*. Paper presented at the annual convention of the American Psychological Association. Chicago, IL.

Voss, J. F., Green, T. R., Post, T. A., & Penner, B. C. (1983). Problem solving skill in the social sciences. In G. H. Bower (ed.), *The Psychology of Learning and*

Motivation: Advances in Research Theory (vol. 17, pp. 165-213). New York, NY: Academic Press.

Voss, J. F., Wolfe, C. R., Lawrence, J. A., & Engle, R. A. (1991). From representation to decision: An analysis of problem-solving in international relations. In R. J. Sternberg and P. A. Frensch (Eds.), *Complex Problem-Solving: Principles and Mechanisms* (pp. 119-158). Hillsdale, NJ: Lawrence Erlbaum.

Walczyk, J. J., & Hall, V. C. (1989a). Is the failure to monitor comprehension an instance of cognitive impulsivity? *Journal of Educational Psychology, 81*, 294-298.

Walczyk, J. J., & Hall, V. C. (1989b). Effects of examples and embedded questions on the accuracy of comprehension self-assessments. *Journal of Educational Psychology, 81*, 435-437.

Weaver, C. A. III (1990). Constraining factors in calibration of performance. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 214-222.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Malwah, NJ: Erlbaum

Yates, J. F. (1990). *Judgment and decision-making*. Englewood Cliffs, NJ: Prentice Hall.

Zimmerman, B. J. (1986). Development of self-regulated learning: What are the key subprocesses? *Contemporary Educational Psychology, 16*, 307-313.

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

Zimmerman, B. J. (1994). Dimensions of academic self-regulation: A conceptual framework for education. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-*

regulation of learning and performance: Issues and educational applications (pp. 3-21). Hillsdale, NJ: Erlbaum.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA: Academic Press.

Zimmerman, B. J., & Schunk, D. H. (1989). *Self-regulated learning and academic achievement: Theory, research and practice*. New York, NY: Springer-Verlag.

Appendix A – Description of Multistructural, Relational and Extended Abstract Levels and Criteria for Assessment of Learning Logs

LEARNING LOGS (60% Due weekly)
6 logs (36%) + weekly self assessment (6%) + growth over semester (8%)

The learning logs consist of 2 parts. They are designed to help you: (a) prepare for class, (b) understand the material, (c) remember the material, and (d) evaluate your understanding

Part 1: Generate and answer questions at each of the following levels. Answers should be submitted to Firstclass® using the assignment form provided in your conference. Questions and answers should be printed and brought to every class because they will be central to class discussions. You will hand in your learning logs for grades 6 times over the semester. Each log will be worth 6 percent of your overall grade, for a total of 36%.

Part 2: For each question, you will include a self-assessment using the form available in your first class conference. This assessment will require you to identify: (a) your interpretation of what you think you are supposed to do, (b) your confidence in your answer, (c) the criteria you think will be used to grade your question and answer, and (d) a letter grade. Collectively these self-assessments will constitute 6 percent of your overall grade

Growth: I do not expect you to master question and answer logs immediately. I will also be looking for growth over the semester. That is, through this course I expect you to learn to ask better questions and provide more sophisticated answers. Growth may be demonstrated in the form of mastery across logs, or revising logs, or writing a reflective piece about your own growth at the end of the semester). This will constitute 8 percent of overall grade.

Learning Log 1: questions I provide	Answer a multistructural, relational, and extended abstract
Learning Log 2:	Generate and answer 1 multistructural + 1 relational question
Learning Log 3:	Generate and answer 1 multistructural + 1 extended abstract
Learning Log 4: question	Generate and answer 1 multistructural + 1 extended abstract
Learning Log 5:	Generate and answer 1 multistructural + 1 relational question
Learning Log 6: abstract question	Generate and answer a multistructural, relational, and extended

Level of Question Depth	Explanation	Examples
Multistructural	Questions that test your knowledge of relationships between ideas in the text. These questions require you to select	• Provide an example of important concepts.

	important ideas that are interrelated. Often the interrelationship between these ideas is made explicit in the text.	• explain cause-and-effect relations. • compare and contrast alternative concepts and claims
Relational	Questions that require you integrate ideas across chapters and/or with prior knowledge. These questions require you do more with the material than is currently stated explicitly in the text, or to make connections that are implicit in the text.	• analyze assumptions underlying a theory or a debate. • summarize models and systems that explain behavior. • evaluate the quality or validity of a hypothesis and give evidence to support your view
Extended Abstract	Questions that require you to extend and apply ideas from the text. These questions require you to go beyond ideas presented in the text while simultaneously demonstrating your understanding of those ideas and the content presented in the text.	• apply principles to real world situations and explain how they work. • invent new hypotheses to explain a situation and justify how those hypotheses might be valid. • generate and evaluate alternative solutions to scientific or social debates.

Criteria for grading learning logs

Logs will be assigned a letter grade as follows.

A	• Excellent thought-provoking questions that match the level of depth • Thoughtful answers that match the level of depth • Correspondence between questions & answers • Demonstrates exceptional understanding of course concepts • Challenges me or gets me thinking • Above and beyond expectations –challenges course content

A-	• Variations of some of the above
B+	• Good questions that match the level of depth • Thoughtful answers that match the level of depth • Correspondence between questions and answers • Demonstrates some understanding of course concepts
B	• Variations of some of the above
C	• Weak questions that do not match level of depth • Weak answers that do not match level of depth • Mismatch between question and answer • Weak or inaccurate understanding of course concepts

Appendix B - Task Analyzer and Performance Evaluator

Note: The Task Analyzer and Performance Evaluator (TAPE) will be administered for each relational and extended abstract question-and-answer set students write in their learning logs. At the end of the course, students will have 6 TAPEs. TAPEs will be typed out in a FirstClass® message and students will submit their TAPEs online. TAPEs will be completed after the students have written the relational or extended abstract question-and-answer set for a particular week.

1. Is your question relational/extended abstract (as the case may be)? Why do you think so?

1. Is your answer relational/extended abstract (as the case may be)? Why do you think so?

1. What grade do you think you will be awarded by X for your relational/extended abstract question and answer set? (Choose one from A+, A, A-, B+, B, B- or C)

1. How confident are you that X will award you the grade you have selected in question 3? Choose from one of the following options:

    (a) Very Confident

    (a) Somewhat Confident

(a) Not so Confident

(a) Not Confident At All

1. In comparison to your classmates, how well do you think you have performed on this relational/extended abstract question-and-answer set?

(a) Much Better

(a) Somewhat Better

(a) Somewhat Worse

(a) Much Worse

1. How useful did you find the worked example/ question stem in producing this question and answer set? Choose from one of the following options:

(a) Very Useful

(a) Somewhat Useful

(a) Not so Useful

(a) Not Useful At All

Note: Question 6 will be asked only in weeks 2, 3, 5 and 6. Questions will be rephrased for subsequent TAPEs, depending on how students respond to them.

Appendix C: Consent Form

CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in a program of research being conducted by Vivek Venkatesh, X and Y of the Z University.

A. PURPOSE

I have been informed that the purpose of the research is as follows:

1) To explore the changing interpretations, assessments, and self-assessments in graduate students' understanding of an academic task (viz., learning logs) over the period of a course.

2) To test and evaluate a dynamic assessment tool used in conjunction with instructional approaches aimed at helping graduate students (a) better understand the academic task (learning logs), (b) better evaluate their own performance, and ultimately (c) improve their academic performance in authentic learning environments.

3) To evaluate the effectiveness of two instructional scaffolds to improve learning log performance.

B. PROCEDURES

- Participation in this research does not involve any additional work beyond course requirements (with the exception of an optional interview described below). All

students, regardless of their participation in the research, will have to meet all the requirements, as described in the course outline, to successfully complete the course.

- You will be assigned to one of two tutorial group for the duration of the course. Your group will be managed by either X or Y. Over the length of the course, both of the groups will be receiving similar instructional scaffolds to help complete the required assignments.

- For students who also agree to participate in the interview portion of the research, interviews will be held three times at your convenience, over the course of 7 weeks. The total time involved will be three hours maximum.

- The research project you are consenting to participate in is concerned with your understanding of and performance in learning log component of the class and your subsequent performance in the examination.

- Only Vivek Venkatesh will be aware of who has consented to participate in the research. X and Y will not be made aware of the identities of any of the participants until after the final submission of grades. This means that if you consent to participate in this project, your materials (i.e., the learning logs, self-assessments, X's assessment of your logs and examination grades and interviews) will be made available to the research team only AFTER the final submission of grades

- Reporting of results will maintain confidentiality. That is, only the researchers will know the identity of the persons participating in the project. Your materials will be used solely for the purpose of the stated research, and no names will be revealed during the course of the writing of the report.

- Names will be removed from the datafile of grades and from assignments themselves. This will ensure that your fellow graduate students do not know your individual grades. One exception to this is that students who are interviewed by Vivek Venkatesh will be discussing their grades as part of the interview.

- If you have any questions or concerns about the research, please direct your enquiries to Vivek Venkatesh before signing this consent form.

- If you wish to withdraw your consent to participating in the research at any time, please contact Vivek Venkatesh (e-mail: vivek.venkatesh@education.concordia.ca, phone: 739-9067) or W (Graduate Program Director). You are free to discontinue at any point in time during the course.

C. CONDITIONS OF PARTICIPATION

Please check off all the boxes that apply to your chosen level of participation in the research:

☐ I consent to release my learning logs, self-evaluations, X's assessments of my learning logs, tests and performance for the purposes of the research.

AND

☐ I consent to participate in interviews three times over the course. These interviews will focus onmy self-assessments and performance on the learning log component of the Learning Theories course

- I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences by contacting either W or Vivek Venkatesh

- I understand that my participation in this study is CONFIDENTIAL (i.e., the researcher will know, but will not disclose my identity)

- I understand that the data from this study may be published.

I HAVE CAREFULLY STUDIED THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND VOLUNTARILY AGREE TO PARTICIPATE IN THIS STUDY.

NAME (please print) _____

SIGNATURE _____

WITNESS SIGNATURE _____

DATE _____

Appendix D – Test of Prior Knowledge

Student ID no. _____

1. When Robert's classmates no longer showed approval of his clowning, his clowning behaviour occurred less frequently. The concept best exemplified by Robert's change in behaviour is

(A) extinction
(A) discrimination
(A) generalization
(A) transfer
(A) learning set

2. A child who is frightened by a dog and develops a fear of dogs is exhibiting which principle of learning

(A) negative transfer
(A) behaviour shaping
(A) stimulas generaliztion
(A) cognitive dissonance
(A) discrimination

3. George Miller's research finding that humans can process seven plus-or-minus two pieces of information applies to which type of memory?

(A) long-term
(A) episodic
(A) sensory register
(A) short-term
(A) implicit

4. Making the amount of time a child can spend playing video games contingent on the amount of time the child spends practicing the piano is an illustration of

(A) primary reinforcement
(B) law of association
(C) aversive conditioning
(D) classical conditioning
(E) operant conditioning

5. Which of the following is a secondary reinforcer?

(A) food
(B) warmth
(C) water

(D) money
(E) sex

6. According to information processing theory, information is progressively processed by

(A) long-term memory,short-term memory,and then sensory memory
(B)sensory memory,short-term memory,and then long-term memory
(C) sensory memory,semantic memory,and then long-term memory
(D) short-term memory,semantic memory,and then long-term memory
(E) short-term memory,long-term memory,and then sensory memory

7. In an approach-avoidance conflict,as the person nears the goal, the levels of attraction and aversion change in which of the following ways?

(A) both increase
(B) both decrease
(C) attraction increases and aversion decreases
(D) attraction decreases and aversion increases
(E) both are extinguished

8. A young child breaks her cookie into a number of pieces and asserts that "now there is more to eat." In Piaget's analysis, the child 's behavior is evidence of

(A) formal logical operations
(B) concrete logical operations
(C) conservation
(D) preoperational thought
(E) sensorimotor analysis

9. Anxiety over performance can positively motivate school achievement in children as long as the degree of anxiety is

(A) very high
(B) high
(C) moderate
(D) low
(E) very low

10. According to Jean Piaget, cognitive development begins with which of the following?

(A) preparations
(B) concrete operations
(C) intuitive thought
(D) sensorimotor activities

(E) formal operations

11. If reinforcement is to be most effective in learning, it should be

(A) provided as sparingly as possible
(B) used on a regularly scheduled basis
(C) used primarily with high achievers
(D) delayed until the end of the learning period
(E) provided soon after the desired behavior occurs

12. The recall memory of an older child is generally better than that of a younger child because the older child

(A) has better perceptual abilities
(B) can organize information better
(C) engages in concrete thinking
(D) does not have to categorize information
(E) recognizes information more easily

13. In situated cognition procedural knowledge (knowing how) and declarative knowledge (knowing that) should be taught in what order?

(A) procedural, and then declarative
(A) declarative, and then procedural
(A) procedural only (i.e., you can't teach declarative knowledge)
(A) declarative only (i.e., you can't teach procedural knowledge)
(A) declarative and procedural simultaneously

14. According to Ausubel, instructional materials should be sequenced from the general, broad topics to the specific pointing out finer distinctions. This is based on his idea of _____.

(A) shaping
(B) progressive differentiation
(C) integrative facilitation
(D) differential reinforcement
(E) prompting and fading

15. Bruner's stage of development that is most similar to Piaget's formal operation stage is the ____.

(A) symbolic stage
(B) abstract stage
(C) iconic stage
(D) operational stage
(E) enactive stage

16. Which of the following educational ideas was developed by Bruner?

(A) learning hierarchies
(B) programmed instruction
(C) spiral curriculum
(D) mastery learning
(E) all of the above

17. Memory schemas, or "schemata," serve as representations of our:

(A) innate knowledge.
(B) specific knowledge.
(C) generic knowledge.
(D) episodic knowledge.
(E) iconic knowledge

18. According to schema theory, having a broader understanding of a situation or story:

(A) always improves memory by providing context.
(B) always hurts memory by confusing new events with old information.
(C) both improves and hurts, for the above reasons.
(D) does little to affect the quality of memory.
(E) does little to affect the quantity of memory

19. Which of the following is not true of constructivist teaching?

(A) The teacher allows the students to have more control over what they do in the learning process.
(B) It is important to respect what students already know before they ever enter the classroom.
(C) Students are encouraged to select projects based on their own interests.
(D) *Effective constructivist teachers spend more time giving information to passive students.*
(E) Students are encouraged to collaborate with others.

20. Bandura calls the learners' confidence in their abilities to learn specific content

(A) self-mastery

(B) confidence ratios
(C) internal reinforcement
(D) self-efficacy
(E) self-concept

21. According to Bandura we can learn in classrooms through observation of others by a process called

(A) indirect integration
(B) vicarious reinforcement
(C) process recall
(D) independent cognition
(E) situated cognition

22. For Vygotsky the zone of proximal development refers to the

(A) student's level of maturation
(B) difference between the student's stage of development and the instruction
(C) area in which the student will learn without reinforcement
(D) area between what the student can learn alone and can learn with help from a teacher or more competent peer
(E) region of the brain that develops when children are able to do formal operations tasks

23. Which of the following is NOT one of the five types of learning trajectories

(A) peripheral
(A) inbound
(A) insider
(A) boundary
(A) proximal

24. According to situated cognition theories, *semiosis* is be defined as:

(A) a process whereby knowledge of the world is mediated through signs, symbols or indexes
(A) a process whereby information is stored by the human processing system
(A) a process whereby knowledge becomes increasingly abstract
(A) a process whereby an individual engages in cognitive apprenticeship

(B) all of the above

25. Vygotsky differs from Piaget in that Vygotsky

(A) has 3 developmental stages rather than 4
(B) thinks teachers can accelerate cognitive growth
(C) sees learning as a passive acquisition of knowledge
(D) accounts for learning solely by maturation
(E) all of the above

26. According to constructivist theory, which method of instruction would NOT promote deep learning

(A) collaborative learning
(B) hypermedia
(C) problem-based learning
(D) role plays
(E) none of the above

27. Which of the following best illustrates metacognition

(A) Memorizing terms and definitions from the textbook
(B) Monitoring one's comprehension while reading
(C) Listening to the radio and studying at the same time
(D) Retrieving information from short-term memory
(E) Retrieving information from long-term memory