FORMATIVE COMPUTER-BASED ASSESSMENTS: THE POTENTIALS AND PITFALLS OF TWO FORMATIVE COMPUTER-BASED ASSESSMENTS USED IN PROFESSIONAL LEARNING PROGRAMS

by

SANDRA T. MILLER

A dissertation submitted to the Faculty of Education

in conformity with the requirements for

the degree of Doctor of Philosophy

Queen's University

Kingston, Ontario, Canada

January, 2009

ABSTRACT

A case study approach is used to examine two formative computer based assessments (CBAs) used in a School of Medicine and Faculty of Education at a Canadian university. Both assessments contained primarily scenario-based, multiple-choice items however the number of items and complexity was different. The formative CBAs were examined in terms of how feedback is provided to students, the effectiveness of different types of feedback, and the characteristics of items used for assessment. Feedback was channelled to students in one case, immediately following responding to an item and in the second case, at the end of a set of items. Feedback given to students immediately following an item was more effective given that the timing captured students' construction of knowledge. The most favoured type of feedback provided additional information. Students also indicated that feedback using a live internet link to direct them to a resource would be acceptable. Although feedback that simply stated a response was correct or incorrect was previously shown to be ineffective, students from the School of Medicine indicated this type of feedback was acceptable for low cognitive items. In both case studies, students reported that more items were required and in one case study, students recommended these items be added at the higher end of the cognitive scale.

ACKNOWLEDGEMENTS

I feel very fortunate to have pursued graduate studies at Queen's University. For this wonderful life-changing opportunity I am truly thankful to all the professors and support staff at the Faculty of Education. For the six years that I have spent at Queen's University working towards a M.Ed. and then my Ph.D. I have come to appreciate and understand the challenges presented in academic inquiry and highly respect and value the contributions each of you have made towards my educational achievements.

I would like to extend a heartfelt expression of gratitude to Dr. Don Klinger and Dr. Lyn Shulha who have supported me throughout the graduate program. The two of you have greatly influenced how I think about studies in education as well as my own cognitive awareness. Sincere thanks to Dr. Denise Stockley and Dr. Lesly Wade-Woolley for their guidance as members of my dissertation committee. I would also like to thank Dr. Robert Wilson for his encouragement in applying to the Ph.D. program; without his advice, I may not have made this journey. A special thanks to Dr. Hugh Munby who helped me understand the process of developing a good argument and painfully reviewing my writing while a committee member for my Master's degree and instructor in the Ph.D. program.

While working on my Ph.D. I had the opportunity to study at the University of Alberta. Sincere thanks to Dr. Todd Rogers and Dr. Mark Gierl for the wonderful opportunity to work with the students at CRAME and hone my knowledge and skills in the field of measurement. Special thanks also go to Dr. Ruth Childs who invited me to work with the students in her lab at OISE during my last year of study. I highly valued

my trips to Toronto every Friday to learn more about IRT and to collaborate with OISE students while conducting research.

The last expression of gratitude goes to my friends and family. To my friends at Queen's: Cheryl Poth (now a faculty member at the University of Alberta), Sue Fostaty-Young , Xiaomei Song, Chris DeLuca, Robin Quantik, and Laura McEwen, thank-you for listening and sharing the trials and tribulations of graduate life. To my friends on Tancook Island, thank-you for always welcoming me back home and reminding me of the simple pleasures of island living. Lastly, to my wonderful family, thank-you for always being a part of this journey.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE

Introduction

The summative assessment paradigm that has driven measurement of student learning for over half a century is slowly giving way to a new paradigm that embraces formative assessment (Stout, 2002). The benefits of formative assessment have been well documented and research indicates that formative assessment practices are associated with improved academic achievement (e.g., Hargreaves, 2005; Hodgen & Marshall, 2005; Wiliam, Lee, Harrison, & Black, 2004). The essential component of formative assessment is feedback that is commonly provided by classroom teachers. As interest in formative assessment has grown, so has experimentation with feedback in new assessment environments, for example, computer-based assessments (CBAs). Formative CBAs embed feedback into on-demand CBAs that have the potential to promote self-regulated learning environments. Formative CBAs go beyond stating whether a response is correct or incorrect by using feedback to clarify learning and direct students to additional learning resources. Although the advantages of formative assessment are well known, the benefits of combining theories of formative assessment with new assessment environments as represented by CBAs are just beginning to be explored (Bell & Crowie, 2001).

Initial research suggests these formative CBAs have been well received by students (Buchanan, 2000, Burrow, Evdorides, Hallam, Freer-Hewish, 2005; Peat & Franklin, 2002;). However, little is known about what aspects of formative CBA students find to be supportive in their learning. It is not known, for example, whether feedback

presented immediately following an item on a formative CBA is more supportive than feedback presented at the end of a set of items. At present, there are no theoretical positions on the manner in which feedback should be presented within the context of formative CBAs. In terms of the actual feedback, little is known about the extent to which students find one type of feedback more supportive than another. For example, how does feedback that indicates whether a response is correct or incorrect compare to feedback that includes additional information to clarify a concept or direct students to a resource? While there is a renewed interest in finding ways to provide more feedback to students, research into the effectiveness of different forms of feedback continues to lag behind the ongoing and evolving practices. Further, contextual factors (e.g., prior technological experience, item construction) can influence the effectiveness of formative CBAs and these influences need to be systematically examined.

Formative CBAs are able to promote self-regulated learning, particularly in higher education (Gipps, 2005). Self-regulated learning environments provide students with a certain amount of freedom to adapt the learning environment to match their own learning goals (Boshuizen, Bromme, & Gruber, 2004). In self-regulated environments, students are empowered to regulate their learning through multiple learning experiences (Kaser, Mundry, Stiles, & Loucks-Horsley, 2006). Thus formative CBAs may be ideal in higher education programs that promote self-regulated learning. In such contexts students can complete formative CBAs when and as frequently as they choose. Some formative CBAs designed in this formative self-regulated framework also allow students to select the cognitive domain they wish to self-assess (Peat & Franklin, 2002), giving students

even more control over their learning. When using formative CBAs to promote self-regulated learning environments it is important to understand how feedback and other factors related to formative CBAs influence students' self-regulated learning. To what extent, for example, are formative CBAs effective instruments in directing students' learning and what factors contribute to using formative CBAs in this context? What are the connections between feedback and CBA in relation to student learning?

In response to the shift in assessment practices from a summative to a formative paradigm (Stout, 2002), formative CBAs have filled a niche in higher education programs that promote self-regulated learning. Research in this field needs to go beyond surveying students' general perceptions of these instruments; the specific factors influencing the effective use of formative CBAs used in these contexts needs to be explored. Prior to any widespread adoption of formative CBAs, the extent to which the formative framework supports student learning when using CBAs in higher education must be explored.

Purpose and Research Questions

The purpose of this dissertation is to examine the use of formative CBAs in higher education and explore the conditions under which formative CBAs might be more effective as instruments to supporting student learning. Specifically, the following research questions focusing on the intentions, structures, and perceived benefits of formative CBAs will be examined:

(a) How is feedback channeled to students and what is the preferred method?

(b) To what extent are particular types of feedback perceived to be effective in supporting student learning? and lastly,

(c) What factors, if any, contribute to students' perceived usefulness of formative CBAs?

Rationale

Formative assessment practices are becoming increasingly common in higher education including professional learning programs such as education, engineering, law, and medicine (Higgins & Bligh, 2006; Peat & Franklin, 2002). The evolving use of computer technology for the purposes of assessment provides a further mechanism to combine formative assessment practices with CBA to support student learning in these programs. However, the extent to which these frameworks support student learning is largely unknown. Given that formative assessment plays such a crucial role in learning (Black & Wiliam, 1999), it is easy to justify research that aims to enhance student learning through the study of formative CBAs used in professional learning programs.

Overview of Dissertation

This dissertation is organized into six chapters. The following sections in this chapter present pertinent terminology and the theoretical framework. Chapter 2 begins with the evolution of CBA and is then followed by CBA guidelines. This chapter concludes with a presentation of research on formative CBAs. Chapter 3 includes a description of the participants in each case study and summarizes the mixed method design used in the case studies. This section also describes the data analysis procedures. In Chapters 4 and 5, the School of Medicine and Faculty of Education findings are presented. These two chapters are organized according to the three research questions posed in this dissertation. Chapter 6 pulls together the findings from the two case studies

and presents a response to each of the three research questions. This chapter concludes with implications for practice and research and includes an update of the formative CBAs currently used in the School of Medicine and Faculty of Education.

Context for the Study

Two different professional learning programs served as the context for this study. The professional learning programs are explored using a case study approach with each case presented in separate chapters. The first professional learning program examined was in a School of Medicine where a formative CBA was used in a first year, 16 hour module on the musculoskeletal (MSK) system. The second professional learning program studied was in a Faculty of Education. In this case study, a formative CBA was used in a nine hour module on classroom assessment in a one-year Bachelor of Education program. Both case studies are described in further detail in subsequent chapters.

Terminology

Some terms used in the field of formative CBAs are interconnected (e.g., linear, formative, and adaptive CBA) while others (e.g., formative assessment) are pivotal to understanding the purpose of formative CBA. Hence it is necessary to define and provide a rationale for the use of key terms related to formative CBA. This section begins with a rationale for using the expression computer-based *assessment* over computer-based *testing.* Next, the term formative assessment is introduced. To fully understand the meaning of formative assessment, summative assessment is also introduced. In the next section, the differences between linear and adaptive CBAs are presented to provide the reader with an understanding of the overlap in these two CBAs. Lastly, the term *cognitive*

complexity is described in terms of the items used in each of the case studies

Computer-based Assessment and Computer-based Testing

In the American educational literature (e.g., American Council on Education, 1995; Association of Test Publishers, 2000; Wainer, 2000) the phrase computer-based testing is used to describe tests administered via the computer. These tests have predominantly been summative instruments used to measure student achievement for the purpose of grading, certification, or ranking. In the British educational literature (e.g., Clariana & Wallace, 2002; Peat & Franklin, 2002), computer-based instruments used to promote learning are often referred to as computer assisted learning or computer-based assessments (CBAs). These instruments tend to be formative in nature. To avoid the summative connotations associated with computer-based testing, computer-based assessment (CBA) is used in this dissertation to describe computer based instruments that could be used in either formative or summative contexts. As such, the expression *formative* CBA used in this dissertation describes CBAs that are used in a formative context.

Formative and Summative Assessment in Education

The Principles for Fair Assessment Practices in Canada document (1993) provides working definitions of the terms formative and summative assessment. Summative assessment is described as a formal approach to formulating and interpreting evidence of student achievement for the reporting process. In contrast, formative assessment is used to support rather than measure student learning. The Principles for Fair Assessment Practices in Canada document further states that “the intent of these [formative]

assessments (e.g., informal observations, quizzes, text-and-curriculum embedded questions, oral questioning) is to inform decisions regarding daily learning, and to inform or otherwise refine the instructional sequence" (*Principles for Fair Assessment Practices for Education in Canada*, 1993, p. 3). Based on this definition, formative assessment refers to any assessment process, including the use of feedback, which makes students aware of the gaps in their learning without imposing any academic penalty. In terms of formative CBA, feedback is provided via computer software to inform students about their learning.

It is important to link formative assessment with a more recent phrase used in assessment; namely, assessment for learning. In an assessment for learning framework evidence of student learning is used to adapt instruction and enhance student learning where: (a) instruction is student-focused and students become active participants in their learning; (b) teachers are required to clarify and share learning objectives and success criteria with their students; (c) feedback is used to support student learning; and (d) peer and self-assessment activities are employed to heighten students' cognition of their learning (Wiliam, 2006). As noted by Wiliam (2006), feedback used in formative assessment is one underlying aspect of the assessment for learning framework.

Linear and Adaptive Computer-based Assessments

In the computer based assessment literature two types of CBAs are referred to: linear and adaptive CBAs. These two CBAs are supported directly by technology in that the items are delivered and assessed using computers. The sequencing of items and the theoretical framework underpinning these instruments is what differentiates one type of

CBA from the other. Although the primary focus of this dissertation is on linear CBAs used in a formative context, computer adaptive assessments are also discussed because they are also used in formative contexts.

Linear Computer-based Assessments

Linear CBAs display items in a preset sequential order. Early linear CBAs began as computerized copies of pencil and paper summative assessments consisting of a series of multiple-choice items (Harper, 2003). In classroom use, linear CBAs were also used in a formative context. These formative instruments typically focused on drill and practice using low cognitive items based on a narrow band of curriculum, for example, practicing addition or subtraction in primary level mathematics (McCormick, 2004). Students were presented with an item and given feedback of *correct* or *try again* based on the response the student provided. Feedback often concluded with telling the student the correct response. More recent linear CBAs allow students to provide their own responses (i.e., numeric or characters). Some newer instruments capitalize on technological features such as scrolling, skipping items and returning to a previous item. These instruments however, may no longer be considered linear CBAs.

Formative computer-based assessments. Formative CBAs go beyond their predecessors by focusing on more descriptive feedback to support student learning. Items, however, are still presented in the same preset sequential or linear order. The descriptive feedback supports student learning by directing students to course material where a concept was introduced or providing a hint through additional information. Some formative CBA software (e.g., Hot Potatoes) allows students to try an item a second time

after receiving some form of feedback. This dissertation will only refer to formative CBAs since the principles for these instruments encompass linear CBAs.

Computer Adaptive Assessments

Computer adaptive assessments were initially designed for summative rather than formative purposes such as credentialing examinations (e.g., Graduate Record Examination, Graduate Management Admissions Test, and the placement and diagnostic assessment system by American College Testing). Instrument developers (e.g., Castle Rock Research, Northwest Evaluation Association) now claim that computer adaptive assessments can support a formative framework in that scores can be interpreted using a reference chart linking the curricular areas. Developers claim this information supports teachers (not students) in their instructional programming. They also espouse that the continuous Rasch scale allows for comparison of student abilities over a period of time thus identifying students who are not progressing and allows for intervention to occur. Computer adaptive assessments barely pass as formative assessments given that the information (a student score) generated from the assessment is directed to teachers for instructional modifications. Students are generally not active partners in this assessment process. Hence computer adaptive assessments are not considered assessment for learning instruments.

Computer adaptive assessments are different from their linear predecessors in that items adapt or are selected based on student responses to previous items. The assessment process typically begins with the algorithm selecting a starting item of moderate difficulty. The next and subsequent items are based on the student's responses to the

previous set of items. If a student responded incorrectly (overall), the next set of items selected is easier. In contrast, if the item set was answered correctly (overall), the subsequent item set would be more difficult. At first, the next set of items selected is significantly harder or easier (depending on the previous set of responses) but after several sets of items, the subsequent item sets are closer in difficulty to the previous item sets and ultimately, the difficulty of the remaining item sets begins to parallel the student's ability. Currently, the item format on computer adaptive assessments tends to be limited to selected response formats.

Although computer adaptive assessments are not frequently used in higher education, they are being used in elementary and secondary schools throughout the United States and have been used in two provinces in Canada (i.e., Alberta and British Columbia). Computer adaptive assessments increasingly have dual roles whereby they are used to measure student learning and as instruments to support student learning.

Cognitive Complexity

Cognitive complexity refers to the cognitive demand required of an item. The cognitive complexity of an item focuses on the item rather than the ability of the student. The cognitive complexity of items presented in the two case studies required students to use more than knowledge and comprehension to correctly answer an item. In the Faculty of Education this was done through well described scenarios. An item at the higher end of the cognitive scale would require students to pull together several pieces of information to correctly respond to the item. In the School of Medicine these items required students to examine x-rays as well as review patient case histories and identify a correct diagnosis.

Hence, the cognitive complexity of items used in both case studies refers to assessment items that draw upon thinking skills such as application, evaluation, and analysis.

Theoretical Framework

Formative CBA employs theories of formative assessment which focus on the use of feedback to promote self-regulated learning environments. In order for the feedback to be more effective, students must be able to use the feedback to support their learning. As Gibbs (2005) noted, when students are able to monitor their learning, a key aspect to self-regulation, they can also control and direct their subsequent learning. Hence theories of formative assessment underpin formative CBAs and can promote self-regulated learning environments.

The central purpose of formative assessment is to support students' learning by providing feedback about the accuracy, adequacy, and quality of their thinking as a consequence of instruction or exposure to new ideas or experiences (Bell & Cowie, 2001; Linn & Gronlund, 2000; Yorke, 2003). Research on formative assessment practices has shown that formative assessment improves student learning (e.g., Black & Wiliam, 1998; Crooks, 1988; Hargreaves, 2005; Hodgen & Marshall, 2005; Kluger & DeNisi, 1996; Wiliam, Lee, Harrison, & Black, 2004) and this form of assessment is valued by educators (Christensen, Nowak, Prinster, & Rinke, 2006; Miller, 2004; Weaver, 2006).

The Importance of Feedback in Formative Assessment

Feedback underpins formative assessment and is conceptualized as "information provided by an agent (e.g., teacher, peer, book, parent, self, experience, [computer]) regarding aspects of one's performance or understanding" (Hattie & Timperley, 2007, p.

81). Feedback is provided in formative CBAs after students have responded to an item. This feedback is related to and supportive of students' understanding or learning. The objective of the feedback is to close the gap between what students understand and what they aim to understand. Formative assessment is claimed to have its greatest positive impact if feedback is provided while students are in the process of constructing knowledge and practicing skills (Bell & Cowie, 2001; Linn & Gronlund, 2000; Yorke, 2003).

Nicol and Mcfarlene-Dick (2006) present seven principles of effective feedback: (a) facilitates the development of self-assessment (reflection) in learning*, (b) encourages teacher and peer dialogue around learning, (c) helps clarify good performance, (d) provides opportunities to close the gap between current and desired performance, (e) delivers high quality information to students about their learning*, (f) encourages positive motivational beliefs and self-esteem*, and (g) provides information to teachers that can be used to help shape their teaching. The three principles marked with an asterisk apply to formative CBA contexts. Although these principles initially referred to feedback provided on performance instruments (e.g., essays or projects), with some modifications they can be applied to feedback used in formative CBAs and serve as the cornerstones guiding the use of feedback on formative CBAs. The justification for applying these three principles to formative CBAs is presented below.

Facilitate self-assessment. Formative CBAs facilitate self-assessment in that feedback is used to clarify and direct learning thus supporting the self-assessment process. In addition, formative CBAs provide students with "the freedom to explore areas

of perceived weakness and to make mistakes without revealing these to those responsible for the final assessment or to peers" (Challis, 2005, p. 534). Formative CBAs that are organized by cognitive and content domains empower students to make decisions about learning in specific areas. Lastly, most formative CBAs (e.g., Hot Potatoes, Moodle) have the option of recording student interaction with the instruments. When desired, this feature can be deactivated allowing students to complete formative CBAs anonymously without exposing their mistakes to their instructors.

Deliver high-quality feedback. Formative CBAs can generate high quality feedback. Feedback addressing common errors can be constructed through the analysis of errors made by prior students on traditional assessments. In a meta-analysis of feedback, Hattie (1999) found the most effective forms of feedback to be cues or reinforcement to learners. This feedback was found to be in the form of video-, audio-, or computer-assisted instructional feedback and was most effective when the feedback provided information on correct rather than incorrect responses (Hattie & Timperley, 2007). Unfortunately, Hattie and Timperley do not provide descriptors or examples describing the effective types of feedback. How, for example, does feedback such as 'correct, this fracture should heal with surgical stabilization' compare to 'correct, review page 52 of the class text for additional reading' where directions to a resource are given? Few studies have explored the effectiveness of different types of feedback. Hence it is unknown whether one type of feedback is of higher quality than another type of feedback. As an example, Miller (in press) grouped feedback on CBAs into four categories (i.e., directing students to a resource, rephrasing a question, providing

additional information, and providing the correct answer) and students from a Faculty of Education showed no preference for one type of feedback over another.

When applying Nicol and Mcfarlene-Dick's (2006) principle of delivering high-quality feedback, instrument developers could follow Sadler's (1989) three components of effective feedback by: identifying students' strengths (acknowledge accurate aspects of their thinking that may have led them to select an answer), identifying students' weaknesses (acknowledge a common misunderstanding), and providing further information to help students bridge the gap in their understanding (i.e., reference to a resource or course material where the concept was introduced). Using these three components for delivering high-quality feedback, formative CBAs can go beyond using ineffective statements such as *good work* or *incorrect* that have been found to be ineffective (Pelton & Pelton, 2006; Hattie & Timperley, 2007).

Encourage positive self-esteem and motivation. Formative CBAs are purported to nurture students' motivation and self-esteem because students can complete formative CBAs at a time and place that is convenient for them (Bennett, 2001) and without revealing their scores to their instructors (Challis, 2005). The self-esteem of students who score low on formative CBAs is protected because they are not required to reveal their scores. The opportunity to re-try an assessment also has the potential to motivate students to excel. In addition, choices offered in formative CBAs such as selecting content areas and cognitive domains has the potential to motivate students because students do not waste time self-assessing content areas or cognitive domains that they have previously mastered.

Self-regulated Learning Environments

Given that students must internalize feedback they receive on formative CBAs and then make decisions about the extent to which they should apply the feedback to improve their learning (e.g., follow-up with feedback that directs students to a resource), formative CBAs can provide a self-regulated learning environment. Self-regulation calls upon an active constructivist approach to learning whereby students become active participants in their learning (Winne, 2005). While operating within the context of their learning environment, students must be able to monitor, regulate, and control their cognition, motivation, and behaviour (Pintrich & Zusho, 2002). This approach to learning facilitates deep learning whereby students are able to process information (e.g., feedback) and engage in higher levels of cognitive thinking that includes analysis, application of concepts, and evaluation. Such thinking provides a conduit for making connections between prior knowledge and new learning. The flexibility offered in formative CBAs allow students to make choices and have control over their learning (Zimmerman & Schunk, 1994). Professional learning programs such as those in education, engineering, law, and medicine that promote life-long learning require students to develop a self-regulated learning approach (Duffy & Holmboe, 2006).

Studies (e.g., Pintrich & Zusho, 2002; Zimmerman, 2000) have found that self-regulated learners are better at exhibiting confidence and perseverance. Self-regulated learners are equipped with the processing skills to engage in deep learning and thereby retain, synthesize, and transfer information at higher rates (Ramsden, 2003). In addition, the learning experience is more enjoyable for self-regulated learners who adopt deep

learning strategies (Tagg, 2003). To help students move from surface level processing strategies that rely on acquisition and recall to deep processing skills that call for critical thinking, self-regulated learning skills needs to be fostered (Tagg, 2003; Zimmerman, 2001). Formative CBAs containing items situated in meaningful scenario-based contexts assessing students' critical thinking have the potential to foster such skills by providing students with opportunities to self-assess when, where, and as frequently as they choose. In addition, formative CBAs employing quality feedback can potentially bridge gaps in students' learning, using feedback to clarify and direct subsequent study and thinking.

Significance of Dissertation

This dissertation provides evidence of the conditions under which formative CBAs were used in professional learning programs to support student learning. A description of the intentions upon which the formative CBAs were used, the structure in which the instruments were created and administered, as well as students' perceived benefits of how formative CBAs are examined to help guide the creation and use of future formative CBAs. In addition, this dissertation suggests areas of future research that can be pursued to advance the use of formative CBAs in professional learning programs.

CHAPTER TWO

Literature Review

This literature review begins with a presentation of the background of computer-based assessment. This section examines how formative CBAs have evolved from instruments grounded in summative contexts to instruments that are being used today to support, rather than measure student learning. Next, I provide a review of the CBA guidelines. The purpose of reviewing the CBA guidelines, originally developed for use in summative CBAs, was to determine whether these guidelines can be applied to formative CBAs. Next, research on the features and challenges of formative CBAs is presented to provide an analysis of the potential for formative CBAs to become an invaluable tool in the field of formative assessment. Lastly, issues related to formative CBAs are presented to highlight the factors influencing the developmental stage of formative CBAs. The issues presented in this section focus on item construction, item format, and item exposure.

The Evolution of Formative CBAs

Although formative CBAs are relatively new, CBAs have been in existence for a number of years. CBAs were initially developed for summative assessments. Psychologists began using computers in the field of assessment in the early 1970s in collaboration with the United States military. The rationale for the use of these CBAs was to control test variability, eliminate examiner bias, and increase efficiency in the assessment of personality, intelligence, and vocational interests (Russell, Goldberg, & O'Conner, 2003). These early CBAs laid the ground work for administering multiple-

choice items in a computer or DOS (disk operating system) format.

The development of CBAs continued and research in the 1970s through the mid 1980s focused on assessment mode effect where performance on pencil and paper assessments was compared to performance on CBAs containing identical items. Results from psychological studies (e.g., Katz & Dalby, 1981) showed no significant differences between the two modes of assessment; however, studies in education (e.g., Lee & Hopkins, 1985) were not as conclusive. Perhaps CBAs in psychology were able to control the many variables that could influence assessment performance whereas in education, this may not have been possible due to the variability in classrooms. Subsequent research has focused on investigating the format, use, and effectiveness of CBAs.

During the 1980s and into the 1990s large-scale assessment programs became increasingly common in educational jurisdictions across North America primarily for the purpose of accountability (Klinger, DeLuca, & Miller, 2008). In response to the high cost associated with administering and scoring large-scale assessments, some educational jurisdictions (e.g., Alberta Public Schools, AB; British Columbia School District 46, BC; Bloomington Public Schools; MN, Oklahoma Public Schools, OK) turned to computers to improve the efficiency of these assessments (Bennett, 2001). Although primarily developed for summative purposes, some developers (e.g., Castle Rock Research) claimed the CBAs would be useful in a formative context if teachers could be provided with diagnostic information to guide instruction. As an example, School District 46 in British Columbia developed CBAs to be administered twice a year. These CBAs were

designed to monitor academic growth over time and reported a student's ability level. Teachers could then use this information to guide instruction in classrooms (Pelton & Pelton, 2006). These CBAs represented a merging of purposes with the assessments serving both summative and formative functions. While pragmatic, this combined use occurred with little exploration of the ability of CBAs to be used effectively in both contexts.

Large-scale assessment initiatives in higher education were also sought by regulating bodies to invoke change in the quality of instruction and learning through the tracking of student outcomes (Lazeron, Wagner, & Shumanis, 1999). This method of accountability was considered an effective way to force faculty to align course expectations, instruction, and assessment. Although organizations of higher education (e.g., National Association of State Universities and Land-Grant Colleges) promoted the use of assessment as a means of supporting student learning and improving programs, faculty resistance tempered the impact of the reform in higher education compared to what occurred in elementary and secondary schools (Lazeron, Wagner, & Shumanis, 1999). By the late 1990s there was little evidence the externally imposed assessment measures were having any effect on student learning in higher education (National Center for Post Secondary Improvement, 1999). As class sizes in higher education gradually increased, the challenge of how to assess large numbers of students became increasingly important. Increased funds were allocated to promote the use of innovative instructional and assessment practices. It was within this context that some universities (e.g., Brock University, Queen's University, University of British Columbia, and University of

Toronto) used a portion of their funds to begin to develop CBAs (National Center for Post Secondary Improvement, 1999). In some cases, the CBAs developed included a function to provide feedback (e.g., Hot Potatoes from University of British Columbia).

The ongoing pressures in higher education have provided fertile ground for continuing research and implementation of CBAs. CBA use has continued to increase over the past 10 years. Hence there has been a growing need to carefully examine the context of CBA use and students' perception of these instruments and their uses (e.g., Buchanan, 2000, Burrow, et. al., 2005; Peat & Franklin, 2002). Not surprisingly, only a small number of studies have explored specific factors influencing students' experiences with CBAs. In the case of the use of feedback, the research is even more sparse. It is not clear what forms of feedback are currently in use or the extent to which students actually use feedback to support their learning. It is also unclear the extent to which different types of feedback (e.g., directing students to a resource, providing additional information, providing the correct response, or simply stating the response is correct or incorrect) are deemed useful by students.

Computer-based Assessment Guidelines

In 1986, the American Psychological Association (APA) published guidelines for CBAs. These guidelines were created in response to the rise of summative CBAs during the 1980s and were called Guidelines for Computer-based Tests and Interpretations (APA, 1986). The APA guidelines stated that summative CBAs should provide the same opportunities as pencil and paper assessments to ensure that: (a) the amount and type of feedback given on pencil and paper assessments should also be given on CBAs; (b) like

pencil and paper assessments, CBAs should permit a student to review and change previous items as well as skip ahead to examine upcoming items; and (c) CBAs and pencil and paper assessments that are equivalent should have the same mean scores, standard deviations, and rankings of individual examinees (APA, 1986). Given that the guidelines focused on summative assessments, they had little connection to formative CBAs. While these guidelines mentioned feedback, it was not referenced in a formative sense.

In 2000, the Association of Test Publishers (ATP) released their guidelines for CBAs with the intention of supplementing the guidelines from the National Council on Measurement in Education (NCME) and the APA. The Association of Test Publishers' guidelines for summative assessment focused on: (a) test construction, evaluation, and documentation; (b) reliability and errors of measurement; (c) test development and revision; (d) scales, norms, and score comparability; (e) test administration, scoring, and reporting; and (f) supporting documentation for tests (ATP, 2000). These guidelines were developed by measurement specialists to guide the use of summative CBAs, namely computer adaptive assessments. Hence the focus of these guidelines was on the summative use of CBAs. Formative CBAs were gaining in popularity during this period but guidelines governing their use had not been developed. This may have stemmed from the low-stakes nature of formative CBAs, given that these instruments were not used for accountability purposes. More recent CBA guidelines produced in 2005 by the International Test Commission (2005) focused solely on the use of summative CBAs. It is possible that since the use of formative CBAs is still in its developmental stage,

guidelines governing their use have yet to be developed. As formative CBAs become more commonplace and valued, guidelines governing their use will likely be developed. Given the current absence of guidelines governing formative CBAs, the manner and the contexts in which formative CBAs are used vary considerably.

Features of Formative Computer-based Assessments

CBAs provide unbiased and accurate scoring as a result of their automated technological aspect. In addition, CBAs can assess students' higher cognitive skills through the development of scenario-based items. Although the above features apply to both formative and summative CBAs, the following features pertain only to formative CBAs and stem from the feedback mechanism in formative CBAs. These features focus on promoting independent learning, using feedback to guide learning, and using scenario based multiple-choice items to assess students' higher cognitive skills.

Feedback provided in formative CBAs has been successfully used to support independent learning (Burrow, Evdorides, Hallam, & Freer-Hewish, 2005; Higgins & Bligh, 2006; Peat & Franklin, 2002; Ricketts & Wilks, 2002). The call for independent learning may stem from programs promoting independent learning or as the solution to large undergraduate programs in which one-to-one feedback with instructors is difficult due to the large number of students. Either way, feedback provided in formative CBAs was reported to be effective in environments promoting independent learning such as professional (e.g., Burrow et al., 2005) and undergraduate programs (e.g., Higgins & Bligh, 2006; Peat & Franklin, 2002, Pitt & Gunn, 2003).

Formative CBAs support student learning by providing on-demand assessments

with immediate feedback to guide students' cognitive development thus allowing students to progress at individual rates (McCormick, 2004; Zakrzewski & Bull, 1999). The on-demand feature of formative CBAs allows students to complete assessments at a time and place that is convenient for them. In addition, the on-demand nature of formative CBAs also facilitates assessment in one-to-one, classroom, or large group settings with immediate results for all users (Bennett, 2004; Raikes & Harding, 2003). For example, Pitt and Gunn (2003) found that students took advantage of the flexibility of formative CBAs and completed the assessments when it was convenient for them. The majority of students in this study completed their formative CBAs outside of regular school hours.

Although tabulating student scores in formative CBAs may not be a focus given the formative nature of these instruments, some formative CBA software (e.g., Hot Potatoes) provides students (not the instructor) with a score showing the number of correctly answered items. These scores provide students with an overall sense of understanding which may be important in the self-assessment process. Regardless of the rationale for providing total scores on formative CBAs, the advantage lies in the process of grading without the cost traditionally associated with pencil and paper assessments (He & Tymms, 2005).

Scenario-based items can present stimulating scenarios with the potential to promote student discussions while tapping into students' higher cognitive skills (Crisp & Ward, 2008). In this sense, the use of scenario-based multiple-choice items that employ feedback to guide learning can be considered an added feature of using formative CBAs.

Scenario-based items begin by describing a situation. This is followed by a set of items exploring various concepts related to the scenario. Students who have used a scenario-based formative CBA have reported positive experiences in terms of: ease in understanding the assessment, authenticity of the scenario, relevance of scenario, and the formative CBA's ability to stimulate learning (Crisp & Ward, 2008). The sample scenarios presented in Crisp and Ward's study (2008) were two to five sentences in length. The scenario was followed with one or two multiple-choice items related to the scenario. These scenarios were relatively simple and straight forward; consequently, the simplicity of these items may have influenced students' positive experiences. It is not clear whether more complex scenarios would have resulted in the same positive experiences. Given that the two case studies examined in this dissertation also utilized scenario-based multiple-choice items, it is important to note students' perceptions of the item format as well as the characteristics of the scenario itself. Once again, it is unclear whether scenario-based multiple-choice items have a direct influence on students' overall experiences with formative CBAs.

Formative CBAs can promote self-regulated learning environments. These environments give the student more control over the timing of the assessment and use of the feedback they receive. In addition, the use of immediate feedback in formative CBAs can better guide students' learning by making them aware of specific strengths and weaknesses. More complex formative CBA items, as found in scenario-based items, can assess more than recall and comprehension, focussing on deeper understanding and use of content. These features likely contribute to the growing popularity of formative CBAs

in higher education, including professional learning programs.

Challenges of Formative Computer-based Assessments

The challenges of formative CBAs have not been as well documented. The first set of challenges applies to CBAs in general and focuses on: construct-irrelevant variance, construct under-representation, the time requirement for developing quality items, and ease of software use. Although these disadvantages apply to both formative and summative CBAs, they are worth noting given their impact on formative CBAs. The last challenge presented pertains to the challenge in developing quality feedback in general and, more specifically, quality feedback for use in formative CBAs.

Construct-irrelevant variance affecting a student's score due to irrelevant constructs such as typing ability or familiarity with the CBA software (Wise, Bhola, & Yang, 2006). When evaluating the presence of construct-irrelevant variance in a mathematics CBA, for example, Gallagher, Bennett, and Rock (2002) reported that some students experienced mechanical difficulties when entering mathematical expressions thus contributing to presence of construct-irrelevant variance. In this study, the challenge of presenting fractions and square root symbols prevented some students from communicating their responses. Similar to pencil and paper assessments, construct-irrelevant variance can also occur when too much unfocused or distracting information is presented in scenario-based items (Ahmed & Pollitt, 2007). Although formative CBAs are not as score oriented as summative CBAs, it is possible that construct-irrelevant variance can occur in formative CBAs whereby construct-irrelevant variance influences students' learning experiences. Given that formative CBAs do not rely heavily on typing

because of the predominant selected response format, it is unlikely that students' typing ability influences their learning using formative CBAs.

The validity of all assessments can be undermined by construct under-representation, including CBAs. Similar to summative CBAs, formative CBAs can require a substantial number of items to support student learning in the construct being assessed. The number of items on a formative CBA is suspected to be dependent on the complexity of the construct(s) being assessed. However, instead of using this idea of determining the number of formative CBA items, Hill, Sanders, Fyfe, Ziman, and Koehler (2008) believed fewer items would attract more students. Hill et al. (2008) reported that when their 30 item formative CBA was reduced to 10 items, the participation rate increased by 40%. The increase was found to be greatest among low achieving students. Researchers acknowledged that the reduction in items resulted in construct under-representation but was considered acceptable given the gains made in participation rates. Although this manner of determining the number of items on a formative CBA resulted in higher participation rates, it may have reduced the amount of learning given that students were exposed to fewer items and by extension, less content.

The time requirement to develop quality items with effective feedback may cause construct under-representation (Crisp & Ward, 2007). Unlike summative CBAs used in high-stakes assessment where teams of item developers create items, formative CBAs are generally created by one or two instructors involved in the instructional program (Fletcher, Kearney, Bartlett, 2005; Miller & Shulha, 2008). The challenge in developing items for formative CBAs can be overwhelming; especially in developing items that go

beyond assessing knowledge based content and use, for example, scenario-based items that tap into students' higher cognitive skills involving critical thinking. As a result of the challenge in developing good quality items, formative CBAs may contain an insufficient number of items, particularly those at the higher end of the cognitive scale. The impact of construct under-representation in one case increased participation rates but the impact on learning is largely unknown. Hence when studying the contextual factors related to formative CBAs, there is a need to note the number of items and students' perceptions of whether the item set was adequate.

Item development in formative CBAs is not solely about developing quality items but also about developing quality feedback to support student learning (Gipps, 2005). The development of feedback is another demanding aspect of using formative CBAs. Developing feedback requires an understanding of the pedagogical issues associated with learning to create effective feedback such as being able to identify common errors in student thinking. In general, developing quality feedback is also viewed as a challenge. This challenge has the potential to cause construct under-representation in formative CBAs. Gipps (2005) attributed high quality feedback on formative CBAs as a key factor in the growing use of these instruments. Although Gipps (2005) did not venture to note that a lack of quality feedback may have negative influences on the growth of these instruments, it is reasonable to anticipate this consequence given that ineffective feedback is not likely to support student learning.

Use of Formative CBA in Higher Education

Research on formative CBAs has focused on four areas: (a) development and evaluation of formative CBAs, (b) timing of feedback, (c) evaluation of different types of feedback, (d) effectiveness of feedback, (d) influences of formative CBA on summative assessments, and (e) the frequency and use of feedback.

Development and Evaluation of Formative CBAs

Given the recent emergence of formative CBAs, research has predominantly focused on the development and evaluation of these instruments in various contexts. Three studies examined formative CBAs (e.g., Course Marker, SAMs, Question Mark Perception) used in higher education programs to address large class sizes (Higgins & Bligh, 2006; Peat & Franklin, 2002; Ricketts & Wilks, 2002). One additional study was conducted in a professional learning program, a graduate course in engineering, which used a formative CBA called TRIADS for the purpose of offering flexible learning while reducing student-instructor contact time (Burrow, Evdorides, Hallam, & Freer-Hewish, 2005).

In each study, questionnaires surveying students' experiences or attitudes towards formative CBAs were used to measure the extent to which students accepted these assessments as a tool to support their learning. Results showed that students' overall experiences were positive including a positive reception of receiving immediate feedback (Burrow et al., 2005; Higgins & Bligh, 2006; Peat & Franklin, 2002; Ricketts & Wilks, 2002). Unfortunately, none of these studies elaborated on the questionnaire used and it is

not known whether the questionnaires examined specific aspects of formative CBAs or potential factors believed to influence students' experiences with formative CBAs.

In an undergraduate engineering program, students also reported favourable experiences and positive learning with their formative CBA (Sambell, Sambell, & Sexton, 1999). These authors further added that students required an initial period of becoming familiar with the instrument prior to reporting positive experiences. While students reported positive experiences with formative CBAs, it is not clear which aspects of the formative CBA environment lead to these positive views. The extent to which the timing of feedback or the effectiveness of feedback influenced students' positive experiences, for example, is still largely unknown.

Timing of Feedback

Although each of these studies reported that students liked immediate feedback, there are at least two methods of providing immediate feedback: immediately after a student responds to an item (Miller, in press) and after a student responds to all of items in a set (Buchanan, 2000). Although the studies cited above did not intend to evaluate these two methods of providing feedback, it is an important factor given that receiving feedback after a student has responded to one item may be more valuable than collating feedback and presenting the feedback to all items after a student has responded to 25 items. When feedback is provided at the end of the formative CBA, students may not remember the context in which the item was presented or the learning that was being constructed. Previous research on timing has focused on providing feedback in regular classroom settings (i.e., not using a formative CBA). A meta-analysis of research in this

area found immediate feedback to be more effective than feedback provided days or weeks following the assessment (Kulik & Kulik, 1988). Although the delay in providing feedback on formative CBAs is not as pronounced (e.g., days or weeks later) as in classroom assessments, the finding suggests that the more immediate the feedback the better.

Formative CBAs have the potential to provide immediate feedback to students not only while they are completing the assessment but also as soon as the assessment is completed. Hence formative CBAs are much more able to meet the requirements associated with the timing of feedback as compared to other forms of formative assessment.

Effectiveness of Feedback

Effective feedback supports student learning. Gipps (2005) acknowledged the value of formative feedback in formative CBAs and noted that the growing popularity of formative CBAs in the UK was due to the use of formative feedback. He cautioned that only formative CBAs that provide effective formative feedback can support student learning. Hence it is important to better understand the features of feedback that would make it more effective. However, this in itself is difficult. Typically, effectiveness is measured through subsequent achievement or on students' or instructors' perspectives. Steven and Hesketh (1999) described two forms of effective feedback that identified common errors and provided a link to additional learning resources. In contrast, simply stating a response was correct or incorrect contributed little, if anything, to furthering students understanding of the concept being assessed.

Burrows et al., (2005) examined a formative CBA (TRIADS – Tripartite Interactive Assessment Delivery System) in a graduate program in engineering that incorporated conventional and innovative item formats including: (a) plotting points and lines; (b) moving objects to create and label diagrams as well place sliders on a scale; (c) using numeric text entry to enter numbers; (d) drawing objects using a standard set of drawing icons; and (e) using combinations of any formats presented above. Feedback included a score, indications of parts of the item that were answered correctly or incorrectly, the correct response to the item, a model answer (full solution), and a list of references to pursue further learning. Although the feedback they provided went beyond stating whether a response was correct or incorrect, it is unknown whether this large amount of feedback actually supported students' learning.

Peat and Franklin (2002) examined formative CBAs in an undergraduate biology course. The feedback included "diagrams and diagrammatic representation of ideas" (p. 517). This feedback was similar to the feedback used by Burrow et al. (2005) that also provided feedback in terms of content rather than focusing on errors in students' thinking. The feedback described in the above two studies did not describe why the incorrect responses were incorrect; however, both studies attempted to provide directions for further learning through the use of diagrams (e.g. Peat & Franklin, 2002) or a list of references to pursue further learning (Burrow et al., 2005). Despite the short-comings in the feedback used in these two studies, a questionnaire completed at the end of the course exploring students' satisfaction with their formative CBA revealed high levels of satisfactions. It is possible that a novelty effect influenced students' perceptions of

formative CBAs when it was first used in the course or that this feedback did provide sufficient information to be viewed as valuable by the students.

Higgins and Bligh (2006) made specific reference to the feedback used in a formative CBA developed for an undergraduate computer science course. These researchers did not provide a detailed description of the feedback that was used in their formative CBA. Higgins and Bligh (2006) commented that the feedback was often too lengthy and focused on student weaknesses (Higgins & Bligh, 2006). They concluded "the method of feedback in CourseMarker [formative CBA used] must be changed to provide less feedback, but of a more motivational nature" (p. 102). Although these researchers did not indicate the purpose of their instrument was to support self-regulated learning, they noted the value of feedback to enhance students' motivation for learning. Previously, Pintrich and Zusho (2002) connected identified motivation as a key feature of self regulation. Combined, these findings suggest that formative CBAs may be able to create a self-regulated learning environment that will enhance students' learning.

Influence of Formative CBAs on Summative Assessment

A number of studies claim that formative CBAs positively influence student results on summative assessments (Pinckey, Mealy, Thomas, & MacWilliams, 2001; Pitt & Gunn, 2003; Steven & Hesketh, 1999; Thelwall, 2000) but the specific aspects of the formative assessments that influence subsequent achievement are still largely unknown. According to Pitt and Gunn (2003), "the introduction of the [formative] CBA package did appear to have a positive influence on examination marks for each module" (p. 10). Nonetheless, they also acknowledged that the nature of the research prevented one from

developing direct unicausal links. Motivation, investment of more time, and familiarity with the formative CBA were other possible explanations for students' increased achievement on subsequent summative assessments.

Henly and Reid (2001) reported that an increase in students' summative assessment results was likely the result of students' willingness to use learning support materials such as the formative CBA. They concluded that academically stronger and more highly motivated students accessed the formative CBA more often than weaker students. Sambell, Sambell, and Sexton (1999) also noted a link between the use of formative CBAs (practice tests) and an increase in summative assessment results. They attributed the improved summative assessment results to students investing more effort in understanding the course material early in the course. In addition, the non-threatening nature of formative CBAs created an assessment process that was a supportive aid for students. Hence, students were more apt to use the formative CBA to prepare for a summative assessment. The willingness, investment of effort, and motivation to use formative CBAs to prepare for summative assessments described in the above studies suggests that these formative CBAs have the features of a self-regulated learning environment.

Charman and Elmes (1998) explored the influences of formative CBA on summative assessments in a first year undergraduate course with similar findings. In particular, they found a substantial gain in achievement by borderline students. When synthesizing this finding with that of Henly and Reid (2001) and Sambell, Sambell, and

Sexton (1999), students (academically strong or weak) who use formative CBAs show an improvement in summative assessment results.

Sly (1999) specifically identified the feedback students received on their incorrect responses as one of two factors influencing their achievement on summative assessments. Sly argued that the feedback helped to clarify misconceptions. The second factor Sly (1999) reported as influencing summative assessment results was students' familiarity with the formative CBA.

While indirect links between the use of formative CBAs and subsequent achievement on summative assessments have been noted, previous research has not been able to establish direct causal links between the use of formative assessments and student achievement. Gretes and Green (2000) compared the summative assessment scores on the basis of those who completed the formative CBA (referred to as a practice test) and those who did not. Students who completed the formative CBAs had higher summative assessment scores. Nonetheless, the voluntary nature of the formative CBA may have resulted in selection bias between the two groups. Subsequently, researchers controlled for students' SAT scores. After controlling for SAT scores, a significant difference was found between the number of formative CBAs taken and students' summative assessment scores.

Instead of comparing student performance on a formative CBA with a summative assessment, Albertelli, Kortemeyer, Sakharuk and Kashy (2003) compared student achievement on a summative assessment with a CBA administered after the summative assessment. This summative assessment held both formative and summative properties.

The CBA was considered summative since students' scores influenced their final score in the course and formative because the CBA was used to support student learning in that students were encouraged to collaborate and explore their learning. Students were provided the opportunity to complete a CBA after completing a summative assessment to improve their summative assessment score. The CBA contained a different set of items related to the original summative assessment. Using CAPA (computer-assessed personalized approach), students could obtain hints, participate in discussion forums, and obtain help from their instructors. The incentive was the potential to increase their mark by 25% of the marks previously missed. Participation rates and academic gains were high when using the formative CBA as a remediation tool.

The influences of formative CBAs on subsequent assessment scores have not always been positive. Henly (2003) compared student performances on three sequential formative CBAs. The mean scores on the formative CBAs decreased over the three assessments and the researcher concluded by stating that students "decided to access the [formative CBA] to guide their revision rather than using it to assess their level of acquired knowledge" (Henly, 2003, p. 119). As a result, students were more likely to guess at items thus lowering their achievement score. This strategy would allow access to the feedback with little effort. It is also possible that the decreasing achievement on the subsequent assessments may also have been due to increased difficulty of assessment items, increased difficulty of concepts, poor satisfaction with the formative CBA, absence of a novelty effect, or poor self-regulation. This impact of decreasing success on student learning and subsequent summative assessments is unknown.

The frequency of using formative CBAs, investment of time to complete formative CBAs, the non-threatening nature of formative CBAs, as well as the use of feedback were cited as aspects of formative CBAs contributing to student learning and subsequent performance. The studies presented here provide evidence of the potential for formative CBAs to directly or indirectly influence student learning by promoting a self-regulated learning environment. While subsequent summative assessment results are a common measure of effectiveness of a formative CBA, it is not known whether formative CBAs are only effective when they are followed with a summative assessment. Is it possible to effectively use a formative CBA for independent learning in the absence of summative assessment? These questions are explored in this dissertation where a formative CBA is followed by a summative assessment in one case and by no subsequent summative assessment in a second case.

Issues Related to Formative Computer-based Assessment

Given that formative CBAs are not high accountability or high-stakes assessments, the issues related to formative CBAs are not as frequently studied as issues related to summative CBAs. The issues related to formative CBAs focus on item construction and format, item banks, and item exposure. These issues pertain to CBAs in general and consequently the literature is sometimes drawn from summative CBA contexts.

Item Design and Format

Designing items is largely an understudied aspect of assessment (Haladyna, Downing, & Rodriguez, 2002). Additionally, there is little, if any, research exploring the

use of multiple-choice items designed to assess critical thinking. Research on traditional multiple-choice items has focused on examining the amount of information items provided. For example, assessments comprised of multiple-choice items have been found to provide more than twice the amount of information compared to an assessment using open-ended items (Jodoin, 2003). It is possible that scenario-based multiple-choice items can also provide large amounts of information similar to traditional multiple-choice items but possibly at the cost of incurring longer response times due to the additional reading and thinking requirement. Since the time required to complete formative CBAs is not as much of an issue as it is with summative assessments, the extra time required to respond to scenario-based items that tap into students' critical thinking may not affect the effectiveness of formative CBAs. However, it is possible that students may consider scenario-based items to be too demanding in comparison to the traditional multiple-choice items they may have become accustomed to using. In addition, the over reliance on multiple-choice items to assess students' lower cognitive abilities such as recall, may have cast a shadow on the use of this item format in formative CBAs (Shepard & Kirst, 1991). As discussed above, it is possible to design multiple-choice items that go beyond simple recognition and recall to assess application and analysis through the use of scenario-based items; however, designing these items is a time consuming and complex task. If formative CBAs are to rely on such items, it will be necessary to strengthen the quality of multiple-choice items in order to improve the effectiveness of formative CBAs.

Other research related to multiple-choice items has focused on the distractors. Good multiple-choice items are determined by the quality of the distractors (Haladyna &

Downing, 1989). Common misunderstandings found in students' responses to the open-ended items in general were used to structure the distractors in multiple-choice items created for use in other assessments.

The number of distractors has also been shown to influence the difficulty of items in summative assessments (Rodriguez, 2005). A meta-analysis exploring the use of pencil and paper multiple-choice items revealed that three distractors in comparison to four or five have minimal effect on decreasing the difficulty of an item (Rodriguez, 2005). However, reducing the number of distractors to two significantly decreased the difficulty of the item. An additional layer of item development inherent in formative CBAs is the creation of feedback for each distractor.

Item Banks

Items used in CBAs are stored in item or data/test banks and can be organized according to content area, cognitive domain, correct response, difficulty level, and other pertinent information (Georgiadou, Triantafillou, & Economides, 2006). In addition, each item bank for formative CBAs contains feedback for each distractor. Item banks in computer-adaptive assessments are very large given that a large number of items are required to narrow in on students' ability levels (see description of computer adaptive assessments in terminology section). In formative CBAs, items banks are not likely to be as large given that items are usually developed by one or two instructors and repetition of items is not viewed as an issue due to the low-stakes nature of the assessment. However, a large item bank could potentially provide several different formative CBAs on the same construct or enable an instructor to develop formative CBA with a very specific focus.

Larger banks of items could also be used to encourage students to better focus their own assessment choices.

Item Exposure

Item exposure rate is the frequency to which an item is presented across all administrations of an assessment. Item banks must be large in order to limit the frequency of exposure (Reckase, 2003). However, the exposure rate of items developed for formative CBAs may not be a concern given that the assessment is low-stakes and student collaboration and discussion on formative CBAs may be encouraged as a method to enhance student learning. If, on the other hand, instructors want students to accurately self-assess their learning then item exposure may be an issue; especially if students repeat formative CBAs. In such cases, a large item bank may be required as well as a mechanism for controlling the exposure of an item. Some formative CBAs (e.g., WebCT) have a randomizing feature to change the order in which items are presented from one student to the next as well as in the order of items presented in repeated attempts with the same formative CBA. When examining the contextual factors of formative CBAs, it is important to note whether or not item exposure was an issue taken into consideration given that item exposure could affect students' experience with the instrument.

Conclusion

Guidelines governing the use of CBAs have primarily focused on CBAs used in a summative context. Although formative CBAs do not carry the importance of a high-stakes assessment, they are growing in use particularly in higher education. The low-stakes nature of formative assessments may explain why guidelines have not been

developed for these instruments. It is also possible that these instruments are still developing and more is to be learned about how formative CBAs support student learning.

Items used in formative CBAs go beyond the assessment of basic knowledge and facts and can tap into students' higher cognitive thinking through the use scenario-based multiple-choice items. Although the challenges of formative CBAs have not been as well studied, they do exist. Issues related to formative CBA include construct-irrelevant variance, construct under-representation, and the challenge and time required to develop quality items with quality feedback. These issues are somewhat interconnected in that the challenge of developing quality items (including quality feedback) may prevent instructors from developing large item sets which in-turn contributes to construct under-representation. Construct-irrelevant variance is a disadvantage of formative CBAs that will need to be monitored as formative CBAs become more technologically savvy where additional technological requirements may create construct-irrelevant variance.

Research in the field of formative CBAs has focused on the development and evaluation of formative CBAs, effectiveness of feedback, and influences of formative CBAs. Overall, formative CBAs have been shown to be effective instruments in supporting student learning so much so that some researchers have claimed that using formative CBAs has positive effects on summative assessments. The least studied area is the evaluation of feedback used in formative CBAs. More research is needed to expand on Steven and Hesketh's finding (1999) exploring the effectiveness of different types of feedback that either identify common errors or link to additional learning resources.

The issues related to formative CBAs focused on item design and format, item banks, and item exposure. Given the absence of guidelines for the development of items used in formative CBAs, particularly the use of items (e.g., scenario-based multiple choice items) that can assess students' higher level thinking, the development of items for formative CBAs can be challenging and possibly lead to the development of fewer items. Developing good quality items to assess higher cognitive thinking for any assessment whether it is a pencil and paper assessment or a CBA can be a challenge. The additional task of including feedback and creating scenarios makes the task even more complex.

Formative CBA is a developing area of assessment that is growing in use as it is developing. Much research is needed to guide the growth of this instrument that has been shown to support student learning through promoting a self-regulated learning environment. A better understanding of the contextual factors influencing students' experiences with this instrument as well as the effective use of feedback will contribute to this area of study.

CHAPTER THREE

Methodology

A multiple case study approach was used to examine two formative CBAs used in different professional learning programs at a Canadian university. The first formative CBA was used in a Musculoskeletal (MSK) Phase IIA Module offered at a School of Medicine and focuses on musculoskeletal concepts for first year M.D. students. The second formative CBA was used in the Assessment Module at a Faculty of Education which introduces classroom assessment practices to teacher candidates enrolled in a Bachelor of Education program. Each case was structured using a mixed-method approach as a means to broaden the scope of the study and to better understand the contextual factors in each case. The methods selected for this dissertation include: (a) instrument analysis of the formative CBAs' capabilities and item formats used in the formative CBAs in each of the two case studies, (b) questionnaire surveying students' preferences for different types of feedback and the potential factors that may influence students' experience with formative CBAs, and (c) focus groups to extend and interpret the findings in the questionnaires. This section begins with a rationale for employing a case study approach and is followed by a description of the participants and context of each case. Next, the three research methods are described and are followed by the respective methods of analysis.

Case Study Research

A case study is viewed as an effective method for empirical inquiries investigating contemporary phenomena within real-life contexts (Yin, 2003). The

phenomenon in question is the use of formative CBAs. The two case studies (Assessment and MSK Phase IIA Modules) were instrumental in learning about the effectiveness of each formative CBA as well as in identifying relationships that may exist between cases to extend current understandings of how formative CBAs can be used in higher education and the conditions under which formative CBAs might deliver their full potential as assessment for learning instruments.

Included in the description of the context is the background information related to the formative CBA, findings from previous course or program evaluations, a description of the formative CBA software and limitations the software may pose, and the process of formative CBA item development. The contextual information provided for the Assessment Module contains more detail given that the researcher was a member of the team responsible for the development and administration of formative CBA items. As a result, the researcher had direct insight into the development process that she does not have in the case of the MSK Module.

Case 1: MSK Participants

A sample of 73 students from the School of Medicine completed the survey. This sample represented 72% of the population of students in the first of their four year medical studies program. These students have completed at least 3 years of an undergraduate program and some have advanced Master's or Doctoral degrees. Students have completed one term of basic science instruction (including anatomy, physiology, and biochemistry) and are embarking upon the study of disease.

Case 2: Assessment Participants

A sample of 308 students from the Faculty of Education completed the survey. This sample represented 46% of the population of students in the fifth and final year of study. Students enter this fifth year either as a concurrent or a consecutive student. Concurrent students begin their teacher education in the undergraduate program at three different universities. Along with courses in the arts or sciences, concurrent students take courses in education and begin practice teaching in their first year of university. Courses in education include topics related to assessment such as the relationship between instruction and assessment, understanding Ministry assessment documents, and developing and using assessment instruments. In the final year of study, concurrent students join together at one institution for intensive studies in education. Consecutive students begin their teacher education after they have completed an undergraduate program and begin practice teaching in year five of their post secondary education. In the final year, students from the two programs take the same courses and spend 13 weeks practice teaching.

Mixed Method Design

Recently, researchers in numerous disciplines have become more willing to pose important questions that go beyond the scope of a single methodological paradigm. (Tashakkori, 2004). The resulting designs are described as the third wave of research methodology (Tashakkori & Teddlie, 2003). This does not mean that long-standing debates related to the incompatibility of paradigms (e.g., objectivism/positivism and objectivism/constructivism) are not continuing. For a number of researchers however

(e.g., Greene & Caracelli, 2003, Tashakkori & Teddlie, 2003), the research method needs to be grounded in decisions related to the phenomena. As a result, different combinations of methods or mixed methods may be necessary.

In this dissertation, a mixed method design employing both qualitative and quantitative research methods was used to answer questions that would not easily be answered using one method (Tashakkori & Teddlie, 2003). Document analyses, questionnaires, and focus groups were used to answer the research questions (see Table 1 matching the research question with the research method). These three methods of research are described below.

Table 1

Research Questions and Research Method

Research Question	Research Method
How is feedback channeled to students and what is the preferred method?	Instrument Analysis, Focus Group
To what extent are particular types of feedback perceived to be effective in supporting student learning?	Questionnaire, Focus Group
What factors, if any, contribute to students' perceived usefulness of formative CBAs.	Questionnaire, Focus Group

Instrument Analysis

An instrument analysis of the formative CBAs used in the two case studies involved a systematic review of the formative CBA capabilities and supporting

documentation. This analysis helped to obtain an understanding of the formative CBA operations and limitations. The instrument analysis provided information to describe the context of each case as well as to examine specifically how feedback is channelled to students. The analysis of the formative CBA software focused in four principal areas: (a) purpose and intended users of the formative CBAs, (b) items used in the formative CBA, (c) the method of feedback (e.g., hint button, automatically provided upon submission of an item or set of items), and (d) the timing of the feedback (e.g., promptly after responding to an item, at the end of the entire set of items). A chart was used to record these findings related to the analysis of the formative CBA software for each case study. When examining the different types of feedback used, a check list was incorporated into the chart documenting the frequency of each type of feedback.

At the item level, individual items were examined in terms of the type of item presented (i.e., true-false, multiple-choice, fill-in-the-blank, or matching). Secondly, the items were examined in terms of the complexity. Guidelines for writing items to assess higher cognitive abilities suggest presenting students with scenarios (cases), graphs, pictures, or tables that require students to apply theories, processes or other types of analysis learned to select the answer (Fanshawe College, 2008; Professional Testing Inc., 2006; University of Texas, 2007). Key words or phrases (e.g., when, where, what) were also used as indicators of lower cognitive skill being assessed. In addition, the number of distractors was recorded for each item given that an item with two distractors is less difficult that an item with three or four distractors (Rodriguez, 2005). Lastly, the type of feedback provided for each item was recorded using the categories: directions to a class

text, directions to class notes, providing additional information, and indicating correct or incorrect). Similar to the analysis of the formative CBA software, the findings from the analysis of the formative CBA items was also recorded in a chart. Given that table of specifications are recommended for recording items according to content and cognitive domains (Professional Testing Inc., 2006), the use of a table of specifications was also noted during this analyses.

Questionnaires

Questionnaires were used to measure the extent to which the two formative CBAs were accepted by students as a tool to support their learning and to identify factors influencing the usefulness of formative CBAs as perceived by students. The questionnaires for both cases were developed in parallel and were both a subset of a larger questionnaire used to evaluate the two courses. A draft of the questionnaire was developed after a thorough review of formative CBA literature. The format of the questionnaires employed both qualitative and quantitative items. Quantitative items were used to extract basic views related to the usefulness of formative CBAs. These items were measured using a five- or seven-point Likert scale anchored on each end with appropriate descriptors (e.g., strongly disagree/strongly agree). The five-point Likert scale was used in the MSK Phase IIA questionnaire because of the existing format presented on the scantron data entry cards. The seven-point Likert type scale was selected for use on the Bachelor of Education questionnaire because it has been shown that fewer respondents rely on the middle response option resulting in greater variance in response patterns (Dawes, 2008). In addition to the Likert scale items, nominal and ordinal scales

were included to explore students' experiences and views of formative CBA use. To explore more complex issues such as what could be done to enhance the usefulness of formative CBAs, one open-ended qualitative item was included.

The questionnaire items were grouped into the following areas of inquiry: (a) demographics; (b) purpose of formative CBA; (c) factors influencing the effectiveness of formative CBAs; (d) item difficulty, format, number, and content; and (e) feedback. All items were arranged in a sequential order of experience. Hence items exploring where students completed the formative CBA and students' perceptions of the purpose of the formative CBA appeared before items exploring students' perceptions of the feedback and whether they viewed the assessment as accurate and fair (see Appendix C for a copy of the questionnaire used at the Faculty of Education and Appendix D for a copy of the MSK Phase IIA questionnaire).

The initial questionnaire was reviewed by the two instructors in the Assessment Module and one expert in the field of measurement. Revisions were made based on these reviews. The final stage of questionnaire development included a think-aloud protocol documenting students' thought processes as they read aloud each questionnaire item (Ericsson & Simon, 1993). Although initially developed for use in psychology, think-aloud procedures are now being used to study human-computer interactions (Cotton & Gresty, 2006). Hence it seemed an appropriate method to refine the questionnaire items and instruction in the electronic questionnaire.

Four students volunteered to complete the think-aloud. Students were given general instructions to simply think-aloud and verbalize their thoughts (Ericsson &

Simon, 1993). All four students completed the think-aloud simultaneously allowing the students to engage in rich discussions surrounding the questionnaire items. The researcher did not interact with the students until after they had read aloud each item and their verbal comments were recorded. Interactions with the students were used to further probe students' thinking and seek clarification of their thoughts. Changes to the questionnaire focused predominantly on the language. The word *item*, for example, was used in one instance and was not understood by students. This word was replaced by *question* which students agreed seemed more appropriate. In addition, students were helpful in identifying typing errors and errors in punctuation.

The questionnaire was revised a second time reflecting the recommendations from the think-aloud participants. The final draft of the Assessment Module questionnaire was reviewed by the two instructors. The last set of revisions involved adapting the questionnaire to the MSK Module. The questionnaire required slight modifications to some items to accommodate (a) the 5-point Likert scale, (b) the paper-based format, and (c) to reflect the language of participants in the MSK Module. The draft of the MSK questionnaire was reviewed by the MSK course instructor and minor revisions were subsequently made. The final draft of the formative CBA questionnaire used in the Assessment Module consisted of 34 items. An additional ten items on the Assessment Module questionnaire and four items on the MSK Module questionnaire, obtained students' demographic information. The additional demographic items on the questionnaire for the Assessment Module were related to students' division, program, and specialization.

The questionnaire for the Assessment Module was administered on-line using the software Survey Monkey. The questionnaire for the MSK Module was administered in a pencil and paper format using scantron cards (specific details are described in the sections below). In both instances, the data were electronically downloaded onto a spreadsheet eliminating the need for data entry. In the latter case, students' responses to the open-ended items were written directly on the questionnaire sheet and were later manually inputted onto the spreadsheet.

MSK module questionnaire administration. The MSK Module questionnaire was administered in class on the last day of the 4 week module. The instructor broke mid-way during class and provided refreshments (coffee and donuts). The questionnaire was distributed during the break and students, on a voluntary basis, were asked to complete the questionnaire. As an incentive to complete the questionnaire, students could enter their name in a draw for 3 - $25 gift certificates to a local book store. Questionnaires were collected and processed by a course administrator for the MSK Module.

Assessment module questionnaire administration. The questionnaire for the Assessment Module was administered over a 3 day period for approximately 5 hours each day. Fourteen laptop computers were set-up in the lobby of the Faculty of Education and students passing by were solicited by the researcher or a research assistant. Although the laptop computers were tucked in a protected nook, the lobby was a high traffic area which enhanced face-to-face contact with participants when soliciting their participation. This method of administering the questionnaire combined the higher response rates often associated with face-to-face contacts in paper and pencil questionnaires with the

advantages offered using on-line questionnaires and overcame the low response rates normally associated with on-line questionnaires (Nulty, 2008). The face-to-face on-line format of the questionnaire positively influenced the response rate such that approximately 20% more students completed the questionnaire than the on-line format offered in the previous year (Miller, in press). Chalk holders were provided as further incentive which helped secure a high response rate.

Technical hurdles included activating what is known as a kiosk function in Survey Monkey. This function permits the questionnaire to be completed several times on one computer. The default setting for this feature is off since the IP (Internet Provider) address is automatically recorded limiting one questionnaire submission per IP address. The second problem occurred when the text boxes from a completed survey did not automatically clear or re-set when the questionnaire was re-started by another student. In this case, a function on the internet toolbar was activated to clear the text boxes when the questionnaire had been submitted. Given that we were using a wireless internet connection in a relatively remote area of the Faculty of Education, a router was required to strengthen the signal and prevent internet disruptions. The last technical hurdle involved disabling other wireless internet connections (outside of the Faculty of Education) to ensure that the wireless connection was made to the internet server located in the Faculty of Education. Fortunately, these technical glitches were quickly overcome and did not hinder the data collection.

MSK Data Cleaning

Given that data were electronically entered, there were no data entry errors. The data were first re-coded from characters (A, B, C, D, E) to numeric responses (1, 2, 3, 4, 5) using the recode feature in SPSS. A review of the response distributions revealed one record in which the response for one item (gender) was outside the acceptable range of values (i.e., 1, 2). In this case, the outlier was removed from the cell because it was not possible to confirm the correct response and the remaining responses were left intact. The questionnaires (n = 79) were analyzed to identify items that contained multiple missing responses that may have been caused by poorly worded or ambiguous items. The closed section items had a few questionnaires with a considerable amount of missing data. It appears that some students chose not to respond to the items exploring the formative CBA. As a result, questionnaires 1, 31, 54, 70, and 76 were removed from the data set because they did not contain any responses to items surveying their formative CBA experiences. In cases where data were randomly missing, cases were excluded from analysis using the pairwise option in SPSS, excluding cases only when they are missing data for a specific analysis. In addition, one questionnaire contained response patterns containing only 1's. Hence it too was removed from the data set leaving a total of 73 questionnaires.

Assessment Data Cleaning

Since this data set was electronically transferred from the online questionnaire to SPSS, the need for data entry and the associated potential for data entry errors were eliminated. The questionnaires (n = 308) were examined to identify items that contained

multiple missing responses that may have been caused by poorly worded or ambiguous items; no bad items were found. In cases where data was randomly missing, cases were excluded from analysis using the pairwise deletion option in SPSS which excludes cases only when they are missing data that is required for a specific analysis.

Data Analysis of Questionnaires

Descriptive statistics were used to describe the characteristics of both samples. Frequencies (number of responses and percent) were calculated for all items. In addition, means and standard deviations were calculated for the continuous items. The skewness and kurtosis were also calculated but only reported if an anomaly was found. To assess the normality of the data, the explore function in SPSS was used to describe the symmetry and distribution of responses. If the data violated the assumptions of normality (symmetrical, bell-shaped curve, which has the greatest frequency of scores in the middle, with smaller frequencies towards the extremes) (Pallant, 2005), the findings were reported.

To determine whether or not all items in the questionnaire (scale) were all measuring the same underlying construct (formative CBA), commonly known as the scale's internal consistency, Cronbach's alpha was calculated. A large alpha coefficient (0.7 or greater) indicates a more reliable scale. There is little agreement where the cut-off point lies; however, alpha coefficients above 0.7 are considered acceptable (Pallant, 2005). Both questionnaires (i.e., MSK and Assessment Module) were considered reliable in the respective samples with high internal consistency (Assessment Module, $\alpha = 0.80$; MSK Module, $\alpha = 0.92$).

Given that the questionnaire was purposely organized by the research questions posed and the subsections within a research question (Question 3), the questionnaire was divided into thematic groups. Second Order Factor Analyses were completed to determine if these thematic groups could be further subdivided into subscales. A principal component factor analysis was conducted within each the thematic groups, using eigenvalues over 1 to determine the number of factors within each thematic group. If more than one component was identified, a Varimax rotation was used to determine those variables that loaded onto each component (subscale). Cronbach's alpha coefficients were also calculated for these subscales and are reported within each of the sections below. Total subscale scores were then calculated for the subscales by summing the responses to the individual items and dividing by the number of items. The purpose of combining items in the subscales was to increase the power of analysis.

Inferential statistical analysis included analyses of variance (ANOVA) to identify group differences in the dependent variables. Age and gender were chosen as the independent variables for group comparisons in the School of Medicine. In the past, gender differences were reported in technological studies (Dorman, 1998) and there has certainly been a common believe that younger members of our society are more comfortable with computers. Given the increased presence of computer technology in society, it was expected that these differences would now be much less. In addition, older students who delay their entry into professional program may have life experiences that differentially influence their learning. The inclusion of gender as an independent variable provided a mechanism to examine this hypothesis. Similarly, the inclusion of age allowed

a similar examination. If gender and age differences were found, these independent variables were used in the subsequent analyses for the Faculty of Education case study. If no there were no differences for age and gender, these variables were not included in the Faculty of Education case study. Independent variables specific to the Faculty of Education included program, division, and assessment knowledge. The rationale for selecting these independent variables is presented in Chapter 5. If a significant difference was found in either of these independent variables, effect sizes were calculated and displayed in the ANOVA tables.

The distribution of students in education programs in general is not consistent across age or gender given that more females enter the teaching profession than males and most students enter the education programs directly upon completion of their undergraduate degree. This pattern of entry into education programs influenced the sample characteristics in the Faculty of Education case study resulting in a sample with a low number of males, particularly older males in the primary/junior division (11%). Based on this sampling anomaly in the Faculty of Education, comparisons of gender and age were only examined if the results from the MSK case study analyses identified important group differences. The most important analyses in the Faculty of Education case study focused on group differences attributed to program, division, and assessment knowledge. Adequate student representation in these groupings allow for statistical analysis that has the potential to influence practice in the Faculty of Education. The Assessment Module could feasibly be customized based on the differences in program (consecutive or concurrent) and division (primary/junior or intermediate/senior). Next,

the independent variable, assessment knowledge, is examined on its own for group differences and is based on the assumption that students enrolled in the specialized course on classroom assessment practices have acquired more assessment knowledge than students whose training in assessment was limited to the nine hour Assessment Module. Analysis is performed only at the intermediate/senior division given that the elective course was only offered in this division. Similar to the independent variables program and division, differences based on assessment knowledge have the potential to inform practice at the Faculty of Education. Lastly, the variable internet experience was not used as an independent variable in either case because 95% of students in both cases rated themselves at the high end of the scale.

In sum, the inferential statistical analysis included a two-way (2 X 2) ANOVA to compare group differences in age and gender in the School of Medicine case study. In addition, a two-way (2 X 2) ANOVA was used to compare group differences in program and division and a one-way ANOVA to examine group differences in assessment knowledge at the Faculty of Education. The assumption of normality was not considered an issue because the sample sizes in both questionnaires exceeded the minimum (i.e., 30+) (Pallant, 2005). Homogeneity of variance was tested using Levene's test of equality of error variance. If Levene's test was significant, a t-test with more rigorous assumptions was calculated.

Focus Groups

A focus group is a group interview involving the interaction of participants based on questions posed by a moderator (Morgan, 1997). Although Frey and Fontana (1993)

have attempted to differentiate between focus groups and group interviews, Catterall and Maclaran (1997) noted the difficulty in making such differentiation. Catterall and Maclaran (1997) described a focus group as those that convene for a discussion on a focused topic and generate data to further an area of inquiry resulting from the group dynamics. Although focus groups can serve as the sole method of inquiry for a study, they often serve as preliminary data in quantitative research generating survey items (Catterall & Maclaran). Alternatively, focus groups can provide follow-up data to interpret and extend the primary findings (Morgan, 1997); as was the case in this dissertation. Krueger and Casey (2000) also noted the use of focus groups in this manner and added that focus groups provide reliable and naturalistic data leading to other insights not revealed in questionnaires.

The focus group protocol used in this dissertation was established by Krueger and Casey (2000). Krueger and Casey (2000) noted the size of focus groups typically ranges from four to twelve participants and further advise researchers to over solicit due to last minute cancellations. The moderator of the focus group must be aware of group dynamics, the focus group process, and content of the discussion (Krueger & Casey, 2000). The researcher's prior experience and knowledge of focus groups as well as familiarity with the content suggested the researcher was well suited for the role of moderator. Although Krueger and Casey (2000) recommended using a moderating team to aid with moderating and note taking, efforts to solicit research assistants were unsuccessful. The researcher conducted the focus group individually and relied on field notes and an audio recording to document group discussions. An interview guide was

prepared for each of the two focus groups and consisted of an introduction outlining the purpose and procedure of the focus group and the questions. At the conclusion of the focus group, participants received remuneration as an expression of gratitude for their time and commitment to this aspect of data collection. Students in the MSK focus group received $20 and students in the Assessment focus group received $15 plus a carrying bag.

MSK focus group. The focus group was held on Tuesday March 25, 2008 from 5:00 to 6:00 p.m. This time was suggested by the course instructor as a time most students would be available. The focus group was held in a meeting room in the medical library accessible to students in the medical program. Students were solicited electronically via their class group e-mail address. Ten students registered for the focus group; eight of those students plus an additional three students arrived to participate. In total, eleven students participated in the MSK focus group. The researcher/moderator was positioned in the middle of a rectangular table and students sat on all sides of the tables.

Assessment focus group. The focus group was held on April 10, 2008 from 5:30 to 6:30 p.m. There were no classes scheduled during this time and the majority of students would have completed their last class prior to 5:30 p.m., hence a large number of students would be available to participate in the focus group. The focus group took place in a classroom at the Faculty of Education. Consequently, students did not travel far or wait to participate in the focus group. Participants were solicited electronically via their class group e-mail address. Thirteen students volunteered for the focus group and twelve

students were in attendance. The seating arrangement was the same as the previous focus group.

Focus Group Analysis

The focus group audio recordings were transcribed to prepare the focus group analyses. Conceptual analysis was used to determine the presence of common words or phrases in the transcriptions such that inferences could be made to reflect the views of the students. For example, the following question was posed in the MSK focus group: Questions were sometimes embedded within the module; others were presented as a quiz. Discuss your preference for these or other formats? The conceptual analysis focused on the frequency of responses that indicated a preference for: (a) embedding questions within the module, (b) embedding questions in the quiz, (c) embedding questions in the module and in the quiz, (d) neither method of embedding questions, or (e) other. The coding of this question involved identification of pre-defined categories. In the event a category was unforeseen, the flexibility of the coding scheme allowed for the addition of such categories. The coding was also flexible in the sense that a number of different phrases could be classified into one category. For example, students who responded to the above question with: “I like it in the quizzes”, “quizzes is better”, and “embedding questions in the quiz is better” would all be coded the same. Information provided in the transcripts that was relevant or unimportant was discarded.

Summary

To obtain a full understanding of each case study, an instrument analysis was first performed to understand the strengths and limitations of the formative CBA software

used. In addition, the items developed for use on the formative CBAs were examined to identify their type (e.g., multiple-choice, true-false, open-response) and cognitive complexity. The primary method of data collection was through the use of a questionnaire. The questionnaires were developed in parallel and items were developed according to the five areas of inquiry. Instructors from both modules were involved in the development of the questionnaires and a field test was done to ensure the items were clearly understood by students. This field test was completed by a sample of four students who used a think-aloud protocol while completing a draft of the questionnaire. The questionnaire for the Assessment Module was administered on-line using the software Survey Monkey, while the questionnaire for the MSK Module was administered on paper and responses were recorded using scantron cards. Questionnaire data analyses included descriptive and inferential statistics. Cronbach's alpha coefficients were calculated to measure the internal consistencies of the scale and factor analyses helped confirm the grouping of items into thematic groups. In addition, ANOVA was used to determine group differences on the independent variables. Lastly, focus groups were used to extend the findings in the two questionnaires. A single focus group was conducted with a sample of students in the each case study. A conceptual analysis was performed of the focus group transcripts to identify the presence of common words or phrases in the transcripts.

CHAPTER FOUR

School of Medicine Musculoskeletal Module

Results from the instrument analysis, questionnaire exploring students' experiences with their formative CBA, and follow-up focus group are drawn together to describe the findings related to each of the research questions. This chapter begins by describing the context of this case study. Next, the findings drawn from the three methods of inquiry are presented. These findings are grouped according to: delivering feedback (Question 1), types of feedback (Question 2), and factors affecting students' experiences with formative CBAs (Question 3). This last section is further organized into the three sections examining the purpose of the formative CBA, the value of the formative CBA, and the item characteristics (difficulty, format, number, and concepts assessed).

The methods used to direct feedback to students were examined through an instrument analysis. The instrument analysis focused on the capabilities and characteristics of the formative CBA software and the characteristics of the items developed for the formative CBAs (e.g., cognitive complexity of item and type of feedback used). Students' preferences for the different types of feedback and the factors affecting students' experiences with formative CBAs were drawn from a 44 item questionnaire distributed at the end of the MSK module. Findings from the questionnaire were extended in the focus group held a few weeks following the completion of the questionnaire. The findings begin with a presentation of the descriptive statistics which are then interwoven with the findings from the focus group. Lastly, the inferential

statistics exploring differences related to gender and age across the dependent variables are presented.

Context of Case Study

The Musculoskeletal (MSK) Module spans four weeks (a total of 16 hours) and is taught using a blended e-learning approach. This instructional approach is distinctly different from students' experiences in preceding courses, which were almost exclusively didactic. The aim of the MSK Module is to introduce students to common and significant injuries and diseases involving the MSK system, as well as related topics such as child abuse and injury prevention. Secondary goals of the module included the promotion of collaboration and self-regulated learning. Given the aims of the module, the purpose of the formative CBAs used in the MSK Module is to support student learning through self-assessment of topics presented in the module. Completion of the formative CBAs aims to build students' competency level.

The lead instructor for the MSK Module began using HotPotatoes CBA software (Half-Baked Software Inc. 2008) in 2006 when the initial draft of items was created. The lead instructor oversaw the creation and collation of these items (see Appendix A for a sample of items) from a number of different course instructors with the help of a graduate student. The items were based on content adapted from the previous didactic versions of the course and feedback was created based on instructors' pedagogical knowledge of the curriculum and prior experience with assessing students using pencil and paper instruments. A Table of Specifications was not used in the creation of these items. The HotPotatoes CBA software was replaced by ThinkingCap Campus in 2007. The MSK

Module is currently supported by a website containing ten units of study that supplements the material presented during class. Students were instructed to complete each unit and had the option of completing the formative CBAs embedded throughout the units of study. Units of study were followed by summative quizzes based on the material presented on the website and in additional readings.

There were 9 units of study and each unit of study contained one or two formative CBAs for a total of 14 formative CBAs. The total number of items presented on the formative CBAs was 51. The cognitive complexity of items in the MSK Module ranged from knowledge based items to items requiring deeper thinking. The first indication of the complexity of an item was determined by the use of key words or phrases such as "how did ...", "what is ... ", and "which of the following." However, upon further examination, an item that begins with "what is ..." did not necessarily indicate a low cognitive item. For example, the first item in the Fractures in Children formative CBA asks "what is the diagnosis." In this item, the identification of a diagnosis was evaluated to be a higher cognitive item given that students must understand the basic concepts related to a fracture and apply them to the context of the scenario. Hence, an examination of the entire item was needed to identify the cognitive complexity. More than half of the items were found to be at the lower end of the cognitive scale (this was subsequently confirmed during the focus group) (see Appendix F for the record of instrument analysis).

MSK Module Demographics

Of the 73 students who completed the questionnaire, males (n = 40, 54.8% of the sample) slightly outnumbered the females (n = 33, 45.2% of the sample). Most of these

students (48, 65.7% of the sample) were in the 20 to 25 age range. Twenty students (27.4%) were in the 26 to 30 age group and six (8.2%) students fell into the oldest age group, 30 and above. Due to the low number of students in the last age group, the age categories were reduced from three categories to two. Students aged 20 to 25 were those students who began their medical studies directly from or soon after an undergraduate program. Students above the age of 26 were considered to be those who had other life experiences that delayed their entry into the program. These categories are important because prior experiences may influence students' learning and use of assessment feedback to support their learning..

The majority of students who completed the questionnaire reported they frequently used meta-research tools (data-base search engines on the internet) (63.0%) or used information located on the internet to complete course assignments (34.2%). It was suspected that this familiarity with the internet likely did not hinder or limit students' experiences with their formative CBAs since students claimed to have reasonable amounts of technological expertise. The two other categories for this item (c. I rarely used the Internet to look for information and d. I had never used the Internet) were each selected by only one student.

Approximately half of the MSK students had prior experience using CBAs (formative or summative) (Item 40). Of the 13 students who provided additional information on the open-ended item, six students reported using WebCT. Other CBAs used included Blackboard, CAPA (Computer Assisted Personalized Approach), and a CBA used in the MCAT (medical college admissions test) prep course. Students' prior

experiences with CBAs tended to be in science courses (e.g., Physics, Chemistry, Biology) while use in Greek, Religious Studies, and Physiology were also reported. In terms of using ThinkingCap (the formative CBA for this course), students reported that the interface did not pose any problems (Item 72: $M = 3.56$, $SD = 1.20$) and assistance using the software was not required (Item 73: $M = 4.18$, $SD = 1.14$).

There was a high quiz completion rate given that 84% of the students reported they completed all formative CBAs. The quizzes were predominantly (68, 93.2%) completed off-campus or a combination of on- and off-campus locations (Item 41). While most students reported they were able to complete the quizzes at a convenient time and place, a small percentage of students (15.1%) indicated this was not possible (Item 42).

The two independent variables (gender and age) were used to compare student responses in the items exploring the types of feedback and factors affecting students' experiences with formative CBAs. Although differences in gender were not anticipated in professional programs, this grouping variable was still included because there has currently been little research comparing male and female uses of formative CBAs. Age was anticipated to pull-apart the population based on the potential differences in life experiences and by extension, experience with formative CBAs. It was hypothesized that the youngest age category might respond more positively due to a greater and more recent exposure to technology and formative assessment practices in education.

Research Question 1: Channeling Feedback

The formative CBA used in the School of Medicine was ThinkingCap Campus version 1.3. This formative CBA is a commercial product designed for use in educational

settings. ThinkingCap Campus 1.3 entered the market of e-learning products in 2004. It is a web-based tool that facilitates on-line learning providing student reports, managing courses, and delivering courses via audio, video, images, and flash in addition to the standard text. Self-directed learning using ThinkingCap is managed using objects such as on-line tutorials or quizzes. Instructors use pre-designed templates to create learning objects that can contain text, images, sound, animations, or videos as well as quizzes using selected and constructed response formats. The items created for the formative CBA in the MSK Module consisted primarily of multiple-choice items with some matching, short-answer, and one word scramble. In addition, a few open-response items were included requiring students to self-assess their response using a model answer that was provided when the check command button was selected.

Feedback in ThinkingCap Campus was provided immediately after a student responded to an item. Students clicked on a "submit" button and a feedback window opened displaying the instructor's feedback to the item. MSK instructors included feedback for all distractors in multiple-choice items as well as for open-ended items. In most cases, the feedback on multiple-choice items consisted of additional information, although, the statement *correct* or *incorrect* was used occasionally.

During the focus groups, students were asked whether they would prefer feedback that was presented at the end of the assessment (as it was in WebCT) or following each item (as it was in ThinkingCap). The discussion focused on weighing the advantages and disadvantages of both methods. It was helpful that five out of thirteen students acknowledged that they had experience with WebCT and were aware of the ways

feedback could be provided. Students in the focus group were all in favour of feedback that immediately followed each item and expressed strong opinions stating, “embedded is better” (MSK FG S1) and “I am totally with Tom [pseudonym], give it right then and there” (MSK FG S3). While students strongly favoured immediate feedback at the item level, one student acknowledged the value of receiving feedback at the end of an assessment in the form of a summary sheet (MSK FG S6). This comment was not opposed by other students in the focus group. In terms of timing, it is evident that students from the School of Medicine preferred feedback that is provided immediately following a response to an item.

Research Question 2: Feedback

Results from the instrument analysis, questionnaire, and focus group are drawn together to describe the findings related to Question 2: To what extent are particular types of feedback effective in supporting student learning? This section begins by describing the type of feedback provided on the formative CBA in the MSK Module. Next, the findings from the questionnaire are presented. The presentation of the questionnaire findings (employed a 5-point Likert scale) begins with the descriptive statistics that are then interwoven with findings from the focus group. Lastly, the inferential statistics examing relationships between the independent variables (i.e., gender and age) are presented.

Descriptive Statistics

Feedback provided in ThinkingCap was created by the instructor responsible for creating the section within the MSK Module. ThinkingCap does not provide a framework

or criteria for developing feedback. In addition, there were no common guidelines provided to the individual instructors for creating feedback. As a result, the format of the feedback used in the MSK Module varied. The feedback in the MSK Module included a variety of statements such as: (a) Incorrect; (b) Correct; (c) Incorrect: crepitus is most commonly associated with osteoarthritis which is usually non-inflammatory; and (d) Correct: agricultural injuries are acute events and this type of data is best expressed using incidence rates. The majority of the feedback provided in each of the formative CBAs was similar to examples (c) and (d) shown above.

On the 5-point Likert scale where 5 represented strongly agree and 1 represented strongly disagree, students reported they reviewed the feedback (Item 66: $M = 3.94$, $SD = 0.99$) and found that it was useful (Item 67: $M = 4.01$, $SD = 0.95$). Feedback indicating whether a student's response was correct or incorrect followed by further information clarifying the response was favoured over feedback which simply stated a response was correct or incorrect (see Table 2). A paired sample t-test revealed a significant difference between the two main types of feedback (i.e., feedback with and without a supplement to clarify/confirm thinking).

Table 2

MSK Module: Effectiveness of different types of feedback

Item	*M*	*SD*
62. Feedback that stated my answer was **incorrect** was effective in supporting my learning.	4.15	0.90
63. Feedback that stated my answer was **correct** was effective in supporting my learning.	4.07	0.92
64. Feedback that stated my answer was **incorrect** AND **provided further information** clarifying the incorrect response was effective in supporting my learning	4.54	0.82
65. Feedback that stated my answer was **correct** AND **supplemented it** with why the answer was correct was effective in supporting my learning.	4.47	0.85

Delving further into student preferences for different types of feedback during the focus group, all students agreed that if an item was assessing a low cognitive ability then feedback statements of correct or incorrect (without any supplement) were acceptable. One student explained this idea stating: “if it is a simple question [for example], how many bones are in the body? Then, yeah” (MSK FG S2). After reviewing all feedback statements in the 10 modules, feedback that stated a response was correct or incorrect, generally did not appear to be used in this manner. For example, an item in the epidemiology module posed the question: What descriptive statistic did the authors calculate in this study? This knowledge or recall based item could warrant stating a student’s response was simply correct or incorrect. However, the feedback included for the three distractors was: (a) Incorrect; (b) Correct: agricultural injuries are acute events

and this type of data is best described using incidence rates; and (c) Incorrect: prevalence is often used when describing chronic diseases and/or injuries. In this example, the instructor provided no supplement for one incorrect response yet supplemented the other incorrect response and correct response with additional information. Although feedback for recall or knowledge based items was not delivered in the manner students recommended, it is important to be cautious in adopting this guideline for delivering feedback because students did not actually experience this manner of providing feedback. Further, an item may be simple recall for one student while another student may still be learning the concept and benefit from feedback beyond stating whether the response was simply correct or incorrect. This focus group discussion provided a potential guideline for when it may be appropriate to use incorrect and correct feedback statements.

If an item was tapping into higher cognitive skills, students unanimously agreed there was a need for a lot of feedback; even when their response was correct. In the case of a correct response, the students indicated a need for feedback to confirm their thinking. In this discussion focusing on feedback (for both correct and incorrect responses) for items assessing higher cognitive skills, two students voiced their views with statements such as, “You need further explanations to validate your thinking” (MSK FG S5) and “I like the one [type of feedback] that says you are correct and then it tells you why” (MSK FG S3).

The students in the focus group all agreed that it was helpful to know why an incorrect answer was wrong. One student commented, “I find it more helpful to know why the answer is wrong” (MSK FG S4). Students believed there was a greater need to

provide feedback for incorrect responses over correct responses. Students cited correct responses achieved by guessing as a reason for not including feedback to correct responses. Two students in the focus group argued that it was not a requirement of a good formative CBA to provide feedback to confirm guesses. Rather, the two students felt students had some responsibility towards their learning. One of the students who raised this issue concluded, “There has to be some self-analysis. Right?” (MSK FG S7) and a third student added, “You kind of know if you are guessing. If it says correct, and you guessed at it, then you have some responsibility to [your learning]” (MSK FG S4). Although students did not feel it was as important to include feedback to correct responses, students all agreed with the need to provide feedback to supplement responses. Towards the end of the discussion one student summed this consensus for feedback with

> I think that if there was some way to incorporate examples three [correct response + supplement] and four [incorrect response + supplement] at the same time that could be more helpful. For example, if you got the correct answer, then you also need to know why the other ones are not correct. Sometimes when I got the correct answer, I also want to click on the incorrect answers to learn why they are incorrect. (MSK FG S5)

Based on this focus group discussion, it appears that feedback explaining why a response is correct or incorrect are both valuable.

The discussion amongst the students indicated that this group of students appeared to exhibit high degrees of self-regulation. They had a definite view that they were responsible and active participants in their learning. For example, one student spoke

against feedback that "short circuit[ed] the rest of the thinking process" (MSK FG S6) by simply providing the correct response. Subsequently, students suggested that feedback should provide a tip for further learning instead of simply providing the correct answer. This evidence of high self-regulation may be influencing the type and amount of feedback students recommend for formative CBAs. Consequently, a student's degree of self-regulation may be a factor influencing the value of feedback on formative CBAs. Students with high levels of self-regulation may have different needs in terms of the type and amount of feedback required to support their learning.

The discussion related to feedback continued in terms of exploring feedback that directed students to various types of resources. Overall, students preferred directions to a resource using an internet link that supported the concept being assessed. Students noted it was necessary that when using an internet link, the link must be active and contain appropriate information to support their learning.

Feedback that directed students to their course notes was not viewed as useful given that, as one student commented,

> Some people are really organized but not everyone has the same organization. I like referring to a course text because in my notes, I may not know where I put it … like I don't know. Sometimes there is a lot of overlap between different lectures so if you say go to malnutrition it may refer to more than one set of course notes. (MSK FG S5)

Directing students to course notes is only valuable if students maintain their course notes. While directions to the course text were preferred over course notes, the majority of the

focus group participants generally preferred an internet link over both the course notes and their course text.

Additional comments related to feedback were obtained from the open-ended section of the questionnaire. One student recommended documenting a correct response on an x-ray when used (e.g., circle or add an arrow to identify the fracture location) (MSK Q S43). Another student recommended using video and audio resources to increase the authenticity of the assessment (MSK Q S19). For example, an item could include an audio recording of a patient describing their symptoms. These comments suggested that students considered their formative CBA an asset and expressed a desire to use the instrument to support learning in new ways. Hence the comments provided by this group of students reflected positive improvements rather than dwelling on negative experiences.

Inferential Statistics

Students were grouped according to their age (20 to 25 or 26+) and gender (Males or Females). Four items exploring feedback surveyed student preferences for different types of feedback (i.e., Items 62 - 65); consequently, these items were not intended to be combined. The remaining two items (Items 66 and 67) exploring feedback surveyed students' attitudes towards the effectiveness of the feedback provided on the formative CBA and were combined to create one factor. Separate two-way (2 X 2) ANOVA were performed using the Bonferroni correction (0.05/5) to correct for Type I errors. The main effects for age and gender and the interaction effect did not reach statistical significance (see Table 3).

Table 3

MSK Module: ANOVA - age and gender x *feedback*

Source	*SS*	*df*	*MS*	*F*	*p*
Feed. incorrect					
Age	0.02	1	0.02	0.02	0.893
Gender	0.00	1	0.00	0.00	0.983
Age x Gender	0.00	1	0.00	0.00	0.989
Error	55.94	1	0.85		
Total	1274.00	66			
Feed. correct					
Age	0.11	1	0.11	0.12	0.726
Gender	0.04	1	0.04	0.05	0.825
Age x Gender	0.00	1	0.00	0.00	0.961
Error	57.31	65	0.88		
Total	1210.00	69			
Feed. incorrect+					
Age	0.48	1	0.48	0.72	0.401
Gender	1.24	1	1.24	1.86	0.177
Age x Gender	0.46	1	0.46	0.70	0.401
Error	43.98	66	0.67		
Total	1499.00	70			
Feed. correct+					
Age	1.55	1	1.55	2.20	0.143
Gender	1.05	1	1.05	1.48	0.228
Age x Gender	0.36	1	0.36	0.51	0.479
Error	45.20	64	0.71		
Total	1415.00	68			
Feed. Attit.					
Age	0.01	1	0.01	0.01	0.939
Gender	0.47	1	0.47	0.57	0.462
Age x Gender	0.00	1	0.00	0.00	0.987
Error	55.27	65	0.85		
Total	1160.00	69			

Research Question 3: Factors Effecting Formative Computer-based Assessments

Items believed to have some effect on students' experiences with formative CBAs were grouped according to students' perceptions of the: purpose (Items: 43, 45, 47 – 51, 59), value (Items: 57, 58, 68 – 71, 74, 75), and item characteristics (Items: 46, 52 – 56, 60, 61). The findings related to these three areas of inquiry are presented separately below.

Purpose of Formative CBA

Descriptive Statistics

Based on student responses, there was a strong overall belief that the purpose of the quizzes was to support student learning (see Table 4).

Table 4

MSK Module: Perceptions of the purpose of the formative CBA

Items	*M*	*SD*
43. The concept of using computer-based quizzes for self-assessment is a good way to direct student learning.	4.03	1.21
45. I collaborated with colleagues when I completed the quizzes	1.84	1.14
47. The purpose of the quizzes was to direct my learning.	3.99	0.99
48. The purpose of the quizzes was <u>not</u> part of my mark in the class.	3.79	1.24
49. The purpose of the quizzes was a diagnostic assessment to help plan instruction.	3.05	1.27
50. The purpose of the quizzes was to identify my strengths and weaknesses.	3.96	1.04
51. I was sure of the purpose of the quizzes.	3.51	1.15
59. The quizzes were helpful in starting discussions with colleagues about the topics presented in each module.	2.92	1.32

Although student collaboration (Item 45) on the quizzes and using the quizzes to start discussions with colleagues (Item 59) were intended to be indicators of students' understanding of the formative nature of the quizzes, these items did not function in the manner intended. On the questionnaire, students generally reported they did not engage in collaboration. Given that this formative assessment was intended to support collaboration or use the formative CBAs to start discussions with colleagues, the outcome of this item was reviewed in the focus group. Students were familiar with this form of learning given

that collaborating and discussing case studies was a major focus of the MSK Module. When students were asked why they chose not to collaborate, one student commented,

> I think it is because we were all doing them [formative CBAs] on our computers alone at home … we weren't in group situations so you would have to write down your question and e-mail it to somebody and that takes time. (MSK FG S7)

Several students nodded in agreement with this statement and no students voiced an opposite view. Based on this comment, it appears that students understood the purpose of the formative CBA; however, the need to collaborate was not a function of the purpose. Another student suggested there was no need to collaborate because, "a lot of the questions were easy anyway" (MSK FG S10). It is possible that the need to collaborate is dependent on the cognitive complexity of the items rather than an understanding of the purpose of the formative CBA. Given that these two items (Items 45 and 59) did not function in the manner intended, they were removed from this thematic group prior to the inferential statistical analysis.

Inferential Statistics

A principal component analysis was performed to determine whether the six items in this subscale could be reduced. The factorability of the correlation matrix was supported by the presence of many coefficients greater 0.3, a Kaiser-Meyer-Oklin value (0.70) greater than the recommended 0.6, and a significant Bartlett's Test of Sphericity. The principal component analysis revealed the presence of two components with eigenvalues exceeding one, explaining 41.4 and 24.0 % of the variance (respectively). Two components were retained and subjected to a principal component analysis with

Varimax rotation. The factors loadings are displayed in Table 5. Item 51 (I was not sure of the purpose of the quizzes) did not load well on either component and was thus removed from this subscale. The factor analysis with Varimax rotation revealed a two factor solution where Items 47, 43, and 50 were combined into one factor: *Supporting Self-directed Learning* ($\alpha = 0.82$). Items 48 and 49 were combined into a second factor: *Supporting Diagnostic Assessment* ($\alpha = 0.55$). Although this alpha coefficient is less than the minimum recommended coefficient of 0.7, smaller coefficients, often associated with scales with less than 10 items, is acceptable (Pallant, 2005)

Table 5

MSK Module: Perceived purpose of formative CBA

Item	Component 1	2
47. The purpose of the quizzes was to direct my learning.	0.87	0.03
43. The concept of using computer-based quizzes for self-assessment is a good way to direct student learning.	0.84	-0.12
50. The purpose of the quizzes was to identify my strengths and weaknesses.	0.81	-0.09
51. I was sure of the purpose of the quizzes.	0. 46	-0.45
48. The purpose of the quizzes was part of my mark in the class.	-0.01	0.82
49. The purpose of the quizzes was a diagnostic assessment to help plan instruction	0.38	0.75

A two-way ANOVA was conducted using the Bonferroni adjustment ($p = 0.05/2$) to explore the impact of gender (male or female) and age (20-25 or 25+) on the two factors. There were no statistically significant differences for the main or interaction effects for the first factor, Supporting Self-directed Learning. A significant difference ($p < 0.05/2$) between the interaction of age and gender was found for the second factor (see Table 6).

Table 6

MSK Module: ANOVA - age and gender x *purpose*

Source	*SS*	*df*	*MS*	*F*	*p*	*Effect Size*
Self-dir.						
Age	1.45	1	4.45	2.12	0.150	0.03
Gender	0.09	1	0.09	0.13	0.718	0.00
Age x Gender	0.00	1	0.00	0.00	0.974	0.00
Error	46.03	67	0.69			
Total	1216.00	71				
Diag. ass.						
Age	0.61	1	0.61	0.70	0.406	0.01
Gender	3.20	1	3.02	3.69	0.059	0.05
Age x Gender	5.07	1	5.07	5.84	0.018	0.08
Error	56.38	65	0.87			
Total	549.00	69				

Based on the mean scores, there appeared to be a difference for males across age (older males ($M = 3.26$, $SD = 1.00$ and younger males: $M = 2.31$, $SD = 1.00$), but a slightly reversed effect for females (older females: $M = 2.0$, $SD = 0.82$ and younger females: $M = 2.46$, $SD = 0.83$). The older males tended to be the most positive while the older females

tended to be the least positive. The causes for this interaction are not clear. Further, given their relatively small values, the findings may have little importance.

Value of the Formative CBA

Descriptive Statistics

Factors such as priority, perception of the task, and personal schedules did not appear to be barriers limiting students' experience with the formative CBA. Table 7 summarizes the findings for these four items. In the survey, each of these items was negatively worded. Students who agreed with the statement were expressing a negative judgement about the quizzes and would respond at the high end of the five-point Likert scale. Prior to analyses, the responses for these items were reverse coded so that their value and meaning would be consistent with other items. Hence the means reported in Table 7 indicate that students were positive about completing the formative CBA quizzes. The other two items intended to reveal how students valued the formative CBAs, namely, Items 57 and 58 also had positive responses (see Table 7). These items were further explored within the context of the focus group to determine what was needed to generate more positive responses. Focus group participants identified specific factors that limited or hindered their experience with the formative CBA. Two students indicated that an overview of the time required to complete the formative CBAs would have helped them prioritize their schoolwork. One of the two students reported, "the module was a little bit big and I didn't get it all done … if I had known how long it would have taken, that would have helped" (MSK FG S1). The second student added "to me it is good to have an overall view of the module" (MSK FG S2). When probing these comments further, all

students agreed that knowing the approximate amount of time it would take to complete each module, including the formative CBAs in each module, would have be useful in planning their schoolwork schedule.

Table 7

MSK Module: Value of formative CBAs - I

Item	*M*	*SD*
68. The quizzes were just 'busy work'.	1.18	1.00
69. The quizzes were not a priority for me.	1.35	0.98
70. I was too busy to complete the quizzes.	1.06	0.94
71. I did not put effort into completing the quizzes.	1.12	1.09
57. The quizzes advanced my understanding of the topics presented in each module.	3.90	0.94
58. The quizzes guided my learning in each module.	3.97	0.93

Another finding from the focus group was that the formative CBAs did not appear, in students' eyes, to prepare them for the ongoing summative quizzes or final examination. This misalignment appeared to decrease the value of the formative CBAs. This finding was supported by comments such as, "I know I would do the questions a lot more if there was like one of the questions given [on the formative CBA] would be on the final exam" (MSK FG S3) and "there was very much a big jump between what I did on the quizzes and module readings and what was questioned on the RATS [summative assessment]" (MSK FG S4). One student aptly described their uncertainty regarding the

alignment by stating “I don’t know whether I feel confident per se that I am going to go in and … get the same score that I scored on the quizzes [formative CBAs]” (MSK FG S5). As a set, these comments suggest that the better the alignment, the higher the value students may have of the formative CBAs. Such an alignment would result in formative CBAs that purposefully prepare students for the summative assessment.

The final analysis of value of formative CBAs explored students’ perceived accuracy and fairness of the instrument (see Table 8). Only 50% of students agreed or strongly agreed that their scores on the formative CBA quizzes were an accurate reflection of their understanding of the concepts presented in each module. Given that students had previously reported that the quizzes were representative of the module content, were completed with effort, and employed a favourable format, this result was further explored in the focus group.

Table 8

MSK Module: Value of formative CBAs - II

Item	*M*	*SD*
74. In retrospect, my results on the quizzes were an accurate reflection of my understanding of the concepts presented in each module.	3.41	0.97
75. The quizzes presented in each module were fair assessment instruments.	3.54	0.94

In discussions about fairness and accuracy, one student responded, “we intuitively know the exam questions are not going to be like what we see on these quizzes” (MSK

FG S3). This sentiment was supported by nods of agreement. Another student further explained stating,

> When we were saying they didn't really match-up with the RATs [summative quizzes] that is probably what people are thinking. I don't feel as irritated right now about the disagreement between the two of them [formative CBAs and summative quizzes] but I remember being really irritated by that because you go through the module doing all the work and then the stupid RATS were out of nowhere. That's probably what that is about. (MSK FG S1)

One student attributed these views towards fairness and accuracy to a misunderstanding about the purpose of the formative CBA. This student explained,

> Quizzes were intended to give students a base understanding of what is to be expected on the RATS; hence, in this vein, they were fair. There was an underlying assumption that the quizzes should prepare students to score well on the summative assessments. Based on that reasoning, students did not feel the quizzes were as fair as they should be because they do what they were intended to do. (MSK FG S5)

The idea of fairness appears to include the issue of aligning the difficulty of the formative and summative instruments. In particular, the formative CBAs were perceived to assess simpler ideas and demand less complex thinking than the summative quizzes and final examination. The variability in responses to Items 74 and 75 may thus reflect some students desire to receive feedback on the full range of their learning. This would

require assessment items that span the full range of difficulty similar to the complexity of items found in students' summative instruments.

Inferential Statistics

The nine items in this subscale (of continuous items) were subjected to a principal component factor analysis. Coefficients of 0.3 and higher, a Kaiser-Meyer-Oklin value of 0.79 and a significant Bartlett's Test of Sphericity support the factorability of the correlation matrix. Principal component analysis revealed the presence of two components with eigenvalues exceeding 1 explaining 40.2 and 22.9% of the variance (respectively). An inspection of the scree plot revealed a clear break after the second component. As a result, two components were retained and Varimax rotation was performed. The rotated solution revealed strong loadings with many items loading in both components. These findings are displayed in Table 9 below. The first component could be described as *Value Completing the Formative CBA* (Items: 71, 68, 70, 69) and the second component as *Value Reviewing the Formative CBA* (Items: 58, 75, 74, 57). The two factors had Cronbach alpha coefficients of 0.88 and 0.74 respectively.

Table 9

MSK Module: Value of formative CBA- III

Item	Component 1	2
71. I put effort into completing the quizzes.	0.87	0.04
68. The quizzes were not busy work.	0.85	0.03
70. I was not too busy to complete the quizzes.	0.84	0.08
69. The quizzes were a priority for me.	0.82	0.16
58. The quizzes guided my learning in each module.	0.32	0.77
75. The quizzes presented in each module were fair assessment instruments	-0.08	0.72
74. In retrospect, my results on the quizzes were an accurate reflection of my understanding of the concepts presented in each module.	0.19	0.71
57. The quizzes advanced my understanding of the topics presented in each module.	0.36	0.68

A two-way (2 X 2) ANOVA was conducted to explore the impact of gender (male or female) and age (20-25 or 25+) across the two factors. The Bonferroni correction (0.05/2) was used to correct for Type I errors. The main effects for gender and age and the interaction effect did not reach statistical significance (Table 10).

Table 10

MSK Module: ANOVA - age and gender x *value*

Source	*SS*	*df*	*MS*	*F*	*p*
Completing					
Age	0.05	1	0.05	0.07	0.797
Gender	0.05	1	0.05	0.08	0.784
Age x Gender	2.68	1	2.68	3.78	0.056
Error	45.33	64	0.71		
Total	1061.75	68			
Reviewing					
Age	0.01	1	0.01	0.01	0.912
Gender	0.00	1	0.00	0.00	0.994
Age x Gender	0.02	1	0.02	0.04	0.845
Error	27.15	61	0.45		
Total	946.88	65			

Items: Difficulty, Format, Number, and Concepts Assessed

Descriptive Statistics

The format of items on the formative CBAs used in the MSK Module included multiple-choice, true-false, fill-in-the-blank, matching, and open responses. Scenarios were embedded in multiple-choice and open response items. The multiple-choice and matching items ranged from low to medium cognitive levels while the open response items tended to tap into students' higher cognitive thinking (see Appendix F for the instrument analysis). Overall, students reported they were generally satisfied with the items (see Table 11). The findings were confirmed in the open-ended questionnaire item with statements like: "I like the questions because I was just reading up on the stuff" (MSK FG S2) or "It was good to make sure I knew everything" (MSK FG S7).

Table 11

MSK Module: Factors Influencing Items

Item	*M*	*SD*
46. The range of items in each quiz reflected the concepts presented in the corresponding module.	3.77	0.92
52. The number of questions in each quiz was appropriate in assessing my understanding of the concepts presented in the corresponding modules.	3.55	1.04
53. The difficulty level of questions was appropriate for these quizzes.	3.83	0.82
54. The topics selected for assessment were appropriate.	3.90	0.83
55. The use of scenarios was appropriate.	4.19	0.83
56. Each scenario used in the quizzes represented a meaningful dilemma.	3.89	0.81
60. Multiple-choice questions were useful in assessing my understanding.	3.94	0.80
61. Matching column questions were useful in assessing my understanding.	3.62	1.04

The characteristics of the items used on the formative CBAs are further discussed according to item difficulty, format, content, and concepts assessed.

Difficulty. Students were somewhat satisfied with the difficulty level of items presented on their formative CBAs (see Table 11, Item 53). To explore what is needed to increase students' level of satisfaction with the items and to confirm that half of the formative CBA items were at the lower end of the cognitive scale, focus group participants were asked to review the cognitive complexity of a few sample items. The general comments from this discussion revealed that the majority of the items were indeed at the lower to middle end of the cognitive scale echoing what was found during

the instrument analysis. The difficulty level of items appeared to be appropriate for some students. One student explained “I personally like stuff [items] that is not too complicated” (MSK FG S3). Another student agreed but extended the view stating,

> I like to have a few easy questions to build my confidence but at the same time I would like one or two difficult questions like the kind we discuss in class or we might see on the exam - so I can start challenging myself to think in that direction and have the same level of discussion that we would have in our group session. (MSK FG S5)

Students’ dissatisfaction with the items iterated a previous finding stemming from the lack of alignment between the summative quizzes and the formative CBAs in terms of item difficulty. One student commented that the summative quizzes often consisted of items assessing “obscure tidbits of facts” (MSK FG S2). Another student stated that the summative quizzes were also based on extra readings and suspected that students may have relied only on the modules and not on the readings, giving students a false sense of preparation for the summative quizzes (MSK FG S5). In sum, students confirmed the presence of a large number of items at the lower end of the cognitive scale. All students agreed with this conclusion and were not particularly opposed to the large number of items at the lower end of the cognitive scale because the formative CBA items and corresponding feedback nurtured their understanding of new concepts. On the other hand, students noted the need to include more items that were better aligned with items on the summative assessments, in particular, items at the higher end of the cognitive scale.

Format. Based on the findings presented in Table 11 above, students reported some degree of satisfaction with the format of multiple-choice and matching items (Items 60 and 61); the most common formats used in the formative CBAs. Using scenarios to situate the items appeared to be well liked by students (Table 11, Items 55 and 56). This positive response to using scenarios may in part be due to the structure of the course where concepts were grounded in medical scenarios. Secondly, the scenarios presented on the formative CBAs were brief synopses of a medical situations containing two to four sentences tapping into the middle of the cognitive scale.

The focus group was used to expand on the usefulness of the open-response format used in the formative CBAs. As previously noted, the open-ended items were self-assessed; the intention was that students would prepare a response and then click on a check button to determine whether their response matched the instructor's response. In the focus group, students were asked if they actually constructed a written response to the open-ended items; all students admitted they did not. Students claimed this item format had value and perceived this format to be challenging. For example, one student commented,

> When I get those open-response questions I recognize those as a challenge to think. Now whether or not I, because of time restraints, go through it, it still sparks in my mind – in my mind I just can't rely on having the answer somewhere there in front of me to know things well enough to think about it. So I think the open-ended questions have value and are challenging. (MSK FG S6)

Another student noted that "if I was running out of time, I would still do the multiple-choice but I wouldn't do the open-response. I would probably just skim this and then just click the box to read what the answer should be" (MSK FG S2). Hence it appears important to provide open-ended items due to students' need to be challenged cognitively. Yet, students may not commit the effort to complete these items given a lack of time. Students' selective use of open-response items that required more time to complete than multiple-choice items suggests that the time to complete a formative CBA is a factor in using formative CBAs.

Students in the focus group described their experience with a word-scramble question (only one item was in this format). This item format was reported to be frustrating and cumbersome given that other arrangements of the words or phrases were possible. As a result, two students (i.e., MSK FG S1 and MSK FG S4) admitted they found the correct answer by a trial and error process, indicating the item did not support their learning.

Number of Items. Slightly under half of the students (60, 45%) agreed or strongly agreed that the number of questions in each quiz was appropriate in assessing their understanding of the concepts in the corresponding module (Table 11, Item 52). However, when confirming students' satisfaction with the number of items in the formative CBA during the focus group, all students agreed that the addition of more items would be an improvement. Student responses from the open-ended item at the end of the questionnaire frequently (14 out of 31 comments) addressed the need for more items. One student commented on the questionnaire that the quizzes were "too short to

help workout problems or judge my true understanding" (MSK Q S12) while another student recommended, "more clinical scenarios" (MSK Q S26). It is evident that not only did students feel they would benefit from more items but there was also a need for items that assessed students' higher cognitive skills.

Concepts Assessed. Students were moderately satisfied with the concepts assessed on the formative CBAs (see Table 11, Item 54). When discussing the content covered on the quizzes during the focus group, once again, several students reported the need for a better alignment of content in terms of the focus of the content on the formative CBAs and that on the summative assessments. This issue was raised in the comment, "I remember thinking the modules were pretty heavy with text and thinking the [summative] quizzes are going to be brutal and then the quizzes didn't cover everything" (MSK FG S3). No students expressed opposing views. Overall, the content covered on the formative CBAs did not appear to be a large factor influencing students' experiences with the formative CBAs; however, the concerns related to the alignment of students' formative and summative assessments was also raised in this discussion.

Inferential Statistics

The eight items in this subscale were subjected to a principal component analysis where the correlation matrix revealed the presence of coefficients greater than 0.3, a Kaiser-Meyer-Oklin value of 0.79, and a significant Bartlett's Test of Sphericity. Two components with eigenvalues exceeding 1, explained 51.5 and 17.5% of the variance respectively, and the scree plot revealed a break after the second component. Two components were retained and Varimax rotation was performed (see Table 12 below).

Table 12

MSK Module: Perceived usefulness of formative CBA items

Item	Component 1	Component 2
54. The topics selected for assessment were appropriate.	0.87	0.04
53. The difficulty level of questions was appropriate for these quizzes.	0.85	0.03
52. The number of questions in each quiz was appropriate in assessing my understanding of the concepts presented in the corresponding modules.	0.84	0.08
46. The range of questions in each quiz reflected the concepts presented in the corresponding module.	0.82	0.16
55. The use of scenarios was appropriate.	0.32	0.77
60. Multiple-choice questions were useful in assessing my understanding.	-0.08	0.72
61. Matching column questions were useful in assessing my understanding.	0.19	0.71
56. Each scenario used in the quizzes represented a meaningful dilemma.	0.36	0.68

The items in the first factor (Items: 54, 53, 52, 46, and 55) represented the factor *Item Characteristics*. This factor had a Cronbach's alpha coefficient of 0.86. The second factor (Items: 60, 61, and 56) represented *Item Type* had a Cronbach's alpha coefficient of 0.80.

A two-way (2 X 2) ANOVA was conducted to explore the impact of gender (male or female) and age (20-25 or 25+) across the these two sub-scales using a Bonferroni

correction (0.05/2) procedure to correct for Type I errors. The main effects for gender and age and the interaction effect did not reach statistical significance (Table 13).

Table 13

MSK Module: ANOVA - age and gender x *items*

Source	*SS*	*df*	*MS*	*F*	*p*
Item Char.					
Age	0.78	1	0.78	1.80	0.185
Gender	0.65	1	0.65	1.51	0.223
Age x	0.06	1	0.06	0.14	0.712
Gender	28.53	66	0.43		
Error	1092.72	70			
Total					
Item Types					
Age	0.14	1	0.14	0.23	0.631
Gender	0.05	1	0.05	0.09	0.766
Age x	0.10	1	0.10	0.17	0.684
Gender	39.12	66	0.59		
Error	1067.89	70			
Total					

CHAPTER FIVE

Faculty of Education Assessment Module

Findings from the instrument analysis, questionnaire exploring students' experiences with their formative CBAs, and follow-up focus group are drawn together to describe the findings related to each of the questions posed in this dissertation. This chapter begins with a description of the context for this case study. Next, the findings drawn from the three methods of inquiry are presented in response to the questions. These findings are grouped according to: delivering feedback (Question 1), types of feedback (Question 2), and factors effecting students' experiences with formative CBAs (Question 3). This last section is further organized into the three areas of inquiry: purpose of the formative CBAs, value of the formative CBAs, and items (difficulty, format, number, and concepts assessed).

The findings exploring how feedback is channelled to students (Question 1) is based on an instrument analysis in which the features of the formative CBA were explored. The findings exploring students' preferences for the different types of feedback used in their formative CBA (Question 2) and the factors affecting students' experiences with formative CBAs (Question 3) are drawn from the questionnaire and then extended in the focus group. Lastly, the inferential statistical summary is presented which explores the relationships based on the independent variables program and division as well as the independent variable assessment knowledge.

Context of Case Study

In response to a recommendation from the Ontario College of Teachers, the Faculty of Education launched a compulsory course divided into four distinct modules (Assessment, Psychology of Learning and Development, Philosophy of Education, and Social Contexts of Education) for all Bachelor of Education students in the fall of 2006. Each of these four modules was scheduled for seven to nine hours of face-to-face instructional time.

The required material for the Assessment Module was presented in a lecture format and contained activities that could be accomplished in an auditorium classroom. Concepts introduced during the classes were supported by a website containing course notes and other supplementary material. Student assessment included a formative CBA and a summative project. A program evaluation of the Fall 2006 Assessment Module was conducted in the spring of 2007. Data collection involved retrieving comments from students' course portfolios and a questionnaire administered six months after the completion of the Assessment Module. Findings related to the formative CBA component of the program evaluation revealed that feedback on the formative CBA items was at least moderately effective in supporting student learning (Miller & Shulha, 2008).

The 2007 Assessment Module added two more hours of instruction resulting in a 9 hour module and an over-hauled website. This Module included formative CBAs, described as a Prior Learning Assessment consisting of four mini formative CBAs each containing three items. The purpose of the Prior Learning Assessment was to give students feedback on their current understanding of assessment principles and provide

opportunities for students to engage in self-regulated learning. Students were encouraged to use their results on the Prior Learning Assessment in setting goals for independent learning in the Assessment Module. The independent learning component required the student to identify a weak area, as reported on the formative CBAs, and pursue further learning in that area using the course website and other related resources.

WebCT was selected as the software for the formative CBAs because (a) the university held a license for the software, (b) the software was supported by the university with class lists and student information could be directly uploaded into the software, and (c) technical support was offered by the university. Although rumours related to the constraints of WebCT surfaced as this formative CBA was developed, there were no other options due to a lack of financial resources and a short timeframe to create the formative CBAs.

The criteria established for item development stipulated that each item would be created based on a scenario and each item would contain at least three distractors. The ideas, connections, and extensions (ICE) model of learning (Fostaty-Young & Wilson, 2000) would serve as the framework for item development. The ICE model is described by Young and Wilson (2000) as an approach to assessment loosely based on Bloom's Taxonomy (1956), which subdivides learning into three broad categories. Learning begins at the Ideas level, where students obtain basic conceptual facts, vocabulary, and details. At the Connections level students begin to make connections among these basic concepts, linking what was learned with prior knowledge and new ideas. At the highest level of learning, students deepen and extend their learning, creating Extensions amongst

what is learned in their classroom with other unique contexts and concepts. The ICE model was used in the Assessment Module to create an assessment framework that would ensure variation in students' thinking to correctly answer the different items. One third of the items were designed at the Extensions level requiring students to think beyond the scenario in order to correctly answer the question. In addition, the items were housed within four formative CBAs, representing the four domains of the module, What does Learning Looking Like, Assessment for Learning, Assessment of Learning, and Large-scale Assessment. A Table of Specifications was created to record the development of items documenting the cognitive complexity and content domains. Items (see Appendix B) were developed by the two course instructors and a research assistant.

Upon meeting to discuss the criteria for developing the items, the course content was divided among the three item developers and each was given the task of drafting three to four items in the assigned content area. Items were then shared with the group. Group discussions were lively as the complexity level for each question was determined and suitable vocabulary and terminology for Bachelor of Education students was crafted. This process of item development was extremely time-consuming due to the need to develop items that would allow students to demonstrate their assessment knowledge as well as less explored concepts. This task was compounded by the difficulty of identifying one level of cognitive complexity for each item. This was not possible in some cases because an item could tap into various cognitive levels or students' own cognitive development could differ. A complex item for one student could be a less complex item for another student who had prior exposure to relevant assessment concepts. For example,

students enrolled in the concurrent program had been exposed to classrooms and exposed to assessment issues much longer than students in the consecutive program. Thus the cognitive complexity identified for each item was determined assuming students had a limited understanding of assessment.

Assessment Module Demographics

The demographic variables describing the population are gender, age, program, division, specialization, and assessment knowledge (enrolled in the elective course on assessment) (Items: 1 to 5, and 7). The distribution of students based on the demographic variables was representative of the students enrolled in the Faculty of Education (see Table 14 below). The demographic variables program, division, and assessment knowledge serve as the independent variables in this case study. The rationale for using these items as the independent variables is described below.

Age and gender. The age range for the majority of participants fell into the 20 to 25 (76.6%) category while a small percentage of students represented the 26 to 30 (13.0%) and 31 and above (10.1%) age categories. The number of females, (242, 78.6%) unsurprisingly out-numbered the males (63, 20.5%) in a four-to-one ratio. Although participants were largely female, this ratio is not atypical of the enrollment in Bachelor of Education programs in the province in which the study was carried out. The outcome of the School of Medicine case study revealed no significant differences related to age or gender. Thus age and gender were not explored for group differences in the Faculty of Education case study.

Program and division. Faculty of Education guidelines control the intake of students in each division. Hence, the number of students in the primary/junior and intermediate/senior divisions was approximately equal and this ratio was reflected in this sample (see Table 14). Fewer students were enrolled in the concurrent program (32.6%) in comparison to the consecutive program (68.2%) and these proportions were also reflected in our sample (see Table 14). Differences in response patterns were anticipated for program and division given that these groups represent very different subpopulations and arrive at the Faculty of Education with different experiences. For example, students in the intermediate/senior division have specialized in a discipline such as mathematics or languages whereas students from the primary/junior division are more likely to come from a general arts program. Likewise, students from the concurrent program arrive with more classroom experience than students from the consecutive program. The specific aspects stemming from differences in students' division and program that may influence students' responses is not known but prior research in this area (e.g., Miller, in press) has identified potential differences in relation to students' program and division.

Specialization. There are only a small number of students at the Faculty of Education who are enrolled in one of four areas of specialization (see Table 14). The small proportion of students in specialized programs is proportionately represented in this sample.

Assessment knowledge. Only 36 (5.2%) students at the Faculty of Education enrolled in the elective course on classroom assessment practices. Twenty-five of these students completed the questionnaire. Given that the elective course on classroom

assessment practices was only offered to students in the intermediate/senior division, statistical analyses focused on the intermediate/senior division only. Significant differences based on assessment knowledge (enrolled in the elective course on assessment) were anticipated due to the additional education in assessment.

Table 14

Comparison of the proportions of B.Ed students in the program and those completing the questionnaire

	2007/2008 Faculty of Education (n=) (%)	Questionnaire raw, (n=308) (%)
Program		
Consecutive	68.2	63.0
Concurrent		
University A	12.8	13.6
University B	18.7	21.4
University C	1.1	1.3
Division		
Primary/Junior	49.4	49.4
Intermediate/Senior	50.6	50.6
Specialization		
Aboriginal Teacher Education	0.8	0
Artists in Community Education	3.3	2.9
Outdoor and Experiential Education	3.1	2.3
Technical Education	4.9	3.9
Classroom Assessment Course	5.2	8.1

Technological background (internet experience). The majority of students in the sample reported they frequently used meta-research tools (180, 58.4%) or used

information located on the internet to complete course assignments (113, 36.7%). Hence over 95% of the students indicated they had a high level of internet experience. Initially, students' technological experience was to be used as an independent variable. However, given the high levels of students' self-reported technological skills, this variable was removed from the list of independent variables.

All of the students completed the formative CBA in the Faculty of Education. While most students reported they completed the Prior Learning Assessment at home (215, 69.8%), approximately 100 students completed the Prior Learning Assessment in the computer labs at the Faculty of Education. Almost 90% of the students (276, 89.6%) were able to complete the Prior Learning Assessment at a convenient time and place.

In terms of students' experiences with other CBAs (formative or summative), 129 (41.9%) students reported they had previously completed a CBA and more specifically, a CBA designed to guide their learning (110, 35.7%). Given that WebCT is also used in a compulsory module on legal issues in education, all students at the Faculty of Education used WebCT in more than one context. A few students indicated they used CBAs in other areas (e.g., WHMIS training) and at other institutions. There was no predominant CBA platform identified. In terms of using WebCT itself, the majority of students (194 or 286 responded with 5, 6, or 7) reported that it was easy to use and they did not require assistance in using this CBA platform (232, 75.3%). Logistical factors related to the use of WebCT such as students knowing their university net identification (required to access WebCT) (M = 6.71, SD = 0.89) or locating the computer lab (where help was available)

(M = 6.74, SD = 0.90) were reported to have little influence on being able to complete the Prior Learning Assessment (where 1 = high impact, 7 = no impact).

During the focus group session, one student commented that although she used WebCT in other courses, it was not used in the capacity of an assessment tool (Assessment FG S4). Rather, her experiences with WebCT were limited to using it as a place to download course notes or as a discussion forum. Students in the focus group also noted that their prior experience with WebCT in the Legal Issues Module was a factor influencing their beliefs about the use of WebCT in the Assessment Module. According to these students, questions on the Legal Issues formative CBA surveyed their understanding of the concepts presented during the lectures (Assessment FG S6). If students responded with 'no' (distractors were yes or no) to an item that asked whether they understood a particular concept, the feedback provided at the end of the assessment directed students to review the module material focusing on a particular concept. One student noted that the legal issues quiz was not a quiz but rather an on-line honour system whereby students could independently seek further learning (Assessment FG S11).

Students in the focus group were asked if their experience with the Legal Issues Module cast a shadow on their experiences with the Prior Learning Assessment given that the feedback provided in the Legal Issues formative CBA did not contain statements to support learning or direct them to other learning resources. Students unanimously agreed that a shadow had been cast. One student commented "it really took away from the credibility of the Prior Learning Assessment" (Assessment FG S1). The extent of impact this unanticipated factor had on students' experiences with the Prior Learning

Assessment is unclear; however, it is evident that it was not conducive to promoting a positive experience with the formative CBA used in the Assessment Module.

Research Question 1: Delivering Feedback

WebCT (CT – course tools) is an online learning tool that facilitates course management similar to ThinkingCap. WebCT was originally developed by the University of British Columbia in 1996 and was purchased by Blackboard in 2005 (Blackboard Inc., 2005). As part of the merger, the WebCT name is scheduled to be phased-out and the Blackboard name will take over. Currently, an internet search for WebCT takes searchers to the Blackboard website advertising the current Blackboard products.

WebCT is still used internationally at a number of universities (e.g., McGill University, University of Gloucestershire, and University of Minnesota). The version of WebCT used at the Faculty of Education is the Campus Edition version 4.1 released in 2003. The team responsible for creating the formative CBAs, found the software difficult to work with due to the number of tabs an instructor must navigate in order to build an assessment instrument. Although the focus of our use was on creating quality items and providing effective feedback, we noted the difficulty navigating through WebCT and the huge challenge to learn how the system operated. Since the release of Campus Edition 4.1, editions 6.0, 6.1, and 6.2 have been released and current claims state the product is more streamlined and user friendly (Blackboard Inc., 2008).

An assessment in WebCT is created by first entering a collection of items into a databank of items. The instructor then selects items from the databank to create a quiz (term used in WebCT) or formative CBA. The item formats available to WebCT (see

Appendix E for the chart document the analysis of the WebCT software) include: multiple-choice, matching (two parallel lists), short answer (students enter a response in a textbox which must match a pre-set answer), and long answer (this item is not computer scored; a large textbox is provided for students to enter their response). The items used in the WebCT formative CBAs in the Faculty of Education used the multiple-choice item format only. The first item in a set of three multiple-choice items presented students with a scenario and the first of three items. The following two multiple-choice items posed further questions related to the scenario. These items are shown in Appendix B.

The items created for each mini formative CBAs contained one item for three different cognitive levels in the ICE Model: ideas, connections, and extensions (Fostaty-Young & Wilson, 2000). The development of these items also included the creation of feedback statements for each distractor and key. The distractors were created based on the instructor's and research assistance's pedagogical understanding of the curriculum. The type of feedback involved providing the student with additional information. The feedback in WebCT is provided to students at the end of the assessment on a summary page. This delay in feedback is a potential short-coming of WebCT when used in a formative CBA context. It was hypothesized that students would not apply the feedback to their learning because the moment for learning the concept had passed. Students may be less likely to read the feedback once the assessment is completed and there is no imminent need to apply the feedback. During the focus group students were ask to compare the method of providing feedback in WebCT with another method that provided

feedback immediately following a response. All students in the focus group agreed that feedback immediately following a response to an item was the preferred method.

Research Question 2: Types of Feedback

Results from the instrument analysis, questionnaire, and focus group are drawn together to describe the findings related to Question 2: To what extent are particular types of feedback effective in supporting student learning? This section begins by describing the type of feedback provided on the formative CBA in the Assessment Module. Next, the findings from the questionnaire are presented and are interwoven with findings from the focus group. Lastly, the inferential statistics examing relationships between the independent variables (i.e., program, division, and assessment knowledge) are presented.

Findings from Instrument Analysis

Feedback was provided for all distractors in each formative CBA item. The type of feedback focused on providing addtional information and was provided for both the incorrect and correct responses. For example, feedback for the correct response to the second item in the set of items exploring *What Does Learning Look Like* (see Appendix B) provided a correct response of: “Correct. The student analyzes the painting in terms of identifying Picasso's style of painting (collage, geometric) as well as extending their response to include an interpretation of the painting based on their own experiences (their family kitchen).” Feedback guiding incorrect responses to this item included: (a) The student's responses go beyond the ideas level. They talk about the style of painting as well as provide an interpretation/connection to the painting and (b) This student is connecting their knowledge of painting with the art work shown; however, they go

beyond connections level. The purpose of the feedback was to provide additional information that either confirmed a correct response or clarified a misunderstanding.

Findings from the Questionnaire and Focus Group

The Bachelor of Education questionnaire contained 44 items and was administered electronically using SurveyMonkey. The questionnaire used a seven-point Likert type scale where seven represented the positive end of the attribute being measured and one represented the negative end. The formative CBAs were referred to as a Prior Learning Assessment; hence the phrase Prior Learning Assessment is used synonymously with formative CBA.

Descriptive Statistics

Feedback in WebCT is provided at the end of the formative CBA after students have responded to all items. The extent to which students reported they reviewed this feedback was not high (Item 39: $M = 3.65$, $SD = 1.53$) and the overall ranking of the usefulness of the feedback in guiding learning fell in the mid-range (Item 41: $M = 3.59$, $SD = 1.64$). Students' responses surveying the different types of feedback are presented in Table 15 below.

Table 15

Assessment Module: Effectiveness of different types of feedback

Item	*M*	*SD*
40. Rank the effectiveness of each specific type of feedback given on the Prior Learning Assessment		
a. Indicated your response was incorrect.	3.85	1.85
b. Indicated your response was correct.	4.14	1.80
c. Indicated your answer was incorrect AND provided further information clarifying the incorrect response.	5.11	1.80
d. Indicated the answer was correct AND supplemented it with why the answer was correct.	5.05	1.80

Based on the mean scores and distributions, more students preferred feedback that provided additional information that either clarified an incorrect response (Item 40c) or supplemented a correct response (Item 40d). An independent sample *t*-test showed that this difference in mean scores was significant ($p < 0.05$). Although these differences in mean scores and effect sizes were small, this significant difference still had practical significance because they iterate the outcome of a prior study examining group differences between division (primary/junior and intermediate/senior) and feedback (Miller, in press).

In the focus group, students were further probed regarding their preferences for feedback. One student in the focus group commented, “if I had the prior knowledge and got the question correct, I don’t need to see again why I was correct” (Assessment FG

S4). In contrast, other students argued that if students were to be exposed to higher cognitive items they need to know why their answers were correct. The questionnaire responses and focus group comments illustrate that students considered each form of feedback to have some value while the feedback that provided further information was valued more.

Further discussion related to feedback centered on using the words correct and incorrect. One student commented that although their response was incorrect, part of their thinking might be correct; hence they did not like, "being told you are wrong especially when I am partly correct" (Assessment FG S4). Other students agreed and one added that possibly levels of achievement could be used instead of simply stating correct or incorrect (Assessment FG S7).

When students in the focus group were asked to focus on examples of feedback that directed them to three different types of resources (i.e., course text, course notes, or active web link) preference was given to the active web link. One comment was raised in opposition of the first example: "I don't like the first one [directions to a course text] because say, I don't have time – I just want to know the topic" (Assessment FG S3) while three students made positive references regarding feedback that directed them to an active web link. These three students commented, "I like all 3 combined – especially if the third type of feedback [incorrect + live link] is live" (Assessment FG S1) "I like the 3rd one especially if it is live link" (Assessment FG S5) and "I like the third one as well – because example two – you are relying on the quality of your course notes. If you have your course notes and got it wrong [the course notes are not correct] then it wouldn't be

good" (Assessment FG S2). Overall, these students seemed to prefer feedback that directed them to an active web link over feedback that directed them to course notes or a text.

Inferential Statistics

Students were grouped according to program (concurrent or consecutive) and division (primary/junior or intermediate/senior). In addition, students were grouped according to whether they took the elective course in classroom assessment. The dependent measures were drawn from the four items focusing on the different types of feedback (i.e., Items 40a to 40d). Since these items were designed to measure different types of feedback, these items were considered separately. The remaining two survey items (Items 39 and 41) surveying students' attitudes towards the effectiveness of the feedback were combined to create a fifth dependent variable representing perceived effectiveness of the formative CBAs. Separate two-way (2 X 2) ANOVAs were performed comparing group differences across program and division. A one-way ANOVA was performed comparing group differences based on assessment knowledge. For each ANOVA, a Bonferroni adjustment (0.05/5) was used to correct for Type I errors.

Program and division. Significant differences were present in all four items exploring the different types of feedback (see Table 16) and division. As reported initially (see Table 15), primary/junior and intermediate/senior students differ in terms of their preference for feedback. Table 17 displays the mean scores comparing the two divisions and program. In all cases, students in the intermediate/senior division responded more

positively towards the effectiveness of each type of feedback. Although the differences between mean scores and effect sizes were small, these significant differences still have practical significance given that these findings iterate the outcome of a prior study revealing that intermediate/senior division students also responded more favourably towards feedback on a formative CBA (Miller, in press). No significant differences ($p > 0.05/5$) were found when comparing the factor representing perceived effectiveness of the formative CBAs (combined Items 49 and 41) with program and division (see Table 18).

Table 16

Assessment Module: Two-way ANOVA - program and division x *feedback*

Source	*SS*	*df*	*MS*	*F*	*p*	*Effect Size*
Incorrect						
Program	0.99	1	0.99	0.30	0.586	0.001
Division	29.67	1	29.67	8.92	0.003	0.035
Prog. x Div.	3.17	1	3.17	0.95	0.330	0.004
Error	818.77	246	3.33			
Total	4559.00	250				
Correct						
Program	2.70	1	2.70	0.86	0.354	0.003
Division	25.32	1	25.32	8.10	0.005	0.030
Prog. x Div.	1.46	1	1.46	0.47	0.495	0.002
Error	809.81	259	3.13			
Total	5347.00	263				
Incorrect + why						
Program	0.05	1	0.05	0.16	0.691	0.001
Division	26.38	1	26.38	8.60	0.004	0.034
Prog. x Div.	3.72	1	3.72	1.21	0.272	0.005
Error	757.86	247	3.07			
Total	7375.00	251				
Correct + why						
Program	3.28	1	3.28	1.07	0.303	0.004
Division	26.27	1	26.27	8.54	0.004	0.033
Prog. x Div.	2.33	1	2.33	0.76	0.385	0.003
Error	774.77	252	3.07			
Total	7385.00	256				

Table 17

Assessment Module: Mean scores for division and program

	Primary/Junior		Intermediate/Senior	
Item	Concurrent	Consecutive	Concurrent	Consecutive
40a	3.29 (1.83)	3.65 (1.85)	4.24 (1.73)	4.14 (1.85)
40b	3.61 (1.85)	3.98 (1.83)	4.41 (1.58)	4.46 (1.76)
40c	4.95 (1.94)	4.61 (1.94)	5.38 (1.64)	5.54 (1.51)
40d	4.98 (1.88)	4.54 (1.86)	5.45 (1.57)	5.41 (1.67)

Table 18

Assessment Module: Two-way ANOVA - program and division x *effectiveness of feedback*

Source	*SS*	*df*	*MS*	*F*	*p*
Program	0.70	1	0.70	0.38	0.534
Division	9.07	1	9.07	5.01	0.026
Prog. x Div.	1.61	1	1.61	0.89	0.347
Error	447.11	247	1.81		
Total	3628.25	251			

Assessment knowledge. When comparing the dependent variables exploring feedback and the independent variable assessment knowledge no significant differences were found in any of the dependent variables (see Table 19).

Table 19

Assessment Module: Assessment knowledge x *feedback*

Source	*SS*	*df*	*MS*	*F*	*p*
Effect. of feed.					
Between	1.09	1	1.09	0.66	0.419
Within	201.03	121	1.66		
Total	1952.50	123			
Incorrect					
Between	4.73	1	4.73	1.49	0.225
Within	384.64	121	3.18		
Total	2504.00	122			
Correct					
Between	5.63	1	5.63	1.98	0.162
Within	355.42	125	2.84		
Total	2848.00	127			
Incorrect + why					
Between	0.10	1	0.10	0.04	0.836
Within	289.37	124	2.33		
Total	4079.00	126			
Correct + why					
Between	0.23	1	0.23	0.09	0.765
Within	316.63	124	2.55		
Total	4030.00	126			

Research Question 3: Factors Affecting Formative Computer-based Assessments

Similar to the School of Medicine case study, items surveying factors believed to have some effect on students' experiences with formative CBAs were grouped according to students' perceptions of the purpose (Items 30, 31, 34c, 36, 42f), value (Items 42, 57-

59, 68 – 75), and items (Items: 32a-c, 33a-b, 35, 42e). The findings from these three areas of inquiry are presented separately below.

Purpose of Formative CBA

Descriptive Statistics

Although the purpose of the Prior Learning Assessment was explained during the lecture as an instrument used to guide students' future learning, 64 (20.8%) students still believed the purpose of the assessment was to provide a mark in the course (Item 30). A further 131 (42.5%) students believed the purpose of the Prior Learning Assessment was to give instructors information about what the students knew. Students frequently referred to the Prior Learning Assessment as a test. They also focused on their score in the four areas of inquiry on the Prior Learning Assessment. One student stated, "we resented the test because how does this show what we know – a zero out of three or one out of three indicates I know nothing" (Assessment FG S3). This language (i.e., test and score) suggests the formative CBAs held summative properties.

Students did not use the Prior Learning Assessment as a tool to direct their learning (Item 31). In this questionnaire item, 128 (41.6%) students disagreed that "the concept of using a Prior Learning Assessment is a good way to direct student learning" (responding with a 1, 2, or 3 where 1 = strongly disagree) (see Table 20). Further, 147 (48%) students reported that the Prior Learning Assessment had no influence on their self-directed learning (see Table 20).

Table 20

Assessment Module: Understanding the purpose of the formative CBA

Item	*M*	*SD*
31. The concept of using a Prior Learning Assessment is a good way to direct student learning. Indicate the extent to which you agree or disagree with the statement.	4.32	1.45
34c. How helpful was the Prior Learning Assessment in starting discussions with colleagues about assessment?	3.32	1.59
36. To what extent did the Prior Learning Assessment influence your self-directed learning?	3.44	1.61
42f. Identify the extent to which the following factors may have influenced your achievement on the Prior Learning Assessment: I was sure of the purpose of the assessment.	4.83	1.92

One student in the focus group suggested it would be better to "go with area of interests" (Assessment FG S5) instead of the intended purpose to improve upon weak areas. Other students supported this idea by verbally agreeing or by nodding. Once again, a misunderstanding of the purpose of the formative CBAs is evident in students' failure to recognize the instrument as a means to direct future learning.

The focus group discussion continued and turned to the creditability of the Prior Learning Assessment. The discussion revealed that the formative CBAs lacked credibility in the eyes of the students. The number of items in each of the four areas assessed (paralleling the four modules in the lectures), the complexity of the items, and the format of items were cited as reasons for the low credibility (these are further discussed below).

One student commented that she did not believe students were the least bit serious when completing the assessment (Assessment FG S5). Another student described the process of responding to the items as very cursory, stating, “you just skim through the question very fast and find the answer that kind of works – and oh well whatever happens” (Assessment FG S7). A third student noted that the Prior Learning Assessment had no future implications for her (Assessment FG S12) while a fourth student commented, “if I answered these questions accurately, it doesn’t mean I have some understanding of assessment” (Assessment FG S5). The reported lack of credibility may be attributed to a misunderstanding of the purpose. If, for example, students believed the instrument was to measure their learning and provide a grade in the Assessment Module, their criteria for evaluating the formative CBAs would be different than the criteria instrument developers used when creating the formative CBAs.

The extent to which students’ collaborated (Table 20, Item 34c) was also believed to be an indicator of students’ understanding of the purpose of the instrument. Students who reported they worked with their peers and discussed solutions to the formative CBAs were more likely to understand the purpose and value of formative assessment practices. On the questionnaire, the amount of reported collaboration was relatively low (Table 20, Item 34c); however, when course instructors and the researcher observed students completing the formative CBAs in various areas and computer labs within the Faculty of Education, students appeared to be engaged in discussions with their colleagues. This observation was described to students during the focus group and students were asked for

their insights. Students replied that they were not discussing the concepts but rather discussing which distractor was the correct response. One student replied,

> The group we worked with really weren't interested in subject matter as with getting the correct answer … my group would go #1 b, b, b, b; #2 a, a, a, a. If anyone said that it was c then we would say does anybody agree – if not then we selected a. It wouldn't be a discussion unless it was a 2 – 2 split. (Assessment FG S3)

Other students in the focus group were asked if they followed the same approach; all agreed that they had. This was disappointing to learn; however, it does provide an explanation to what was viewed during the completion of the Prior Learning Assessment and the extent students reported they collaborated. Given that this item (Item 34c) did not function in the manner intended it was not used in further analysis.

Lastly, difficulty comprehending the assessment language used in the module may have also contributed to students' misunderstandings of the purpose of the Prior Learning Assessment. During the focus group session, several students indicated they had a difficult time understanding the meaning of formative assessment and assessment for learning. One student described the onset of the Prior Learning Assessment as "a lot of mass confusion initially with it being formative - people are asking if this counts for my grade? If I fail, will I fail the course?" (Assessment FG S3) This confusion was attributed to not understanding the meaning of formative assessment. It is possible that students have had minimal exposure to formative assessment practices in their secondary or post secondary education. Consequently, it may have been too early in the assessment module

to not only expect students to understand the theory supporting new methods of assessment but also engage in these practices.

It is evident that students did not have a clear understanding of the purpose of the formative CBAs even though the purpose was articulated during the first couple of lectures. It is likely that a lack of exposure to formative assessment practices (including the one-time exposure with this Prior Learning Assessment) as well as limited exposure to assessment terminology contributed to students' misunderstandings of the purpose of the formative CBAs.

Inferential Statistics

The three continuous items exploring students' perceived purpose of the formative CBAs were subjected to a principal component factor analysis, identifying one component with an eigenvalue exceeding 1 and explaining 52.96% of the variance. The factor loadings are shown in Table 21. The three items were combined to represent students' perceived purpose of the formative CBA.

Table 21

Assessment Module: Factor analysis - perceived purpose of the formative CBA

Item	Component
31. The concept of using a Prior Learning Assessment is a good way to direct student learning.	0.84
36. To what extent did the Prior Learning Assessment influence your self-directed learning?	0.83
42f. I was sure of the purpose of the assessment.	0.44

Program and division. A two-way (2 X 2) ANOVA was conducted to explore group differences in program and division. No significant differences were found for either of the independent variables (see Tables 22).

Table 22

Assessment Module: Two-way ANOVA - program and division x *purpose*

Source	*SS*	*df*	*MS*	*F*	*p*	Effect Size
Program	0.45	3	0.15	0.23	0.879	0.003
Division	0.48	1	0.48	0.72	0.397	0.003
Prog. x Div.	0.14	3	0.05	0.07	0.975	0.001
Within (error)	178.02	267	0.67			
Total	4553.33	275				

Assessment knowledge. A separate one-way ANOVA was performed comparing group differences in assessment knowledge. Similar to the findings examining group differences for program and division, no significant differences were found (see Table 23).

Table 23

Assessment Module: One-way ANOVA - Assessment knowledge x *purpose*

Source	*SS*	*df*	*MS*	*F*	*p*	Effect Size
Between	0.22	1	0.22	0.38	0.537	0.003
Within	75.78	135	0.56			
Total	2351.44	137				

Value of the Formative CBA

The value of the formative CBAs was examined from three different perspectives. The first was from the perspectives of students in terms of: the priority given to the formative CBAs (Item 42a), perception of the task (Item 42b), personal schedule (Item 42c), and effort to complete the formative CBAs (Item 46). The second perspective focused on the instrument in terms of whether students felt it was helpful and guided their learning (Items 34a and 34b). The last perspective was from an assessment perspective in terms of whether students' perceived the instrument to be fair and accurate (Items 44 and 45). The findings in each of these three areas are presented below.

Descriptive Statistics

Value (student perspective). When exploring the variables priority, perception of the task, personal schedules thought to influence students' experiences with the formative CBA, and effort, students' responses were neutral (see Table 24). These items were extended in the focus group to explore what limited students' views towards the Prior Learning Assessment. All students agreed there was a need to improve the items (discussed below) as well as a need to first expose students to assessment theory and terminology. One student explained "you need to break it into two components – do the theory first and then get us to apply it. We didn't understand the theory so we couldn't apply it" (Assessment FG S9).

Table 24

Assessment Module: Value (Student Perspective)

Item	*M*	*SD*
42. Identify the extent the following factors may have influenced your achievement on the Prior Learning Assessment		
a. Not a priority for me given other faculty expectations	4.59	1.9
b. I thought it was just 'busy work'	4.27	1.9
c. I was too busy	4.72	1.8
46. Identify the degree of effort (seriousness) that you put forth in completing the Prior Learning Assessment	4.23	1.5

Value (instrument). Students were asked on the questionnaire whether the Prior Learning Assessment was helpful in advancing their understanding of assessment (Item 34a) and whether the instrument guided their learning during the module (Item 34b). It was believed that if the formative CBAs were helpful and guided student learning, students would value the instrument. In each of these areas, students generally reported the Prior Learning Assessment played a minor role in guiding their learning (see Table 25). Given these findings, it is not surprising that students also reported the overall usefulness of the Prior Learning Assessment (Item 41) to be less than moderate (see Table 25).

Table 25

Assessment Value: Value (instrument perspective)

Item	*M*	*SD*
34a. How helpful was the Prior Learning Assessment in advancing your understanding of assessment?	3.21	1.45
34b. How helpful was the Prior Learning Assessment in guiding your learning during the module?	3.46	1.52
41. Rank the OVERALL usefulness of feedback provided on the Prior Learning Assessment in guiding your learning.	3.59	1.64

Value (accuracy and fairness). The last two factors believed to influence students' perceptions of the value of the formative CBA were related to the accuracy and fairness of the instrument. These two dichotomous items surveyed students' perceptions of the accuracy and fairness of the formative CBA. Approximately half (52.6%) of the students reported the assessment was accurate and the same number of students reported the assessment was fair. During the focus group students were asked what could be done to make the Prior Learning Assessment more accurate and fairer. Other than the comments previously mentioned (i.e., improve the items), students did not add anything new. One student commented that it would have been useful to have sample scenario-based items on the website that they could attempt to self-assess their understanding of assessment in each of the four content areas of the module (Assessment FG S8). The student felt the opportunity to practice scenario-based items would model the assessment activities in the

course. It is possible that the requirement for all students to complete the formative CBAs in the Assessment Module imposed a summative overtone to the instrument.

Another student considered fairness from the perspective of getting value for their tuition fee. This student stated,

> This is pointing to the flaws in on-line learning and regulated learning for sure. Which are becoming more and more obvious – I don't want to stare at a computer screen. It started back during the first week of the course – we didn't know it was going to be self-regulated learning and there was no information to that until we stepped into the lecture. And then it was 'and by the way' you are expected to do this – you kind of panic. We assumed it was going to be a lecture where we were going to be taught. It was a wild scramble because we didn't know this was going to be a component of the course. It was disappointing to walk into that atmosphere – we have a lot of type A people who like to talk and engage with others. It is really disappointing to come to an institution and have to do that. Part of me says, 'I can't believe, I dropped 500 dollars for that'. That doesn't help with the fairness and accuracy. (Assessment FG S3)

The concept of using formative CBAs as a self-regulated learning instrument may have presented too large a gap between students' state of readiness for self-regulated learning and understanding of assessment principles called for in the Prior Learning Assessment. The gap may have been further exacerbated given that the feedback intended to bridge this gap did not appear to have been read (discussed below).

Inferential Statistics

The seven continuous items exploring students' perceived value of the formative CBA were subjected to a principal component factor analysis. Coefficients of 0.3 and higher, a Kaiser-Meyer-Oklin value of 0.723 and significant Bartlett's Test of Sphericity support the factorability of the correlation matrix. Principal component analysis revealed the presence of two components with eigenvalues exceeding 1 explaining 37.71 and 28.69% of the variance, respectively. An inspection of the scree plot revealed a break after the second component and two components were retained for further investigation. A Varimax rotation was performed and the rotated solution revealed strong loadings on each component with only one item loading on both components. These findings are displayed in Table 26 below. The first component could be described as the perceived usefulness of the formative CBA (34b, 34a, 41, 46) and the second component is described as student factors limiting the usefulness of the formative CBAs (42c, 42a, 42b). The two factors had Cronbach's alpha coefficients each of 0.77. The variables within each component as identified through factor analysis were combined to create two factors representing the two components.

Table 26

Assessment Module: Factor analysis - perceived value of formative CBA

Item	Component 1	2
34b. How helpful was the Prior Learning Assessment in guiding your learning during the module?	0.89	
34a. How helpful was the Prior Learning Assessment in advancing your understanding?	0.86	
41. Rank the OVERALL usefulness of feedback provided on the Prior Learning Assessment in guiding your learning.	0.84	
46. Identify the degree of effort (seriousness) that you put forth in completing the Prior Learning Assessment	0.42	0.34
42c. Identify the extent to which the following factors may have influenced your achievement on the Prior Learning Assessment: I was too busy.		0.86
42a. Identify the extent to which the following factors may have influenced your achievement on the Prior Learning Assessment: not a priority for me given other faculty expectations		0.84
42b. Identify the extent to which the following factors may have influenced your achievement on the Prior Learning Assessment: I thought it was just 'busy work'		0.75

Program and division. A two-way (2 X 2) ANOVA was conducted to explore the impact of program (concurrent or consecutive) and division (primary/junior or intermediate/senior) on the two factors (see Table 27). The Bonferroni correction (0.05/2) was used to correct for Type I errors.

Table 27

Assessment Module: Program and division x *value*

Source	*SS*	*df*	*MS*	*F*	*p*	*Effect Size*
Usefulness						
Program	0.45	1	0.45	0.34	0.559	0.001
Division	11.18	1	11.18	8.52	0.004	0.030
Prog. x Div.	0.42	1	0.42	0.32	0.571	0.001
Error	356.68	272	1.31			
Total	3990.50	276				
Student factors						
Program	9.03	1	9.03	3.67	0.057	0.013
Division	0.84	1	0.84	0.34	0.560	0.001
Prog. x Div.	0.12	1	0.12	0.05	0.823	0.000
Error	682.38	277	2.46			
Total	6450.22	281				

The mean scores for division and the first factor, students' perceived usefulness of the formative CBA, revealed that intermediate/senior division students found the formative CBA more useful than primary/junior division students (primary/junior: $M = 3.97$, $SD = 1.11$; intermediate/senior: $M = 4.26$, $SD = 1.00$).

Assessment knowledge. A one-way ANOVA was used to explore the impact of students' assessment knowledge on the two factors. The analysis did not reach statistical significance ($p > 0.05$) (see Table 28).

Table 28

Assessment Module: One-way ANOVA - assessment knowledge x *value*

Source	*SS*	*df*	*MS*	*F*	*p*
Perceived Usefulness of formative CBA					
Between	0.01	1	0.01	0.01	0.940
Within	178.01	133	1.34		
Total	2163.69	135			
Student factors limiting the usefulness of formative CBAs					
Stu. fact. limit.	0.71	1	0.71	0.29	0.594
Between	336.90	136	2.48		
Within	3135.11	138			
Total					

Item: Format, Difficulty, Number, and Concepts Assessed

The item characteristics explored included format, difficulty, number, and concepts assessed. Questionnaire Items 32a – 32c, 33a, 33b, 35, and 42e survey this area of inquiry. The descriptive findings related to each item characteristics are described separately below. Table 29 displays the means and standard deviation for these items and the discussion surrounding these areas of inquiry is presented following the table.

Table 29

MSK Module: Factors influencing items

Item	*M*	*SD*
32a. In terms of the Prior Learning Assessment, to what extent was the number of questions in each of the 4 units appropriate in assessing your understanding of assessment?	3.59	1.36
32b. In terms of the Prior Learning Assessment, to what extent was the difficulty level of questions appropriate for this assessment?	3.94	1.34
32c. In terms of the Prior Learning Assessment, to what extent were the topics selected for assessment appropriate?	4.16	1.38
33a. To what extent do you agree or disagree with the following statement: reading and responding to a scenario is a good way to assess your thinking about assessment.	4.65	1.46
33b. To what extent do you agree or disagree with the following statement: each scenario in the Prior Learning Assessment represented a meaningful assessment dilemma.	3.97	1.44
35. To what extent were the multiple-choice questions on the Prior Learning Assessment useful in assessing your understanding of assessment?	3.20	1.47
42e. Identify the extent to which the following factors may have influenced your achievement on the Prior Learning Assessment: the scenarios were realistic.	5.32	1.65

Descriptive Statistics

Item format. The format of items on the formative CBAs were all multiple-choice. The multiple-choice items were scenario-based in that students were presented with a dilemma in classroom assessment and were then required to respond to three items

related to the dilemma. Students were not overly convinced (see Table 29, Item 35) that this format was useful in assessing their understanding of classroom assessment practices. The reason for this dissatisfaction with multiple-choice items was discussed in the focus group; in fact, it was the first issue students wanted to discuss at the onset of the focus group. According to students, multiple-choice meant multiple-guess. One student commented "the multiple-choice thing was – multiple guess. You could do well by guessing" (Assessment FG S1). Other students in the focus group supported this comment. Other students thought it was ironic that they were being assessed using multiple-choice items when the instructors advised them against using this item format in their classrooms. Interestingly, the appropriate use of the multiple-choice item format was not actually discussed in the lectures. On a more positive note, one student commented in a subsequent question that "I don't think it was a problem with the format – multiple-choice … I think the problem is with the actual questions. I had a really, really difficult time with the questions" (Assessment FG S12).

A second issue connected to item format was related to the rules for creating multiple-choice items. Students felt instructors violated the guidelines for creating multiple-choice items citing the first question in the Assessment of Learning component of the Prior Learning Assessment (see Appendix B). Students claimed the size of the stem was problematic in that they had to re-read the scenario several times to understand the question. In this case, one student commented,

> Like I said, they went against everything we were taught about multiple-choice questions - I mean, just the size of that first paragraph alone. It was all just one

giant blob [Assessment for Learning, Question 1]. I had to go back and re-read. (Assessment FG S5)

Traditionally, good multiple-choice items have short stems; however, when multiple-choice items are presented in a scenario format it seems reasonable that the stem be expanded to allow for the presentation of the contextual information. It is evident that students from the Faculty of Education learned about the traditional format and uses of multiple-choice items, such as those used on standardized provincial assessments and not about how they can be used in different context to assess higher cognitive thinking.

Another interesting finding connected to the format of multiple-choice items was the use of creativity to situate the context of the scenario. One student in the focus group noted that the name of the teacher (i.e., Waddlebug) in the first scenario (see Appendix B, Assessment for Learning, Question 1) was distracting (Assessment FG S5); all students in the focus group agreed with the statement. This student stated "just to be honest – a name like Waddlebug distracts you every time you read it. It makes it difficult to take seriously. A name like Jones would be a lot better" (Assessment FG S5). Whether or not there is place for wit in scenario-based multiple-choice items is beyond the scope of this study. However, this finding is worth noting especially since it may have distracted students from the seriousness of the formative CBAs.

Item format (scenarios). Students moderately agreed (Table 29, Item 33a) that reading and responding to scenario-based items was a good way to assess their thinking. Students were less convinced that the scenarios represented meaningful assessment

dilemmas (Table 29, Item 33b). Yet, students reported that the scenarios were not a factor influencing their experience on the Prior Learning Assessment (Table 29, Item 42e).

During the focus group students were asked to re-visit a few of the items used on the Prior Learning Assessment to explore what prevented them from feeling more positive towards the use of scenarios. As noted above, all students did not like the first scenario in the Assessment of Learning section (see Appendix B, *Assessment for Learning*, Question 1), primarily due to the length of the item in that students believed the item was too long and complicated. The scenario in the *What does Learning Look Like* (see Appendix B, *What does Learning Look Like*, Question 1) section received mixed reviews. One student thought the first sentence was a filler while another student disagreed and thought the scenario was good. This student commented "I remember reading it [the scenario] and wondering why all questions can't be like this – it's a short paragraph. Not that I knew the answer, but at least I knew what it was asking" (Assessment FG S4). It is interesting that this student also referenced the length of the scenario suggesting that the shorter scenario (one paragraph) was easier to comprehend. The discussion on scenarios continued focusing on the complexity of the scenario. Another student commented, "I thought the question was assessing two things: whether we knew what Ideas, Connections, and Extensions [ICE model] were and whether we knew what level of learning it was" (Assessment FG S6). This was precisely how the question was structured; however, the student preferred more scaffolding with the inclusion of questions such as: "What do you think ideas are? What are connections? And then link them to a definition and then apply them" (Assessment FG S6). It appears there

is a wide gap between students' entry level of assessment knowledge and the cognitive level assessed in the Prior Learning Assessment. Following-up with this need for support, students were specifically asked if a formative CBA that would allow them to access a help button that, for example, described or provided a definition for what ideas represented in the ICE model, would be helpful. Four students responded positively to this type of formative CBA (e.g., HotPotatoes or ThinkingCap). As an alternative to using a help button to assist with learning about new terminology or concepts, one student suggested (i.e., without disagreement from other students) to simply focus the items on what they knew (Assessment FG S8). Students appeared to struggle with using the formative CBAs to explore what they knew about assessment and secondly, to direct their future learning. It is possible that students' lack of experience in using formative CBAs in this context or their reluctance to engage in this type of self-regulated learning may have influenced their views.

The last scenario reviewed by students was the scenario presented in the *Large-scale Assessment* section of the Prior Learning Assessment (see Appendix B, *Large-scale Assessment*, Question 1). This item presented a replica of a student report produced by the Ontario Ministry of Education for parents and students. The student report contained a pictorial representation of a table and a box-and-whisker-plot displaying a sample student's performance on the provincial Grade 6 Mathematics Assessment. Students in the focus group responded negatively towards this scenario claiming it was not representative of what they knew about large-scale assessment and they had never viewed

a graph like the one presented in the item. Hence it was unreasonable to ask them to interpret the graph. One student commented,

> I found this really intimidating for people who are not mathematically inclined. I am mathematically inclined and I thought what does that mean? The black part must be the central part and the grey part must be the spread of the data but like I took statistics. (Assessment FG S5)

Another student added, "Yeah, who can understand that. It looks like they accidentally put some grey blocks on the side" (Assessment FG S2). Another student added,

> This question isn't representative of what I know about large-scale assessment. You are throwing a graph at me that I have never seen in my life and you are asking me to read it. Given that I am not a math person at all – I just guessed. (Assessment FG S4)

It is likely that this formative CBA item contributed to students' low ratings of what they considered a meaningful assessment dilemma and this focus group finding reflect a misunderstanding of the purpose of the formative CBAs. This position towards meaningful assessment dilemmas used on the formative CBAs was shared by all students in the focus group. Based on these comments, students did not find the scenario-based items meaningful because they knew little about interpreting results on the provincial report card. If students understood the purpose of the formative CBAs as a tool to identify weak areas and direct future learning, they might have acknowledged the complexity of provincial reports. Consequently, this scenario may not be a poor scenario given the context in which the formative CBA was intended.

Students were probed further to explore whether scenario-based items were at the root of the confusion related to the multiple-choice item format. Several students indicated that the scenario or case study approach was agreeable. However, they felt there should be a separation from the description of the scenario and the actual item, thereby reducing the size of the stem. This discussion supports Item 42e (see Table 29) where students reported the realistic scenarios did not negatively influence their formative CBA experience.

Item difficulty. Students were not convinced that the difficulty level of items was appropriate for this assessment (Table 29, Item 32b). During the focus group, students were specifically asked if the items on the Prior Learning Assessment were too easy, too difficult, or just right. The first comment was "too hard" (Assessment FG S2) but then another student quickly interjected, "I don't know whether too hard is the right word – too complicated – they assumed a lot" (Assessment FG S5). Another student added that they found the items "too cumbersome" (Assessment FG S7). Students appeared to be hesitant to indicate the items were too difficult and were more comfortable with describing the items as too complex, requiring more effort than traditional multiple-choice items. If complexity is a factor of difficulty, where a student must process the context of a scenario to extract the pertinent information in order to respond to an item, then in all likelihood, the items were more difficult than students anticipated. Given that the Prior Learning Assessment was a non-traditional instrument using scenario-based multiple-choice items to tap into students' higher cognitive skills, it is possible that students' unfamiliarity with this item format influenced their thoughts on item difficulty.

When students were asked what level of cognitive complexity new items should be added, students asked for clarification on the meaning of cognitive complexity. When the question was rephrased to the achievement chart categories (the terminology used in the elementary/secondary curriculum) students had a better understanding of the question and replied that a question may be easy for one person and difficult for another person. Hence students unanimously agreed that questions assessing a range of cognitive complexities should be added. This insight into evaluating item difficulty is interesting because it reveals some understanding towards the different levels of assessment knowledge that students arrive with; however, students did not connect this thinking to the purpose of the formative CBAs in directing their learning.

Another aspect connected to item difficulty was the manner in which items were presented. One student noted that embedding large scenarios with the actual item was difficult for students with learning disabilities and commented that the instructors should be modeling what will be required of students in their classrooms. This student stated,

> This [scenario] is a terrible way to present it. It needs to be double spaced, or less wordy, again separating it from the question would be good. Again, we are modelling how we would do it for our students and if a student who had any reading or comprehension difficulties they would have a terrible time. (Assessment FG S5)

Other students in the focus group did not object to this comment. It is interesting that students in the focus group associated themselves with elementary or secondary students similar to the students they would eventually be teaching and not as professionals in a

professional learning program. As a result, this student may be assuming that instructors should model the same assessment practices that Bachelor of Education students are expected to apply in their classrooms. This finding related to item difficulty in formative CBAs may only be isolated to this particular case study given that the discipline (i.e. assessment) overlaps with the assessment instrument. In this incidence where a master (i.e., the instructor) is modeling their discipline, it may be especially important to clarify the expectations of the learner and the context of the assessment instrument.

Item content. As noted previously, the topics selected for assessment were grouped into four main areas: *What does Learning Look Like, Assessment for Learning, Assessment of Learning,* and *Large-scale Assessment*. These topics were selected based on a needs assessment conducted prior to the development of the Assessment Module. The needs assessment involved a survey of teacher candidates who were near the completion of the Bachelor of Education program and associate teachers. Although the topics selected for the Prior Learning Assessment were previously identified by Bachelor of Education students as important concepts to know prior to commencing their first practicum, students in this case study were moderately satisfied with the topics (Table 30, 32c).

*Number of items***.** On the questionnaire, students reported they were not satisfied with the number of items (Table 29, Item 32a). This finding was supported in the focus group where the low number of items was frequently identified as a problem. In this area of inquiry, 12 different comments were made during the focus group referencing the insufficient number of items.

Inferential Statistics

The seven items exploring students experience with the formative CBAs were subjected to a principal component analysis where the correlation matrix revealed the presence of coefficients greater than 0.3, a Kaiser-Meyer-Oklin value of 0.85, and a significant Bartlett's Test of Sphericity. One component with an eigenvalue exceeding 1 explained 54.15% of the variance and the scree plot revealed a break after the second component. Factor loadings are shown in Table 30. These items representing the formative CBAs item characteristics were combined into a single factor with Cronbach's alpha coefficient of 0.84.

Table 30

Assessment Module: Factor analysis – item characteristics

Item	Component 1
32b. In terms of the Prior Learning Assessment, to what extent was the difficulty level of questions appropriate for this assessment?	0.84
32c. In terms of the Prior Learning Assessment, to what extent were the topics selected for assessment appropriate?	0.82
32a. In terms of the Prior Learning Assessment, to what extent was the number of questions in each of the 4 units appropriate in assessing your understanding of assessment?	0.81
35. To what extent were the multiple-choice questions on the Prior Learning Assessment useful in assessing your understanding of assessment?	0.77
33a. To what extent do you agree or disagree with the following statement: reading and responding to a scenario is a good way to assess your thinking about assessment.	0.60
42e Identify the extent to which the following factors may have influenced your achievement on the Prior Learning Assessment: the scenarios were realistic.	0.44

Program and division. A two-way ANOVA was conducted to explore the impact program (concurrent or consecutive) and division (primary/junior) had on this factor. The main effect for division and the interaction effect reached statistical significance while the main effect for program did not reach statistical significance (Table 31).

Table 31

Assessment Module: Two-way ANOVA - program and division x *items*

Source	*SS*	*df*	*MS*	*F*	*p*	*Effect Size*
Program	0.15	1	0.15	0.16	0.694	0.001
Division	9.60	1	9.57	9.81	0.002	0.035
Prog. x Div.	6.47	1	6.47	6.63	0.011	0.024
Error	261.47	268	0.98			
Total	4920.55	272				

The notable difference in mean scores (Primary/Junior: $M = 3.88$, $SD = 1.07$ and Intermediate/senior: $M = 4.37$, $SD = 0.92$) and small effect size confirms the practical significance of this finding. This finding reiterates previous findings indicating that intermediate/senior students had more positive formative CBA experiences than the primary/junior division students.

Assessment knowledge. A one-way ANOVA of variance was performed to explore differences between item characteristics and assessment knowledge. A statistical difference ($p > 0.05$) was not found (see Table 32).

Table 32

Assessment Module: One-way ANOVA - assessment knowledge x *item characteristics*

Source	*SS*	*df*	*MS*	*F*	*p*
Ass. Know	1.33	1	1.33	1.280	0.259
Error	268.54	259	1.04		
Total	4718.76	261			

CHAPTER SIX

Discussion

The purpose of this chapter is to provide a constructive and critical analysis of the research findings in relation to the questions posed in the study and the literature underpinning formative computer based assessments (CBAs). The findings from both case studies are drawn together to address each of the three research questions posed in this dissertation.

Research Question One: Channeling Feedback

How is feedback channeled to students and which is the preferred method?

Given the different contexts under which each case operated, it is not surprising that feedback was channeled to students using different methods. In the School of Medicine case study, student feedback was provided following each response. In contrast, student feedback provided in the formative CBA used in the Faculty of Education was provided after the student had responded to all items on the assessment. This difference in channeling feedback is attributed to the different assessment software packages used in the two case studies. The software used in the School of Medicine was called ThinkingCap whereas the Faculty of Education used WebCT software. Feedback in WebCT was presented in the form of a chart, showing the correct response, the student's response and the associated feedback (if available) for each item. Feedback in ThinkingCap Campus was provided by clicking a command button (called 'check') located beside each item. This command button would then display the feedback for that particular item. The primary difference between the two software systems used was the

timing in which feedback was delivered; ThinkingCap provides feedback following a response to an item and WebCT provides feedback after responding to all items in the assessment.

It was hypothesized that feedback following each response would be the preferred method for receiving feedback given that students would be cognitively focused on the concept being assessed. This method of channeling feedback captures students' thinking and either reinforces a correct response or clarifies a misunderstanding. During the focus groups, students in both case studies were queried regarding the timing of feedback. Students were asked whether they would prefer feedback that was presented at the end of the assessment (as it was in WebCT) or following each item (as it was in ThinkingCap). Students from the School of Medicine all preferred feedback immediately following a response to an item. One student added that feedback provided at the end of the assessment would have some value if student results and feedback were presented in the form of a summary chart. Students from the Faculty of Education expressed similar views in response to the need for feedback immediately following an item. Based on the findings from students in the Faculty of Education, feedback provided at the end of the formative CBA was ineffective. This method of channeling feedback risks being unnoticed or is too late to support learning given that the moment for processing each item had passed or the student had completed the task (responded to the items) and was ready to move onto other tasks regardless of whether their responses were correct or incorrect.

Research Question One Summary

In sum, feedback was channelled to students in the two ways: following a response to an item and secondly, at the end of a set of items. Students in the School of Medicine and at the Faculty of Education overwhelmingly preferred feedback that was provided after each response. The medical students also acknowledged the potential value of receiving additional feedback in the form of a summary chart to highlight their strengths and weaknesses in terms of the content and cognitive complexity assessed similar to the chart used in WebCT. However when a summary chart is the sole source of feedback as was the case in WebCT used by students at the Faculty of Education, it did not appear to support student learning given that it went unnoticed or the opportunity for incorporating feedback into learning had passed.

Research Question Two: Types of Feedback

To what extent are particular types of feedback effective in supporting student learning?

As previously found in related studies, feedback indicating whether a response was simply correct or incorrect was the least favoured type of feedback. Feedback that supplemented a correct or incorrect response by providing additional information was found to be the most supportive. When directing students to a resource, (i.e., course text, course notes, internet link), both groups strongly favoured feedback that directed them to a live internet link over other resources. Additional findings discovered in this area of inquiry included the suggestion to reference diagrams used in items and the impact that students' readiness for self-regulated learning had on their preferences for feedback. The discussion related to each type of feedback is presented below.

Correct/Incorrect Feedback

Similar to a prior study (Steven & Hesketh, 1999), feedback that stated a response was correct or incorrect was found to support student learning the least in both case studies. However, students in the School of Medicine focus group suggested that this type of feedback would be suitable for items that assessed lower cognitive skills such as knowledge and recall. One student cited an example of an item which asked how many bones were in the body; in this example, all students agreed that it was not necessary to provide feedback beyond stating whether a students' response was correct or incorrect. The key factor in this finding could be attributed to the extent to which student are ready to engage in self-regulated learning. Students in the School of Medicine focus group believed a learner held some ownership or responsibility for learning hence it was not necessary to provide a hint for items assessing knowledge or recall. Further research is needed to determine whether students' readiness for self-regulated learning is a factor underpinning their preference for this feedback guideline and by extension, a factor underpinning formative CBAs.

Feedback Clarifying a Concept

In both formative CBAs examined in this dissertation, the type of feedback predominantly used was in the form of additional information used to clarify a concept. In most cases, a concept was clarified by describing why an incorrect/correct response was incorrect/correct (e.g., Correct. Although this teacher's comments were well intended, they only focused on the student's weakness or next steps) or by providing more

information in the form of a hint (e.g., More than 4 measurement errors were made). Students from both case studies preferred this type of feedback.

Considerable discussion in both focus groups occurred surrounding the need to clarify correct responses. Both groups came to the same conclusion, noting there was a need to support student learning by clarifying correct responses to items assessing higher cognitive skills. Based on these findings, the criteria for including additional feedback can be determined by the cognitive skills being assessed. When assessing lower cognitive skills (i.e., knowledge), it may not be necessary to include feedback beyond stating that the response was correct. However, students felt it was beneficial to include additional feedback to confirm or clarify their thinking to items requiring higher cognitive skills.

However, the findings suggest that students varied in their use and desire for feedback. Consistent with notions of self regulated learning, some of the students spoke of the need to include questions that would further challenge and motivate them. These students would talk about their own responsibility to use the formative CBA to guide their learning. It was not always necessary to confirm correct responses to knowledge and recall items since the students reported that it was their responsibility to know this information. As students become more self-regulated in their learning, they will likely be able to better judge their needs for descriptive and guiding feedback. There is a need to develop methods to nurture such self-regulation and to study the relationship between self-regulation and the usefulness of formative assessment feedback.

Another finding stemming from the focus group at the Faculty of Education was students' sensitivity to the term incorrect. Students did not feel it was appropriate to use

the term, *incorrect*, given that part of their thinking may be correct. This may be related to these students' readiness for self-regulated learning. As an alternative, students from the Faculty of Education suggested using levels to identify the extent of accuracy in their thinking. The logistics of indicating the degree of accuracy in of a response/distractor in formative CBAs was not discussed but may be worth further research, especially if the formative CBA is to be used with students whose self-regulated learning skills are developing.

There appears to be two aspects of self-regulated learning when considering formative CBAs. The first aspect appears in students' readiness for self-regulated learning. To use formative CBAs effectively students need to be ready for this type of learning environment. Such students demonstrate a certain amount of self-regulated learning skills whereby they are willing to engage in independently learning. The second aspect lies in the potential for formative CBAs to promote self-regulated learning environments. In this respect, formative CBAs can be used as an instrument to nurture students' self-regulated learning skills. These two aspects of self-regulated learning may be cyclical in that one aspect spurs the moment of the other.

Feedback Directing Students to Resources

The feedback provided on the formative CBAs used in both case studies did not include directions to additional resources. As a result, the findings exploring the effectiveness of this type of feedback were based on future preferences rather than direct experiences. Of the three types of resources surveyed (i.e., directing students to their class notes, course text, and Internet links) students from both case studies preferred

directions to a resource using a live internet link. The least favoured resource was students' class notes. In both groups, students felt their class notes were not a reliable source of information. In addition, students in the School of Medicine noted that some topics overlapped and feedback that directed students to their epidemiology notes, for example, might be confusing giving that they may have notes referring to epidemiology in a number of different sections in their notes.

Although students' class text was considered a reliable source, a live internet link was believed to be more supportive given the immediate accessibility. Students from the School of Medicine stipulated some criteria for using internet links. Firstly, the link must be live or active and, secondly, it must be a reliable source of information.

In sum, students favoured feedback that supplemented a correct (high cognitive items) and incorrect response. When feedback directs students to a resource, they preferred a live internet link over feedback that directed them to their course notes or course text. The ease and efficiency of accessing information in this manner were cited as reasons supporting this preference.

There were no significant differences found between the independent variables (age and gender) and the items exploring feedback in the School of Medicine case study. Similarly, no significant differences were found between two independent variables (program and assessment knowledge) and the items exploring feedback in the Faculty of Education case study. A significant finding between the independent variable division and five dependent variables (i.e., four types of feedback, 1 factor that focused on the

student attitudes towards the effectiveness of feedback) confirmed the hypothesis indicating students do differ based on their division.

Based on this finding, students from the intermediate/senior division responded more positively in terms of the effectiveness of different types of feedback, the value of formative CBA, and item characteristics. It is possible that, as hypothesized, the differences in undergraduate experiences may be producing post secondary graduates with different skills. At this point, the factors contributing to the differences based on division are largely unknown. It is suspected that intermediate/senior students' may be more prepared for self-regulated learning given that they responded more positively towards using the formative CBAs to direct learning. However, further research is needed to explore these and other factors that may be contributing to the differences in the two divisions.

Research Question Two Summary

In sum, the different types of feedback do not equally support student learning. Feedback which simply stated a response was correct or incorrect was judged the least effective. However, this type of feedback may be considered sufficient for low cognitive items when used with students who exhibit high self-regulated learning skills. Feedback which supplemented a correct or incorrect response was reported to be the most effective in supporting learning. Feedback statements that directed students to a resource, in particular, a live internet link, were judged to be more effective than directions to course notes or a course text.

Research Question Three: Factors Affecting Formative Computer-based Assessments

What factors if any, contribute to students' perceived usefulness of formative CBAs?

The factors suspected of contributing to students' perceived usefulness of formative CBAs were organized according to: perceived purpose of formative CBAs, value of formative CBAs, and item characteristics such as difficulty, format, number, and content. The findings from these three areas of inquiry are discussed separately below and are followed by a discussion of the significant differences in response variations based on the independent variables examined.

Purpose of Formative CBA

A purpose of the formative CBAs in both case studies was to support student learning by creating a self-regulated learning environment. Students from the School of Medicine generally believed their formative CBAs were to be used in this manner; however, a relatively larger proportion of students from the Faculty of Education believed their formative CBAs held summative characteristics. The discussion surrounding these findings is presented below and includes a possible explanation for the differences in the two findings.

Students from the School of Medicine were positive about using their formative CBAs to support their learning in terms of using the instrument to self-assess, direct learning, and identify strengths and weaknesses. In comparison, students from the Faculty of Education were generally not as positive about using the formative CBAs to support their learning. Based on the findings from the questionnaire and focus group, it was evident that several Bachelor of Education students held some misunderstanding about

the purpose of the formative CBAs in that students believed the formative CBAs held summative properties. Just over 20% of students believed the purpose of the assessment was to provide instructors with a mark in the course and others believed the purpose was to provide information about student knowledge (42.5%). During the focus group, students frequently referred to the formative CBAs as a test and continually referenced their score on the formative CBAs over their learning of new concepts. Believing the formative CBAs were being used to measure rather than support learning provides evidence of a misunderstanding of the purpose of the instrument.

This finding is disturbing given that the purpose of the formative CBAs was verbally explained at the beginning of the Assessment Module. It is possible that students may have been too overwhelmed with the procedures and routines of starting a new program and consequently had difficulty with the concept of formative assessment in general. However, when students completed the questionnaire following the completion of the Assessment Module and participated in the focus group at the beginning of the following semester, it would seem reasonable to expect students to have acquired a better understanding of the meaning of formative assessment from this module, their curriculum courses, and practicum experience. It is possible that students who have not been frequently exposed to formative assessment practices in their own learning find it difficult to grasp the purpose of the formative CBAs. Lastly, the short exposure to formative assessment (i.e., formative CBA) in the Assessment Module may not have been enough exposure to cause a paradigm shift in students' thinking about the benefits and uses of formative assessment.

Collaboration. It was believed that evidence of collaboration would be an indicator of students' understanding of the formative aspect of the formative CBAs since discussing the scenarios with colleagues was permissible and encouraged. However, most students from the School of Medicine reported they completed the formative CBAs individually and off-campus. Hence collaboration was not a feasible means of completing the assessment. Students from the School of Medicine also indicated that most items on their formative CBA were relatively easy hence there was no need to collaborate. As a result, measuring the extent to which students collaborated was not a good indicator of their engagement with formative CBAs.

Students from the Faculty of Education also reported low levels of collaboration. Informal observations of students completing the formative CBAs in the Faculty of Education in a number of different labs and in wireless areas, suggested that these students were collaborating. The focus group participants suggested these observations missed an important aspect of the observed collaborations; students were not collaborating about the scenarios but rather were simply sharing the correct responses.

No significant differences were found between the grouping variables (age and gender) in the School of Medicine or with the grouping variables (program, division, and assessment knowledge) in the Faculty of Education and the independent variables exploring students understanding of the purpose.

Value of formative CBA

The value students held for their formative CBAs was examined from three different perspectives: (a) student (priority given to the formative CBA, perception of the

task, personal schedule, and effort); (b) instrument (helpful, and guided learning); and (c) assessment (fair and accurate). Each of these perspectives is discussed separately below.

Value (Student). Students from the School of Medicine ranked the priority of the formative CBA moderately and slightly higher than students from the Faculty of Education. Likewise, both groups were relatively neutral on those items exploring students' value of the formative CBAs (formative CBAs were busy work, too busy to complete formative CBA, and effort put forth). Given these neutral responses, students in both groups were asked what was needed to move their responses more towards the positive end of the scale. Students from the School of Medicine indicated that being aware of the number of items in each formative CBA and the estimated time required to complete the formative CBA would have been helpful. It is unlikely that improvements in these areas would change how students prioritize the formative CBAs but, for example, knowing the approximate amount of time it would take to complete a formative CBA may help students work these tasks into their schedule. Conversely, knowing that a formative CBA would take more time than students could allocate, may deter some students from beginning the task. Regardless, including the number of items and the estimated time to complete a formative CBA at the onset of the assessment can be viewed as a feasible improvement.

The second point raised by students from the School of Medicine is more likely to influence students' perception of formative CBAs. The relationship between formative CBAs and subsequent summative assessments was voiced as a factor that has the potential to change how students perceive formative CBAs. Based on comments from

students in the School of Medicine, if the formative CBAs were more aligned with their summative assessment in terms of item difficulty and content, they believed students would be more apt to complete the formative CBAs and with greater effort. For this group of students, the difficultly level of items on the formative CBAs, was less than what they experienced on their summative assessments. This resulted in a misalignment between their formative and summative assessments. On one hand, students acknowledged that their formative CBAs may have only been intended to provide an introductory level of assessment but on the other hand, students spoke very strongly about the benefits of being able to use the formative CBAs to provide them with an accurate measure of how they could expect to perform on subsequent summative assessments.

Although the formative CBAs used in the Faculty of Education was followed by a summative CBA, the formative and summative CBAs had different purposes. The context of the formative CBAs used in the Faculty of Education was that of a Prior Learning Assessment; consequently, there were no concerns related to the alignment between the formative CBAs and summative CBA. Concerns raised by students from the Faculty of Education focused on improving the items (discussed below) and a better method of delivering the course material. Students suggested that theory and terminology should be taught prior to the application of theory, in particular, the application of theory on formative assessment. Students from the Faculty of Education did not generally like using formative CBAs as a Prior Learning Assessment that exposed what they did not

know about the field of assessment. This focus on the lack of knowledge may have negatively impacted students' experiences and perceptions.

Value (instrument). Similar to the responses above, both groups of students reported that the formative CBAs were moderately helpful and guided their learning. In the focus group discussion following the presentation of these findings, students did not provide any further insights other than those discussed above about the alignment between the formative and summative assessments.

Value (assessment). Approximately half of the students in both case studies reported that their formative CBAs were fair and accurate. However, the reasons for these findings were different. Students from the School of Medicine responded in terms of the alignment between the formative CBAs and their subsequent summative assessments. Given that the items on their formative CBAs were reported to be less difficult than items on students' summative quizzes and were anticipated to be much less difficult than items on their final exam, a large number of students did not believe the formative CBAs were fair. This finding highlights the previously reported lack of certainty surrounding the purpose of the formative CBAs used in the School of Medicine in terms of whether the formative CBAs were to be used as an introductory instrument to allow students to familiarize themselves with new concepts or whether the instrument was to prepare them for subsequent summative assessments.

Students from the Faculty of Education focused their discussion on improving the items and using a better feedback mechanism. The idea of adding practice scenario items to the Assessment Module website was raised as a possible means of making the

formative CBAs more fair and accurate. In this context, a non-threatening way of familiarizing students with formative CBAs was needed. The desire for practice assessment items prior to the formative CBAs provides evidence of the perceived summative undertones associated with the formative CBAs used in the Faculty of Education.

No significant differences were found between the two factors exploring students' perceived value of the formative CBA and the independent variables age and gender (School of Medicine). Similar to previous findings in the Faculty of Education, a significant difference was found between the first factor (i.e., perceived usefulness of the formative CBA) and division. Intermediate/senior students responded more positively towards the perceived usefulness of the formative CBA.

Item Characteristics

Difficulty. Students from both case studies recognized the need for a large number of items that assessed the full range of cognitive skills. Indirectly, both groups identified a value for items that specifically assessed lower cognitive skills. One student explained the purpose of these items was to "build my confidence." At the other end of the difficulty scale, students, particularly those from the School of Medicine, wanted a few items that assessed higher cognitive thinking; the type of thinking they could anticipate on their final examination. When formative CBAs are followed by subsequent summative assessments, students believed it was important that the difficulty of items on the formative CBAs paralleled the difficulty of items on summative assessments. Students

stated that such an alignment would make the formative CBAs more valuable and there would be more incentive to complete all items on all formative CBAs.

The last point raised was the manner in which items were presented. Students from the Faculty of Education felt that the large scenario-based items they experienced were made more difficult by the presentation format of the item. Students suggested that bodies of text should be double spaced and the scenario should be separated from the item. Although the presentation of items in the formative CBAs used in the two case studies was controlled by the assessment software, this finding is important to acknowledge for future developments of formative CBAs. When possible, the layout or presentation of the item should be considered when creating formative CBAs to ensure that items are pleasing to the reader and do not hinder readability.

Format. The multiple-choice format was problematic for students from the Faculty of Education primarily due to how these students were introduced to the theory surrounding multiple-choice items in the Assessment Module. According to students who participated in the focus group, multiple-choice items were not promoted as good items for classroom assessment. In addition, multiple-choice items were to contain a short stem. When students were presented with multiple-choice items used in a non-traditional format (i.e., with expanded stems to describe a scenario), students had difficulty accepting and using this new style of multiple-choice item. This gap in students' knowledge about how multiple-choice items can be used effectively in classrooms, in particular, classrooms in professional learning programs, needs to be addressed to help students better understand the complexity and theory behind multiple-choice items.

It is interesting that students from the School of Medicine did not voice any reservations related to the scenario-based multiple-choice format. This finding may be because the formative CBA items used in the School of Medicine were used in a relatively traditional format in that the scenarios were brief (i.e., two to four sentences) and, according to students, assessed lower to mid cognitive skills.

Other item formats used in the School of Medicine included a word scramble, matching columns, and open-ended items. Students liked the word-scramble and matching columns item formats the least because these items simply focused on knowledge and recall. Students from the School of Medicine appeared to be less concerned with format and more with the cognitive level being assessed. Consequently, students preferred more challenging multiple-choice items over all the other formats including the open-ended format. Students admitted they liked the open-ended item since it was more apt to assess higher cognitive skills; however, none of the students actually constructed a response to the open-ended item. Students stated that they thought about a response prior to checking their thinking but did not actually construct their own response. Students reported they were more likely to pass over this item if their time was limited. Based on this finding, it may be important to consider a balance between open-ended items and multiple-choice items (assessing higher cognitive skills) when creating formative CBAs.

Number of items. Although the number of items in the formative CBAs varied between the two case studies where the School of Medicine had more items on average than used at the Faculty of Education, students from both focus groups strongly

recommended increasing the number of items on their formative CBAs. Students believed more items would give them a more accurate reflection of their understanding. In particular, students from the School of Medicine requested more items assessing higher cognitive skills.

Content. Content covered on the formative CBAs did not appear to be a major factor influencing students' experiences with the formative CBAs. Similar to the issues related to the alignment of difficulty between the formative CBAs and subsequent summative assessments, students from the School of Medicine believed there was a need for more thorough content coverage with at least one item for each major concept.

No significant differences were found between the two factors exploring item characteristics and the independent variables age and gender (School of Medicine). A significant difference with practical significance was found between the independent variable division and the one factor representing item characteristics. Once again, students from the intermediate/senior division were more satisfied with formative CBA items than students in the primary/junior division. This finding provides clear evidence that students from the intermediate/senior division not only responded differently than students in the primary/junior division but more positively, indicating a more positive experience with formative CBAs. More research is needed to explore the characteristics students entering the Bachelor of Education program carry with them such that one group has different experiences using formative CBAs than another group. This finding has direct implications for the Bachelor of Education program. One recommendation would be to postpone the use of formative CBAs in the primary/junior division until further

research can determine why students in the primary/junior division do not experience formative CBAs as positively as their counterparts.

Research Question Three Summary

In sum, understanding the purpose of formative CBAs and a well developed set of formative CBA items appears to influence students' experiences with formative CBAs. Firstly, a clear understanding of the purpose of the formative CBA has a direct impact on helping students view the formative CBAs as instruments to support rather than measure their learning. Once students understand the purpose of the instrument, the findings related to how students value the instrument are likely to improve. Secondly, a well developed set of items will better support student learning. The predominant multiple-choice format did not appear to present any issues for students; however, the word-scramble and matching format were seen to be least desirable because they assessed low cognitive skills. The use of scenarios was generally viewed as a positive aspect of formative CBAs providing that the presentation of the scenario is readable and contextually relevant. If the formative CBAs are intended to prepare students for a summative assessment then the formative CBA items should reflect the content, style, and cognitive complexity of items on the summative assessment. The ideal number of items in a formative CBA is still unknown.

Conclusion

Implications for Research and Practice

This section connects the findings in each of the areas explored with previous studies and theory related to formative CBAs. The purpose was to conclude this dissertation by bridging this research with previous research and provide recommendations for the use of formative CBAs and directions for further research.

The findings from this dissertation confirm and extend what is known about channelling feedback on traditional pencil and paper assessments where immediate feedback was shown to be more effective than delayed feedback (Kulik & Kulik, 1988). When using formative CBAs as the instrument of choice, immediate feedback was also shown to be more effective than delayed feedback. What is important in this finding is that feedback on formative CBAs is considered immediate when compared to pencil and paper assessments; however, there are variations in the amount of time that elapses between responding to an item and when feedback is provided on formative CBAs. This time lapse or delay in channelling feedback in WebCT, for example, was short (i.e. at the end of the assessment) when compared to the delay in feedback provided by a classroom teacher on a pencil and paper assessment. However, this delay impacted students' perception of learning rendering the feedback ineffective; students either missed the feedback or intentionally ignored it. As previously noted by Hattie and Timperley (2007), feedback has its greatest impact when it is provided while students are in the process of constructing knowledge. This applies to feedback on formative CBAs as well. Feedback provided during the learning process, as was the case with the formative CBAs used at

the School of Medicine, was reported to be more effective than feedback provided at the end of the assessment in WebCT. Although feedback provided at the end of the assessment did not support learning at the moment knowledge was being constructed, this delay in providing feedback could provide a summary of students' strengths and weaknesses in various strands and areas of cognitive development. In sum, effective feedback on formative CBAs requires that feedback to students be provided immediately after the student has responded to an item.

Feedback on formative CBAs was previously shown to be a key factor in the growing use of these instruments (Gipps, 2005). However, little was previously reported about which aspects of feedback contributed to the growing use of formative CBAs. There are a number of different ways that feedback used to support students' learning can be stated. This dissertation explored feedback that: simply stated whether a response was correct or incorrect, clarified a concept, or directed students to additional resources. Although a previous study found that feedback that stated a response was correct or incorrect did not support learning (Pelton & Pelton, 2006), this research identified an application for this type of feedback. Such feedback may be acceptable for low cognitive items. This finding needs to be further explored, comparing this form of feedback with other forms of feedback provided for low cognitive items.

The most favoured type of feedback found in both case studies was feedback that clarified a concept. Students also considered feedback that directed them to additional resources useful, especially if it directed them to a trusted internet source. Since this

finding was based on students' self-reported beliefs, it is necessary to examine the effectiveness of this type of feedback in an operational formative CBA.

Overall, the feedback used in the two formative CBAs adhered to the principles of feedback outlined by Nicol and Mcfarlene-Dick (2006) whereby the instrument facilitated self-assessment by providing high quality directions for further learning. The opportunity to complete the formative CBAs without students revealing their scores had the potential to promote positive self-esteem. However, the perceived summative undertones voiced by students from the Faculty of Education, may have undermined their experience with this instrument. Hence it is recommended that instructors continually work with students to help them understand the purpose of formative CBAs by providing numerous formative CBAs throughout the duration of the module.

The factors influencing the usefulness of formative CBAs have not been previously well studied. Consequently, much was learned about the influence of students' understanding of the purpose of the formative CBAs, how students valued the instrument, and the items used in the formative CBAs.

Similar to previous studies (e.g., Boshizen, Brome, & Gruber, 2004; Kaser, Mundy, Stiles, & Loucks-Horsley, 2006), the purpose of the two formative CBAs examined in this dissertation was primarily to support students' learning. At the same time, the formative CBAs were used to address large class sizes and provide students with a certain amount of freedom to adapt and regulate their learning environment. It appears these intended purposes of the formative CBAs may not have been well understood or adopted by students given their ubiquitous prior experiences with

summative assessments. Hence it was hypothesized that given a limited exposure to formative assessment opportunities in general, some students would not have understood or had the opportunity to adapt to a new method of assessment. This certainly was found with students from the Faculty of Education. Conversely, the School of Medicine students' understanding of the purpose of their formative CBAs was more commonly aligned with the intended purposes. This was possibly due to the repeated access these students had to different formative CBAs presented in the MSK Module over the four week span of the module. In future formative CBAs, it is recommended that students be provided with the opportunity to experience formative CBAs over a period of time to familiarize themselves with the nature of the instrument (Sambell, Sambell, & Sexton, 1999) and also to acquire an understanding of the instrument's purpose. Following these recommendations will provide opportunities for students to begin to acquire an appreciation and understanding of formative CBAs.

Students' perceived value of formative CBAs was connected to their understanding of the purpose of the instrument as well as the validity of the instrument. If students have a good understanding of the purpose of formative CBAs and the instrument includes appropriate items, there is a good chance that students will value the instrument.

The findings also reveal the importance of recognizing that students have busy schedules that may not allow time to complete all formative CBAs. Students' decisions to complete as many or as little of the formative CBA items should not be seen as a failure or limitation of formative CBAs. This creates an intriguing problem for future formative CBA development because the students from both case studies recommend adding more

items. If new items are developed to cover a range of cognitive levels and curriculum areas, students could potentially decide whether to narrow their formative CBAs by focusing on a specific cognitive level or content area. Alternatively, students could complete all areas of the formative CBA, if time permits.

Students value formative CBAs that inform them about their subsequent performance on summative assessments. Students also value formative CBAs that allow them to assess their learning acquired through independent learning opportunities. Thus there are at least two distinct classes of formative CBAs. One class of formative CBAs helps prepare a student for a summative assessment. If this is the case, there needs to be alignment between the two instruments such that the formative CBA(s) prepares students for the subsequent summative assessment. Students who do not feel the two instruments are well aligned will rate the value and fairness of formative CBAs low. The second class of formative CBAs is not aligned with a summative assessment. In this case, the purpose of the instrument is to provide students with a self-assessment to guide learning in an area of independent study.

Given the reliance of formative CBAs on multiple-choice items, it is critical that research in this area continue to focus on the quality of items; in particular, the quality of multiple-choice items assessing higher cognitive skills. In such items, scenarios are likely to be used in the stem, as was the case in both case studies examined in this dissertation. These scenarios differed in terms of the size and by extension, the amount of information provided to situate the scenario. In one case study (i.e., Faculty of Education) the scenarios were often viewed as overbearing and too complex; this was not the case for

the smaller scenarios used in the formative CBAs at the School of Medicine. Much is still to be learned about the amount of information that is needed in scenarios and whether scaffolding of large scenarios is required to present the scenario in smaller, easier to comprehend sections. It is possible that as students become more familiar with using scenario-based multiple-choice items, the scenarios could become more complex thus introducing a hierarchical aspect to scenarios that parallel students' experience with this item format. This too needs further research to explore these factors and other factors underpinning the successful use of scenario-based multiple-choice items on formative CBAs.

In terms of the number of items a formative CBA has, more is generally better. In both case studies, students indicated there were not enough items and as a result of this construct under-representation, the validity of the instrument was undermined as was predicted by Crisp and Ward (2007). The validity of the formative CBAs used in the Faculty of Education was affected more due to the small set of items which is highly suspected of negatively influencing students' formative CBA experience. Although there is a need for a large number of items, it is important to acknowledge the time commitment to develop formative CBA items, in particular, scenario-based multiple-choice items and recognize that formative CBAs are evolving. As each formative CBA is used, it is hoped that instructors will improve upon existing items and add more items to the data set. What is the minimum number of items a formative CBA can have and what is the balance of items in terms of cognitive complexity or difficulty? These are not easy questions given the context for each formative CBA is likely to vary considerably. Based

on my experience as a researcher in the field of formative CBAs and a teacher of 15 years, I would recommend a minimum of 8 to 10 items when piloting a formative CBA as was the case with the School of Medicine. The range of cognitive skills would depend on students' level of learning where, for example, a set of items designed for students beginning a program (such as students from the Faculty of Education) might have more items at the lower to middle end of the cognitive scale with a couple items tapping into students higher cognitive skills. Students who are well into a program, such as students from the School of Medicine, need more challenging items; hence items for this group of learners would contain more items at the higher end of the cognitive scale. The range of difficulty and to some extent the number of items is very much contextually based and, as a result, may differ from one context to another.

It is necessary to reflect on the technological factors that can influence students' experience with formative CBAs. A lack of familiarity with computers and the formative CBA software was previously cited as a potential disadvantage of using formative CBAs due to construct-irrelevant variance (Wise, Bhola, & Yang, 2006). This was not the case in the formative CBAs used in these two case studies given students' high familiarity rating with the software and computers. Further, given the advancing technological presence in society combined with user-friendly interfaces, construct-irrelevant variance related to technology is not likely to be a factor in formative CBAs used in other contexts.

Together, these factors, the timing of feedback, and the directions given to support student learning have influenced students' experience with their formative CBAs.

It is not possible to state that one factor in particular is responsible for influencing students' experiences more so than another factor the interplay of these factors is likely to continue to influence students' experiences with formative CBAs.

Limitations

Although the questionnaires completed in the two case studies represented the corresponding populations, by design, the small number of participants in each of the focus groups may not have represented the views or thoughts of the entire population. This may be a greater concern for the Faculty of Education case study because students were grouped by program and division. In hindsight, it would have been more informative to have conducted focus groups according to these groupings. In addition, it would have been beneficial to group students according to the extent they engaged in self-regulated learning given that self-regulation appeared to be a key component influencing students' experiences with formative CBAs. The second limitation related to focus groups applies to all focus groups; biasing may have been introduced in phrasing questions and steering focus group discussions.

Commonly, factor analysis is performed on an entire set of questionnaire items to determine the component structures. Due to the number of students in the MSK module, it was not possible to complete a first order factor analysis of all of the items to determine if the thematic groupings were viable. The sample size from the Faculty of Education was larger but still was below the recommended 10 participants per item. Thus only second order factor analyses were conducted. Further internal consistency measures (Cronbach's alpha) were used to ensure the items within each subscale were providing a reliable

measure. The last limitation is related to the case study methodology. As with all case studies, when focusing on one or two populations, caution should be exercised when generalizing the findings to the greater population. Given that the findings from this dissertation were drawn from two case studies, generalizing these findings to inform the use of formative CBAs in other higher education programs may be appropriate.

Next Steps for the Two Case Studies

ThinkingCap Campus has recently been abandoned by the School of Medicine due to high cost, difficulties accessing technical support, and incompatibility with some browsers. As of the 2008/2009 school term, the School of Medicine will be using a product known as eXe. This CBA is the result of the New Zealand government post secondary education commission electronic collaboration (eCollaboration) fund which was led by the University of Auckland, Auckland University of Technology, and Tairawhiti Polytechnic.

WebCT is still being used in the Assessment Module at the Faculty of Education primarily due to a lack of funds to purchase new formative CBA software. New multiple-choice scenario-based items were developed. Feedback for these items focused on identifying parts of students thinking that was correct and extending thinking towards the correct response. In addition, the Assessment Modules were reduced from four to three areas of inquiry and students' preparation for the formative CBA included discussions on professional learning, self-regulated learning, and the purpose of the formative CBAs.

REFERENCES

Ahmed, A., & Pollitt, A. (2007). Improving the quality of contextualized questions: An experimental investigation of focus. *Assessment in Education, 14*(2), 201–232.

Albertelli, G., Kortemeyer, G., Sakharuk, A., & Kashy, E. (2003). *Personalized examinations in large on-campus classes*. Paper presented at Frontiers in Education: Boulder, CO. Retrieved December 12, 2008 from http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/8925/28250/01263321.pdf?arnumber=1263321

American Council on Education. (1995). *Guidelines for computerized adaptive test development and use in education*. Washington, DC: Author.

American Psychological Association (1986). In J. Olsen (Ed.), *Guidelines for computer based testing*. Retrieved January 22, 2007, from http://www.isoc.org/oti/articles/0500/ olsen.html

Association of Test Publishers. (2000). *Guidelines for computer-based testing*. Washington, DC: Author.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education, 85,* 536–553.

Bennett, R. E. (2001). How the internet will help large-scale assessment reinvent itself. *Educational Policy Analysis Archives, 9*(5), 1–26.

Bennett, R. E. (2004). Reinventing assessment: How the internet will help large-scale assessment reinvent itself. *Education Policy Analysis Archives*, *9*(*5*), 1–26.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139-148.

Black, P., & Wiliam, D. (1999). Assessment for learning: Beyond the black box. *Assessment Reform Group.* Retrieved January 22, 2007, from http://arg.educ.cam.ac.uk/AssessInsides.pdf

Blackboard Inc. (2005, October 12). *Blackboard and WebCT announce agreement to merge* [Press release]. Retrieved April 16, 2008, from http://investor.blackboard.com/phoenix.zhtml?c=177018&p=irol-newsArticle&ID=767025

Blackboard Inc. (2008). *WebCT campus edition: Content presentation*, Retrieved April 26, 2008, from http://www2.blackboard.com/ce6/viewpage?name=ce6_ content_presentation

Bloom, B. S. (1956). *Taxonomy of Educational Objectives: Handbook 1: The cognitive domain*. New York: Longmans, Green.

Boshuizen, H. R., Bromme, R., & Gruber, H. (2004). (Eds.) *Professional learning: Gaps and transitions on the way from novice to expert*. Dordrecht: Kluwer Academic Publishers.

Burrow, M., Evdorides, H., Hallam, B., & Freer-Hewish, R. (2005). Developing formative assessments for postgraduate students in engineering. *European Journal of Engineering Education, 30*, 255–263.

Catterall, M., & MacLaran, P. (1997). Focus group data and qualitative analysis programs: Coding the moving picture as well as snapshots. Sociological

Research Online. Retrieved September 2, 2008, from http://www.socresonline.org.uk/2/1/6.html

Challis, D. (2005). Committing to quality learning through adaptive online assessment. *Assessment and Evaluation in Higher Education, 30*(5), 519–527.

Charman, D., & Elmes, A. (1998). A computer-based formative assessment strategy for a basic statistics module in geography. *Journal of Geography in Higher Education, 22,* 381–385.

Christensen, D., Nowak, E., Prinster, M., & Rinke, L. (2006). *How do educational leaders implement formative assessment to improve student learning?* Unpublished graduating paper, Educational Administration and Leadership program, Department of Educational Studies, Faculty of Education, University of British Columbia, Vancouver, B.C. Retrieved February 4, 2008, from http://slc.educ.ubc.ca/Masters/2006/ChristensenAbstract.pdf

Cotton, D., & Gresty, K. (2006). Reflecting on the think-aloud method for evaluating e-learning. *British Journal of Educational Technology, 37,* 45–54.

Crisp, V., & Ward, C. (2008). The development of a formative scenario-based computer assisted assessment tool in psychology for teachers: The PePCAA project. *Computers and Education, 50,* 1509–1526.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58,* 438-481.

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point, and 10-point scales. *International*

Journal of Market Research, 50, 61–77.

Dorman, S. (1998). Technology and the gender gap. *Journal of School Health, 68*, 165-166.

Duffy, F. D., & Holmboe, E. S. (2006). Self-assessment in lifelong learning and improving performance in practice. *Journal of American Medical Association, 296*, 1–5.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: Bradford Books/MIT Press.

Fanshawe College (2008). *Writing multiple choice items to assess higher cognitive processes*. Retrieved February 21, 2008, from http://gs.fanshawec.on.ca/cdr/writing_multiple_choice_items_to.htm

Fletcher, M., Kearney, J., & Bartlett, B. (2005). *Revising the vision: Learning with computer-based assessment*. Retrieved August 18, 2008, from http://www98.griffith.edu.au/dspace/bitstream/10072/1449/1/19702_1.pdf

Fostaty-Young, S., & Wilson, R. J. (2000). *Assessment and learning: The ICE approach*. Winnipeg, MB: Portage & Main press.

Frey, J. H., & Fontana, H. (1993). The group interview in social research. In D. Morgan (Ed.), *Successful focus groups: Advancing the state of art* (pp. 20–34). Newbury Park, CA: Sage.

Gallagher, A., Bennett, R. E., & Rock, D. A. (2002). Validity and fairness in technology-based assessment: Detecting construct-irrelevant variance in open-ended, computerized mathematics tasks. *Educational Assessment, 8*, 27–41.

Georgiadou, E., Triantafillou, E., & Economides, A. (2006). Evaluation parameters for computer adaptive testing. *British Journal of Educational Technology*, *37*(2), 261–278.

Gipps, C. V. (2005). What is the role of ICT – based assessment in university? *Studies in Higher Education*, *30*, 171–180.

Greene, J. C., & Caracelli, V. J. (2003). Making paradigmatic sense of mixed methods practice. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioural research* (pp. 91–110). Thousand Oaks, CA: Sage.

Gretes, J. A., & Green, M. (2000). Improving undergraduate learning with computer-assisted assessment. *Journal of Research on Computing in Education*, *33*, 46–54.

Hattie, J. (1999). *Influences on student learning*. Inaugural professorial address. Unpublished manuscript. Retrieved January 4, 2008, from http://www.arts.auckland.ac.nz/edu/staff/jhattie/Inaugural.html

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

Haladyna, T. M. & Downing, S. M. (1989). Validity of a taxonomy of multiple-choice item writing rules. *Applied Measurement in Education*, *2*, 51–78.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15, 309–334.

Half-baked Software Inc. (2008). *Hot potatoes version 6*. Retrieve September 3, 2008, from http://hotpot.uvic.ca/

Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education, 35*(2), 213–224.

Harper, R. (2003). Correcting computer-based assessments for guessing. *Journal of Computer Assisted Learning, 19*, 2–8.

He, Q., & Tymms, P. (2005). A computer-assisted test design and diagnosis system for use by classroom teachers. *Journal of Computer Assisted Learning, 21*, 419–429.

Henly, D. (2003). Use of web-based formative assessment to support student learning in a metabolism/nutrition unit. *Dental Education, 7*, 116–122.

Henley, D., & Reid, A. E. (2001). Use of the web to provide learning support for a large metabolism and nutrition class. *Biochemistry and Molecular Biology Education, 29*, 229–233.

Higgins, C., & Bligh, B. (2006). *Formative computer-based assessment in diagram based domains.* Paper presented at the Annual Joint Conference Integrating Technology into Computer Science Education, Proceedings of the 11th annual SIGCSE conference on innovation in technology in computer science education. Retrieved January 25, 2007, from http://portal.acm.org/citation.cfm?id=1140123.1140152&coll=&dl=ACM&CFID=15151515&CFTOKEN=6184618

Hill, J., Meyer, J. Sanders, K., Fyfe, G. Ziman, M., & Koehler, N. (2008). *Taking baby steps: The impact of test length on first year student engagement with online formative assessments in human biology*. Proceedings of the 17th Annual Teaching Learning Forum, 30-31 January 2008. Perth: Curtin University of

Technology. Retrieved on August 23, 2008, from http://lsn.curtin.edu.au/tlf/tlf2008/refereed/hill.html.

Hodgen, J., & Marshall, B. (2005). Assessment for learning in English and Mathematics: A comparison. *The Curriculum Journal*, *16*(2), 153–176.

International Test Commission. (2005). *International guidelines on computer-based and internet delivered testing*. Retrieved January 23, 2007, from http://www.intestcom.org/

Jodoin, M. (2003). Measurement efficiency of innovative item formats in computer-based testing. *Journal of Educational Measurement, 40*, 1–15.

Kaser, J. S., Mundry, S. E., Stiles, K. E., & Loucks-Horsley, S. (2006). *Leading everyday: 124 actions for professional leadership*. Thousand Oaks, CA: Corwin press.

Katz, L., & Dalby, J. T. (1981). Computer and manual administration of the Eysenck Personality Inventory. *Journal of Clinical Psychology*, *37*, 586–592.

Klinger, D., DeLuca, C., & Miller, T. (2008). The evolving culture of large-scale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy*, *76*, 1–34. Retrieved on December 12, 2008, from http://www.umanitoba.ca/publications/cjeap/articles/klinger.html

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254-284.

Krueger, R., & Casey, M. (2000). *Focus groups: A practical guide for applied research*

(3rd ed.). Newbury Park, CA: Sage.

Kulik, J., & Kulik, C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*, 79–97.

Lazeron, M., Wagner, U., & Shumanis, N. (1999). *What makes revolution: Teaching and learning in higher education, 1980–2000*. National Center for Post Secondary Improvement, School of Education, Stanford University. Retrieved on February 5, 2008, from http://www.stanford.edu/group/ncpi/documents/pdfs/5-11_revolution.pdf

Lee, J. A., & Hopkins, L. (1985). The effects of training on computerized aptitude test performance and anxiety. In M. Russell, A. Goldberg, & K. O'Conner (Eds.), *Computer-based Testing and Validity: A Look Back and into the Future*. Retrieved February 12, 2008, from http://www.bc.edu/research/intasc/PDF/ComputerBasedValidity.pdf

Linn, R., & Gronlund, N. (2000). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Merrill.

Mertens, D. M. (2005). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (2nd ed.). Thousand Oaks, CA: Sage.

McCormick, R. (2004). ICT and pupil assessment. *The Curriculum Journal*, *15*(2), 115–137.

Miller, T. (2004). *Assessment practices in grade 9 academic and applied stream mathematics*. Unpublished master's thesis, Queen's University, Canada.

Miller, T. (in press). Formative computer-based assessment in higher education: The effectiveness of feedback in supporting student learning. *Assessment and Evaluation in Higher Education*.

Miller, T., & Shulha, L. (2008). Formative computer-based assessment in higher education. In K. McFerrin (Ed.), *Proceedings of Society for Information Technology and Teacher Education International Conference 2008* (pp. 105–111). Chesapeake, VA: AACE.

Morgan, D. L. (1997). *Focus groups as qualitative research*. Thousand Oaks, CA: Sage.

National Center for Post Secondary Improvement. (September/October, 1999). Revolution or evolution? Gauging the impact of institutional student assessment strategies. *The Landscape in Change*, 53–56.

Nicol, D. (2007). E-assessment by design: Using multiple-choice tests to good effect. *Journal of Further and Higher Education*, *31*, 53–64.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218.

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment and Evaluation in Higher Education*, *33*, 301–314.

Ontario Ministry of Education and Training. (2000). *The Ontario curriculum, grades 9 to 12: Program planning and assessment*. Toronto, ON: Queen's Printer for Ontario.

Pallant, J. (2005). *SPSS survival manual*. Maiden head, Philadelphia: Open University

Press.

Peat, M., & Franklin, S. (2002). Supporting student learning: The use of computer-based formative assessment modules. *British Journal of Educational Technology, 33,* 515–523.

Pelton, T., & Pelton, L. F. (2006). Introducing computer-adaptive testing system to a small school district. In S. L. Howell & M. Hricko (Eds.), *On Line Measurement: Case Studies from Higher Education, K-12 and Corporate.* Idea Group Inc.

Pinckey, R. D., Mealy, M.J., Thomas, C. B., & MacWilliams, P. S. (2001). Impact of a computer-based auto-tutorial program on parasitology test scores of four consecutive classes of veterinary medical students. *Journal of Veterinary Medical Education, 28,*136–139.

Pintrich, P. R., & Zusho, A. (2002). Student motivation and self-regulated learning in the college classroom. In J. C. Smart & W.G. Tierney (Eds.), *Higher Education: Handbook of Theory and Research, XVII*, New York: Agathon Press.

Pitt, S. J., & Gunn, A. (2004). The value of computer-based formative assessment in undergraduate biological science teaching, *Bioscience Education Electronic.* Retrieved February 5, 2008, from http://bio.ltsn.ac.uk/journal/vol3/beej-3-1.htm

Principles for Fair Assessment Practices for Education in Canada. (1993). Edmonton, Alberta: Joint Advisory Committee. Retrieved May 12, 2008 from http://www.education.ualberta.ca/educ/psych/ crame/files/eng_prin.pdf

Professional Testing Inc. (2006). *Building high quality examination programs*. Retrieved February 23, 2008 from

http://www.proftesting.com/test_topics/steps.shtml

Raikes, N., & Harding, R. (2003). The horseless carriage stage: Replacing conventional measures. *Assessment in Education, 10*(3), 267–277.

Ramsden, P. (2003). *Learning to teach in higher education*. London: Routledge-Falmer.

Reckase, M. D. (2003). *Item pool design for computerized adaptive testing*. Paper presented at the annual meeting of the National Council of Measurement in Education. Chicago, IL. Retrieved December 12, 2008 from http://www.psych.umn.edu/psylabs/catcentral/pdf%20files/cat07reckase.pdf

Ricketts, C., & Wilks, S. J. (2002). Improving student performance through computer-based assessment: Insights from recent research. *Assessment and Evaluation in Higher Education, 27,* 475–479.

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice, 24*(*2*), 3–13.

Royal Commission on Learning. (1994). *For the love of learning report: Report of the royal commission on learning*. Toronto, ON: Queen's Printer for Ontario.

Russell, M., Goldberg, A., & O'Conner, K. (2003). Computer-based testing and validity: A look back into the future. *Assessment in Education, 10*(3), 279–293.

Sadler, D.R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education, 5*(1), 77–84.

Sambell, K., Sambell, A., & Sexton, G. (1999). Student perceptions of the learning benefits of computer-assisted assessment: A case study in electronic engineering.

In S. Brown, P. Race, & J. Bull (Eds.), *SEDA Staff Educational and Development Series: Computer Assisted Assessment*. pp. 179–191. London, UK: Kogan Page.

Shepard, L., & Kirst, M. W. (1991). Interview on assessment issues with Lorrie Shepard. *Educational Researcher, 20*, 21–27.

Sly, L. (1999). Practice tests as formative assessment improve student performance on computer-managed learning assessments. *Assessment and Evaluation in Higher Education, 24*, 339–343.

Steven, C., & Hesketh, I. (1999). Increasing learner responsibility and support with the aid of adaptive formative assessment using QM designer software. In S. Brown, P. Race, & J. Bull (Eds.), *SEDA Staff Educational and Development Series: Computer Assisted Assessment* (pp. 179–191). London, UK: Kogan Page.

Stout. W. (2002). Psychometrics: From practice to theory and back. 15 years of nonparametric multidimensional IRT, DIF/Test equity, and skills diagnostic assessment. *Psychometrika, 67*, 485–518.

Tagg, J. (2003). *The Learning Paradigm College*. Boston, MA: Anker.

Tashakkori, A. (2004). AERA Symposium 2004. Retrieved February 14, 2008, from http://convention.allacademic.com/aera2004/AERA_papers/AERA_sess_1040_10492a.PDF

Tashakkori, A., & Teddlie, C. (2003). *Handbook on mixed methods in the behavioural social sciences*. Thousand Oaks, CA: Sage Publications.

Thelwall, M. (2000). Computer-based assessment: A versatile educational tool. *Computers and Education, 34*, 37–49.

University of Texas (2007). *Instructional assessment resources*. Retrieved October 20, 2008, from http://www.utexas.edu/academic/diia/assessment/iar/students/plan/method/exams-mchoice.php

Wainer, H. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Mahway, NJ: Lawrence Erlbaum Associates.

Weaver, M. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment and Evaluation in Higher Education, 31*, 379–394.

Wiliam, D. (2006). Assessment for learning: Why, what, and how? *Orbit, 36*(2), 2–7.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy and Practice, 11*(1), 49–65.

Winne, P. (2005). A perspective on state of the art research on self-regulated learning. *Instructional Science, 33*, 559–565.

Wise, S., Bhola, D., & Yang, S. (2006). Taking the time to improve the validity of low-stakes tests: The effort-monitoring CBT. *Educational Measurement: Issues and Practice, 14*, 21–30.

Yin, R. K. (2002). *Case study research: Design and methods*. (3rd ed.). Thousand Oaks, CA: Sage Publications

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogical practice. *Higher Education, 45*, 477–501.

Zakrzewski, S., & Bull, J. (1999). The mass implementation and evaluation of computer-based assessments. *Assessment and Evaluation in Higher Education,*

23, 141–152.

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: Academic Press.

Zimmerman, B. J. (2001). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*, 64–72.

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

APPENDIX A

Sample Items from the School of Medicine MSK Module

1. A 65-year old man slips and falls on ice resulting in a closed fracture of his tibia. The fracture line is short and oblique involving the middle third of the diaphyseal portion of the bone. He has a history of heart disease (triple-vessel bypass two years ago) and diabetes. He is a non-smoker with no prior history of fractures. Identify factors from the described scenario which may cause this patient to have a slow rate of fracture healing.

a) ☐ Healed fracture

b) ☐ Delayed union

c) ☐ Hypertrophic non-union

d) ☐ Atrophic non-union

2. What is the significance of a 15^{o} scoliosis in a pre-menarchal 10-year old girl?

3. What investigation or intervention will be most useful in establishing a diagnosis?

4. You evaluate a 16 year old girl because of a newly noted spinal deformity. She has a 25 degree thoracolumbar curve that appears to be idiopathic by history, physical and review of radiographs. Her bone age is equal to her chronological age. She reports that she had her first period at age 11.

If untreated, what is the most likely outcome from this patient's scoliosis?

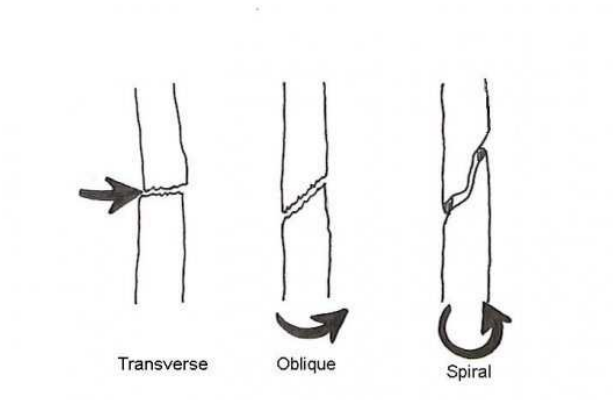
a) ☐ No problems

b) ☐ Mild back pain

c) ☐ Severe back pain

d) ☐ Cardiopulmonary complications in mid-life

5. The fracture pattern is dictated by the type and direction of force that the bone has been subjected to. High energy injuries (motor vehicle accidents) often cause several fracture lines resulting in **comminuted** fractures. Under these circumstances, there is often considerable soft tissue disruption and extensive devascularization of bone which may lead to delayed healing. The direction of the fracture line may be described as transverse, oblique or spiral.



Which fracture type is most likely to be stable in a cast?

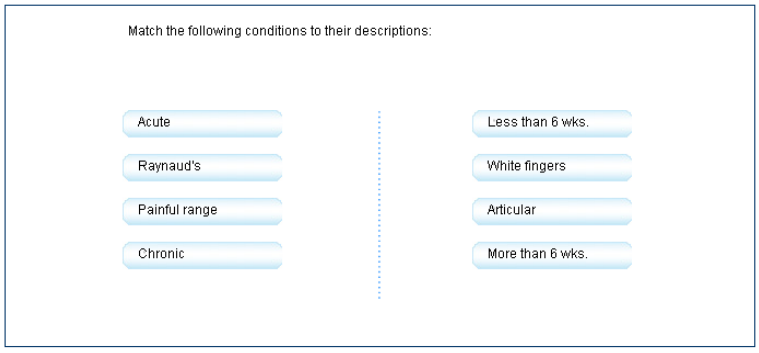
a) Transverse

b) Oblique

c) Spiral

d) Comminuted

6.


Match the following conditions to their descriptions:
Acute
Raynaud's
Painful range
Chronic
Less than 6 wks.
White fingers
Articular
More than 6 wks.

APPENDIX B

Prior Learning Assessments

1. What does learning look like?
Question 1 (1 point)

Students in an Art class were shown a painting (The Kitchen) by a famous artist previously studied in class and asked two questions related to the painting. The first question the teacher asked is: Identify the artist of this painting and describe the period of this work. What level of learning is the teacher assessing with this question?

Pablo Picasso The Museum of Modern Art, New York

- a. Ideas
- b. Connections
- c. Extensions

Save answer
Bottom of Form

Question 2 (1 point)

Continued from the previous scenario. The teacher also asked students to: Analyze the painting and describe the message in the painting. Here is one student's response: The collage of circles resembling eyes and geometric lines resembling the kitchen space and items found in the kitchen makes me think of the chaos in my grandmother's kitchen. Our entire family would gather – some would be cooking, others would be washing dishes, and the kids were usually eating or running around. What level of learning is the student demonstrating in this response?

- a. Ideas
- b. Connections
- c. Extensions

Save answer

Question 3 (1 point)

When an assessment question is created for a particular cognitive level, a student's response to the question:

- a. indicates the cognitive level of thinking for that student
- b. does not necessarily demonstrate the student's cognitive level of thinking
- c. must parallel one of the four levels of cognition as shown in the curriculum planner

Save answer

2. Assessment for Learning

Question 1 (1 point)

A student in a history class wrote the following essay: U.S. History When the French and Indian war ended, British expected Americans to help them pay back there wear debts. That would be a reasonable request if the war was fought for the colonies, but it was fought for English imperialism so you can't blame them for not wanting to pay. The taxes were just the start of the slow turn toward rebellion another factor was when parliament decided to forbid the colonial government to make any more money. Specie became scarcer than ever, and a lot of merchants were pushed into a "two way squeeze" and faced bankruptcy. If I had the choice between being loyal or rebelling and having something to eat, I know what my choice would be. The colonist who were really loyal never did rebel, and 1/3 support the revolution. The main thing that turned most people was the amount of propaganda, speeches from people like Patrick Henry, and organizations like the "Association". After the Boston Massacer and the issuing of the Intolerable acts, people were convinced there was a conspiracy in the royal government to extinguish America's liberties. I think a lot of people also just were going with the flow, or were being pressured by the Sons of Liberty. Merchants who didn't go along with boycotts often became the victims of mob violence. Overall though, people were sick of getting overtaxed, and walked on and decided let's do something about it. The history teacher noted the following comments: i) your topic sentence is weak ii) more

factual detail would improve your essay iii) note spelling and grammar errors Ministry policy asks that when teacher's give feedback to students that they focus on students' strengths, weaknesses, and next steps. Which of the statements above focused on the student's strength?

- a. i)
- b. ii)
- c. iii)
- d. none of the above
- e. all of the above

Save answer

Question 2 (1 point)

Another teacher assessed the same student's work. This teacher wrote: The greatest strength of this essay is its outstanding effort to grapple thoughtfully with the questions, why did the colonist rebel? Keep thinking personally, "What if I were there"? It is a great place to start. To make the essay work, however, you need to refine your organization strategies significantly. Remember that your reader is basically ignorant, so you need to express your view as clearly as you can. Remember when we talked about this is class? In the beginning, tell what side you're on: What made the colonists rebel – money, propaganda, conformity? In the middle, justify your view. What factors support your idea and will convince your reader? In the end, remind your reader again about your point of view. Do another edit and let me read it again. The primary focus of this teacher's response was ... (multiple answers)

- a. Strengths
- b. Weaknesses
- c. Next Steps

Save answer

Question 3 (1 point)

The instrument (history essay) presented in the above question is used as ...

- a. assessment FOR learning
- b. assessment OF learning
- c. combination of assessment FOR and assessment OF learning

Save answer

---

3. Assessment Of Learning

Question 1 (1 point)

Mrs. Waddlebug has allocated 40 minutes for her science test. The test is out of 50 and marks for each question are indicated in brackets at the end of each question. The test contains enough space so that students can write their response immediately below each question. A variety of question formats such as multiple-choice, fill-in the blank, short answer, and one longer answer are included on the test. Students in the front of each row are given the tests for each person in their row. Mrs. Waddlebug tells students to begin the test once they receive it. Ten minutes into the test, a student brings a typing error to Mrs. Waddlebug's attention. She informs students of the correction and writes the correction on the whiteboard. When the 40 minutes is up, students at the back of each row are instructed to pass their tests forward where Mrs. Waddlebug collects the tests from the first person in each row. That evening Mrs. Waddlebug creates an answer copy in which the answers and marking scheme are shown. She has decided that question 4 should be out of 5 instead of 3 (there was more to the answer than she originally thought). After marking all 28 tests, Mrs. Waddlebug realized that the majority of students did not score very well on question 12, including the students in her class that she anticipated would receive full marks for the question. She decided to drop question 12 (worth 3 marks) reducing the test total to 49marks. As she prepares to enter the grades in her grading file she realizes that a grade out of 49 is more difficult to interpret (convert to a grade out of 100) so she decides that anyone who attempted question 12 would get 1 mark. This is fair since everyone attempted the question. How many places has the potential for measurement error been introduced?

- a. 4
- b. 5
- c. 6
- d. 7

Save answer

Question 2 (1 point)

At the end of the year when the school principal reviewed Mrs. Waddlebug's final grades for her students he noted two grades just below 50%? Thomas had an overall score of 49 and Cecilia had an overall score of 47. The principal counselled Mrs. Waddlebug to 'bump' these scores to a 50. The possible reasons for this practice include:

- a. There is significant evidence that students who fail will achieve significantly better by repeating the material.
- b. It is likely that these students have achieved as well as students whose grade is 50%-53%.
- c. The principal did not want to put the teacher in a situation where she has to defend how she got these grades to parents.

Save answer

Question 3 (1 point)

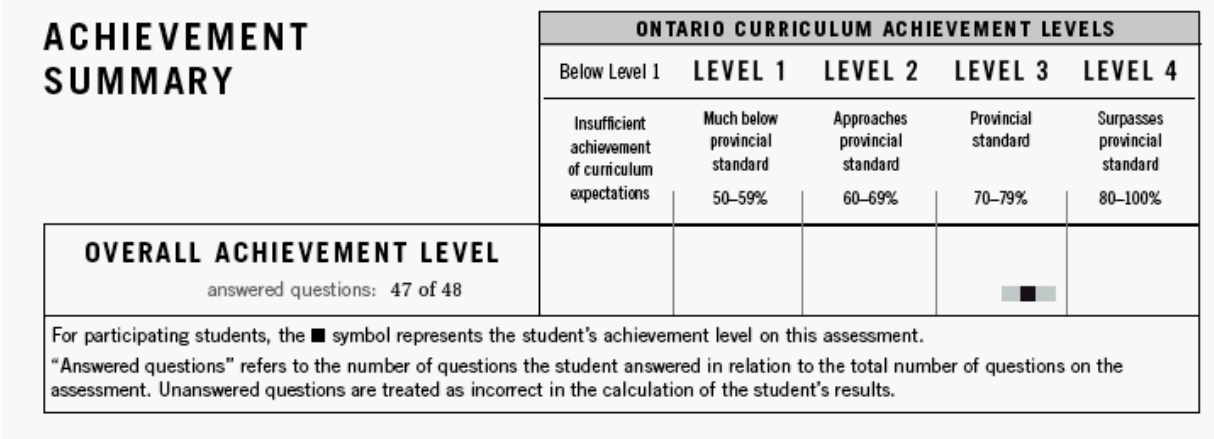
Based on the description of Mrs. Waddlebug's assessment practices, what is Mrs. Waddlebug's premise for this assessment?

- a. Measure student's understanding.
- b. Obtain a grade for this section of the curriculum.
- c. Create a fair assessment instrument.

Save answer

4. Large-scale Assessment

Below is a screen clip of a grade 9 assessment of mathematics report. Identify the range in which this student score falls.



ACHIEVEMENT SUMMARY

	ONTARIO CURRICULUM ACHIEVEMENT LEVELS				
	Below Level 1	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4
	Insufficient achievement of curriculum expectations	Much below provincial standard	Approaches provincial standard	Provincial standard	Surpasses provincial standard
		50–59%	60–69%	70–79%	80–100%
OVERALL ACHIEVEMENT LEVEL answered questions: 47 of 48				■	

For participating students, the ■ symbol represents the student's achievement level on this assessment.

"Answered questions" refers to the number of questions the student answered in relation to the total number of questions on the assessment. Unanswered questions are treated as incorrect in the calculation of the student's results.

- a. 70% to 79%
- b. 76% to 78%
- c. No measurement error (ME) was reported

Save answer

Question 2 (1 point)

What is the expected performance in mathematics for grade 9 students?

- a. Given that the 'cut-score' changes with every test, it is not possible to identify the expected performance.
- b. 70% to 79%
- c. A credit in mathematics is granted with a performance of 50% and above which counts as one of the four compulsory mathematics credits. Hence an expected performance of 50% and above is acceptable in grade 9 mathematics.

Save answer

Question 3 (1 point)

Prior to administering one of Ontario's large-scale assessments, some teachers will halt regular instruction and classroom activities a month prior to the assessment in order to focus on the specific content that will be on the assessment. This practice ...

- a. is common in most schools given the need to prepare students to write assessment.
- b. goes beyond what is required to prepare students to write the large-scale assessment given that the large-scale assessment is based on the curriculum.
- c. No class time should be allocated for preparing students to write the large-scale assessment given that the assessment is based on the curriculum.

Save answer

APPENDIX C

MSK Descriptive Findings

37. Identify your gender
    a. Male 40(54.7)
    b. Female 33(45.2)

38. Identify your age range.
    a. 20 – 25 48(65.0)
    b. 26 – 30 20(27.4)
    c. 30+ 6(8.2)

39. Indicate your experience in using the Internet to obtain information (prior to arriving at the School of Medicine
    a. I frequently used meta research tools (data-base search engines on the Internet) 46(63.0)
    b. I have used information located on the Internet to complete course assignments 25(34.2)
    c. I rarely used the Internet to look for information 1(1.4)
    d. I had never used the Internet 1(1.4)

40. Prior to the MSK course, have you ever completed a computer-based assessment? If yes, indicate the name of the software (if known) and the course or program where the computer-based assessment was used (on the narrative answer sheet).
    a. 1 – Yes 39(53.4)
    b. 2 – No 34(46.5)

41. Where did you complete the quizzes?
    a. – Home (off-campus) 57(78.1)
    b. – University 5(6.8)
    c. – Home and University 11(14.7)
    d. - Other _____ 1(1.4)

Please use the 5-point scale for the remaining questions:

A: **Strongly disagree**
B: **Disagree**
C: **Neither agree nor disagree**
D: **Disagree**
E: **Strongly disagree**

	SD	D	N	A	SA	*M*	*SD*
42. I completed the quizzes at a convenient time and place.	11(14.7)	0	3(4.0)	23(30.7)	31(41.3)	3.93	1.41
43. The concept of using computer-based quizzes for self-assessment is a good way to direct student learning.	6(8.0)	4(5.3)	4(5.3)	28(37.3)	32(42.7)	4.03	1.21
44. I completed all quizzes presented in all modules.	4(5.3)	3(4.0)	8(10.7)	19(25.3)	39(52.0)	4.18	1.14
45. I collaborated with colleagues when I completed the quizzes.	38(50.7)	21(28.0)	6(8.0)	4(5.3)	4(5.3)	1.84	1.14
46. The range of questions in each quiz reflected the concepts presented in the corresponding module.	2(2.7)	4(5.3)	17(22.7)	37(49.3)	14(18.7)	3.77	0.92
47. The purpose of the quizzes was to direct my learning.	3(4.0)	3(4.0)	9(12.0)	36(48.0)	23(30.7)	3.99	0.99
48. The purpose of the quizzes was part of my mark in the class. R	4(5.3)	7(9.3)	19(25.3)	12(16.0)	30(40.0)	3.79	1.24
49. The purpose of the quizzes was a diagnostic assessment to help plan instruction.	12(16.0)	10(13.3)	23(30.7)	18(24.0)	10(13.3)	3.05	1.27
50. The purpose of the quizzes was to identify my strengths and weaknesses.	4(5.3)	3(4.0)	8(10.7)	36(48.0)	23(30.7)	3.96	1.04
51. I was not sure of the purpose of the quizzes. R	3(4.0)	10(13.3)	24(32.0)	16(21.3)	18(24.0)	3.51	1.15
52. The number of questions in each quiz was appropriate in assessing my understanding of the concepts presented in the corresponding modules.	3(4.0)	10(13.3)	15(20.0)	34(45.3)	11(14.7)	3.55	1.04
53. The difficulty level of questions was appropriate for these quizzes.	2(2.7)	3(4.0)	10(13.3)	47(62.7)	10(13.3)	3.83	0.82
54. The topics selected for assessment were appropriate.	2(2.7)	1(1.3)	13(17.3)	42(56.0)	14(18.7)	3.90	0.83
55. The use of scenarios was	2(2.7)	0	(9.3)	37(49.3)	27(36.0)	4.19	0.83

appropriate.							
56. Each scenario used in the quizzes represented a meaningful dilemma.	2(2.7)	2(2.7)	10(13.3)	47(62.7)	12(16.0)	3.89	0.81
57. The quizzes advanced my understanding of the topics presented in each module.	3(4.0)	3(4.0)	8(10.7)	42(56.0)	16(21.3)	3.90	0.94
58. The quizzes guided my learning in each module.	3(4.0)	2(2.7)	8(10.7)	40(53.3)	19(25.30	3.97	0.93
59. The quizzes were helpful in starting discussions with colleagues about the topics presented in each module.	15(20.0)	13(17.3)	16(21.3)	21(28.0)	8(10.7)	2.92	1.32
60. Multiple choice questions were useful in assessing my understanding.	2(2.7)	1(1.3)	10(13.3)	45(60.0)	14(18.7)	3.94	0.80
61. Matching column questions were useful in assessing my understanding.	4(5.3)	6(8.0)	15(20.0)	35(46.7)	12(16.0)	3.62	1.04
62. Feedback that stated my answer was **incorrect** was effective in supporting my leaning.	2(2.7)	2(2.7)	6(8.0)	35(46.7)	27(36.0)	4.15	0.90
63. Feedback that stated my answer was **correct** was effective in supporting my learning.	2(2.7)	3(4.0)	6(8.0)	37(49.3)	23(30.7)	4.07	0.92
64. Feedback that stated my answer was **incorrect** AND **provided further information** clarifying the incorrect response was effective in supporting my learning.	2(2.7)	0	3(4.0)	19(25.3)	48(64.0)	4.54	0.82
65. Feedback that stated my answer was **correct** AND **supplemented** it with why the answer was correct was effective in supporting my learning.	2(2.7)	0	4(5.3)	21(28.0)	43(57.3)	4.47	0.85
66. I thoroughly reviewed the feedback I was given.	2(2.7)	4(5.3)	13(17.3)	30(40.0)	23(30.7)	3.94	0.99
67. **Overall,** the feedback provided on the quizzes was very useful.	3(4.0)	1(1.3)	10(13.3)	35(46.7)	22(29.3)	4.01	0.95
68. The quizzes were just 'busy work'. R	2(2.7)	6(8.0)	13(17.3)	32(42.7)	18(24.0)	3.82	1.00
69. The quizzes were not a priority for me. R	1(1.3)	9(12.0)	18(24.0)	30(40.0)	14(18.7)	3.65	0.98
70. I was too busy to complete the quizzes. R	2(2.7)	4(5.3)	9(12.0)	37(49.3)	19(25.3)	3.94	0.94
71. I did not put effort into completing the quizzes. R	3(4.0)	5(6.7)	14(18.7)	26(34.7)	24(32.0)	3.88	1.09

72. The software interface (lay-out of screen) did not interfere with my test-taking ability.	6(8.0)	10(13.3)	8(10.7)	34(45.3)	14(18.7)	3.56	1.20
73. I needed assistance in using the assessment software to complete the quizzes.	40(53.3)	16(21.3)	8(10.7)	5(6.7)	3(4.0)	1.82	1.14
74. In retrospect, my results on the quizzes were an accurate reflection of my understanding of the concepts presented in each module.	3(4.0)	9(12.0)	21(28.0)	32(42.7)	6(8.0)	3.41	0.97
75. The quizzes presented in each module were fair assessment instruments.	3(4.0)	7(9.3)	13(17.3)	40(53.3)	5(6.7)	3.54	0.94

76. What could be done to enhance the usefulness of the quizzes used in the MSK course (use Feedback Research Study Narrative answer sheet)?

APPENDIX D

BEd Descriptive Findings

Background Information

1. Gender.
   63 (20.5)Male
   242 (78.6)Female

2. Identify your age range.
   236 (76.6) 20 – 25
   40 (13.0) 26 – 30
   31 (10.1) 31+

3. Indicate the program you are registered in.
   194 (63.0) Consecutive
   42 (13.6) University A [pseudonym] Concurrent
   66 (21.4) University B [pseudonym] Concurrent
   4 (1.3) University C [pseudonym] Concurrent

4. Identify the division you are registered in.
   152 (49.4) Primary/Junior
   156 (50.6) Intermediate/Senior

5. Indicate if you are in one of the following specialized programs.
   Aboriginal Teacher Education
   9 (2.9) Artists in Community Education
   7 (2.3) Outdoor and Experiential Education
   12 (3.9) Technological Education
   273 (88.6) Not specialized

6. If you are an intermediate/senior teacher candidate, indicate your teachable subjects. Open.

7. Are you currently taking EDST409: Classroom Assessment?
   25 (8.1) Yes
   271 (88.0) No

8. Indicate your experience in using the Internet to obtain information (prior to arriving at the Faculty of Education).
   180 (58.4) I frequently used meta research tools (data-base search engines on the Internet)
   113 (36.7) I have used information located on the Internet to complete course assignments
   10 (3.2) I rarely used the Internet to look for information
   I had never used the Internet

9. Prior to the Prof 150/155 course, have you ever completed a computer-based assessment?
   129 (41.9) Yes
   177 (57.5) No

10. Other than in the Prof 150/155 Classroom Assessment module, have you completed a computer-based assessment designed to guide your learning?
    110 (35.7) Yes
    197 (64.0) No

Prior Learning Assessment

28. Where did you complete the Prior Learning Assessment?
    215 (69.8) Home (off-campus)
    65 (21.1) University

29. Were you able to complete the Prior Learning Assessment at a convenient time and place?
    276 (89.6) Yes
    8 (2.6) No

30. What do you believe were the purposes of the Prior Learning Assessment/
    205 (66.6) To guide my decisions about learning in the module
    64 (20.8) To provide the first mark in this course
    131 (42.5) To give instructors information about what I know
    215 (69.8) To help identify my strengths and weaknesses in selected assessment topics

31. The concept of using a Prior Learning Assessment is a good way to direct student learning. Indicate the extent to which you agree or disagree with this statement.

Strongly Disagree	2	3	4	5	6	Strongly Agree	*M*	*SD*
20(6.5)	40(13.0)	68(22.1)	77(25.0)	53(17.2)	16(5.2)	11(3.6)	4.32	1.45

32. In terms of the Prior Learning Assessment, to what extent …

	Not Appro.	2	3	4	5	6	Very Appro.	*M*	*SD*
Was the number of questions in each of the 4 units appropriate in assessing your understanding of assessment?	17(5.5)	44(14.3)	77(25.0)	73(23.7)	57(18.5)	11(3.6)	7(2.3)	3.59	1.36
Was the difficulty level of questions appropriate for this assessment?	11(3.6)	33(10.7)	55(17.9)	86(27.9)	69(22.4)	24(7.8)	7(2.3)	3.94	1.34
Were the topics selected for assessment appropriate?	9(2.9)	23(7.5)	59(19.2)	70(22.7)	80(26.0)	29(9.4)	13(4.2)	4.16	1.38

33. To what extent do you agree or disagree with the following statements?

	Strongly Disagree	2	3	4	5	6	Strongly Agree	*M*	*SD*
Reading and responding to a scenario is a good way to assess your thinking about assessment	5(1.6)	19(6.2)	38(12.3)	61(19.8)	82(26.6)	48(15.6)	33(10.7)	4.65	1.46
Each scenario in the Prior Learning Assessment represented a meaningful assessment dilemma	12(3.9)	36(11.7)	52(16.9)	88(28.6)	56(18.2)	30(9.7)	12(3.9)	3.97	1.44

34. How helpful was the Prior Learning Assessment in …

	No Impact	2	3	4	5	6	High Impact	M	SD
advancing your understanding of assessment?	43(14.0)	51(16.6)	70(22.7)	63(20.5)	46(14.9)	9(2.9)	4(1.3)	3.21	1.45
guiding your learning during the module?	30(9.7)	54(17.5)	66(21.4)	55(17.9)	54(17.5)	19(6.2)	6(1.9)	3.46	1.52
starting discussions with colleagues about assessment?	42(13.6)	58(18.8)	55(17.9)	58(18.8)	42(13.6)	24(7.8)	5(1.6)	3.32	1.59

35. To what extent were the multiple choice questions on the Prior Learning Assessment useful in assessing your understanding of assessment?

Not Useful	2	3	4	5	6	Very Useful	M	SD
43(14.0)	50(16.2)	71(23.1)	64(20.8)	41(13.3)	8(2.6)	6(1.9)	3.20	1.47

36. To what extent did the Prior Learning Assessment influence your self-directed learning?

No Influence	2	3	4	5	6	Sign. Influence	M	SD
40(13.0)	46(14.9)	61(19.8)	64(20.8)	43(14.0)	19(6.2)	11(3.6)	3.44	1.61

37. What was your experience in using WebCT to answer questions on the Prior Learning Assessment?

Very Difficult	2	3	4	5	6	Very Easy	M	SD
5(1.6)	18(5.8)	29(9.4)	40(13.0)	43(14.0)	39(12.7)	112(36.4)	5.32	1.73

38. If you required assistance in using WebCT technology to complete the Prior Learning Assessment, where did you receive this assistance (Check all that apply)

13 (4.2)Computer Labs
IT Support Services
23 (7.5)Colleagues
10 (3.2)Instructors
232 (75.3)Did not need assistance

39. To what extent did you review the feedback provided on the Prior Learning Assessment

Not at all	2	3	4	5	6	Thor-oughly	M	SD
26(8.4)	40(13.0)	49(15.9)	62(20.1)	43(14.0)	37(12.0)	Nil	3.65	1.53

40. Rank the effectiveness of each specific type of feedback given on the Prior Learning Assessment.

	N/A	Not Very Effective	2	3	4	5	6	Very Effecti ve	M	SD
Indicated your response was incorrect	34(11.9)	31(10.1)	41(13.3)	41(13.3)	40(13.0)	44(14.3)	31(10.1)	24(7.8)	3.27	2.35
Indicated your response was correct	18(5.8)	25(8.1)	34(11.0)	37(12.0)	46(14.9)	58(18.8)	37(12.0)	28(9.1)	3.81	2.14
Indicated your answer was incorrect AND provided further information clarifying the incorrect response	33(10.7)	10(3.2)	19(6.2)	25(8.1)	30(9.7)	41(13.3)	52(16.9)	76(24.7)	4.40	2.58
Indicated the answer was correct AND supplemente d it with why the answer was correct	28(9.1)	13(4.2)	17(5.5)	23(7.5)	36(11.7)	42(13.6)	55(17.9)	72(23.4)	4.46	2.48

41. Rank the OVERALL usefulness of feedback provided on the Prior Learning Assessment in guiding your learning.

Not Useful	2	3	4	5	6	Very Useful	M	SD
39(12.7)	41(13.3)	48(15.6)	63(20.5)	65(21.1)	14(4.5)	13(4.2)	3.59	1.64

42. Identify the extent to which the following factors may have influenced your achievement on the Prior Learning Assessment: (All Items Reversed)

	Large Factors	2	3	4	5	6	Not a Factor	M	SD
Not a priority for me given other faculty expectations	24(7.8)	21(6.8)	32(10.4)	66(21.4)	39(12.7)	36(11.7)	68(22.1)	4.59	1.90
I thought it was just 'busy work'	33(10.7)	24(7.8)	45(14.6)	52(16.9)	48(15.6)	30(9.7)	54(17.5)	4.27	1.94
I was too busy	16(5.2)	20(6.5)	43(14.0)	48(15.6)	44(14.3)	46(14.9)	66(21.4)	4.72	1.83
It occurred too early in the term	37(12.0)	26(8.4)	24(7.8)	33(10.7)	36(11.7)	49(15.9)	80(26.0)	4.66	2.13
The scenarios were unrealistic	13(4.2)	8(2.6)	19(6.2)	35(11.4)	51(16.6)	74(24.0)	82(26.6)	5.32	1.65
I was not sure of the purpose of the assessment	25(8.1)	17(5.5)	26(8.4)	46(14.9)	38(12.3)	60(19.5)	70(22.7)	4.83	1.92

43. To what extent did the following logistical factors influence your completion of the Prior Learning Assessment?

	No Impact	2	3	4	5	6	High Impact	M	SD
Did not know my Queen's net ID (e.g., 1aeg@Queensu.ca)	246(79.9)	13(4.2)	6(1.9)	7(2.3)	9(2.9)			1.29	0.89
Could not find the computer lab	253(82.1)	6(1.9)	6(1.9)	9(2.9)	2(0.6)	4(1.3)		1.26	0.90

44. In retrospect, were your results on the Prior Learning Assessment an accurate reflection of your understanding of assessment at the time?
    162 (52.6) Yes
    124 (40.3) No

45. Was the Prior Learning Assessment a fair assessment instrument?
    155 (50.3) Yes
    129 (41.9) No

46. Identify the degree of effort (seriousness) that you put forth in completing the Prior Learning Assessment.

No Effort	2	3	4	5	6	Great Effort	M	SD
11(3.6)	34(11.0)	43(14.0)	64(20.8)	74(24.0)	42(13.6)	17(5.5)	4.23	1.5

APPENDIX E

School of Medicine Analysis of formative CBA software

Name of Software	ThinkingCap Campus			
Stated Purpose of Formative CBA	- Self-assessment of concepts presented in MSK Module - self-directed learning			
Intended Users	First year MD students			
	MC	T/F	Matching	Blank
Item Types Available	✓	✓	✓	✓
	MC	T/F	Matching	Blank
Item Types Used	✓	✓	✓	✓
	After Each Item	At end of Ass.		
Method for Channelling Feedback	✓			

Note: A word-scramble item was included as was the open-response format

APPENDIX F

School of Medicine Analysis of formative CBA items used

Legend:

Item Formats

m – matching
o – open
c – multiple-choice
TF – true false
SA – fill in blank
W – word scramble

Visual

s – scenario
t – table
g – graph
p - picture

Feedback

I – for incorrect responses
C – for correct responses
.

	Visual	Ideas	Connections	Extensions	Key words/phrase	Class Notes	Class Text	Additional Info	Correct/ Incorrect	Item Type	# Distracters (MC only)
Item #	High Complexity				Evidence of complexity	Feedback					
Epidemilogy											
1			✓		How did …? Why is …?			b		o	
2		✓			What …?			b		c	3
3		✓			What is …?			b		o	
4		✓			What does …?			b		o	
5		✓			What does …?			b		o	
Imagining Quiz 1											
6		✓			Fill in the blank …		i	c		s	
7		✓			Which of the following		i	c		c	4
8		✓			Which has …		i	c		t	
9		✓			Fill in the blank …		i	c		s	
10		✓			Which of the following		i	c		c	4
Imaging Quiz 2											
11	p		✓		Identify …		i	c		*	
12	p		✓		Identify …		i	c		*	
13	p		✓		Identify…		i	c		*	
14	p	✓			What is					w	
15					Which of		i	c		c	3

16		✓			Match		i	c		m	
17			✓		Evaluate the x-ray. Which …		i	c		s	
Bone Morphology Quiz 1							i	c		c	2
			✓		This type is found …		i	c		c	2
			✓		Which of the following is correct		i	c		c	5
Bone Morphology Quiz 2											
		✓			Check that you recall …			c		w	
	p	s			Identify factors		i	c		c	5
Fractures in Children Quiz 1											
	p			s	What is the diagnosis?		i	c		c	5
	p		s		Describe 2 complications …		i			o	
Fractures in Children Quiz 2											
	p			s	What to recommend to avoid deformity		i			o	
		✓			Match		i			m	
Approached to Arthritis											
		✓			Which of the following …		i	c		c	4
		✓			Match …					m	
	s		✓		Which describes …					c	3
Most Common Types of Arthritis											4
		✓			Which of the following …		i	c		c	4
		✓			Match …			c			
		✓			Which of the following …		i	c		c	3
Limping Quiz 1											
		✓			What features …		i	c		c	4
	s		✓		Which statement is correct …		i	c		c	2
	s			✓	Are parents correct in concern …			c		o	
	s			✓	What is the diagnosis …			c		o	
Limping Quiz 2											
			✓		Describe …			c		o	
			✓		What features …			c		o	
	s			✓	Characterize the problem …			c		o	
Bone Pain											
	s	✓			What is …			i	c	c	4
	s			✓	Diagnosis?			i	c	c	4
	s			✓	What would you expect …				c	o	
	s		✓		What is characterized …			i	c	c	3
	s			✓	Intervention most useful				c	o	
Deformity Quiz 1											
	s			✓	Diagnosis …			i	c	c	5
	p		✓		Which is …			i	c	c	3
Deformity Quiz 2											
		✓			Which of the following …			i	c	c	4
		✓			What is the significance …				c	o	
	s		✓		What is the outcome …			i	c	c	4

* Item required students to point at an object.
Table of Specifications Yes or ✓No

APPENDIX G

Faculty of Education Analysis of formative CBA software

Name of Software	WebCT			
Stated Purpose of Formative CBA	Online virtual learning tool marketed at the college and university level for discussion boards, mail and live chat rooms, and assessment			
Intended Users	College and university students			
	MC	T/F	Matching	Blank
Item Types Available	✓	✓	✓	✓
	MC	T/F	Matching	Blank
Item Types Used	✓			
	After Each Item	At end of Ass.		
Method for Channelling Feedback		✓		

APPENDIX H

Faculty of Education Analysis of formative CBA items Used

	Scenario (S)	Table (T)	Graph (G)	Picture (P)	Item	Class Notes	Class Text	Additional Info	Correct/ Incorrect	Item Type	# Distracters (MC only)
Item #	High Complexity				Low complexity	Feedback					
What does learning look like?											
1. Students in an Art class were shown a painting (The Kitchen) by a famous artist previously studied in class and asked two questions related to the painting. The first question the teacher asked is: Identify the artist of this painting and describe the period of this work. What level of learning is the teacher assessing with this question? <picture of painting here> Scenario Graphic Multiple-choice 3 Distracters											
	S			P				✓	✓	M	3
2. Continued from the previous scenario. The teacher also asked students to: Analyze the painting and describe the message in the painting. Here is one student's response: The collage of circles resembling eyes and geometric lines resembling the kitchen space and items found in the kitchen makes me think of the chaos in my grandmother's kitchen. Our entire family would gather – some would be cooking, others would be washing dishes, and the kids were usually eating or running around. What level of learning is the student demonstrating in this response?											
	S							✓	✓	M	3
3. When an assessment question is created for a particular cognitive level, a student's response to the question:											
	S							✓	✓	M	3
Assessment for Learning											
4. A student in a history class wrote the following: U.S. History When the French and Indian war ended, British expected Americans to help them pay back there wear debts. That would be a reasonable request if the war was fought for the colonies, but it was fought for English imperialism so you can't blame them for not wanting to pay. The taxes were just the start of the slow turn toward rebellion another factor was when parliament decided to forbid the colonial government to make any more money. Specie became scarcer than ever, and a lot of merchants were pushed into a "two way squeeze" and faced bankruptcy. If I had the choice between being loyal, or rebelling and having something to eat, I know what my choice would be. The colonist who were really loyal never did rebel, and 1/3 support the revolution. The main thing that turned most people was the amount of propaganda, speeches from people like Patrick Henry, and organizations like the "Association". After the Boston Massacre and the issuing of the Intolerable acts, people were convinced there was a conspiracy in the royal government to extinguish America's liberties. I think a lot of people also just were going with the flow, or were being pressured by the Sons of Liberty. Merchants who didn't go along with boycotts often became the victims of mob violence. Overall though, people were sick of getting overtaxed, and walked on and decided let's do something about it. The history teacher noted											

the following comments:

i) your topic sentence is weak
ii) more factual detail would improve your essay
iii) note spelling and grammar errors

Ministry policy asks that when teacher's give feedback to students that they focus on students' strengths, weaknesses, and next steps. Which of the statements above focused on the student's strength?

	S							✓	✓	M	5

5. Another teacher assessed the same student's work. This teacher wrote: The greatest strength of this essay is its outstanding effort to grapple thoughtfully with the questions, why did the colonist rebel? Keep thinking personally, "What if I were there"? It is a great place to start. To make the essay work, however, you need to refine your organization strategies significantly. Remember that your reader is basically ignorant, so you need to express your view as clearly as you can. Remember when we talked about this is class? In the beginning, tell what side you're on: What made the colonists rebel – money, propaganda, conformity? In the middle, justify your view. What factors support your idea and will convince your reader? In the end, remind your reader again about your point of view. Do another edit and let me read it again. The primary focus of this teacher's response was ... (multiple answers)

	S								✓	M	3

6. The assessment question presented above was used as ...

	S							✓	✓	M	3

Assessment of Learning

7. Mrs. Wattle has allocated 40 minutes for her science test. The test is out of 50 and marks for each question are indicated in brackets at the end of each question. The test contains enough space so that students can write their response immediately below each question. A variety of question formats such as multiple-choice, fill-in the blank, short answer, and one longer answer are included on the test. Students in the front of each row are given the tests for each person in their row. Mrs. Wattle tells students to begin the test once they receive it. Ten minutes into the test, a student brings a typing error to Mrs. Wattle attention. She informs students of the correction and writes the correction on the whiteboard. When the 40 minutes is up, students at the back of each row are instructed to pass their tests forward where Mrs. Wattle collects the tests from the first person in each row. That evening Mrs. Wattle creates an answer copy in which the answers and marking scheme are shown. She has decided that question 4 should be out of 5 instead of 3 (there was more to the answer than she originally thought). After marking all 28 tests, Mrs. Wattle realized that the majority of students did not score very well on question 12, including the students in her class that she anticipated would receive full marks for the question. She decided to drop question 12 (worth 3 marks) reducing the test total to 49 marks. As she prepares to enter the grades in her grading file she realizes that a grade out of 49 is more difficult to interpret (convert to a grade out of 100) so she decides that anyone who attempted question 12 would get 1 mark. This is fair since everyone attempted the question.

How many places has the potential for measurement error been introduced?

	S								✓	M	4

8. At the end of the year when the school principal reviewed Mrs. Wattle's final grades for her students he noted two grades just below 50%?. Thomas had an overall score of 49 and Cecilia had an overall score of 47. The principal counselled Mrs. Wattle to 'bump' these scores to a 50. The possible reasons for this practice include:

	S							✓	✓	M	3

9. Based on the description of Mrs. Wattle's assessment practices, what is Mrs. Wattle's premise for this assessment?

	S									M	3

Large-scale Assessment

10. Below is a screen clip of a grade 9 assessment of mathematics report. Identify the range in which this student score falls.

<picture of report card here>											
	S			P				✓	✓	M	3
11. What is the expected performance in mathematics for grade 9 students?											
	S							✓	✓	M	3
12. Prior to administering one of Ontario's large-scale assessments, some teachers will halt regular instruction and classroom activities a month prior to the assessment in order to focus on the specific content that will be on the assessment. This practice ...											
	S							✓	✓	M	3

Table of Specifications ✓Yes or No