Research Article

Uncovering Relationships between Task Understanding and Monitoring Proficiencies in Postsecondary Learners: Comparing Work Task and Learner as Statistical Units of Analyses

Vivek Venkatesh and Kamran Shaikh

Department of Education, Centre for the Study of Learning and Performance & Learning for Life Centre, Concordia University, 1455 de Maisonneuve West, Montreal, Quebec, Canada H3G 1M8

Correspondence should be addressed to Vivek Venkatesh, vivek.venkatesh@education.concordia.ca

Educational psychologists have researched the generality and specificity of metacognitive monitoring in the context of college-level multiple-choice tests, but fairly little is known as to how learners monitor their performance on more complex academic tasks. Even lesser is known about how monitoring proficiencies such as discrimination and bias might be related to key self-regulatory processes associated with task understanding. This quantitative study explores the relationship between monitoring proficiencies and task understanding in 39 adult learners tackling ill-structured writing tasks for a graduate "theories of e-learning" course. Using learner as unit of analysis, the generality of monitoring is confirmed through intra-measure correlation analyses while facets of its specificity stand out due to the absence of inter-measure correlations. Unsurprisingly, learner-based correlational and repeated measures analyses did not reveal how monitoring proficiencies and task understanding might be related. However, using essay as unit of analysis, ordinal and multinomial regressions reveal how monitoring influences different levels of task understanding. Results are interpreted not only in light of novel procedures undertaken in calculating performance prediction capability but also in the application of essay-based, intra-sample statistical analysis that reveal heretofore unseen relationships between academic self-regulatory constructs.

1. Introduction

Research on academic writing and self-regulatory processes has focused on the empirically-validate strategies for instructional development of cognitive and metacognitive behaviors in a variety of learners (e.g., Bereiter and Scardamalia [1]; Bruning and Horn [2]; Graham and Harris [3]; Harris and Graham [4]; Langer [5]; Zimmerman and Risemberg [6]). However, there is a paucity of work that focuses on a fine-grained analysis of how individual components of self-regulation interact and, possibly, influence one another, especially in academic writing tasks at the postsecondary level (Venkatesh and Shaikh [7]). Essay-writing is considered to be the "default genre" for measurement of understanding, and dare we say, higher-order cognitive processing in higher education in developed nations (Andrews [8]). It would be

to our distinct advantage to continue to develop higher-order cognitive skills in postsecondary learners, given that think tanks such as the Canadian Council on Learning [9] has recently advocated that graduates from postsecondary institutions not only learn to adapt to the shifting landscape of the job market in an increasingly international context, but also to innovate, create, and transfer knowledge on their jobs. Such a tall order would necessitate that our future workforce learn to be creative as well as self-regulated, and thereby transfer their postsecondary skills to their jobs (Simard et al. [10]). Ill-structured essay-writing has been empirically demonstrated as an activity that stimulates advanced cognitive processing (Andrews [8]; Lindblom-Ylänne and Pihlajamäki [11]) and self-regulation (Tynjälä [12]; Venkatesh and Shaikh [7]) in postsecondary learners. The empirical research reported in this paper explores

the development of self-regulatory processes, specifically monitoring and task understanding, in graduate learners as they engage in essay-writing activities.

2. Theoretical Overview: Pinning Down Monitoring and Task Understanding as Components of Academic Self-Regulation

2.1. Monitoring as a Component of Self-Regulation. The work reported in this paper 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, taking into account their emotions as well as motivational states within a specific learning context or environment (Pintrich [13]; Winne and Hadwin [14]; Zimmerman [15, 16]). A critical component of academic SRL is *monitoring*, or learners' abilities to evaluate their performance and learning while engaging in an academic task (Dunlosky and Bjork [17]; Dunlosky and Metcalfe [18]). While monitoring has been described as an eccentric phenomenon, with variations from one individual to the next (Nietfeld et al. [19]; Schraw et al. [20]), research on monitoring or calibration proficiencies in college students taking multiple-choice tests has revealed both domain-specific and domain-general monitoring abilities in students (Nietfeld et al. [21]; Schraw [22]; Schraw and Nietfeld [23]). There is, however, a paucity of research on the development of monitoring skills in graduate learners in the context of writing tasks requiring higher-order thinking (Venkatesh and Shaikh [7]), as well as whether adults use their monitoring skills in similar ways when tackling different types of academic activities (Nietfeld et al. [24]; Nietfeld and Schraw [25]; Winne [26]). The present study explores the development of graduate learners' monitoring proficiencies (i.e., how well or poorly graduate learners monitor their learning) while tackling ill-structured writing tasks. Such research is necessary in order to further develop the theoretical notion of monitoring in different academic contexts, as well as to explore whether monitoring in graduate learners can be characterised as task-general or content-specific.

2.2. Task Understanding. Another critical phase in self-regulation, *task understanding*, (Winne and Hadwin [27]) is influenced not only by learners' perceptions of the nature and assessment criteria related to an academic task, but also by their "knowledge of self-as-learner", which includes prior knowledge, as well as individual motivational, affective, and emotional states. The cyclical nature of self-regulation demands that learners revisit and redefine their task understanding over the time spent engaging with an academic activity (Venkatesh and Shaikh [7]). Unfortunately, there is a lack of research on how to improve learners' task understanding during their engagement in ill-structured academic tasks (see Venkatesh et al. [28] for a detailed explanation of task understanding). Specifically, there exists a pressing need to investigate how to better attune learners' perceptions of the assessment criteria for

a task to match those of the instructor (Perry et al. [29]; Venkatesh and Shaikh [7]; Venkatesh et al. [28]). The research reported herein attempts to bridge some of the gaps in the understanding of adult learners' monitoring skills while engaging in tasks that require higher-order processing skills. In addition, this research will provide much-needed empirical support for employing instructional interventions in ameliorating the understanding of assessment criteria for writing assignments that demand discussions of practicality, yet can be characterized as recondite. For the purposes of this paper, task understanding refers to learners' perceptions of the assessment criteria for an academic task.

3. Exploring the Relationship between Monitoring, Task Understanding, and Performance

There is a paucity of empirical evidence of the relationship between learners' accuracy in monitoring and their performance. Studies supporting the relationships between accuracy of monitoring and performance are few and far between (Nietfeld et al. [21, 24]; Pressley and Schneider [30]). In fact, there are instances of studies that point out the contrary, that is, that improved performance in test-taking situations is related to less accurate monitoring (e.g., Begg et al. [31]), or that improved performance cannot be attributed to improved monitoring (e.g., Dunlosky and Connor [32]). An additional point of concern is drawn by Maki [33], who, in reviewing instructional attempts to improve monitoring accuracy, reveals that research efforts have produced less than stellar results.

In prior research, Thiede et al. [34], as well as Thiede and Anderson [35], have 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. Their work demonstrates 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. More recent work by Thiede et al. [36] has demonstrated that providing an increased amount of time for learners to reflect on their comprehension on a piece 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 future 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 and his colleagues' studies.

In classroom-based investigations, when using repeated measures analyses across instantiations of a task, intermeasure correlations between psychological constructs (such as monitoring and task understanding) generally reveal insignificant relationships due to a lack of statistical power (Shaffer and Serlin [37]). Interestingly, a prior qualitative investigation (Venkatesh and Shaikh [7]) using interviews with graduate volunteers who were writing essays over the course of a semester's worth of instruction revealed a distinct essay-specific relationship between perceptions of assessment criteria (an element of task understanding) 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 a self-assessment tool led to a more comprehensive understanding of the complex essay-writing task they were assigned. In the present study, this cryptic 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.

4. 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 repeated instantiations of an ill-structured writing task; (b) to shed light on the task-specificity and/or content-specificity of these learners' monitoring skills using inter-and intrameasure correlational procedures; (c) to explore the statistical relationship, if any, between graduate 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) to discuss the theoretical and methodological implications of investigating monitoring (or calibration) of performance and task understanding in the context of graduate essay-writing assignments.

5. 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 first 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, 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 neocorpus of learner essays written for 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' [38, 39] Structured Objectives for Learning Outcomes (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. 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 [40]). Each essay was accompanied by the self-assessment Task Analyzer and Performance Evaluator (TAPE) tool (see Appendix A), designed using Maki's [41] principles of discrepancy reduction, with the objective of (a) helping students articulate, in written form, their justifications for meeting the assessment criteria, and (b) eliciting 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 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 assessment criteria. Consent forms 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 the first author as well as the need to explore self-regulatory constructs in naturalistic settings, students' 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 feedback, grades, and measures of performance prediction and confidence in predictions were stored electronically in a password-protected hard drive.

6. Data Sources and Design

Pretest measures of content knowledge specific to the course offered, as well as pretest 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 B for pretest 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 performance assessment (score range: 0 to 100; converted grade range: C to A+) and 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) performance assessment (score range: 0 to 100; converted grade range: C to A+); (b) students' performance predictions (range: 0 to converted grade range: C to A+), (c) students' confidence in predictions (range: 0 to 100), and (d) 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; and (b) bias (range: −100 to 100), which measured the degree to which students were over or under-confident in their predictions.

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 illumination of trends in the measures of performance and monitoring of interest in this study.

6.1. Calculation of Discrimination and Bias. Procedures for calculating discrimination and bias for the present study were different from those described by Schraw [22] insofar as Schraw has never factored the theoretical notion of performance prediction capability into his conceptions of monitoring.

6.1.1. Discrimination. For each essay written, the measures of 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.

6.1.2. 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.

6.2. Work Task as Unit of Analysis. In an attempt to better explicate the relationship between a singular facet of task understanding, namely, 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 explained, in great detail, in Shaffer and Serlin's [37] landmark piece on intrasample statistical analysis (ISSA). In the present study, there is sufficient qualitative evidence (Venkatesh and Shaikh [7]) 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.). Additionally, in recent studies conducted by Thiede and his colleagues [34–36], 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 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 and Shaikh [7] 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 [37]. 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. In addition, while treating an individual essay as unit of

analysis, after taking into account all possible measured factors, including the writer 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, writer of the essay, session in which essay was written, numerical sequence (i.e., essay number 1, 2, 3, 4, 5, 6 or 7), performance assessment, performance prediction, confidence in performance prediction, and the calculated measures of discrimination, bias, and absolute accuracy (i.e., the unsigned difference between the prediction and assessment for each essay).

7. Results

7.1. Pretest Equivalence and Interrater 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. The raters received the essays in the same order as they were submitted to the instructor; the weekly sequence of submission for the course was adhered to as was the sequence in which the four sessions were held. Initial meetings between the raters and the author 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 interrater reliability coefficient, was calculated to be 0.87. All 24 discrepancies in rating were resolved through discussion.

7.2. Descriptive Statistics. Descriptive statistics (see Appendix C for a complete table) for essays 1 through 6 showed that the average performance assessments ranged from 77.84 to 90.84 (range of SDs: 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 SDs: 5.79 to 6.73); (b) their confidence in predictions ranged from 74.03 to 81.50 (range of SDs:

TABLE 1: Intrameasure 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: Intrameasure 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**	—

$^{*} p < .05, ^{**} p < .001.$

9.91 to 17.84); (c) they were more prone to negative discrimination, that is, they assigned higher confidence to inaccurate predictions than accurate ones (range of Ms: −29.52 to −52.13, range of SDs: 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 Ms: −9.52 to −16.63, range of SDs: 10.54 to 19.93); and (e) average absolute accuracy (i.e., the unsigned difference between the performance prediction and performance assessment for each essay) ranged from 7.58 to 8.47 (range of SDs: 5.60 to 7.51).

7.3. Repeated Measure Procedures. Repeated measures analyses were conducted using 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 pretest scores for content knowledge and essay-writing ability as covariates.

The repeated measures analyses revealed that 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. However, performance assessments across the six essays 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. Pairwise comparisons between performance assessments (range of Ms: 77.59 to 90.84, range of SDs: 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.

7.4. Correlational Procedures. Intraitem correlational procedures revealed that the 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 intraitem correlations (see Tables 1 and 2 for the intraitem correlations for the monitoring measures). Accuracy as well as absolute accuracy, on the other hand, showed insignificant relationships across the essays.

Partial intracorrelations between confidence scores across the essays improved when variance explained through correlations between confidence and performance assessments was controlled. On the other hand, partial intracorrelations between performance prediction scores across the essays did not show remarkable differences when variance accounted for through correlations between confidence and performance predictions was 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 intermeasure nonparametric correlations between learners' task understanding (i.e., feedback on student's self-assessment) and each of the monitoring proficiencies produced insignificant results.

8. Results from Using Essay as Unit of Analysis

8.1. Multiple Regression Procedure. When considering essays as unit of analysis, performance assessment was parametrically regressed on the essay-specific measures of 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 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.

8.2. NonParametric Regression Procedures. A nonparametric ordinal regression procedure was used to evaluate the predictors of performance assessment (as a grade). The omnibus

TABLE 3: Results of multinomial regression procedure using 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 $p < .01$.

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 feedback on self-assessment improves from partially correct to completely correct, but it also decreases when confidence or absolute accuracy increases, 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 increases, provided other variables remain constant.

9. Discussion and Educational Significance

9.1. 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, 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 Schraw et al. [20], Schraw and Nietfeld [23], as well as Nietfeld et al. [21].

The results also suggested that learners' prediction confidence scores on any one essay was not necessarily bound to their performance assessments on that essay. 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.

9.2. 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 performance assessments and students' prediction confidence scores. Prior statistical investigations by Schraw and his colleagues [20, 23] did not deal with the notion of students' performance predictions and how these predictions might be related to their actual performance and confidence; these studies 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 most prior studies reviewed, students implicitly predicted perfect performance, and monitoring proficiencies were calculated using performance and confidence scores alone [22]. In the present study, the theoretical notion of performance predictions adds new theoretical and methodological dimensions to measuring monitoring proficiencies. Both the calculated measures of monitoring proficiencies in the present study, 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.

9.3. Performance Prediction, 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.

One reason why both the performance assessments and students' performance predictions did not seem to have an effect on 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 [36] 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 assessment of performance. Put simply, an increase in performance assessment or performance prediction did not necessarily prompt an increase in prediction confidence. The significant intracorrelations between performance prediction measures suggest that performance prediction behaved very differently from the performance assessments. While performance assessments were essay-specific, findings suggest that performance predictions developed as a stable pattern across the essays.

9.4. Discrimination in Predictions. Results suggest that students showed an increased ability with regard 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; that is, 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. One possible reason for the generality of the discrimination measures lies in its' calculation procedures, which take into account (a) the progressive nature of the learning essay task, (b) students' performance predictions, (c) performance assessments, and (d) students' prediction confidence scores. The existence of a pattern of discrimination in students engaged in an 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 (as seen in prior studies conducted by Schraw and his colleagues) 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. 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 performance assessments, 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.

9.5. Bias in Predictions. Results of analyses on bias scores revealed that students were, for the most part, underconfident 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.

9.6. 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. 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, namely, students' abilities to explicitly express how they met the 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, 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 nonparametric 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

TABLE 4: Descriptive statistics for 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.

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.

9.7. Contribution to Theory. Traditional modular theories have viewed cognitive skills as domain-specific (e.g., Glaser and Chi [42]), while information-processing theorists have proposed and found support for the existence of more domain-general skills (e.g., Borkowski and Muthukrishna [43]). Results of studies by Schraw, Nitefeld and their colleagues [20, 21, 23, 24] have supported the existence of both domain-specific and domain-general types of monitoring skills in college learners tackling multiple-choice questions. 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 [44]; Schraw and Impara [45]); 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. In addition, researchers should also investigate the relationship between the newly derived measures of discrimination and bias, and whether these two proficiencies coexist 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 [13]; Thiede and Anderson [35]; Thiede et al. [34, 36]; Winne and Hadwin [14, 27]; Zimmerman [15, 16]). 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.

10. Conclusion

The nonexistence of intraitem correlations between the students' performance scores juxtaposed against the strong intraitem correlations between the monitoring-related measures gives credence to the content and task-generality of monitoring skills. However, the lack of intermeasure 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 we contend 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.

Appendices

A. Task Analyzer and Performance Evaluator (TAPE)

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) In no more than 100 words, explain how you think your essay meets the instructor's assessment criteria.

(Q2) How many <u>marks</u> do you think the instructor will award you for your essay (0 to 100)?

(Q3) How <u>confident</u> are you that you will receive the marks you predicted (0 to 100)?

B. Pretest

(1) Describe, in your own words, the term "e-learning", "metadata", "blended learning"?

(2) What do the acronyms LCMS, CMS, LMS, and SCORM stand for?

(3) What is the difference between asynchronous and synchronous online communication?

(4) 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.

(5) What are "scenario-based e-learning" and "simulation-based e-learning"?

(6) 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. 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 connections made to an underlying theory.

C. Descriptive Statistics

(See Table 4).

References

[1] C. Bereiter and M. Scardamalia, *The Psychology of Written Composition*, Erlbaum, Hillsdale, NJ, USA, 1987.

[2] R. Bruning and C. Horn, "Developing motivation to write," *Educational Psychologist*, vol. 35, no. 1, pp. 25–37, 2000.

[3] S. Graham and K. R. Harris, "Self-regulation and writing: where do we go from here?" *Contemporary Educational Psychology*, vol. 22, pp. 102–114, 1997.

[4] K. R. Harris and S. Graham, *Making the Writing Process Work: Strategies for Composition and Self-Regulation*, Brookline, Cambridge, Mass, USA, 1996.

[5] J. A. Langer, "Beating the odds: teaching middle and high school students to read and write well," *American Educational Research Journal*, vol. 38, no. 4, pp. 837–880, 2001.

[6] B. J. Zimmerman and R. Risemberg, "Caveats and recommendations about self-regulation of writing: a social cognitive rejoinder," *Contemporary Educational Psychology*, vol. 22, no. 1, pp. 115–122, 1997.

[7] V. Venkatesh and K. Shaikh, "Investigating task understanding in online repositories equipped with Topic Map indexes: implications for improving self-regulatory processes in graduate learners," *International Journal of Technologies in Higher Education*, vol. 5, no. 3, pp. 22–35, 2008.

[8] R. Andrews, "The end of the essay?" *Teaching in Higher Education*, vol. 8, no. 1, pp. 117–128, 2003.

[9] Canadian Council on Learning [CCL], "Post-secondary education in Canada: strategies for success," 2009, http://www.ccl-cca.ca/pdfs/PSE/2009/PSE2008_English.pdf.

[10] A. Simard, J. Broome, M. Drury, B. Haddon, B. O'Neil and, and D. Pasho, "Understanding knowledge services at Natural Resources Canada," 2007, http://dsp-psd.pwgsc.gc.ca/collection_2007/nrcan-rncan/M4-45-2006E.pdf.

[11] S. Lindblom-Ylänne and H. Pihlajamäki, "Can a collaborative network environment enhance essay-writing processes?" *British Journal of Educational Technology*, vol. 34, no. 1, pp. 17–30, 2003.

[12] P. Tynjälä, "Writing, learning and the development of expertise in higher education," in *Writing as a Learning Tool: Integrating Theory and Practice*, P. Tynjälä, Ed., pp. 37–56, Kluwer Academic Publishers, Amsterdam, The Netherlands, 2001.

[13] P. R. Pintrich, "The role of goal orientation in self-regulated learning," in *Handbook of Self-Regulation*, M. Boekaerts, P. R. Pintrich, and M. Zeidner, Eds., pp. 451–502, Kluwer Academic Publishers, San Diego, Calif, USA, 2000.

[14] P. H. Winne and A. F. Hadwin, "The weave of motivation and self-regulated learning," in *Motivation and Self-Regulated Learning: Theory, Research, and Applications*, D. H. Schunk and B. J. Zimmerman, Eds., pp. 297–314, Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2008.

[15] B. J. Zimmerman, "Attaining self-regulation: a social cognitive perspective," in *Handbook of Self-Regulation*, M. Boekaerts, P. R. Pintrich, and M. Zeidner, Eds., pp. 13–39, Academic Press, San Diego, Calif, USA, 2000.

[16] B. J. Zimmerman, "Investigating self-regulation and motivation: historical background, methodological developments,

and future prospects," *American Educational Research Journal*, vol. 45, no. 1, pp. 166–183, 2008.

[17] J. Dunlosky and R. A. Bjork, *Handbook of Metamemory and Memory*, Psychology Press, New York, NY, USA, 2008.

[18] J. Dunlosky and J. Metcalfe, *Metacognition*, SAGE, Beverly Hills, Calif, USA, 2009.

[19] J. L. Nietfeld, C. K. Enders, and G. Schraw, "A Monte Carlo comparison of measures of relative and absolute monitoring accuracy," *Educational and Psychological Measurement*, vol. 66, no. 2, pp. 258–271, 2006.

[20] G. Schraw, M. E. Dunkle, L. D. Bendixen, and T. D. Roedel, "Does a general monitoring skill exist?" *Journal of Educational Psychology*, vol. 87, no. 3, pp. 433–444, 1995.

[21] J. L. Nietfeld, L. Cao, and J. W. Osborne, "The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy," *Metacognition and Learning*, vol. 1, no. 2, pp. 159–179, 2006.

[22] G. Schraw, "A conceptual analysis of five measures of metacognitive monitoring," *Metacognition and Learning*, vol. 4, no. 1, pp. 33–45, 2009.

[23] G. Schraw and J. Nietfeld, "A further test of the general monitoring skill hypothesis," *Journal of Educational Psychology*, vol. 90, no. 2, pp. 236–248, 1998.

[24] J. L. Nietfeld, L. Cao, and J. W. Osborne, "Metacognitive monitoring accuracy and student performance in the postsecondary classroom," *The Journal of Experimental Education*, vol. 74, no. 1, pp. 7–28, 2005.

[25] J. L. Nietfeld and G. Schraw, "The effect of knowledge and strategy training on monitoring accuracy," *Journal of Educational Research*, vol. 95, no. 3, pp. 131–142, 2002.

[26] P. H. Winne, "Students' calibration of knowledge and learning processes: implications for designing powerful software learning environments," *International Journal of Educational Research*, vol. 41, no. 6, pp. 466–488, 2004.

[27] P. H. Winne and A. F. Hadwin, "Studying as self-regulated learning," in *Metacognition in Educational Theory and Practice*, D. J. Hacker, J. Dunlosky, and A. C. Graesser, Eds., pp. 277–304, Erlbaum, Malwah, Calif, USA, 1998.

[28] V. Venkatesh, K. Shaikh, and A. Zuberi, "Topic Maps as indexing tools in the educational sphere: theoretical foundations, review of empirical research and future challenges," in *Cognitive Maps*, K. Perusich, Ed., pp. 1–12, In-Tech, Vienna, Austria, 2010, http://www.intechopen.com/articles/show/title/topic-maps-as-indexing-tools-in-the-education-al-sphere-theoretical-foundations-review-of-empirical-r.

[29] N. E. Perry, K. O. VandeKamp, L. K. Mercer, and C. J. Nordby, "Investigating teacher-student interactions that foster self-regulated learning," *Educational Psychologist*, vol. 37, no. 1, pp. 5–15, 2002.

[30] M. Pressley and W. Schneider, *Introduction to Memory Development During Childhood and Adolescence*, Erlbaum, Mahwah, NJ, USA, 1997.

[31] I. M. Begg, L. A. Martin, and D. R. Needham, "Memory monitoring: how useful is self-knowledge about memory?" *European Journal of Cognitive Psychology*, vol. 4, pp. 195–218, 1992.

[32] J. Dunlosky and L. T. Connor, "Age differences in the allocation of study time account for age differences in memory performance," *Memory and Cognition*, vol. 25, no. 5, pp. 691–700, 1997.

[33] R. H. Maki, "Predicting performance on text: delayed versus immediate predictions and tests," *Memory and Cognition*, vol. 26, no. 5, pp. 959–964, 1998.

[34] K. W. Thiede, M. C. M. Anderson, and D. Therriault, "Accuracy of metacognitive monitoring affects learning of texts," *Journal of Educational Psychology*, vol. 95, no. 1, pp. 66–73, 2003.

[35] K. W. Thiede and M. C. M. Anderson, "Summarizing can improve metacomprehension accuracy," *Contemporary Educational Psychology*, vol. 28, no. 2, pp. 129–160, 2003.

[36] K. W. Thiede, J. Dunlpsky, T. D. Griffin, and J. Wiley, "Understanding the delayed-keyword effect on metacomprehension accuracy," *Journal of Experimental Psychology*, vol. 31, no. 6, pp. 1267–1280, 2005.

[37] D. W. Shaffer and R. C. Serlin, "What good are statistics that don't generalize?" *Educational Researcher*, vol. 33, no. 9, pp. 14–25, 2004.

[38] J. B. Biggs, "Student learning in the context of school," in *Teaching for Learning: The View from Cognitive Psychology*, J. B. Biggs, Ed., pp. 7–29, The Australian Council for Educational Research Ltd., Hawthorn, Australia, 1991.

[39] J. B. Biggs, "Enhancing teaching through constructive alignment," *Higher Education*, vol. 32, no. 3, pp. 347–364, 1996.

[40] W. Reitman, *Cognition and Thought*, John Wiley & Sons, New York, NY, USA, 1965.

[41] R. H. Maki, "Accuracy of metacomprehension judgments for questions of varying importance levels," *American Journal of Psychology*, vol. 108, pp. 327–344, 1995.

[42] R. Glaser and M. T. Chi, *The Nature of Expertise*, Erlbaum, Hillsdale, NJ, USA, 1998.

[43] J. Borkowski and N. Muthukrishna, "Moving metacognition into the classroom: 'working models' and effective strategy instruction," in *Promoting Academic Competence and Literacy in School*, M. Pressley, K. Harris, and J. Guthrie, Eds., pp. 477–501, Academic Press, New York, NY, USA, 1992.

[44] A. L. Brown, "Metacognition, executive control, self-regulation, and other more mysterious mechanisms," in *Metacognition, Motivation, and Understanding*, F. Weinert and R. Kluwe, Eds., pp. 65–115, Erlbaum, Hillsdale, NJ, USA, 1987.

[45] G. Schraw and J. C. Impara, *Issues in the Measurement of Metacognition*, Buros Institute of Mental Measurement, Lincoln, Neb, USA, 2000.