Quantitative explorations of graduate learners' monitoring proficiencies and task understandings in the context of ill-structured writing assignments: From learner to work task as unit of analysis

Vivek Venkatesh

A Thesis

in

The Department

of

Education

Presented in Partial Fulfillment of the Requirements
for the Degree of Doctorate of Philosophy in Educational Technology at
Concordia University
Montréal, Québec, Canada

July, 2008

Your file Votre référence
ISBN: 978-0-494-45719-1
Our file Notre référence
ISBN: 978-0-494-45719-1

ABSTRACT

Quantitative explorations of graduate learners' monitoring proficiencies and task understandings in the context of ill-structured writing assignments: From learner to work task as unit of analysis

By Vivek Venkatesh, Ph. D.
Concordia University, 2008

Research has debated the degree of domain generality of monitoring skills through the theoretical lens of self-regulated learning, largely in the context of studies involving college/undergraduate-level objective, multiple-choice tests. The present quantitative study sheds some much-needed light on the nature of monitoring skills in 39 adult learners tackling ill-structured writing tasks for a graduate-level e-learning theory course in the domain of educational technology. Performance prediction and confidence in predictions were collected through a theoretically-grounded self-assessment tool termed TAPE (Task Analyzer and Performance Evaluator). Monitoring proficiencies were calculated using the instructor's assessment of performance and the TAPE-related measures.

Using "learner" as unit of analysis, repeated measures procedures reveal improvements in the instructor's assessment of performance but not in any monitoring proficiencies. While the task-generality of the monitoring skills of discrimination and bias is confirmed through correlational analyses, facets of their specificities stand out due to the absence of intra-monitoring measure correlations. Subsequently, using the 247 instances of the writing task as unit of analysis, parametric multiple regression procedures demonstrate that 39% of variance in individual essay performance is predicted by combined variances in absolute prediction accuracy, discrimination, performance prediction and self-assessment scores. In addition, non-parametric ordinal and multinomial regression procedures reveal that individual essay performance can be predicted from the monitoring measures of bias, prediction confidence and absolute prediction accuracy, as well as from the self-assessment scores.

The dual levels of analyses allow not only the quantitative description of learners' content-specific calibration of performance on a writing task, but also contextualized, essay-specific insight into how individual performance on an instance of the writing task is influenced by measures of monitoring and task understanding. Results are interpreted in light of the novel procedures undertaken in calculating monitoring measures like bias using the theoretical notion of performance prediction capability. Findings are also discussed with respect to the "work task as unit of analysis" approach which enables not only the generalization to the tasks completed for the specific course described in this study, but also the interchangeability of the tasks when treating variables such as time, class session, individual student and gender as fixed effects in the various regression approaches adopted for analyses.

ACKNOWLEDGEMENTS

I thank my supervisor, Steven Shaw, for his friendship and mentoring, as well as his loyalty and commitment to enriching my research and teaching experiences at Concordia. Steven is also a rather solid companion for rock music concerts, which makes him even more indispensable!

I am indebted to my committee member, Robert Bernard, for his friendship, his belief in my abilities to undertake the proposed analysis of the data as well as for encouraging me to explore new perspectives in quantitative analyses. I tip my hat to my committee member, Dennis Dicks, for his friendship and insight on instructional design-related features of my study.

I thank my friend, colleague and mentor, Richard Schmid, for always being there.

Horns up to my friend and colleague, Stef Rucco, for his sage-like demeanour and his even sager advice on how to tackle the annals of academia.

I thank my friend and colleague, Gretchen Lowerison, who has always been a great sounding board for ideas throughout my graduate experiences at Concordia.

I thank our program administrator, Anne Brown, for her invaluable friendship and for keeping an eye out for me; for making sure that I met all the deadlines with respect to my defense, thesis submission and other program requirements. More importantly, I thank Anne for all our wonderful conversations centering around the merits and demerits of single malt Scotches and the pitfalls of addictions to collecting compact discs and vinyl.

I must acknowledge the role played by the incessant din of the heavy metal and avant-garde jazz soundtracks emanating from (and hopefully remaining within) my office, in both inspiring and distracting me from finishing my dissertation. Special mention to the doom drone masterpieces created by Sunn0))) and Miles Davis' On The Corner six compact disc box set, which helped me retreat into a noise-filled meditation when I needed most.

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

I thank my children, Maya Jade Alexandrine Detière-Venkatesh, and Marek Thierry Valério Detière-Venkatesh, who at four and two years of age respectively, have kept me grounded and taught me humility.

I dedicate this thesis to my wife and the love of my life, 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 need it the most. Perhaps, most importantly, I thank Laurence for helping me figure out how to calculate the monitoring proficiency measure of 'bias' when I was writing my master of arts thesis; late one autumn evening in 2002 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

Chapter 1 - Theoretical Framework	1
Monitoring of Learning, Performance and Comprehension	2
Self-Regulation and Instructional Design	13
Task Understanding	14
Exploring the relationship between Monitoring, Task Understanding and Performance	22
Objectives of Research	25
Chapter 2 – Methodology	26
Context and Procedure	26
Data Sources and Design	28
Chapter 3 - Results	35
Pre-test Equivalence and Inter-rater Reliability	35
Descriptive Statistics and Cross Tabulations	36
Repeated Measure Procedures	37
Correlational Procedures	38
Results from using Essay as Unit of Analysis	40
Chapter 4: Discussion and Educational Significance	43
Exploration of the Nature of Graduate Learners' Monitoring Proficiencies Tackling Ill-Structured Writing Tasks	43
Demystifying the Relationship between Task Understanding and Monitoring	49

Contribution to Theory 51

Chapter 5 - Conclusion 54

References 56

Appendix A - Course Outline, Reading List and Consent Form 67

Appendix B - Task Analyzer and Performance Evaluator 91

Appendix C - Pre-Test 92

Appendix D - Descriptive Statistics 93

Appendix E - Figures Describing Means of Collected and Calculated Measures of Performance and Monitoring 94

Appendix F – Distribution of Task Understanding Measure Across Essays 100

LIST OF TABLES

Table 1 - Intra-measure correlations for performance predictions (upper triangle) and confidence in performance predictions (lower triangle) from essays 2 through 6, *n*=38. 39

Table 2 - Intra-measure correlations for bias (upper triangle) and discrimination (lower triangle) from essays 2 through 6, *n*=38. 39

Table 3 - Results of multinomial regression procedure using instructor's performance assessment (grade) as predicted variable with monitoring and task understanding measures as predictors, as well as learner and session as fixed effects (*n* of essays=247) 42

Graduate Learners' Monitoring and Task Understanding in Ill-Structured Writing Tasks

Chapter 1 - Theoretical Framework

The work reported herein is rooted in a platform of research on self-regulated learning (SRL). Academic SRL involves the strategic application and adaptation of learners' cognitive and metacognitive thought processes in influencing their own behaviors while tackling academic tasks (Zimmerman, 1990, 1994, 2000), 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 study, is the exploration of learners' *task understanding* and *monitoring proficiencies*, in the context of ill-structured writing tasks.

A critical component of academic self-regulation is *monitoring*, or learners' abilities to evaluate their performance and learning while engaging in an academic task (Nelson & Narens, 1990; Pintrich, Wolters, & Baxter, 2000; Schraw, 1994, 1997, 1998). While monitoring has been described as an eccentric phenomenon, with variations from one individual to the next (Schraw, Dunkle, Bendixen & Roedel, 1995), research on monitoring proficiencies in college students taking multiple-choice tests has revealed both domain-specific and domain-general monitoring abilities in students (Schraw & Nietfeld, 1998). Apart from an initial, exploratory master's thesis study by the author (Venkatesh, 2002), there is a paucity of research on the nature and development of monitoring skills in graduate learners in the context of writing tasks requiring higher-order thinking skills. Also of concern is the lack of research exploring whether adults use their monitoring skills in a content-general or task-specific manner while engaged in completing ill-structured essay-writing tasks.

Monitoring of Learning, Performance and Comprehension

Metacognition and Monitoring

Monitoring falls under the general umbrella term of metacognition, which, in turn, has been discussed within the theory of SRL. For example, the self-regulatory processes of monitoring, controlling and regulating are related to, and dependent on, the metacognitive knowledge about self and cognition (Garcia & Pintrich, 1994; Pintrich et al., 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 recognize and distinguish between three aspects of metacognition: (a) metacognitive knowledge, (b) metacognitive judgments and monitoring, as well as (c) self-regulation and control of cognition (e.g., Pintrich et al., 2000). 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). Of interest in this study are the processes associated with *learning, performance* and *comprehension monitoring* in graduate learners in the context of ill-structured 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 not only to 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 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 et al., 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 conclusions, relevant to the present study, 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.

The present study focuses on learners' calibration of their performance (Glenberg, Sanocki, Epstein & Morris, 1987; Schraw, et al., 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, the general term *monitoring* is used throughout this paper, as it is more familiar to readers than the term *calibration*.

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 Pressley & Ghatala, 1990 and Schraw & Moshman, 1995, for 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 and

Epstein (1987) found a negative relationship between expertise and monitoring, Morris' (1990) research demonstrated 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) found, interestingly, 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 and 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 and 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, and 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 NS Willert (1990) found that students who generated missing information for text provided in a test, monitored more accurately than those who did not.

Characteristics of Monitoring

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 et al. (1995) observed, monitoring in test-taking situations is best characterized as an "idiosyncratic phenomenon" (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 in, or be learned by, 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 of 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 of these do not use their metacognitive 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 focused primarily within the domain of reading comprehension in a school-based population. The present study extends work reported in Venkatesh (2002) and further explores the development of graduate learners' monitoring proficiencies while tackling ill-structured writing tasks. Such research is necessary in order to further develop the notion of monitoring proficiencies in various academic contexts, as well as to explore whether monitoring can be characterised as task-general or content-specific.

Domain Specificity versus Domain Generality of Monitoring

The literature seems to be divided in its description of the nature of metacognition. The term metacognition has been defined, 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. 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. 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 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 studies to test the domain-specific and domain-general assumptions by assessing students' performance and confidence in correctly answering eight sets of multiple-choice tests. Each of the multiple-choice tests reflected a different domain of knowledge, and mainly required students to recall semantics from their long-term memories of 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 (Lindeberg, 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 study 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 study 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 study 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 study 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 study, Schraw and 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 and 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 and Nietfeld's 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).

Exploration of Adult Learners' Monitoring in Ill-Structured Writing Tasks

While Schraw and his colleagues' work demonstrates the difficulties in conceptualizing the nature of monitoring proficiencies in multiple-choice as well as fluid and crystallized tasks (see Schraw & Nietfeld, 1998), calibration of learner's performance in more complex learning tasks remains a phenomenon that is rather recondite. In addition, the generality of monitoring across repeated instantiations of a task requiring more than simple recall from semantic memory might look very different than with those tasks explored by Schraw and his colleagues. A study conducted by the author as part of a thesis project (Venkatesh, 2002), investigated monitoring proficiencies in 17 graduate learners', who completed six essay tasks over the course of a semester in an education-based learning theories course. Apart from collecting data about the instructor's assessment of performance, learners' predictions of performance, confidence in predictions and freshly derived measures of discrimination, bias and accuracy (which took into account performance predictions) were also calculated. Results indicated that performance was a task-specific phenomenon due to the level of difficulty in content covered and that performance on one essay was, for the most part, unrelated to performance on other essays. Results also revealed that monitoring measures displayed a propensity towards a general ability, manifesting themselves as one or more unique patterns across a set of logs. In addition, inter-correlation values between monitoring measures were insignificant, indicating some task and content-specific monitoring qualities in the learners. The present project extends the pilot work conducted in

Venkatesh (2002) to a larger sample and improves upon the methodology of inquiry by providing more fine-grained measurements of performance, performance predictions, confidence in performance predictions and hence, monitoring proficiencies.

Self-Regulation and Instructional Design

A secondary purpose of the present study is to explore a specific facet of learners' task understandings in the context of essay-writing tasks. Specifically, through the use of empirically supported instructional design (ID) principles, this study explores the possibilities of attuning learners' perceptions of the assessment criteria for a 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 the author leading a project on the 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, the review 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 one can better instructionally promote *task structuring* in the context of ill-structured essay-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). While the above theorization of task understanding is not new, it provides a different approach to view the distinctions offered by Winne and Hadwin (1998), who distinguish between the task and cognitive conditions that influence students' comprehension of an academic task. According to Winne and Hadwin (1998), *task conditions* refer to the nature and assessment criteria of a task, whereas *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 the 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 the 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 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 acknowledging the importance of task understanding as a critical, first phase of SRL, 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, in accordance with Winne and Hadwin (1998), it is important to acknowledge 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

The aforementioned review (Venkatesh & Hadwin, 2002) 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, the 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".

However, the review 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.

Task Analyzer and Performance Evaluator – A Tool to Improve Task Understanding

In the present study, a self-assessment tool, the Task Analyzer and Performance Evaluator [TAPE, originally conceived in Venkatesh (2002)], is used to enable learners to develop a more accurate understanding of the assessment criteria for an essay-writing task. In prior investigations (Venkatesh, 2002, 2005), the TAPE tool was shown to have helped 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 the Venkatesh and Hadwin (2002) review, 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. The TAPE tool does not elicit the motivational, affective or emotional states of students while they performed their self-assessments. A second function of the TAPE tool is to assess the development of monitoring proficiencies of learners as they tackled the writing task; to this end, the TAPE tool is used to collect measures of performance prediction and prediction confidence for the writing task that learners will engage in over the course of instruction. The TAPE tool is also, therefore, built on principles espoused in discrepancy-reduction models of SRL (see Maki, 1995), which espouses the theoretical platform that persons capable of monitoring their learning proficiently will more effectively regulate their performance on an academic task.

Exploring the relationship between Monitoring, Task Understanding and Performance

Apart from the studies conducted by Schraw and his colleagues, there is little empirical evidence of the relationship between learners' accuracy in monitoring and their performance. For that matter, Pressley and Schneider's (1997) review of literature points out that studies supporting the relationships between prediction (monitoring) accuracy and performance are few and far between. In fact, there are instances of studies that point out the contrary, i.e., that improved performance in test-taking situations is related to less accurate monitoring (e.g., Begg, Martin & Needham, 1992), or that improved performance cannot be attributed to improved monitoring (e.g., Dunlosky & Connor, 1997; Dunlosky & Hertzog, 1997; Kelly, Scholnick, Travers & Johnson, 1976). An additional point of concern is drawn by Maki (1998), who, in reviewing instructional attempts to improve monitoring accuracy reveals that research efforts have produced less than stellar results.

Recently, Thiede, Anderson and Therriault (2003), as well as Thiede and Anderson (2003) have both speculated that, perhaps, the reason that researchers have not had success in observing a relationship between monitoring and performance is the lack of control in the experimental designs employed. They contend that if learners are allowed to allocate time to use the results of their monitoring to regulate their performance, one might better observe a causal relationship between monitoring and performance. In a recent investigation, Thiede et al. (2003) experimentally manipulated levels of monitoring accuracy and observed its differential effects on 66 undergraduate learners' generation of keywords and comprehension after they read (or during their reading of – depending on the experimental condition assigned) six pieces of expository texts. Monitoring accuracy, measured as Goodman-Kruskal gamma correlation between learners' comprehension rating and pre-reading test performance, was found to be greater for a delayed-keyword group (i.e., students who wrote keywords after a delay) than for a group that wrote keywords immediately following the reading and a group that wrote no keywords at all. In addition, the delayed-keyword group's performance on a reading comprehension test (composed of multiple-choice questions) was significantly greater than those of the other two groups. More recent work by Thiede, Dunlosky, Griffin and Wiley (2005) has demonstrated that providing an increased amount of time for learners to reflect on their comprehension on a pieces of text before generating keywords does not necessarily lead to improved monitoring proficiency: in effect, simply providing the opportunity to explicitly reflect on the text and generate keywords after any delay of time was sufficient for improved accuracy in metacomprehension ability.

While Thiede and his colleagues' experimental designs allow researchers to compare performances among students with variable monitoring proficiencies in well-structured tasks, the question still remains as to how one can better design instructional tools to help learners regulate their performance on more complex and consequential academic tasks. In the context of this study, preparing graduate learners for the educational technology-related workforces includes helping these knowledge workers to become better judges of their own performance on ill-structured written tasks, thereby increasing the efficiency with which such tasks can be accomplished.

While acknowledging the importance of the results of the experimental investigations of the differential effects of monitoring on performance, it remains to be seen how one can implement instructional tools, based on these causal relationships, to help learners attain higher levels of self-regulation. It is therefore still necessary to observe and explore how monitoring proficiencies develop in naturalistic environments, where learners avail of feedback on their performance, explicitly monitor their performance and task understanding, and in turn, try and ameliorate their performance on less-structured and graded academic tasks than those experimentally investigated in Thiede et al.'s (2003, 2005) studies.

Results from inter-measure correlational procedures employed by the author's prior study (Venkatesh, 2002) on the relationship between the monitoring and task understanding revealed a complex, but insignificant relationship. However, qualitative investigations (Venkatesh, 2005) using interviews with participants from the initial masters study (Venkatesh, 2002) revealed a distinct essay-specific relationship between perceptions of assessment criteria and monitoring of performance. In their interviews, learners were generally convinced that, from one essay to the next, the simple act of explicitly monitoring their own performance using the TAPE tool led to a more comprehensive understanding of the complex essay-writing task they were assigned. In the present study, this complex and recondite relationship is further explored using statistical methods that view the data, both from the lens of the learner as well as the essay as respective units of statistical analyses.

Objectives of Research

The objectives of the present study are: (a) to explore statistically the development of graduate learners' monitoring proficiencies as they engage in six instantiations of an ill-structured writing task; (b) to shed light on the task-specificity and/or content-specificity of adult learners' monitoring skills using inter and intra-measure correlational procedures; (c) to explore the statistical relationship, if any, between learner's self assessment of meeting assessment criteria (a facet of their task understanding) with their monitoring proficiencies using both learner and essay as unit of analysis; and (d) discuss the theoretical and practical implications of investigating monitoring (or calibration) of performance and task understanding in the context of graduate essay-writing assignments.

Chapter 2 – Methodology

Context and Procedure

Thirty-nine student volunteers, 15 of whom were male, were recruited from a total of four sessions of a graduate, classroom and laboratory-based "theories of e-learning" course given by the author at a large North American university. The sessions took place consecutively between January 2006 and June 2007. Each session of the course included a total of 13 classroom-based tutorials and five to six laboratory-based storyboarding and usability-testing activities. Each tutorial and laboratory-based activity lasted between 90 to 120 minutes. Tutorials included group-based discussions of assigned readings (see Appendix A for course outline and reading list used for all four sessions), while laboratory-based sessions included storyboarding of a clients' sales-based training needs as well as a usability test of an indexing mechanism for a neo-corpus of learner essays written for the instructor's previous instantiations of the "theories of e-learning" course.

In preparation for the tutorials, students were expected to complete an ill-structured essay-writing assignment, on subject(s) of their choice, based on topics covered in the assigned readings and/or laboratory activities. Assessment criteria used to grade the essays were developed using Biggs' (1991, 1996) SOLO taxonomy. Essays that received a top grade needed to (a) make valid links between practical e-learning related issues and learning theories, (b) extend discussions from the readings to application-based scenarios and (c) provide a clear balance between the pros and cons of adopting a specific theoretical perspective. Criteria were made explicit to all students before the writing of the first essay (see Appendix A for complete description of criteria). This essay-writing assignment was classified as ill-structured because (a) the goals of the essay were not well defined, (b) the constraints imposed by contextual factors were not readily apparent, (c) the solution to the essay-writing problem was not easily known and (d) there were multiple perspectives on both the solution and the solution path (Reitman, 1965; Voss, 1998; Voss & Post, 1988). Each essay was accompanied by the self-assessment tool described previously, the Task Analyzer and Performance Evaluator [TAPE, Venkatesh (2002); see Appendix B], designed to (a) help students articulate, in written form, their justifications for meeting the instructor's assessment criteria, and (b) elicit learners' predictions of performance and their confidence in these predictions. One essay was written for every two tutorial sessions. Essays were submitted and graded online using the FirstClass® conferencing software tool. Feedback from the instructor was embedded and the assignments were returned electronically to the student within 72 hours of submission along with comments on the portion of the TAPE that dealt with students' justifications of having met the instructor's assessment criteria. Consent forms (see Appendix A) were prepared and all data were collected in accordance with principles outlined by the American Psychological Association; ethical approval was obtained from the university's Ethics Committee. While all participants were aware of the research program of their instructor, their consent forms and performance prediction-related data were only made available to the author after final grades for the courses were submitted to the university. All essays and accompanying instructor comments, grades and measures of performance prediction and confidence in predictions were stored electronically in a password-protected hard drive.

Data Sources and Design

Pre-test measures of content knowledge specific to the course offered, as well as pre-test scores on an essay-writing assignment based on the SOLO taxonomy, were collected from each student during the first tutorial for each of the four sessions (see Appendix C for pre-test questions). All essays for the sessions of the course were written by learners, individually, at their convenience, between the second and thirteenth tutorial (note that the laboratories were held immediately following select tutorials). For the first essay, only the instructor's assessment (score range: 0 to 100; converted grade range: C to A+) and the instructor's feedback on the student's self assessment were recorded (0=incorrect, 1=partially correct, 2=correct). For all subsequent essays, the following measures were obtained: (a) instructor's assessments of student essays; and TAPE-related scores, which included (b) students' performance predictions (range: 0 to 100; converted grade range: C to A+), (c) students' confidence in predictions (range: 0 to 100), and, (d) the instructor's feedback on students' self-assessment. Also collected, from essay number 2 onwards, were theoretically derived measures, including (a) discrimination (range: -100 to 100), which measured students' abilities to assign an appropriate level of confidence to their predictions (based on initial work by Schraw et al., 1995; modified and piloted in Venkatesh, 2002); and (b) bias (range:-100 to 100), which measured the degree to which students were over or under-confident in their predictions (based on initial work by Schraw et al., 1995, and Schraw & Nietfeld, 1998; modified and piloted in Venkatesh, 2002).

Fifteen of the 39 students wrote seven essays over the duration of the course; twenty-three others wrote a total of six essays, while one student wrote four and subsequently dropped the course. For all individual-based analyses, measures collected and calculated from the first six essays written by each of 38 participants who completed the course (one student dropped out) were used in a one-shot case study-based repeated measures design. In combination with correlation procedures, the design enables the uncovering of trends in the measures of performance and monitoring of interest in this study.

Calculation of Discrimination and Bias

Procedures for calculating discrimination and bias for the present study (initially piloted in the author's master's thesis study, Venkatesh, 2002) were different from those employed by Schraw and his colleagues insofar as Schraw and his team never factored the theoretical notion of performance prediction capability into their theoretical and practical conceptions of monitoring.

Discrimination. For each essay written, the measures of instructor's performance assessments (both grade and score), student's performance predictions (both grade and score) and student's prediction confidences were used to calculate two measures of monitoring proficiency. The first measure of monitoring proficiency calculated is *discrimination*, which, in the context of this study, measures the degree to which learners assign an appropriate level of confidence to their predictions of the grade for each essay.

Discrimination was cumulatively calculated by taking the signed difference between the average prediction confidence scores for accurate predictions and the average prediction confidence scores for inaccurate predictions for all essays written up to a specific point in time. Discrimination scores were calculated for each essay. The value of discrimination ranged from –100 to +100. 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 +100) are very proficient in monitoring as it suggests that they can assign a high value of confidence when accurately predicting their grades on the essay assignment. The closer the value of discrimination to 100, the more accurate was a student's monitoring.

Performance predictions were deemed accurate if the grade predicted by the student was the same as the grade assigned by the instructor. For example, a performance prediction score of 86 (i.e., a grade of A) is accurate if and only if the instructor's performance assessment score lies between 85 and 89 (i.e., the range of scores describing the grade of A). For essay 1, if the students' performance prediction grade was equal to the performance assessment grade, 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 essays, discrimination was calculated by taking the average of the signed, converted prediction confidence score (using the same procedures as described for essay 1) and the previous essay's discrimination score. This means that the score of discrimination for essay 2 represents the student's ability to discriminate, based on predictions from both essays 1 and 2. Discrimination scores for essay 6 provide a measure of the students' abilities to discriminate, based on predictions from all six essays.

Bias. The second measure of monitoring proficiency calculated is *bias*. Bias measured the extent to which a learner's capacity to predict performance is commensurate with their prediction confidence. In other words, bias measured the degree to which individuals are over or under-confident for each TAPE self-evaluation made. Bias was calculated by taking the signed difference between performance confidence and prediction capability. Like the discrimination score, bias ranged in value from –100 to +100. A negative value of bias indicated under-confidence, whereas positive values indicated overconfidence in predicting scores; 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 scores. 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. For example, a 75% prediction subtracted from 75% prediction confidence yields the ideal bias value of 0.

Bias was calculated independently for each essay. Prediction capability was calculated by taking the percentage of the ratio of the values of performance prediction and performance assessments, with the smaller of the two values in the numerator of the ratio. Prediction capability hence measured how well the student had predicted a grade for a particular essay. For example, if a student predicted a score of 50 and received a 50 from the instructor, the value of prediction capability would be calculated as the ratio of 50 to 50, yielding a score of 1, suggesting 100% prediction capability. If the student overestimates performance and predicts a score of 90 for the essay, but in fact receives a 60, then prediction capability is calculated as the ratio of 60 to 90, yielding prediction capability of 66.67%. This suggests that the student was able to receive only 66.67% of the grade predicted. If the student underestimates performance by predicting, for example, a score of 80, but receiving a perfect score of 100 from the instructor, the prediction capability is calculated as the ratio of 80 to 100, which gives a percentage score for prediction capability as 80%. This suggests that the student was able to predict only 80% of the final grade received.

Work Task as Unit of Analysis

In an attempt to better explicate the relationship between a singular facet of task understanding, viz., learners' perceptions of the ill-structured writing assignment's assessment criteria and their variable monitoring proficiencies, an attempt has been made to consider the essays themselves as a statistical unit of analysis. The theoretical basis for conducting this procedure is explicated, in great detail, in Shaffer and Serlin's (2004) landmark piece on intra-sample statistical analysis (ISSA). In the present study, there is sufficient qualitative evidence (Venkatesh, 2005) suggesting that learners' perceptions of the assessment criteria are related to their perceived proficiencies in monitoring (e.g., their confidence in predicting their grades, their grade predictions themselves, etc.). Quantitatively, however, this relationship has revealed itself to be of a complex, but insignificant variety as evidenced by the rather low inter monitoring measure correlation values, which tend to fluctuate in both directions (Venkatesh, 2002). Additionally, in recent studies conducted by Thiede and his colleagues (Thiede & Anderson, 2003; Thiede et al. 2003; Thiede et al. 2005), the suggestion has been put forth that not enough experimental control is exerted for researchers to be certain how monitoring affects task performance or even academic self-regulation. Finally, when confronted with data organized and analysed by learner as unit of analysis, it is not uncommon to notice that the lack of a large sample combined with the repeated measure procedures (as employed in a prior pilot investigation, Venkatesh, 2002) leaves very little room for powerful statistical results. Treating the work task, or in this case, the essay, as unit of analysis, would enable the harnessing of powerful, multivariate statistical procedures, with a relatively larger sample, so as to confirm some of the qualitative observations made in Venkatesh (2005) and provide fodder for future theoretical and research considerations in the area of exploring the development of monitoring proficiencies.

Two major issues taken into consideration before commencing the essay-based analyses were those of generalizability and exchangeability/interchangeability (Shaffer & Serlin, 2004). All essay-based analyses are generalized to all essays that could possibly have been written by the set of 39 learners registered in the four session of the "theories of e-learning" course offered by the instructor. In addition, while treating an individual essay as unit of analysis, after taking into account all possible measured factors, including the author of the essay, session in which it was written, and the numerical sequence in which the essay was written (i.e., essay 1 through 7), essays can be considered exchangeable or interchangeable with one another. The notion of exchangeability demands that one treats individual learners as fixed effects in any multivariate model so as to contextualize the results to the sample of individuals from which the essays were drawn.

In treating the work task as unit of analysis, a total of 247 essays were used (i.e., 15 learners who wrote seven essays each, 23 who wrote six essays each, and one learner who wrote four essays and later dropped the course). Each essay was described by the following variables: unique identification code, author, session in which essay was written, numerical sequence (i.e., essay number 1, 2, 3, 4, 5, 6 or 7), instructor's performance assessment, author's performance prediction, author's confidence in performance prediction, and the calculated measures of discrimination, bias and absolute accuracy (i.e., the unsigned difference between the prediction and instructor's assessment for each essay).

Chapter 3 - Results

Pretest Equivalence and Inter-rater Reliability

Pretest scores of content knowledge and essay-writing ability showed no statistical differences across the four sessions, gender, or prior relevant work experience, thereby justifying the collapsing of the graduate participants into one group of 38 (excluding the one learner who wrote four essays and dropped the course). All 247 essays (from the 39 participants) were scored by two independent raters who were chosen based on their past university teaching experience, excellent command of the English language, high levels of prior content knowledge, and experience in writing essays using the SOLO taxonomy for prior instantiations of the "theories of e-learning" course offered by the instructor. The raters received the essays in the same order as the instructor received them; the weekly sequence of submission for the course was adhered to as was the sequence in which the four sessions were held. This ensured that the raters viewed the essays in precisely the same order as the instructor. Initial meetings between the raters and the instructor were held to enable training and clarification of doubts concerning the criteria for the essay-writing assignment. Subsequent to this training, meetings were held after raters had completed the grading for an entire session's worth of essays. Fleiss' Kappa, an inter-rater reliability coefficient, was calculated to be 0.87. All 24 discrepancies in rating were resolved through discussion.

Descriptive Statistics and Cross Tabulations

Descriptive statistics (see Appendix D for a complete table, Appendix E for related figures) for essays 1 through 6 showed that the instructor's average performance assessments ranged from 77.84 to 90.84 (range of *SD*s: 7.47 to 9.46). For essays 2 through 6, descriptives for the monitoring-related measures were as follows: (a) the learners' average performance predictions across the essays ranged from 80.47 to 84.03 (range of *SD*s: 5.79 to 6.73); (b) their confidence in predictions ranged from 74.03 to 81.50 (range of *SD*s: 9.91 to 17.84); (c) they were more prone to negative discrimination, i.e., they assigned higher confidence to inaccurate predictions than accurate ones (range of *M*s: -29.52 to -52.13, range of *SD*s: 42.92 to 61.28); (d) they were generally underconfident in their predictions (i.e., they demonstrated negative bias) across the duration of writing essays 2 through 6 (range of *M*s: -9.52 to -16.63 , range of *SD*s: 10.54 to 19.93); and (e) average absolute accuracy (i.e., the unsigned difference between the performance prediction and instructor's assessment for each essay) ranged from 7.58 to 8.47 (range of *SD*s: 5.60 to 7.51) . Finally, the distribution of categories for the instructor's feedback on students' self-assessment of meeting the criteria for the essay (see Appendix F) showed chance variation from essays 1 through 6 according to results of Friedman's non-parametric test of related samples.

Repeated Measure Procedures

Repeated measures analyses were conducted using instructor's performance assessments, students' performance predictions, students' confidence in predictions and the monitoring proficiencies of discrimination, bias and absolute accuracy as dependent measures while session, gender, student status (full-time versus part-time) were designated as independent variables. In addition, the multivariate models included pre-test scores for content knowledge and essay-writing ability as covariates.

The analysis revealed that (a) the collected monitoring measures of students' performance predictions, confidence in predictions and calculated monitoring measures of discrimination, bias and absolute accuracy fluctuated with chance across the essays and showed no interactions with any of the independent variables or covariates; (b) instructor's performance assessments yielded a statistically significant value of .51 for Pillai's trace, [omnibus $F(5,17)=3.46$, $p=.02$, *partial* $\eta^2=.51$, $ES=1.02$] and showed no interactions with any of the independent variables or covariates; and (c) pairwise comparisons between instructor's performance assessments (range of *M*s: 77.59 to 90.84, range of *SD*s: 6.80 to 9.56), corrected by Bonferroni's adjustment showed certain significant improvements across time ($p<.05$). Specifically, essays written in the first week scored significantly lower than essays written in the fourth, fifth and sixth week; those written in the second week were significantly poorer than those from the fifth and sixth week; and finally, those written in the third week scored significantly lesser than those written in the sixth week.

Correlational Procedures

Intra-item correlational procedures revealed that instructor's assessment of scores on student essays fluctuated largely due to chance across the essays, while the monitoring-related measures of performance prediction, confidence, bias and discrimination showed statistically significant intra-item correlations (see tables 1 and 2 for the intra-item correlations for the monitoring measures). Accuracy as well as absolute accuracy, on the other hand, showed insignificant relationships across the essays.

Partial intra-correlations between confidence scores across the essays improved when variance explained through correlations between confidence and performance assessments were controlled. On the other hand, partial intra-correlations between performance prediction scores across the essays did not show remarkable differences when variance accounted for through correlations between confidence and performance predictions were controlled. In addition, partial intra-correlation scores across essays, for both discrimination and bias, showed improved values when variability explained by performance assessments was controlled for. The inter-measure non-parametric correlations between learners' task understanding (i.e., instructor's feedback on student's self-assessment of meeting assessment criteria) and each of the monitoring proficiencies produced insignificant results.

Table 1: Intra-measure correlations for performance predictions (upper triangle) and confidence in performance predictions (lower triangle) from essays 2 through 6, n=38.

	Essay 2	Essay 3	Essay 4	Essay 5	Essay 6
Essay 2	-	.56**	.54**	.31	.33*
Essay 3	.52**	-	.59**	.43**	.36*
Essay 4	.62**	.72**	-	.80**	.77**
Essay 5	.52**	.50**	.59**	-	.74**
Essay 6	.60**	.48**	.67**	.73**	-

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

Table 2: Intra-measure correlations for bias (upper triangle) and discrimination (lower triangle) from essays 2 through 6, n=38.

	Essay 2	Essay 3	Essay 4	Essay 5	Essay 6
Essay 2	-	.32*	.50**	.37*	.49**
Essay 3	.58**	-	.64**	.41**	.21
Essay 4	.23	.60**	-	.45**	.45**
Essay 5	.15	.57**	.75**	-	.57**
Essay 6	-.08	.29	.48**	.64**	-

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

Results from using Essay as Unit of Analysis

Multiple Regression Procedure

When considering essays as unit of analysis, the instructor's assessment of performance was parametrically regressed on the essay-specific measures of instructor's feedback on self-assessment (i.e. task understanding), performance predictions, confidence in predictions, absolute prediction accuracy, discrimination, and bias, while treating gender, time (i.e., the numerical sequence in which the essays were written) and individual student as fixed effects through the use of dummy variables (p to enter $< .05$, p to remove $> .10$). Overall, a statistically significant amount of variance in the performance assessment (39%) was explained by a combination of the variance in measures of absolute accuracy ($\beta = .61$), discrimination ($\beta = .31$), performance prediction ($\beta = .29$) and instructor's feedback on self-assessment ($\beta = .20$), R^2=.39, $F(4, 203)$=34.31, p<.001. A further 13.5% of variance was predicted by fixed effects, including six individual learners and two instances of time.

Non-parametric Regression Procedures

A non-parametric ordinal regression procedure was used to evaluate the predictors of the instructor's performance assessment (as a grade). The omnibus model included the predictors of essay-specific bias, absolute accuracy, confidence in prediction, and performance prediction, while treating individual learner, gender, session, feedback on self-assessment, and time as fixed effects. If all variables in the model are held constant while the manipulations are carried out in each of the following cases, ordinal regression procedures revealed that (a) if the essay-specific bias were to increase by one unit, then the log-odds estimate of improving performance would decrease by a factor of -7.60; (b) if the essay-specific absolute accuracy were to increase by one unit, then the log-odds estimate of improving performance would increase by a factor of 8.51; (c) if the essay-specific confidence were to increase by one unit, then the log-odds estimate of improving performance would increase by a factor of 7.60; and (d) if the essay-specific performance prediction were to increase from a B to a B+, then the log-odds estimate of improving performance would increase by a factor of 6.31. Two individual learners were also revealed as predictors of performance.

A follow-up multinomial regression procedure provides specific models for predictors of individual performance assessment grades, relative to the grade of A+ (see table 3). The log-odds estimate of scoring an A- or A grade (relative to A+) increases as the essay-specific bias increases by one unit or when the instructor's feedback on self-assessment improves from partially correct to completely correct, but it also decreases when confidence or absolute accuracy increase, provided all other variables in the model remain constant. Similarly, the log-odds estimate of scoring a B+ grade (relative to A+) increases as the essay-specific bias increases by one unit, but decreases when confidence or absolute accuracy increase, provided all other variables in the model remain constant.

Table 3 – Results of multinomial regression procedure using instructor's performance assessment (grade) as predicted variable with monitoring and task understanding measures as predictors, as well as learner and session as fixed effects (*n* of essays=247)

Predictor	Regression coefficient for logit of achieving grade (relative to A+)		
	B+	A-	A
Bias	15.57	15.27	4.03
Confidence	-15.54	-15.25	-4.00
Absolute Accuracy	-17.20	-17.47	-4.72
Partially Correct Self-Assessment	*not significant*	6.25	5.81

Note: for all cells, except for those denoted with the entry *not significant*, regression coefficients had Wald statistic with a significance detected at $p<.01$

Chapter 4 - Discussion and Educational Significance

Exploration of the Nature of Graduates Learners' Monitoring Proficiencies Tackling Ill-Structured Writing Tasks

Evidence of General Monitoring Ability

The results of this study point to some interesting facets of graduate learners' monitoring proficiencies in the context of an ill-structured writing task. While the performance assessments were, in large part, an essay-specific phenomena, with performance on one essay mostly unrelated to performance on another essay, prediction confidence scores were strongly related to one another, over and above the performance assessments. This provides support for the presence of a general confidence ability, which 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).

The results also suggested that learners' prediction confidence scores on any one essay was not necessarily bound to their performance assessments on that essay, 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 essay was related neither to performance assessment on the previous essay nor to performance assessment on essays 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 essays, in any meaningful way.

The results also suggest that prediction confidence develops as a unique pattern across successive essays when feedback was available for the earlier essay; 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 Performance Predictions in Calculating Monitoring Proficiencies

An important aspect of this study is the introduction 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' monitoring-related statistical 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 implicitly predicted perfect performance. Further, in Schraw and his team's studies, monitoring proficiencies were calculated using performance and confidence scores. In the present 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. Results demonstrate that monitoring of performance in the context of ill-structured writing activities needs to take into account students' performance predictions. 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 this study indicate that as the essays progressed, students' consistently predicted higher grades and had greater confidence in their predictions. However, the relationship between prediction confidence and performance predictions was highly essay-specific, with no discernable patterns across essays.

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 (see also Thiede et al.'s, 2005 contention that more experimental work needs to be conducted in exploring monitoring by controlling for difficulty levels of content). In fact, this difficulty factor might have played a large role in the essay-specificity of the instructor's assessment of performance (c.f., Venkatesh, 2002 where a similar phenomenon occurred). Despite the fact that the students were performing the same task (essay writing) over the semester, 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. The significant intra-correlations between performance prediction measures suggest that performance prediction behaved very differently from the instructor's performance assessments. While performance assessments were essay-specific, findings suggest that performance predictions developed as a stable pattern across the essays.

Discrimination in Predictions

Results suggest that students showed an increased ability with regards to discrimination, that is, as the essays progressed students were better able to assign an appropriate level of confidence to their performance predictions. Findings reveal the possible existence of a discrimination pattern across essays 3, 4, 5 and 6; i.e., regardless of the content of the readings, class discussions and their essays, learners tended to discriminate in a distinct pattern between essays 3 and 6. These results conflict, somewhat, with those that are found in Schraw et al.'s (1995) study, where discrimination was a content-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 study. The calculation procedures for the measure of discrimination in this study take into account (a) the progressive nature of the learning essay task, (b) students' performance predictions, (c) instructor's performance assessments, and (d) students' prediction confidence scores. The existence of a pattern of discrimination in students engaged in a ill-structured 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 the results support the existence of two essay-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.

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 essays. 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 essays, 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.

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 exists across the essays. This notion of a general bias ability is supported by the increased association between significantly correlated bias scores when variation due to the performance assessments is 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, in the present study, bias and confidence did not show strong intercorrelations.

Investigating Monitoring Proficiencies in the Context of Ill-structured Writing Tasks

The above discussion provides a picture of how monitoring proficiencies developed in 38 graduate learners across six essays. The exploratory procedures employed in the analyses provide preliminary evidence that learners' monitoring proficiencies showed a propensity towards being a general phenomenon across the essays, as opposed to being specific to each essay. While the measures of prediction confidence, performance prediction, discrimination and bias, each revealed intra-correlated patterns that spanned across a set of essays, successful performance, as gauged by the instructor, was the only variable that retained an essence of being specific to each learning essay. Support for a task-general hypothesis would have been strongest if performance measures were uncorrelated, and confidence, discrimination and bias were uniformly correlated across the essays. Such a pattern of correlations would mean that student prediction confidence and monitoring abilities were related across the essays despite performance being a unique phenomenon to each essay. While each of the monitoring measures show intra-measure correlations across the essays, no inter-monitoring measure correlations were found, making it difficult to view a general monitoring ability and yet, at the same time, opening the door for an explanation via a theory of content or task-specific monitoring.

Demystifying the Relationship between Task Understanding and Monitoring

Not surprisingly, when viewed through the perspective of student as unit of analysis, the intercorrelations between the measures of task understanding and each of the monitoring proficiencies did not produce significant findings, reflecting what was observed in Venkatesh (2002). While part of the reason for this can be accorded to the fact that task understanding is a complex phenomenon, and that this study looked at a specific facet of the same, viz., students' abilities to explicitly express how they met the instructor's assessment criteria, it is encouraging to see that the essay-based analyses begin to scratch the surface of how task understanding, monitoring and performance seem to interact with one another.

Keeping in mind that the essay-based procedures can only be generalized to all possible essays that could have been written within the context of the course being given by the instructor, the results provide an exceptional opportunity for future research to better investigate the slippery phenomena of task understanding and monitoring.

The multiple regression procedure reveals that essay-specific performance can be significantly predicted by four combined measures of task understanding and monitoring (the variance accounted for by the four measures was 39%). This relationship holds true even in the face of using individual learners and time as fixed factors; in fact, these fixed factors accounted for no more than 12% of the variance in performance. In addition, the models resulting from the non-parametric regressions reveal precisely how the measures of task understanding and monitoring engage in a complex battle to influence how essay-specific performance might fluctuate in the context of the ill-structured writing assignment assigned for the four sessions of the "theories of e-learning" course described. Specifically, when one views the details of the models proposed by the multinomial regression procedures, it is interesting to note how increased confidence and inaccurate predictions reduce the likelihood of improved performance. However, an increase in essay-specific bias and the ability to improve task understanding seemed to influence performance positively. It remains to be seen how future research can conceptualise these seemingly conflicting directions that seem to pull apart the self-regulatory mechanisms that guide how learners perceive their comprehensions of tasks and how they calibrate their performance.

Contribution to Theory

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, Borkowski & Schneider, 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. The present study explores monitoring proficiencies in the context of a more ill-structured 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, tempered by an essay-specific monitoring ability which manifests itself as unrelated discrimination, bias and absolute accuracy measures.

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 the latter. Future research in the investigation of monitoring of learning and performance in ill-structured writing tasks should, therefore, investigate which types of domain-specific monitoring abilities are, in fact, present and are utilised by learners in such contexts. Future research should also investigate the relationship between the newly derived measures of 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; Thiede & Anderson, 2003; Thiede et al., 2003; Thiede et al., 2005; 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 has been 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, as has been proposed by Schraw and Impara (2000). 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, Chan and Muthukrishna (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 ill-structured writing activities should explore whether and how monitoring proficiencies become more domain-general or domain-specific in nature. To follow a train of thought initiated by Schraw and Impara (2000), the results of the present study indicate that, for example, if one subscribes to the good information processing model, then learners who are engaged in writing learning essays 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.

Chapter 5 - Conclusion

The non-existence of intra-item correlations between the students' performance scores juxtaposed against the strong intra-item correlations between the monitoring-related measures gives credence to the content and task-generality of monitoring skills. However, the lack of inter-measure correlations for the monitoring-based variables shows that graduate learners' engaged in ill-structured essay tasks tend to adapt their method of calibration in a different way than is seen for more objectively oriented tasks. Students might therefore possess a general monitoring ability across essays in addition to essay-specific knowledge and regulatory skills. These findings lend strong support to the content-general hypothesis of monitoring, and yet provide fodder for discussions related to the task-specificity of these same monitoring skills. The inclusion of prediction capability in the calculation of bias and discrimination in the present study should impact the way researchers and practitioners conceive of, measure and apply interventions to improve adult learners' monitoring proficiencies. The lack of relationship between measures of monitoring and performance, when viewed from the lens of individual as unit of analysis, also represents a reality faced by researchers of SRL-related constructs in that the individual components of SRL may sometimes not work in concert towards development of what the author contends is a still esoterically-defined trait. The use of essay as unit of analysis enables the fine-grained dissection of how task understanding and monitoring might work in concert and against one another in predicting essay-specific performance. While the results from the essay-based analyses cannot be generalized to a context outside of the one explored in the present study, they encourage and fuel the cycle of building theoretical hypotheses which can be tested in a future research program. Finally, from a practical perspective, trend analyses, longitudinal correlation-based research, and work task-related perspectives on key self-regulatory processes in academic settings unveils both the context-specific and context-general instructional features that need to be integrated into learning environments to better promote monitoring and task understanding among graduate learners tackling fairly difficult writing tasks.

References

Begg, I. M., Martin, L. A., & Needham, D. R. (1992). Memory monitoring: How useful is self-knowledge about memory? *European Journal of Cognitive Psychology, 4*, 195-218.

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* (4^{th} 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.

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition, 25*, 691-700.

Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for a restudy during multitrial learning. *Journal of Gerontology: Psychological Science, 52*, 178-186.

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 to 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). Hillsdale, 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.

Kelly, M., Scholnick, E. K., Travers, S. H., & Johnson, J. W. (1976). Relations among memory, memory appraisal, and memory strategies. *Child Development, 47*, 648-659.

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. (1995). Accuracy of metacomprehension judgments for questions of varying importance levels. *American Journal of Psychology, 108*, 327-344.

Maki, R. H. (1998). Predicting performance on text: Delayed versus immediate predictions and tests. *Memory & Cognition, 26*, 959-964.

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.

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 metacogniion and self-regulated learning. In G. Schraw & J. C. Impara (Eds.), *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., & 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., & Schneider, W. (1997). *Introduction to memory development during childhood and adolescence*. Mahwah, NJ: Erlbaum.

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.

Reitman, W. (1965). *Cognition and thought.* New York: John Wiley.

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. (2000). *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. (1997). Social origins of self-regulatory competence. *Educational Psychologist, 32*, 195-208.

Shaffer, D. W., & Serlin, R. C. (2004). What good are statistics that don't generalize? *Educational Researcher, 33* (9), 14-25.

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

Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28*, 129-160.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95* (2), 66-73.

Thiede, K. W., Dunlosky, J., Griffin, T. D., and Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31* (6), 1267-1280.

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. (2002). *Exploring the development of graduate students' monitoring proficiencies and task understandings in the context of a complex writing task.* Unpublished masters thesis. Department of Education, Concordia University: Montreal, QC.

Venkatesh, V. (2005, April). *Exploring the development of graduate learners' task understandings in ill-structured writing tasks.* Paper presented at the Annual Convention of the American Educational Research Association: Montreal, QC.

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. (1998). On the representation of problems: An information-processing approach to foreign policy decision making. In D. A. Sylvan & J. F. Voss, *Problem Representation in Foreign Policy Decision Making* (pp. 8-26). Cambridge, UK: Cambridge University Press.

Voss, J. F., & Post, T. A. (1988). On the solving of ill-structured problems. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The Nature of Expertise* (pp. 261-285). 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. (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.

Appendix A – Course Outline, Reading List and Consent Form

Course Outline

Computer Assisted Instruction

ETEC XYZ

Course Instructor: Vivek Venkatesh

Course Description and Objectives

This session of ETEC XYZ, *Computer-Assisted Instruction*, is designed with three major purposes in mind. First, this course intends to engage graduate learners in Educational Technology in a discussion of the current trends in computer-assisted instruction (CAI) applications and learning technologies, including those relevant in (a) school, college and university-based educational environments, as well as (b) human performance technology-related industrial and business settings.

Learners are strongly encouraged to explore relevant topics on the Web and participate actively in the classroom discussions. Tutorial sessions will focus mainly on a critique and mindful discussion of the weekly assigned readings. Sessions will include prepared debates, instructor-led discussions of case studies and learner-led discussions of essays.

Readings are compiled from databases managed by the Association for Advancement of Computing in Education (AACE), Association for Educational Communications and Technology (AECT), online editions of relevant educational technology and e-learning-based journals, and web-based white papers. All digital articles will be made available to learners online – distribution/reproduction of these articles for monetary gain or non-monetary purposes is strictly prohibited by copyright law.

Readings will focus on the following topics: e-learning overview; e-learning design issues and strategies; learning technology standards and meta-data tagging; re-usable learning objects; human resource issues and competency models; learning content management systems and learning management systems; organizational impact of learning technologies; usability; content management strategies; blended learning; and best practices in e-learning.

Connections to the World Wide Web will be made available during tutorial discussions; learners who wish to demonstrate applications or discuss web-sites that are related to a particular week's topic are strongly encouraged to do so, upon discussion with the instructor. Class attendance is highly advisable; please note also that attendance in this course is synonymous with *verbal participation* in class discussions.

Six times over the period of the course, you will be required to individually write an 800 word *essay* and an accompanying *self-assessment* based on the assigned readings – which together contribute up to 65% of your grade for this course. These logs can take the form of an opinion piece, wherein you should be able to thoughtfully *discuss* and *extend* some key concepts and notions covered in the readings. Logs may also take the form of a description of a CAI-related construct *applied* to real-life, with sound *connections* and *links* made to an *underlying theory*. You will have access to a neo-corpus, web-based environment to prepare your essays; this environment contains anonymous, annotated logs from previous learners who have taken this course.

The second purpose of this course is to provide graduate learners in Educational Technology an opportunity to individually experience the role of a *developer* in the generation of an instructional episode or course. Courses will be storyboarded using PowerPoint™ . All laboratory sessions will be facilitated by the instructor and will focus on the development of the technical skills necessary to adeptly develop storyboards in PowerPoint™ . All laboratory work will be completed individually. Successful completion of the storyboard will be rewarded with up to 15% of the course grade. The content of your courses will be determined by the instructor. As such, the instructor will act as the client (or subject matter expert), and will be available for an initial interview session as well as one review session to finalise your storyboards. These interview and review sessions will be held during class and/or lab time.

The third purpose of this course is to engage graduate learners in Educational Technology in the art of conducting usability tests for web-based instructional environments. As stated earlier, you will have access to a web-based, neo-corpus environment to help you prepare for the essay assignment. The instructor will conduct an individual interview with you for the purposes of testing the usability of the abovementioned environment. You will use these interviews to individually prepare and write up a usability report.

In summary, each learner in the class will prepare the following individual assignments – there are no group assignments for this course:

<u>Essays & Self-Assessment (65%, due BEFORE scheduled class time)</u>:

- Essays: 6 total, each worth 10% of the total grade for the course – Total 60%
    - Essays may be written as opinion pieces, extending and discussing key concepts and issues illuminated in the weekly readings, backed up with solid, logical arguments. Essays may also be written as applications of theory presented in the readings to real-life applications with sound connections made to the underlying theory being discussed. Essays may also be used as a platform to pursue an issue over the course of the seven week duration of the course; that is to say, you may write your essays on the same general topic, but use each of your six essays to eke out a better understanding of your topic using a fresh perspective. You will avail of a web-based neo-corpus environment to help prepare your essays. The content of the essays can span several weeks' worth of readings. Each essay must be written within 550 (minimum) to 800 (maximum) words.
- Self-assessments: Total 5% for completing 6 self-assessments
    - Each essay is accompanied by a self-assessment of (a) how you met the criteria for writing the essay (maximum 100 words), (b) a prediction of your performance on the essay, and (c) how confident you are of your prediction. This self-assessment is used to help you keep track of your performance on the logs over the period of the course. The instructor will

ensure that he will only be reading and providing weekly feedback to your assessment of how you met the criteria for writing the essay. The performance and confidence predictions will not be made available to him so as to ensure that your grade is not influenced by the self-assessment you make. These predictions are intended to help you better keep track of your performance from one essay to the next; it is to your distinct advantage to thoughtfully complete these self-assessments as they will serve as a running record of how well you can gauge your performance at writing logs of this nature. You will receive 5% of the grade for your course for completing these self-assessments; they will NOT be graded. The instructor will provide you with details of how to submit your weekly logs and self-assessments online.

<u>Development of storyboard for e-learning course (total of 15% due July 1, 2007)</u>

- Storyboards must be created using PowerPoint™ and must be conducted under the supervision of the instructor. You will be assessed on your ability to use principles of e-learning-based instructional design in creating distinct pieces of training material. Examples and instruction will be provided during the lab sessions.

<u>Usability report (total of 20% due July 1, 2007)</u>

- The usability assignment will focus on the web-based environment you will be using to read annotated versions of previous learners' logs in preparation for your own essays. Usability interviews will be conducted individually with the instructor during laboratory sessions. The final usability report will include your impressions of the environment, its ability to live up to its objectives as well as suggestions for improvement.

The following pages detail the *reading list*, which are mostly in the form of PDF and MHTM (archived HTML) files; these are *all* available on the FirstClass® course folder. Learners are required to download digital copies of the readings from FirstClass® and are responsible for making one copy of each for personal use.

PLAGIARISM

The instructor takes a serious stance towards learners who insist on plagiarizing in writing their opinion pieces. All instances will be immediately reported to the University, as per regulations. Please see the following document to understand what constitutes plagiarism:

http://secretariat.concordia.ca/policies/academic/en/Code%20of%20Conduct-Academic.pdf

Also, please see the following site for an overview of how to avoid plagiarism:

http://cdev.concordia.ca/CnD/learnerlearn/Help/handouts/WritingHO/AvoidingPlagiarism.html

Grading Scheme for Essays

A+, A, A-	**B+**
With variations to account for the three grades • EXCEPTIONAL in all respects • contains original creative thought • very well organized and expressed • sound critical evaluation skills • clear command of techniques and principles of the discipline • consistently exceeds expectations • high level of synthesis, often across sources • new understandings and hypotheses explained clearly • extension of course content and true abstraction of content to real applications	• EXCELLENT • well organized with few errors • shows clear understanding of concepts • evidence of critical thought • ability to discriminate & interpret issues • analytic treatment of content • application of ideas • synthesis—connections among disparate details or ideas • manipulation and interpretation of data • near perfect abstraction of content to real applications
B	**B-**
• VERY GOOD • meets extension of ideas and	• ADEQUATE to GOOD • constitutes baseline for graduate work

discussion criteria for assignment, but fails to go far beyond • shows basic competence in synthesis • critical thinking • logically organized	• shows comprehension of course content • coherent, understandable • descriptive treatment of content • contains key elements, basic facts/knowledge • little extension, abstraction or integration of concepts
C WEAK minimally meets requirements, just passes	

- Each learning essay will be graded as follows:
  - 10% for choice of issues explored in essay
  - 20% for opinion expressed, real-life application used, hypothesis or theory espoused, or how new understandings are presented
  - 50% for validity of writer's opinion, acceptable linkage to theory and extent to which argument is logical
  - 20% for overall quality of essay, grammatical correctness, and fluidity in language

Reading List

Note that all readings are available in the online course conference

E-learning Overview

Clear (2002). E-learning: A vehicle for e-transformation or Trojan Horse for enterprise? : Revisiting the role of public higher education institutions. *IJEL*

Landis et al (2002). An e-learning manifesto. *IJEL*

Trentin (2002). From distance education to virtual communities of practice: The wide range of possibilities for using the Internet in continuous education and training. *IJEL*

Wesley (2002). A critical analysis on the evolution of e-learning. *IJEL*

Greenagel (2002). The illusion of e-learning: Why we are missing out on the promise of technology. http://www.league.org/publication/whitepapers/0802.html

Online Learning Environment (OLEs) and Instructional Design (ID)

Bishop & Cates (2001). Theoretical foundations for sound's use in multimedia instruction to enhance learning. *ETR&D.*

Helic et al. (2004). Delivering relevant training objects to personal desktop with modern WBT-systems. *IJEL*

Jegan & Eswaran (2004). Patterns for e-learning content development. *JILR*

Kinshuk et al. (2004). Adaptivity through the use of agents in web-based student modeling. *IJEL*

Pahl (2004). Data mining for the evaluation of learning content interaction. *IJEL*

Clark (2002). Six principles of effective e-learning: What works and why. *ELDJ*

Cyrs (1997). Competence in teaching at a distance. *New Directions for Teaching and Learning*

Horton & Horton (2002). Bring top classroom features online-no more boredom. *ELDJ*

Kindley (2002). The power of simulation-based e-learning (SIMBEL). *ELDJ*

Longmire, Tuso & Wagner (2000). Learning without limits vol. 3 (special issue on competency modeling and ID)

Marks (2002). Improving online sales education: Learning styles and streaming media. *ELDJ*

Koohang & Du Plessis (2004). Architecting usability properties in the e-learning instructional design process. *ELDJ*

Hill et al. (2004). Exploring research on internet-based learning: From infrastructure to interactions. From *The Handbook of Educational Technology*

Koohang (2004). A study of users' perceptions toward e-learning courseware usability. *IJEL*

Moallem (2003). An interactive online course: A collaborative design model. *ETR&D*

Nadolski et al. (2001). A model for optimizing step size of learning tasks in competency-based multimedia practicals. *ETR&D*

Sautter et al. (2004). Assessing students in online courses. *IJEL*

Willging (2003). A model for the development of online instruction. *IJEL*

Williams (2004). Teachnology: Web-based instruction's dual environment. *IJEL*

Hobbs (2002). A constructivist approach to web course design: A review of the literature. *IJEL*

Hirumi (2002). The design and sequencing of e-learning interactions: A grounded approach. *IJEL*

Christiansen & Anderson (2004). Feasibility of course development based on learning objects. http://www.itdl.org/Journal/Mar_04/article02.htm

Muirhead (2004a). Research insights into interactivity. http://www.itdl.org/Journal/Mar_04/article05.htm

Muirhead (2004b). Encouraging interaction in online classes. http://itdl.org/Journal/Jun_04/article07.htm

Santally & Senteni (2004). A cognitive approach to evaluating web-based distance learning environments. http://www.itdl.org/journal/Feb_04/article04.htm

Orrill (2002). Learning objects to support inquiry-based, online learning. From *The Instructional Use of Learning Objects*

Martinez (2002). Designing learning objects to personalize learning. From *The Instructional Use of Learning Objects*

Athabasca University (2004). *Theory and Practice of Online Learning*. Online edition

Venkatesh & Shaw (2003). Raising the bar. *Knowledge Management*

E-learning Quality Assurance (QA)

Quilter (2004). Quality assurance for online teaching in higher education. *IJEL*

Lesniak (2002). Putting it to the test: Quality control for e-learning courses. *ELDJ*

Learning Management Systems (LMSs) and Learning Content Management Systems (LCMSs)

Collis & Strijker (2002). New pedagogies and re-usable learning objects: Toward a different role for an LMS. *AACE*

De Pietro & Appratto (2004). Advanced technologies for contents sharing, exchanging, and searching in e-learning systems. *IJEL*

Garvin (2004). Introduction to SAIDE, a Simplified Authoring for Instructors using Distance Education. *AACE*

Graesser et al. (2004). Modules and information retrieval facilities of the Human Use Regulatory Affairs Advisor (HURAA). *IJEL*

Guralnick (2004). Tools for large scale e-learning development. *AACE*

Kunz (2004). The next generation of learning management system: Requirements from a constructivist perspective. *AACE*

Lee & Geller (2003). A collaborative and shareable web-based learning system (COLS). *IJEL*

McGee (2004). Cognitive tools in web-based learning environments: Implications for design and practice. *AACE*

Sheridan et al. (2004). Detecting cheaters using a learning management system. *AACE*

Shaw & Hudson (2004). Distributed content models – A new innovation in learning content management systems and strategies. *AACE*

Shaw (2004). *Dynamic content: Connecting performance and learning*. White paper from www.eedo.com

Shaw & Venkatesh (2005). *The missing link to enhanced course management systems: Adopting learning content management systems in the educational sphere.*

Szabo (2002). CMI theory and practice: Historical roots of learning management systems. *AACE*

Tantono (2004). A web-based framework for e-learning: A model for online presentation authoring. *AACE*

Venkatesh et al. (in press, 2007). Topic Maps: Adopting user-centered indexing technologies in course management systems. *Journal of Interactive Learning Research.*

Learning Objects

Hannafin et al. (2002). Designing resource-based learning and performance support systems. From *The Instructional Use of Learning Objects.*

Najjar (2004). User behavior in learning object repositories: An empirical analysis. *AACE*

Collis & Strijker (2003). Re-usable learning objects in context. *IJEL*

Downes (2004). The Buntine Oration: Learning Objects

http://itdl.org/Journal/Nov_04/invited01.htm

Downes (2004). Editorial: The Rise of Learning Objects.

http://www.itdl.org/Journal/Mar_04/editor.htm

Wiley (2002). Learning Objects: Difficulties and Opportunities. From

http://wiley.ed.usu.edu/docs/lo_do.pdf

Hodgins, (2002). The future of learning objects.

Hill & Hannafin (2001). Teaching and learning in digital environments: The resurgence of resource-based learning. *ETR&D*

Jonassen & Churchill (2004). Is there a learning orientation in learning objects? *IJEL*

Bannan-Ritland et al. (2002). Learning object systems as constructivist learning environments: Related assumptions, theories and applications. From *The Instructional Use of Learning Objects*

McGreal et al. (2004). EduSource: Canada's learning object repository network. *ITDL*

McGreal (2004). Learning objects: A Practical definition. *ITDL*

Parrish (2004). The trouble with learning objects. *ETR&D*

Polsani (2003). Use and abuse of reusable learning objects. *Journal of Digital Information 3,* (4). http://jodi.ecs.soton.ac.uk/Articles/v03/i04/Polsani/

Shi et al. (2004). Open learning objects as an intelligent way of organizing educational material. *IJEL*

Verbert & Duval (2004). Towards a global component architecture for learning objects: A comparative analysis of learning object content models. *AACE*

Wiley (2002). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. From *The Instructional Use of Learning Objects*

Williams (2002). Evaluation of learning objects and instruction using learning objects. From *The Instructional Use of Learning Objects*

Littlejohn & Buckingham Shum (Eds., 2003). Reviewing online resources (special issue) *Journal of Interactive Media in Education*.

Metadata & Standards

CETIS – www.cetis.ac.uk

Barker (2004). Adopting SCORM 1.2 Standards in a Courseware Production Environment. *IJEL*

Bental (2004). Generating user-tailored descriptions of online educational resources. *IJEL*

Duval (2001). Standardized metadata for education: A status report. *AACE*

Duval & Hodgins (2003). A LOM Research Agenda. Paper presented in WWW2003 - Twelfth International World Wide Web Conference, 20-24 May 2003, Budapest, Hungary.

Lytras et al. (2002): Dynamic e-learning settings through advanced semantics: The value justification of a knowledge management oriented metadata schema. *IJEL*

Recker et al. (2002). Collaboratively filtering learning objects. From *The Instructional Use of Learning Objects*

Yahya et al. (2002). The development of the learning object standard using a pedagogic approach: A comparative study. *AACE*

Schatz, S. (2002). Paradigm Shifts and Challenges for Instructional Designers: An Introduction to Meta Tags and Knowledge Bits. *IMS*

Competency Models

Mirabile (1997). Everything you wanted to know about competency modeling. *Training & Development*

McLagan (1997). Competencies: The next generation. *Training & Development*

Zemke & Zemke (1999). Putting competencies to work. *Training*

Gottfredson (2002). Rapid Task analysis™ - The key to developing competency-based e-learning. *ELDJ*

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 of the Department of Education at Concordia University.

A. PURPOSE

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

1) To explore how graduate learners use a web-based environment designed to help in completing and better performing at an essay task.
2) To explore the changing interpretations of graduate learners' understanding of an essay task over the period of a course.
3) To provide opportunities for learners to evaluate their own learning and assess their academic performance.
4) To test and evaluate a self-assessment tool used in conjunction with instructional approaches aimed at helping graduate learners (a) better understand the essay task (b) better evaluate their own performance, and ultimately (c) improve their academic performance in authentic learning environments.

B. PROCEDURES

- Participation in this research does not involve any additional work than the course-work assigned to you in the class, as described in the course outline provided by Mr. Venkatesh. All learners, 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. No additional time is required of you for this research project. All learners, regardless of whether the participate in the research are expected to devote time to preparing for each class, writing essays and conducting self-assessments, as well as preparing for the course-related usability project.

- You will be required to complete class readings, write one essay, and complete self-assessments for the essay on a weekly basis. You will avail of a web-based environment which will enable you navigate previous learners' essays. The research project you are consenting to participate in is concerned with your understanding of and evaluation of performance in the essay writing assignment.
- Mr. Venkatesh will not be aware of who has consented to participating in the research until after the final submission of grades. This means that if you consent to participating in this project, your materials (i.e., the essays, self-assessments, Mr. Venkatesh's assessment of your essays, usability interviews, tracking logfiles) will be made available to Mr. Venkatesh only AFTER the final submission of grades of the session. All consent forms will be sealed and handed to Gretchen Lowerison, a doctoral candidate in the Educational Technology program and will only be opened after the session is completed and grades submitted.
- All reporting of results will remain confidential, that is, only Mr. Venkatesh 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.

- If you have any questions or concerns about the research, please direct your enquiries to Mr. Venkatesh before signing this consent form. If you have concerns after signing this form, please see Gretchen Lowerison (g_lowerison@education.concordia.ca), so that she may relay your queries anonymously to Mr. Venkatesh. You may also contact Mr. Venkatesh's supervisor, Dr. Steven Shaw (shaws@vax2.concordia.ca) to address any concerns you might have with the course.

C. CONDITIONS OF PARTICIPATION

- I consent to providing access to my usability interviews, logfiles tracking my use of the web-based environment, my weekly essays, self-assessments, and Mr. Venkatesh's assessments of my essays for the purposes of the research.

- I understand that ALL course assignments are a compulsory aspect of the course, regardless of my decision to participate or not participate in Mr. Venkatesh's research project.

- I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences.*

- I understand that my participation in this study is CONFIDENTIAL (i.e., the researcher, Mr. Venkatesh, 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 ____

* If you would like to withdraw consent, please contact Gretchen Lowerison or Dr. Steven Shaw. You must NOT inform Mr. Venkatesh of your decision to

discontinue participation, as he will be unaware of the identity of any of the participants until after the final grades for the Fall 2006 are submitted.

Contact information

Vivek Venkatesh

vivek.venkatesh@education.concordia.ca

(514) 848-2424 ext 8936; (514) 739-9067; (514) 992-0225

Gretchen Lowerison

glowerison@education.concordia.ca

Steven Shaw, Ph. D.

shaws@vax2.concordia.ca

(514) 848-2424 ext. 2044

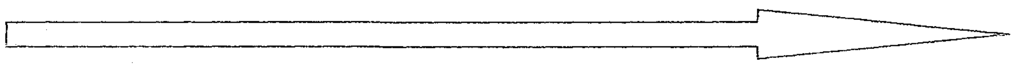
Appendix B – Task Analyzer and Performance Evaluator

Welcome to the assessment of performance and confidence for your log

Make sure you complete this assessment AFTER having completed your log and the accompanying self-assessment of meeting the evaluation criteria for the essay

Q1. How many <u>marks</u> do you think the instructor will award you for your log?

Minimum Maximum

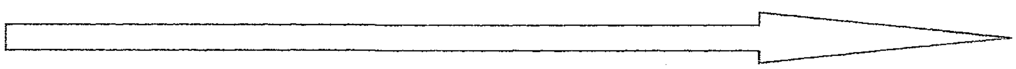


0 10 20 30 40 50 60 70 80 90 100

A1. Your prediction of marks:

Q2. How <u>confident</u> are you that you will receive the marks you predicted above in question 1?

Minimum confidence Maximum confidence



0 10 20 30 40 50 60 70 80 90 100

A2. Your prediction of confidence:

Appendix C – Pre-test

1. Describe, in your own words, the term "e-learning"?

2. What is the meaning of the term "metadata"?

3. What do you understand by the term "blended learning"?

4. What does the acronym LCMS stand for? What does the acronym CMS stand for? What does LMS stand for? What are the major differences these three technologies?

5. Do you know what SCORM is?

6. What is the difference between asynchronous and synchronous online communication? Could you provide examples of both these types of communication?

7. What is a competency model?

8. Imagine you are conducting a usability test of an online course? What types of questions would you ask participants during such a test – list them all.

9. What is the meaning of the term "scenario-based e-learning"?

10. What is the meaning of the term "simulation-based e-learning?"

11. Please write a short essay describing how you would convert the Learning Theories course offered here in the Educational Technology program to an e-learning course. If you have not taken the Learning Theories course as yet, you may choose another course you have taken as a basis for your discussion. Your essay should be between 550 and 800 words. Your essay may be written as an opinion piece, extending and discussing key concepts and issues related to the topic of e-learning. Essays may also be written as applications of a theory to real-life applications with sound connections made to the underlying theory being discussed. No references are needed for this essay.

Appendix D – Descriptive Statistics

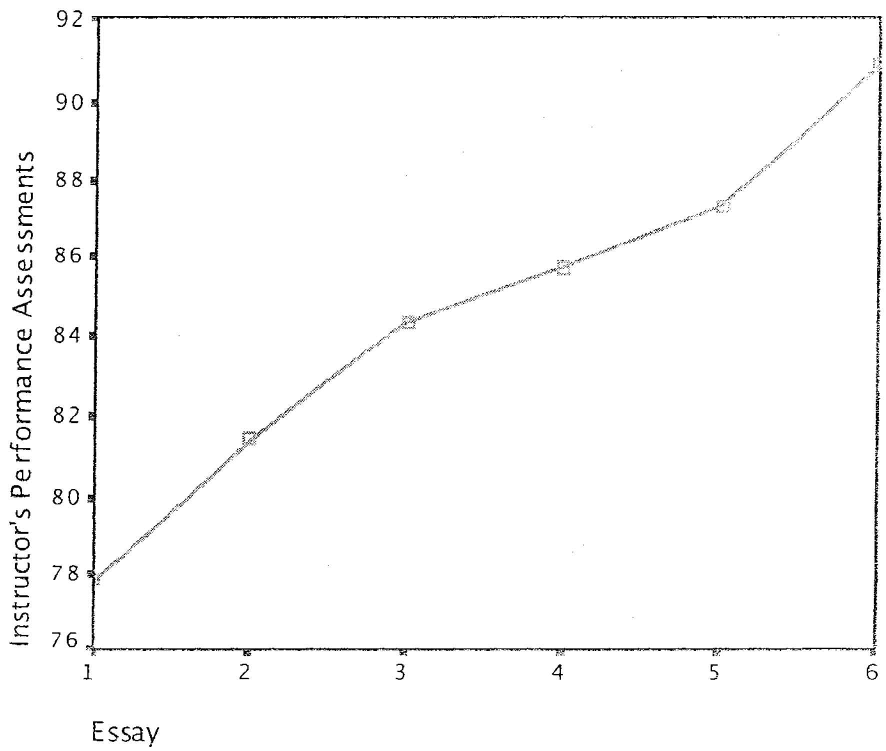
Descriptive statistics for Instructor's Performance Assessments, Students' Performance Predictions, Prediction Confidence, Bias, Discrimination and Absolute Accuracy

Essay	Performance Assessments (maximum score: 100)		Performance Predictions (maximum score: 100)		Prediction Confidence (maximum score: 100)		Discrimination (range: -100 to 100)		Bias (range: -100 to 100)		Absolute Accuracy (range: 0 to 100)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	77.84	7.55	-	-	-	-	-	-	-	-	-	-
2	81.39	6.75	81.24	5.86	79.21	12.53	-52.13	61.28	-12.07	13.50	7.58	5.60
3	84.32	7.92	80.47	6.38	74.03	17.84	-42.41	42.92	-16.63	19.93	8.42	7.37
4	85.68	9.46	83.08	6.30	79.47	10.77	-31.52	48.47	-11.47	12.63	8.18	5.65
5	87.29	7.81	84.03	6.73	80.20	11.61	-35.52	50.73	-10.64	11.41	7.58	5.27
6	90.84	7.47	83.79	5.79	81.50	9.91	-29.52	51.22	-9.52	10.54	8.47	7.51

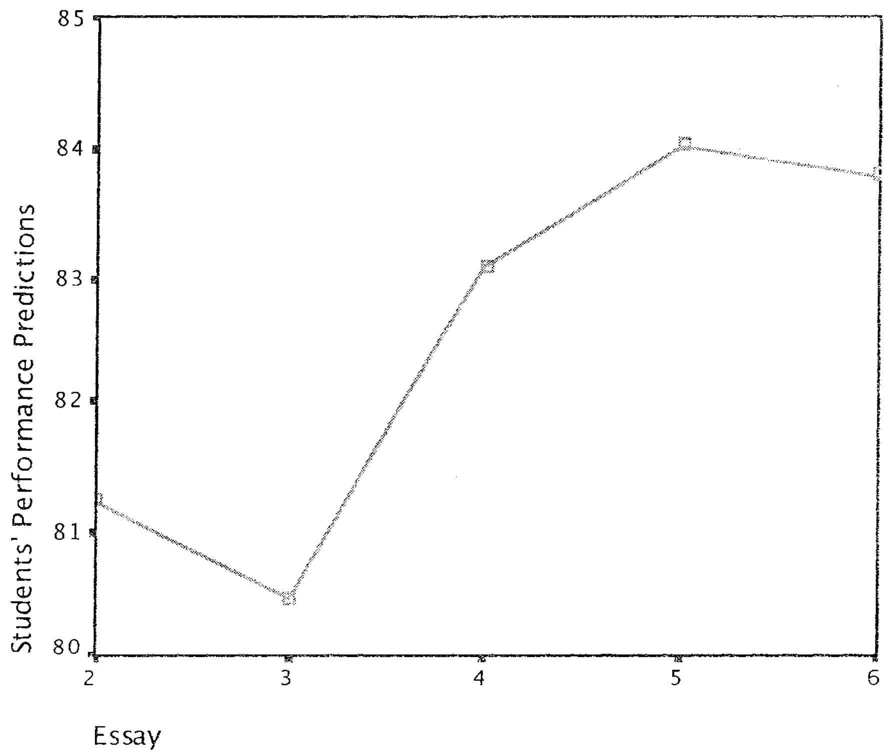
Note. *n* for all essays was 38; empty cell (indicated with a dash) represents that data was not collected/calculated for that particular essay

Appendix E – Figures Describing Means of Collected and Calculated Measures of Performance and Monitoring

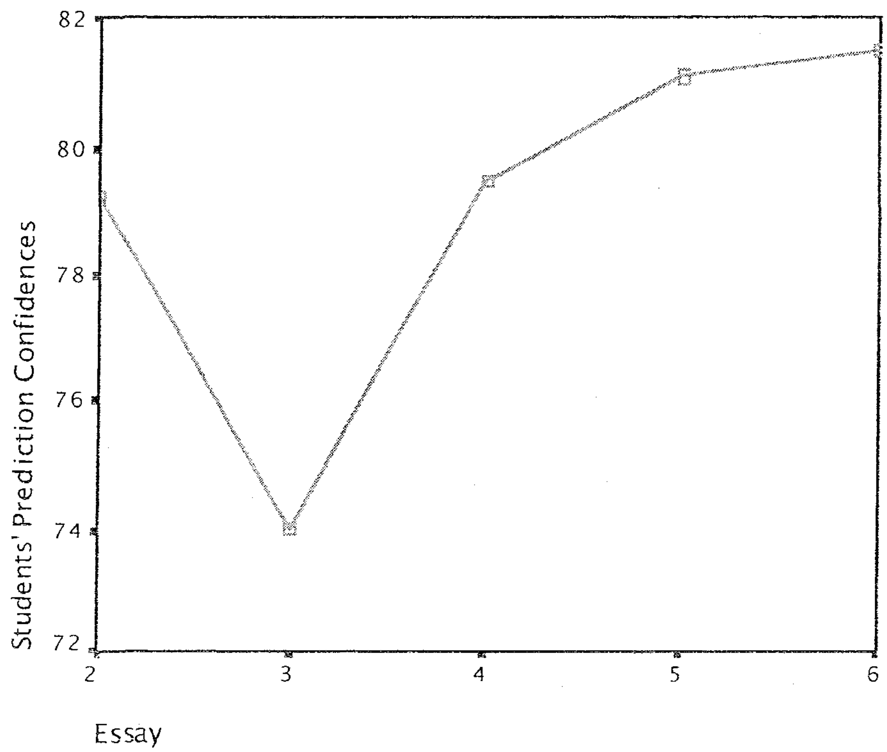
Mean Scores of Instructor's Performance Assessments Across Essays (n=38)



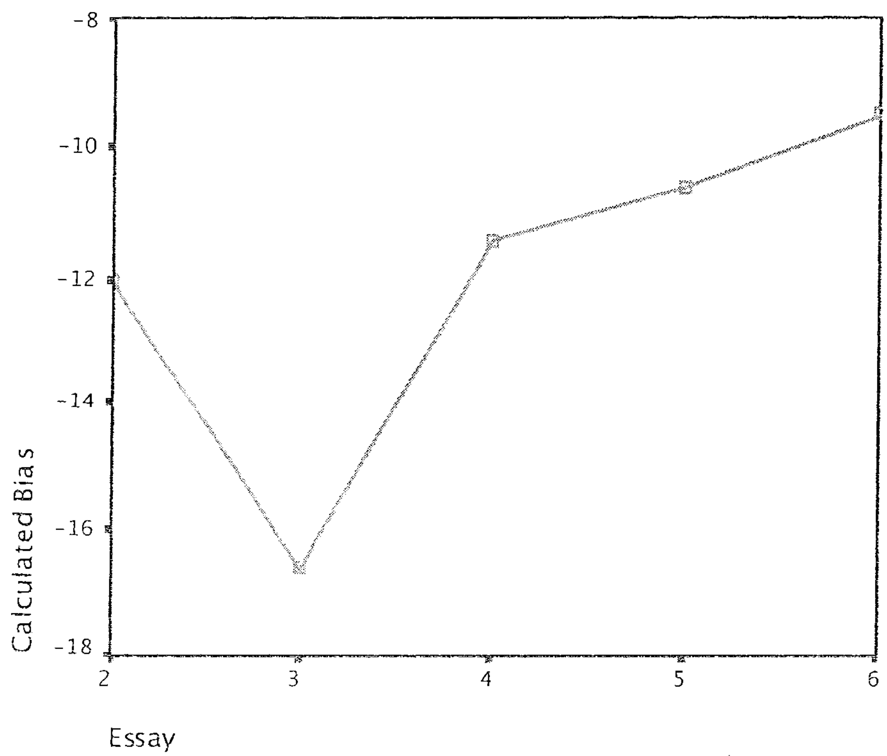
Mean Scores of Students' Performance Predictions Across Essays (*n*=38)



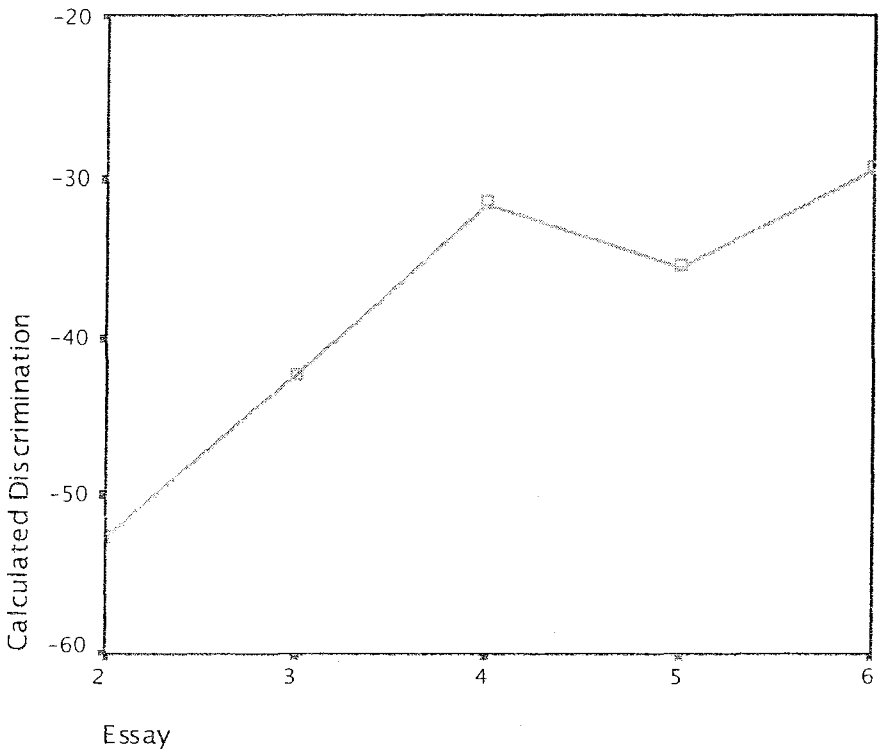
Mean Scores of Students' Prediction Confidences Across Essays (*n*=38)



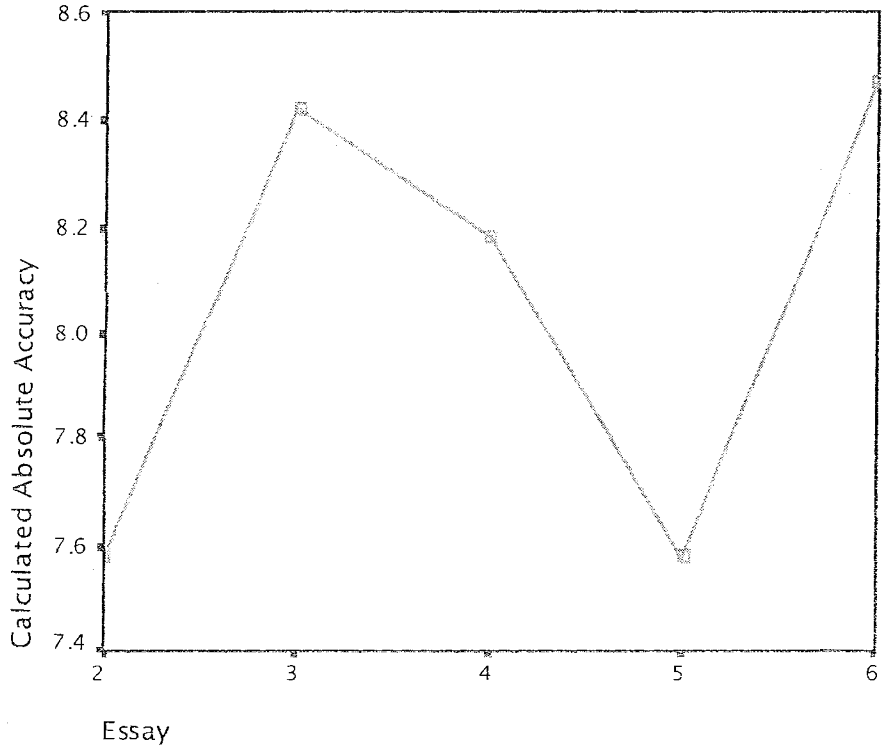
Mean Scores of Calculated Bias Across Essays (n=38)



Mean Scores of Calculated Discrimination Across Essays (*n*=38)



Mean Scores of Calculated Absolute Accuracy Across Essays (n=38)



Appendix F – Distribution of Task Understanding Measure Across Essays

Frequency of distribution of instructor's judgments of learners' self assessments

Essay	Instructor's Judgment of Learners' Self Assessment of Meeting Criteria for Essay Assignment		
	Incorrect Justification	Partially Correct/Incomplete Justification	Correct Justification
1	11	13	15
2	11	14	14
3	12	16	11
4	7	19	13
5	7	18	13
6	8	13	17

Note: n for all essays 1 through 4 = 39; n for essays 5 and 6 = 38