Investigating Multiple-Choice Assessment: Secondary School Teachers’ Perceptions and Practices

Derek Oei
Department of Graduate and Undergraduate Studies in Education

Submitted in partial fulfillment of the requirements for the degree of Master of Education

Faculty of Education, Brock University
St. Catharines, Ontario

© Derek Oei 2013
Abstract

Multiple-choice assessment is used within nearly all levels of education and is often heavily relied upon within both secondary and postsecondary institutions in determining a student’s present and future success. Understanding why it is effective or ineffective, how it is developed, and when it is or is not used by teachers can further inform teachers’ assessment practices, and subsequently, improve opportunities for student success.

Twenty-eight teachers from 3 secondary schools in southern Ontario were interviewed about their perceptions and use of multiple-choice assessment and participated in a single-session introductory workshop on this topic. Perceptions and practices were revealed, discussed, and challenged through the use of a qualitative research method and examined alongside existing multiple-choice research. Discussion centered upon participants’ perspectives prior to and following their participation in the workshop. Implications related to future assessment practices and research in this field of assessment were presented. Findings indicated that many teachers utilized the multiple-choice form of assessment having had very little teacher education coursework or inservice professional development in the use of this format. The findings also revealed that teachers were receptive to training in this area but simply had not been exposed to or been given the opportunity to further develop their understanding. Participants generally agreed on its strengths (e.g., objectivity) and weaknesses (e.g., development difficulty). Participants were particularly interested in the potential for this assessment format to assess different levels of cognitive difficulty (i.e., levels beyond remembering of Bloom’s revised taxonomy), in addition to its potential to perhaps provide equitable means for assessing students of varying cultures, disabilities, and academic streams.
Acknowledgements

There are many people I would like to thank for their significant contributions throughout my studies: the district school board for allowing me to conduct my study within their schools; all the teachers and administrators for their time, participation, professionalism, and facilitation of my study; Dr. Lionel LaCroix and Dr. Louis Volante for being members of my internal committee and for supporting and critiquing my thesis over the last year; Dr. Don Klinger for being my external examiner and again lending me support and critique of my thesis; my friends (you know who you are!) and family—mom, dad, and sister—for their unceasing support, patience, and impatience, with me all these years as I have completed this Master’s; and last, Dr. Joe Engemann, for taking me on as one of his many thesis students, for encouraging me continuously over the last few years, for spending countless hours editing my thesis, guiding me, travelling to meet with me, and for accommodating my ever-changing life. Thank you all!
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td><strong>CHAPTER ONE: INTRODUCTION TO THE PROBLEM</strong></td>
<td>i</td>
</tr>
<tr>
<td>Background of the Problem</td>
<td>2</td>
</tr>
<tr>
<td>Statement of the Problem Context</td>
<td>8</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>10</td>
</tr>
<tr>
<td>Research Questions</td>
<td>11</td>
</tr>
<tr>
<td>Rationale</td>
<td>11</td>
</tr>
<tr>
<td>Study Framework</td>
<td>15</td>
</tr>
<tr>
<td>Scope and Limitations of Study</td>
<td>15</td>
</tr>
<tr>
<td>Outline of Remainder of the Document</td>
<td>16</td>
</tr>
<tr>
<td><strong>CHAPTER TWO: REVIEW OF RELATED LITERATURE</strong></td>
<td>19</td>
</tr>
<tr>
<td>Professional Learning Workshops</td>
<td>19</td>
</tr>
<tr>
<td>The Role and Emergence of Multiple-Choice Assessment</td>
<td>21</td>
</tr>
<tr>
<td>Multiple-Choice Assessment Benefits</td>
<td>24</td>
</tr>
<tr>
<td>Multiple-Choice Assessment Concerns</td>
<td>25</td>
</tr>
<tr>
<td>Item Writing</td>
<td>27</td>
</tr>
<tr>
<td>Modifications to Conventional Multiple-Choice Assessment</td>
<td>29</td>
</tr>
<tr>
<td>Optimal Number of Response Options</td>
<td>30</td>
</tr>
<tr>
<td>Response Option Modifications</td>
<td>31</td>
</tr>
<tr>
<td>Modified Marking Schemes</td>
<td>33</td>
</tr>
<tr>
<td>Feedback</td>
<td>37</td>
</tr>
<tr>
<td>Bloom’s Revised Taxonomy</td>
<td>38</td>
</tr>
<tr>
<td>Summary</td>
<td>44</td>
</tr>
<tr>
<td><strong>CHAPTER THREE: RESEARCH METHODOLOGY</strong></td>
<td>47</td>
</tr>
<tr>
<td>Research Design</td>
<td>47</td>
</tr>
<tr>
<td>Selection of Site and Participants</td>
<td>52</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>52</td>
</tr>
<tr>
<td>Data Collection</td>
<td>53</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>55</td>
</tr>
<tr>
<td>Limitations</td>
<td>56</td>
</tr>
<tr>
<td>Ethical Considerations</td>
<td>57</td>
</tr>
<tr>
<td>Summary</td>
<td>58</td>
</tr>
<tr>
<td><strong>CHAPTER FOUR: PRESENTATION OF RESULTS</strong></td>
<td>59</td>
</tr>
<tr>
<td>Findings</td>
<td>60</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>CHAPTER FIVE: DISCUSSION</td>
<td>91</td>
</tr>
<tr>
<td>Summary of the Study</td>
<td>92</td>
</tr>
<tr>
<td>Discussion</td>
<td>94</td>
</tr>
<tr>
<td>Implications</td>
<td>113</td>
</tr>
<tr>
<td>Final Thoughts</td>
<td>119</td>
</tr>
<tr>
<td>References</td>
<td>121</td>
</tr>
<tr>
<td>Appendix A: Interview Questions</td>
<td>140</td>
</tr>
<tr>
<td>Appendix B: Survey</td>
<td>141</td>
</tr>
<tr>
<td>Appendix C: Workshop Handout</td>
<td>144</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
</tr>
<tr>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>1. Participants</td>
<td>55</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bloom’s Revised Taxonomy</td>
<td>46</td>
</tr>
</tbody>
</table>
CHAPTER ONE: INTRODUCTION TO THE PROBLEM

The multiple-choice form of assessment has been in use since the early 1900s (Swartz, 2006) in a variety of educational contexts. Since its large-scale use in measuring the abilities of World War One U.S. recruits, multiple-choice assessment has grown in its popularity, and in the process, fallen under much scrutiny (Epstein et al., 2002; Williams, 2006). In discussion with my peers as both a student and teacher, I have come to observe that multiple-choice assessment is at times highly valued, yet in many other instances, strongly criticized.

The impetus for my study was to advise and keep educators current on the practical benefits, weaknesses, and overall value of well-developed multiple-choice assessments while also contributing to the literature. This, coupled with the widespread use of the multiple-choice format within education (Betts, Elder, Hartley, & Trueman, 2009; Bleske-Rechek, Zeug, & Webb, 2007; Peyton, 2010), only adds further motivation to this study.

A primarily qualitative approach was used to investigate perceptions, opinions, and the level of familiarity that educators possess in regard to their use of the multiple-choice form of assessment. An opportunity for professional development (PD) was presented in the form of a workshop. Though the success of such workshops has often been questioned, PD workshops are still frequently used. The workshop outlined guidelines and instructions for creating appropriate multiple-choice questions and presented current research that detailed methods for effectively implementing this form of assessment in the classroom.
This research was informed principally by the work of Thomas Haladyna (e.g., Haladyna, Downing, & Rodriguez, 2002). Findings from Haladyna’s research not only shaped the elements of the workshop, but the analytical tools used throughout the study (e.g., surveys and interviews). In addition, Bloom’s revised taxonomy (Anderson et al., 2001) was used as a means of reference towards ascertaining various intended levels of cognition among multiple-choice questions. In an attempt to explore the merit of multiple-choice assessment, relevant literature alongside teachers’ experiences and personal insights were qualitatively examined within an Ontario secondary school context. The results of this study may not only add to the literature on the use of multiple-choice questions as a method of assessing student learning, but may have immediate practical implications for educators as well.

**Background of the Problem**

As educators, we are in a continual search for effective and equitable forms of assessment. While most often used as a tool in measuring student achievement, testing can be dangerously tied to the belief that assessment *is* testing (Marsh, Roediger, Bjork, & Bjork, 2007). Marsh et al. (2007) suggest that assessment is more than simply an endpoint in learning. It can and should be viewed as a learning tool that will impact future student learning both positively and, unfortunately, sometimes negatively. Assessment should encourage active student reflection, include teacher and peer feedback, and promote self-evaluation (Keppell & Carless, 2006) and most important, play a direct role in improving *any* student’s learning (Ontario Ministry of Education, 2010). While this type of learning may more often be associated with formative forms of assessment (i.e., assessment that is ongoing (Volante & Beckett, 2011) and may be
considered a process (Popham, 2009) and supportive rather than conclusive (Morrissette, 2011), this should not count as reason to omit attempts at incorporating these attributes into the more often seen forms of summative assessment (i.e., tests, exams, large-scale assessment). This is particularly relevant given the Ontario Ministry of Education’s push for increased student-centered, ongoing, and meaningful assessments (Ontario Ministry of Education, 2010). Nevertheless, the emphasis on the more traditional evaluative aspects of assessment remain strong—particularly through the use of large-scale assessment—in part due to standards driven education systems worldwide (Volante & Ben Jafar, 2008) and an intensifying demand for accountability within education (Davidson, 2009; Linn, 2005; Ontario Ministry of Education, 2010).

As both an educator and a student, I often sense a disconnect between curriculum expectations and the knowledge and understanding that students are able to express within a testing environment. Disagreements between students and teachers about valid means of assessment often reflect this reality. Assessment has become increasingly high stakes since the 1970s (Linn, 2005). Consequently, high-stakes testing may detract educators from achieving the full range of desired curriculum objectives and may also undermine students achieving their potential (Reese et al., 2004; Volante & Ben Jafar, 2010). This may lead to teaching approaches that are unethical (i.e., teaching to the test in environments designed to reflect and score students’ capabilities in relation to others) and, in turn, may fail to uphold the integrity of a genuine learning experience (Chanock, Clerehan, Moore, & Prince, 2004; Volante & Ben Jafar, 2010). As an end result, students may develop poor learning habits by simply “studying” or preparing for the correct or expected responses and, in doing so, may neglect the important process
through which they might develop a genuine understanding of the material (Gulikers, Bastiaens, Kirschner, & Kester, 2008).

Through my learning and teaching experiences, I have observed that society as a whole—from small communities to large banks to governmental institutions—frequently has much to say in regards to the importance and delivery of education, along with the role it should play in young people’s lives (e.g., Moussaly-Sergieh & Vaillancourt, 2009; Toronto Dominion Bank Financial Group, 2007). According to Baker (1978), education is viewed as one of the means through which the transmission of a generation’s culture can be passed. Education appears to be taking on an ever-increasing role in the lives of students. Furthermore, it often plays a meaningful role in determining success within many societies, charged with the responsibility of maintaining and growing the skilled workforce while representing and building a society’s power, intellect, and financial well-being (Hall & Matthews, 2008). As such, whether it is from families and students, industry, political bodies, or society at large (Stewart, 1998), everyone has a stake and wants a say in how education is delivered, along with a share in what is produced.

In 1996, the province of Ontario began implementing a standardized assessment in Grades 3, 6, 9, and 10 and subsequently began publishing these test results through a government-funded body known as the Education Quality and Accountability Office (EQAO). EQAO testing possesses similar structural similarities to other large-scale assessment measures used both at the provincial (e.g., New Brunswick’s English Language Proficiency Assessment) and national (e.g., Pan Canadian Assessment Protocol (PCAP) levels, in that they are all largely and consistently comprised of multiple-choice questions. In general, multiple-choice assessment is considered to be one of the most
commonly used testing formats in education today (Moreno, Martinez, & Muniz, 2006).

As compared to constructed forms of assessment (i.e., written/open responses), the objective nature of multiple-choice questions (Torres, Lopes, Babo, & Azevedo, 2011), the ease and consistency with which these questions can be marked their potential for greater reliability (Briggs, Alonzo, Schwab, & Wilson, 2006; Wainer & Thissen, 1993), and their nearly unavoidable presence within higher education (Nicol, 2007)—in particular within popular undergraduate courses (Heyborne, Clarke, & Perrett, 2011)—are documented reasons why teachers at varying levels of education have used and continue to use multiple-choice tests (Tasdemir, 2010).

Alongside its use in high-stakes testing within various subject areas (e.g., Palmer, Duggan, Devitt, & Russell, 2010; Vogler & Burton, 2010), multiple-choice assessment has caused some teachers to alter the manner in which they approach their instructional practice. In some cases, teachers spend excessive amounts of time presenting information similar to that which is found on previous standardized tests. This is often referred to as “teaching to the test”. This form of teaching often does not produce long-lasting understanding and authentic learning. And, under these kinds of circumstances, the results of multiple-choice assessment on a standardized test may, therefore, lend a false sense of security when “good” grades are produced. This, unfortunately, undermines the intent to accurately identify student learning that has taken place (Yildirim, 2004). However, as Vogler (2002) suggested, state-mandated tests in Massachusetts, similar to Ontario’s own EQAO tests, have actually yielded improved instructional practices deemed by educational researchers to be “best practices.”
With such frequent use and important implications, it would appear that the multiple-choice format and assessment, in general, should be carried out with much deserved care and expertise. However, research suggests that, for example, in the United States, a large majority of teachers feel inadequately prepared for the assessment of student learning (Mertler, 2009). In fact, Klinger (2009) indicated out of 18 teacher education programs in Ontario, only two offered a standalone course in classroom assessment. Unsurprisingly, Mertler (2009) contended that, even after completing teacher training, many teachers still lack confidence in their abilities to develop or implement varying forms of assessment. Volante and Fazio (2007) affirm that within most American states and Canadian provinces, relatively little emphasis is placed on developing and improving teachers’ assessment competency. In particular, beginning teachers indicate a lack of understanding in how to develop diverse methods of both summative and formative forms of assessment. Furthermore, they also often demonstrate an unclear understanding of the many purposes of assessment, along with how and when various assessments can and should be used (Volante & Fazio, 2007). Predictably, beginning teachers’ appear to have relatively low levels of self-efficacy in relation to their assessment abilities. Consequently, teachers who are often left to create their own assessments from scratch may find the perceived ease of multiple-choice questions a tempting form of assessment to use without understanding the thinking that goes into its effective use.

Ineffective use of multiple-choice assessment can be seen through observations of student attitudes and student feedback. Hammann, Phan, Ehmer, & Grimm (2008) suggest that the possible grab-bag nature of responses (e.g., ranging from disdain to
favour) that are often accompanied by either subpar performances, due to large-scale assessment test anxiety and confusion, or better than expected performances accompanied by little or no studying, indicate that the multiple-choice format is potentially inconsistent in its nature and a reason why it is often perceived to have reduced validity when compared to its constructed-response assessment counterparts. This inconsistency does not make the creation of appropriate and effective multiple-choice test items any easier, as this process is already considered difficult as is and is often compared to creative writing. Item writing is, at the very least, as much an art as it is science (Crehan, Haladyna, & Brewer, 1993; Downing, 2005; Droegemueller, Gant, Brekken, & Webb, 2005). Not surprisingly, item-writing experts are, therefore, often relied upon to create certain high-stakes professional tests (Southgate et al., 2001). In addition, expertly written guidelines offering tips for writing good multiple-choice questions can lend educators ideas and suggestions immediately supporting them in the creation of their tests (Haladyna et al., 2002; Moreno et al., 2006). However, it would appear that these guidelines are potentially either not widely known, accessible, or considered to be worthwhile by the teachers who are aware of them.

Lost within multiple-choice assessment practice is often the pedagogical intent of the actual test. It is important that the intent of the assessment be clearly defined and understood by teachers and students (Moreno et al., 2006). The often perceived ease of use that many teachers attribute to multiple-choice assessment, combined with the very real challenge of generating a variety and multitude of questions, is a recipe that leads to poorly designed test items (Bush, 2006). This increases the probability of yielding invalid and unreliable results that are not only less effective in distinguishing between
low- and high-achieving students (Tarrant & Ware, 2008) but that also potentially stray from the criterion and expectations set forth by educators. This, in itself, can do great damage to the academic progress of students.

The multiple-choice form of assessment has been around for decades and is often characterized as an ineffective form of assessment. Yet—for better or for worse—it appears to be increasing in popularity, particularly at higher levels of education (Bleske-Rechek et al., 2007; Heyborne et al., 2011). Given the prominence of multiple-choice assessment within our current education system, it is imperative that this unique form of assessment is appropriately employed. Research has established the merit of effective multiple-choice assessment use—this research now needs to be presented and implemented.

Statement of the Problem Context

It is rare to encounter someone who has gone through any form of formalized schooling who has not come into contact with the multiple-choice form of assessment. Ironically, few other forms of assessment have come under more scrutiny or criticism from assessment experts in regards to their authenticity and ability to assess higher order thinking (Williams, 2006). The attitudes of both teachers and students towards multiple-choice questions are often an unpredictable mix of both positive and negative opinions. Some research has shown that multiple-choice assessment effectively measures a test-taker’s understanding of required materials, while other studies have indicated that this is not the case (Kuechler & Simkin, 2010). In addition, according to various research results, there may or may not exist a relationship between constructed-response and multiple-choice assessments (Kuechler & Simkin, 2010), leaving the merit and level of
efficacy of this format somewhat unclear. Nevertheless, multiple-choice assessment has become increasingly visible throughout most of the upper echelon of the academic world (Nicol, 2007).

Multiple-choice assessment has many features that appeal to educators. First, the multiple-choice format can assess different levels of cognition when developed properly (Nicol, 2007; Tarrant & Ware, 2008). Second, it is one of the most objective forms of assessment available (Torres et al., 2011). Third, multiple-choice assessment is both time and cost efficient and thereby increasingly important given the reality of our time and resource constraints (whether for small- or large-scale assessment purposes; Swartz, 2006). This is particularly evident at the postsecondary level given the continued reliance upon the multiple-choice form of assessment (Bleske-Rechek et al., 2007) and can be attributed to increasingly larger postsecondary class sizes, fewer resources available to teachers (Paxton, 2000), and the formation of an increasingly competitive postsecondary landscape (Wangenge-Ouma & Langa, 2010).

However, multiple-choice tests are still too often trivialized and considered limited to lower cognitive levels tied to a more rigid or rote style of learning by educators (Swartz, 2006). As a result, questions are often poorly created in an attempt to render this “simple” test more difficult. For instance, many multiple-choice tests are formatted in an overly complicated manner in an attempt to increase the level of content difficulty (e.g., insertion of excessive verbiage, trick materials, and exceedingly lengthy questions; Haladyna, 1992). Jensen et al. (2006) further indicate that accompanying the sheer growth of multiple-choice assessment is its misuse, which is even found within large-scale standardized testing such as the SAT. All educators should be mindful of the
difficulties encountered in using multiple-choice questions. If not, the quality of the assessment, along with the students who are undertaking these tests, can suffer greatly. Nevertheless, this boom in multiple-choice implementation continues and, in combination with accountability initiatives such as EQAO testing, appears to be an unavoidable reality as a part of our education system today.

**Purpose of the Study**

The purpose of this study is to gain insight into secondary teachers’ perceptions of multiple-choice assessment through the presentation and discussion of expertly created and empirically validated guidelines. As a result, this study aims to contribute to a practical and growing understanding for both educators and researchers of how multiple-choice assessment can and should be effectively used in classrooms today.

Workshops are typically held to further support ongoing professional learning in varying fields. Studies have shown that workshops can have a lasting impact on the professional development of those who participate and can create an environment highly conducive towards current and future faculty learning (Persellin & Goodrick, 2010). Nonetheless, the effectiveness of workshops as learning tools is often questioned. Varying factors such as the presence of incentives, the cost of the workshop, the perceived transferability to immediate practical use, and the time commitment involved in participating in a workshop can influence the effectiveness of workshops considerably. In particular, the amount of time necessitated by a workshop is of concern in light of our time-constrained reality, which often requires that workshops be contained within a single day format (Johnson, 2009). While this limited amount of time devoted to learning is far from ideal, one-day workshops have been found to be particularly effective if some
form of periodic follow-up is implemented (Johnson, 2009). Therefore, it seems most practical, given the nature of the participants, that the workshop conducted as part of this study be run as a single session with some follow-up.

**Research Questions**

The following are the research questions that underpinned this study:

1. What are secondary teachers’ perspectives and experiences in regards to their assessment training and, in particular, their multiple-choice assessment training?
2. What are secondary teachers’ perceptions of multiple-choice assessment?
3. What are secondary teachers’ perspectives about the value of a 1-day multiple-choice training and assessment workshop?

**Rationale**

From a researcher’s vantage point, there are numerous reasons why an investigation into multiple-choice assessment is warranted. First of all, there exists a general void in current research concerning multiple-choice assessment and, more specifically, research in assessing language and grammar (Gergely, 2007). A lack in the availability of suggested guidelines, an underestimation of the usefulness of existing research results by educators, differing perspectives towards continual professional development, and the layout of the multiple-choice format may be reasons for the lack of research in this area.

Multiple-choice assessments frequently encounter much criticism and have a bad reputation as a test format that promotes nontransferable skills (Paxton, 2000), recognition-oriented studying habits, guessing, and the honing of test-taking skills as opposed to the learning of course content (Rogers & Harley, 1999). An awareness of
these criticisms could also lead to the presumption that those administering multiple-choice tests are doing simply so out of indifference or due to a lack of a concerted effort to assess learning in other ways (Jensen et al., 2006). While many educators perceive multiple-choice questions as being ineffective for drawing out creative and meaningful written responses from test-takers, Nicol (2007) has shown that multiple-choice questions are capable of assessing much higher levels of cognition than most would expect.

Multiple-choice questions can and should go beyond the simple testing of students’ abilities to recall information. In fact, these questions should be developed in a manner that would assess students’ abilities to understand, apply, analyze, and evaluate subject content. Furthermore, contrary to popular belief, it has been demonstrated that multiple-choice questions can assess levels of cognition above simple recall or recognition, such as higher level cognitive reasoning that is reached utilizing constructed response questions (Simkin & Kuechler, 2005). It has also been shown that the development of multiple-choice questions can be accomplished without overcomplicating the nature and format of the test with unnecessary frills and test-taking complexities (Haladyna, 1992).

The increase in the desire to have multiple-choice assessment take on a greater cognitive-oriented look is not coincidental and can be linked to the rising influence of cognitive psychology (Robins, Gosling, & Craik, 1999). The research of Benjamin Bloom effectively reflects the emerging influence of cognitive psychology and successfully translates these concepts into an education-oriented design, so that it might be integrated into instructional goals for both teaching and learning processes (Anderson et al., 2001). This application is now known as Bloom’s taxonomy, although for this
study, I have used the updated revised Bloom’s taxonomy, which reflects recent advances made in the field of cognitive psychology (Hanna, 2007).

The benefits of the multiple-choice format can also be examined from a larger scope in terms of (a) its ability to create financial savings for both educators and administrators, (b) its versatility, (c) its nonbiased method of evaluation across nationalities (Gatfield & Larmar, 2006), and (d) its function as an objective grading tool, often found lacking in other measurement options (Williams, 2006). With the advent of multiple-choice aiding technology, such as Scantron© cards, this format has become even more appealing in terms of the time saved in scoring its questions, the cost-effectiveness of utilizing multiple-choice assessment, and the efficiency of its use for educators, regardless of the level of schooling or the subject matter (Nicol, 2007). This appeal is evidenced by the fact that researchers continue to investigate the attributes and effectiveness of multiple-choice assessment across a multitude of levels and different subjects from the elementary level (e.g., Hammann et al., 2008), the secondary level (e.g., Peyton, 2010; Vogler, 2002; Yildirim, 2004), and all levels of postsecondary and professional schooling (e.g., Fischer, Herrmann, & Kopp, 2005; Heyborne et al., 2011; Majerich, Stull, Jansen Varnum, Gilles, & Ducette, 2011; Tarrant & Ware, 2008).

As previously mentioned, I cannot say conclusively why current guidelines and training for designing multiple-choice assessment items are not more often used. Moreno et al. (2006) however hypothesize that multiple-choice assessment guidelines are often obscure and somewhat convoluted to the average educator, appearing more researcher friendly than classroom teacher friendly. If guidelines are not field tested, practical, and
research based, it is difficult to imagine that teachers would give these guidelines more than a second glance.

Within a small Ontario context, my research aims to assess the value of and need for further multiple-choice assessment training. As it stands, there exists a lack of preservice programs that contain dedicated assessment courses as a part of their consecutive teacher education programs (Volante, 2010). I hope to identify teachers’ perceptions of the multiple-choice format. In turn, this might assist me in ascertaining the reasons for which these individuals or groups might wish to use this format and the educational contexts and subject areas in which this form of assessment might be most beneficial, relevant, and realistic. At the same time, I hope that my study will contribute to the mandate found within the Ontario College of Teachers (OCT) Foundations of Professional Practice—to commit, not only to the success of student learning but to one’s ongoing professional learning (Ontario College of Teachers, 2006).

This study is important to me as a researcher and a teacher because it seeks to reveal an important understanding of the potential efficacy of multiple-choice assessment while at the same time potentially clearing doubts regarding the myths about multiple-choice assessment. This research also demonstrates the viability of the workshop format in supporting a teacher’s professional development. Furthermore, as a vast majority of multiple-choice tests created today do not typically promote appropriate studying or test-taking habits (Rogers & Harley, 1999), it was hoped that the outcomes of this study would inform future research on multiple-choice assessment.

Overall, this research aims to survey and address misconceptions that surround multiple-choice assessment by identifying appropriate assessment training techniques and
allowing multiple-choice assessment to fill its niche role in educational assessment. In doing so, I hoped, through this study, to change one of contemporary educational assessment’s enduring standards—the multiple-choice test—from an oft-questioned weakness to an effective and research-supported strength.

**Study Framework**

The cognitive levels and multiple-choice guidelines used within this study were based upon research conducted by Haladyna et al. (2002) and Haladyna (2004) and Anderson and Krathwohl (2001). Haladyna and his colleagues have conducted research in order to produce a list of guidelines useful for writing successful multiple-choice questions. From the revised Bloom’s taxonomy (Anderson et al., 2001), six levels of thinking were used to describe intended levels of cognition for multiple-choice questions, along with guidelines and suggestions for improving the development of this form of assessment.

**Scope and Limitations of Study**

This study was conducted with secondary school teachers who volunteered to participate in the workshop and in the research. As such, the scope of participants is fairly limited, leading to a sample of educators who might already practice good assessment habits as indicated by their willingness to participate in this study. Because the participants were drawn from the secondary school panel, the results of this study may not be generalized to elementary school teachers or even necessarily all secondary school teachers. Participants were from a school board in southern Ontario and are thus likely to be representative of the demographic only within that area. Participants who
volunteered to be a part of this study were allowed to participate regardless of their prior work or academic experience, limiting the generalizability of the results.

Time limitations due to the demanding schedule of secondary school teachers left the participants with little opportunity to discuss and/or attempt to implement any changes in their approach towards multiple-choice assessment. Ideally, professional development with repeated focus group interviews and workshops would take place throughout the year. In fact, over the course of several years, long-term professional development would provide for a more comprehensive and effectual understanding of how this form of assessment might be best designed and used by teachers.

Within the central scope of this study was the need to inform participants about existing research on multiple-choice assessment and to provide them with opportunities to reflect, comment, and discuss their thoughts and experiences with multiple-choice assessment. Additionally, my hope as a researcher is that the feedback and results from this study will both extend and spotlight relevant information that could be used in forthcoming research and assessment development across all areas of education.

**Outline of Remainder of the Document**

Chapter 2 includes a review of the literature pertaining to the current and evolving role and use of multiple-choice assessment in education as well as the effectiveness of a workshop format on teachers’ professional learning. The strengths and weaknesses of multiple-choice assessment are discussed alongside more detailed descriptions of different aspects of response items, structures, and marking schemes used in multiple-choice assessment today. The chapter concludes with a summary of Bloom’s revised taxonomy (Anderson et al., 2001) in regard to the manner in which knowledge and
cognitive processes are defined as well as a breakdown of Bloom’s alignment assessment objectives.

Chapter 3 outlines the research methodology of the study. The specific aspects of the methodology covered in this chapter include the research design, selection of site and participants, instrumentation, data collection, and data analysis. In addition, the limitations and ethical considerations of the methodology used in this study are examined.

Chapter 4 presents the research findings. The results revealed teachers’ experiences with assessment training and, in particular, training that those teachers received in the area of multiple-choice assessment. These results illuminate the lack of assessment training most teachers possess, the accompanying level of uncertainty, and the desire for hands-on training regarding specific variations and uses of the multiple-choice format. Responses suggest that while all teachers employ multiple-choice assessment, it is yet used to quite varying degrees. In addition, teachers’ comfort levels in implementing multiple-choice tests and their general perception of this assessment format are also reported. Following this, detailed and anecdotal teacher insights are presented in relation to perceived multiple-choice assessment strengths and weaknesses, along with the participants’ perceptions about student attitudes towards their use of the format. Finally, teachers’ opinions of multiple-choice assessment following their participation in the workshop are presented, documenting any changes in attitudes towards both the multiple-choice format and its accompanying training. These results help to reveal the current reality, practitioner benefits, and overall value that multiple-choice assessment can present to secondary school teachers.
Chapter 5 presents the importance of the research findings. This section restates the purpose of the study, followed by a brief summary of the research used and the findings from the study. Next, the findings are discussed in the context of everyday secondary school pedagogy. In light of existing research, this discussion surrounds teachers’ perspectives of their multiple-choice assessment training alongside their views as they pertain to both their teaching practice and their students’ levels of success. Teachers’ perspectives on multiple-choice assessment are then briefly described following their participation in a multiple-choice workshop. Finally, this chapter concludes with a discussion of the implications that these findings may have on secondary school teaching and assessment practices and on future research on multiple-choice assessment.
CHAPTER TWO: REVIEW OF RELATED LITERATURE

While multiple-choice research appears to make up a large portion of assessment research done over the years, there appears to be limited research pertaining to how multiple-choice is both perceived and used in a secondary school setting. Chapter 2 summarizes research on the use of the multiple-choice assessment format across its wide range of assessment functions. Throughout this chapter, the role, benefits, apprehensions, and strategies surrounding the multiple-choice format are also investigated. The effects of professional learning workshops on educators’ views of multiple-choice assessment are also discussed. The chapter concludes with a review of recent revisions to Bloom’s taxonomy regarding the types of knowledge and levels of cognition pertinent to purposeful and successful assessment.

Professional Learning Workshops

In an ever-evolving educational and societal landscape, educators continually face the challenge of keeping abreast of new educational standards and knowledge (Yang & Liu, 2004). In order to keep up with these continual changes, many educators attend workshops and conferences in order to keep their knowledge and skills current. For example, in a workshop-related study, Lamiani and Furey (2008) demonstrated that workshops were successful in teaching nurses how to more effectively educate their patients. In general, workshops can promote new professional knowledge and understandings required by various professional industries (Bulik & Shokar, 2007; Cronin & Connolly, 2007). For teachers, this is traditionally accomplished through the use of inservice professional development courses (Kelly & McDiarmid, 2002).
Researchers surmise that teachers’ participation in workshops will lead to the expansion of their knowledge and skills, their growth, and their effectiveness with students (Guskey, 2002). Not surprisingly, teachers believe they gain valuable knowledge and skills during a workshop, and often come out of workshops feeling more prepared and confident in applying their newly acquired skills (Persellin & Goodrick, 2010; Shriner, Schlee, Hamil, & Libler, 2009). Johnson (2009) notes, however, that while workshops, over extended periods of time, can be beneficial, they are characteristically of little interest to teachers or they simply do not fit into the busy schedules of educators. Johnson further iterates that single session or one day workshops are more common than longer term workshops.

The shortcomings of the single-session workshops are often cited. These workshops are commonly blamed for (a) their costly nature in terms of travel inconveniences and other expenses, (b) the artificialness of their setting, since most workshops take place outside of the classroom, and (c) their lack of long-term effectiveness and continued support following the workshop (Yang & Liu, 2004). Additionally, given the time constraints of a teaching schedule, it is unlikely that teachers would be able to remain up to date with the rapidly changing landscape and knowledge available through the occasional workshop session (Minott, 2010). However, in research conducted by Johnson (2009) and Lydon and King (2009), single- and double-session professional learning workshops for teachers were shown to be beneficial and effective. According to Johnson, the use of incentive-laden postworkshop assignments, within which participants were asked to develop, implement, and reflect upon workshop ideas, proved to be the catalyst for success. Similarly, researchers agree that the means for
effective professional development stem from reflective teaching strategies that must be in place, followed by the execution of self-directed, collaborative, and hands-on interactive sessions occurring amongst colleagues (Alber & Nelson, 2002; Shagrir, 2010; Simon, Campbell, Johnson, & Stylianidou, 2011).

Therefore, it is likely that short-term workshops can be effective when cleverly planned alongside components spread over a period of time and conducted within an appropriate context that is relevant to teachers’ desires and needs (Bulik & Shokar, 2007; Cronin & Connolly, 2007; Johnson, 2009; Lee, 2004-2005). Furthermore, Lee (2004-2005) reiterates that professional development sessions are most effective when teachers receive opportunities to take what they have learned and immediately put this learning into action in their classrooms. The importance of having teachers take part in follow-up sessions to both measure and ensure their level of success cannot be understated. It is also important to note that discrete workshops of brief lengths are sometimes the only realistic media in which continued professional learning and growth can take place for teachers.

The Role and Emergence of Multiple-Choice Assessment

In recent decades, there has been an increased focus towards the success of all students. This educational trend has played a role in the emergent need for greater transparency and accountability in all areas of education, both within the United States (Linn, 2005) and in Canada (Daniel, 2005; Graham & Jahnukainen, 2011). Researchers have suggested that in order to achieve these goals, educational policy must insist upon the frequent use of assessment—in particular large-scale assessment—wherein administrations, schools, staff, and students are all held accountable (Linn, 2005;
Haviland, Shin, & Turley, 2010; Reese et al., 2004). Unsurprisingly, the effect of accountability practices is likely to have a large impact on everyone, whether this contact be through direct (e.g., students, teachers, and administrators) or indirect (e.g., parents) involvement (Daniel, 2005).

While the use of large-scale assessments in Canada has increased both provincially/territorially and nationally, and at the same time become increasingly high stakes (Klinger & Luce-Kapler, 2007), it is the varying interpretations of the purpose of these assessments and the varying applications of its results that emphasize and drive accountability (Volante & Ben Jaafar, 2008). The emphasis and execution of accountability measures is, however, considered inconsistent and, perhaps, unsustainable by some educators, researchers, and even the general population—given the varying purposes and interpretations that come with each policy level. Nonetheless, it is important to give Canada some time to develop these measures, noting the complexity and care that are required to balance the internal pressure of satisfying and respecting Canada’s diverse and ever-changing culture while pushing to have an increasingly homogenized state of assessment that would also legitimize Canada’s system internationally (Volante & Ben Jaafar, 2008).

From a more direct pedagogical standpoint, researchers have in the meantime noted that provincial- or state-wide assessments have become limited in terms of accurately measuring intended targets (Marsh et al., 2007), while the efficacy of these tests in serving as comprehensive forms of assessment has also decreased (de la Torre, 2009; Nicol, 2007). In some instances, rather than ensure the quality of the learning and teaching experience, educators have too often been caught up with the results of
standardized assessments (Scharf & Baldwin, 2007; Volante, 2004). However, studies by Paxton (2000) and de la Torre (2009) have shown that educators, in essence, have opposed this school of thought, espousing the same more time-honoured belief that assessment should be used “to educate and improve student performance, and not merely audit it” (Wiggins, 1998, p. 11).

While the multiple-choice format has been in use since the early 1900s (Swartz, 2006), the growth of this format, primarily in the United States, has been attributed to many factors (Williams, 2006). These factors include a surging number of students entering higher education, a reduction of teaching resources at multiple levels of education, modularization (the breaking down of subject areas or problems into subproblems and subtasks), and the emergence of modern technology and the ease it affords in administration of multiple-choice testing (Majeric et al., 2011; Nicol, 2007). In many instances, educators at both the secondary and postsecondary levels are unable to increase the amount of resources needed to keep pace with the growing student population and corresponding workload (Heyborne et al., 2011). As a result, teachers are often left with the option of either accepting a decrease in the quality of their work or coming up with ways of reducing the necessitated workload. The continuing modularization of subjects favours the use of multiple-choice questioning in the areas of knowledge recall and recognition (Nicol, 2007). Consequently, it is not uncommon to find multiple-choice assessments frequently used in secondary schools (Briggs et al., 2006; Peyton, 2010;), within large-scale assessment environments at provincial- or state-wide levels (Rogers & Harley, 1999), at college and university levels (Heyborne et al.,
2011; Holtzman, 2008; Parmenter, 2009), and in high-stakes environments such as medical and health sciences education (Fischer et al., 2005; Tarrant & Ware, 2008).

**Multiple-Choice Assessment Benefits**

While efficient and easy to implement, research suggests that educators also embrace the multiple-choice format owing to its simplicity and objectivity in assessing student knowledge (Marsh et al., 2007; Swartz, 2006). Many school textbook publishers include a testbank of multiple-choice questions from which educators can select questions for use in their classroom assessment (Simkin & Kuechler, 2005), further increasing this format’s ease of access and use in addition to the amount of time and resources saved. Researchers also indicated that it is possible to test cognitive abilities such as recognition, application, and analysis using multiple-choice testing (Haladyna, 1992; Simkin & Kuechler, 2005; Tarrant & Ware, 2008; Torres et al., 2011).

Furthermore, the work of Marsh et al. (2007) indicated that the general and more frequent use of multiple-choice tests, to the benefit of both students and educators, is capable of providing students with noticeable memory-enhancing benefits such as additional study opportunities, knowledge retrieval practice, and cues in the form of response options. Subsequently, when constructed well, multiple-choice tests can be used as both a testing and learning tool. As a learning tool, a multiple-choice test can cover the same content area as traditional studying methods and also closely simulate the format and experience encountered on a test.

Nicol (2007) indicated that the utility of multiple-choice questions extends beyond typical classroom assessment and has created its own niche within modern teaching tools such as the Electronic Voting System (EVS) as well as in metalearning
tools such as the Multiple-Choice Item Development Assignments (MCIDA)—whereby students are asked to develop their own multiple-choice questions. These types of tools capitalize on the marking accessibility and simplicity of the multiple-choice format with their unique representation of subject content (Jensen et al., 2006; Scharf & Baldwin, 2007). Multiple-choice assessment can also further be used as a form of formative assessment that enhances learner autonomy through unique learning experiences similar to the EVS (e.g., Gupta, 2010; Majerich et al., 2011) and the MCIDA (e.g., Pittenger & Lounsbery, 2011). Budding technological assessment aids and changes in assessment context (e.g., questions written by students) have provided unique opportunities for self-assessment and reflection, group discussion, and two-way feedback between both students and their teachers (Nicol, 2007). Holtzman (2008) also demonstrated that, through focused goals, multiple-choice tools are capable of presenting new methods and opportunities for deeper learning that go beyond the levels of recall and recognition. Therefore, when linked to clear pedagogical goals that include test context and content and timely and appropriate feedback (Parmenter, 2009), the use of the multiple-choice format can initiate effective learning. In short, multiple-choice assessment should be given due consideration as a formidable assessment format that educators can use to instruct, engage, and dialogue with their students (Nicol, 2007).

Multiple-Choice Assessment Concerns

Despite the benefits of and optimism towards multiple-choice assessment and its applications, the limitations and negative aspects of this format are well known (e.g., Fellenz, 2004; Paxton, 2000; Williams, 2006). Examinations employing multiple-choice questions are often viewed as assessing surface learning rather than deep learning (Betts
et al., 2009; Rogers & Harley, 1999). Williams suggests that the successful completion of a multiple-choice test is not often seen as equivalent to mastering real-world skills—skills that hold great value for educators. In essence, recognizing the correct response on a test does not provide an indication of whether a student can construct that response independently. Thus, the multiple-choice format is sometimes viewed as lacking in authenticity.

From the perspective of many students and teachers, multiple-choice questions succeed only in evaluating factual recall and lower—level abilities (Haladyna, 1992; Scharf & Baldwin, 2007). Questions are seen as restrictive in terms of independent thought and creativity (Cappel, 2007; Williams & Clark, 2004) and lacking in their ability to assess higher order critical thinking (Schultz, 2011) and/or communication skills (Paxton, 2000). In addition, some educators believe that the multiple-choice format promotes test-taking strategies, student guesswork, “test-wiseness,” and poor studying habits over actual learning (Jensen et al., 2006; Scouller, 1998; Williams & Clark, 2004).

While it may be possible to test distinctive levels of cognition (e.g., recognition, application, analysis), this would depend heavily on the way in which a test is created (Haladyna, 1992; Tarrant & Ware, 2008). It is important to note that assessment development frequently absorbs a substantial amount of a secondary school teacher’s planning time (Jensen et al., 2006) while also necessitating a combination of both skill and experience on the teacher’s part in order to develop multiple-choice questions that are both well written and capable of discriminating between higher and lower achieving students (Bush, 2006; Taylor & Smith, 2009). Subsequently, caution is warranted, as the
successful creation of questions measuring different levels of cognition remains an area of uncertainty and concern.

**Item Writing**

Designing effective test questions is no easy task (Bush, 2006). In order to create well-written multiple-choice questions, existing guidelines suggest the following generic principles: (a) items should be content-specific, clear, and concise; (b) questions should not be answerable without targeted content knowledge, with the exception of limited guessing while employing plausible distractors; and (c) educators should avoid using absolute or vague terms, as well as “trick” questions (Haladyna et al., 2002; Niemi, Vallone, Wang, & Griffin, 2007; Taylor & Smith, 2009). These principles appear to be commonly established threads among many researchers and, while appropriate, do not provide clear instruction for educators on how to effectively construct multiple-choice questions. In addition, Haladyna et al. (2002) suggest that due to the lack of a validated taxonomy of cognition, there is little upon which to base an item-writing guideline that targets specific levels of cognition. Consequently, it is difficult to gain headway in addressing the cognitive assessment difficulties related to multiple-choice testing. In essence, the aforementioned guidelines appear to be no more than a generic list of items for analyzing and evaluating multiple-choice questions that have little effectiveness in aiding teachers with the creation of cognitive level-discerning multiple-choice questions.

Overall, item writing is considered by some researchers (e.g., Crehan et al., 1993; Downing, 2005; Droegemueller et al., 2005) to be more of an art form than a science, while others (e.g., Haladyna et al., 2002) consider item writing to be an immature science in need of further empirically validated guidelines. Ebel (1951) stated:
Just as there can be no set formulas for producing a good story or a good painting, so there can be no set of rules that will guarantee the production of good test items. Principles can be established and suggestions offered, but it is the item writer’s judgement in the application (and occasional disregard) of these principles and suggestions that determines whether good items or mediocre ones will be produced. (p. 185)

When teachers lack creativity, have a “writer’s block,” or for whatever reason are struggling to design new questions, it can be quite onerous to prepare a large bank of quality questions. This, in turn, can predictably cause flaws to appear in the questions (Bush, 2006). In a study by Tarrant and Ware (2008), question flaws were identified and analyzed within high-stakes nursing assessments and were found to impact both higher and lower achieving students. Tarrant and Ware discovered that not only do these flaws reduce the scores of higher achieving students, but that the scores of lower achieving students increase, resulting in a regression towards the mean. Similarly, Haladyna et al. (2002) and Downing (2005) discovered that these flaws, or construct-irrelevant factors within questions, yield very similar outcomes. Such flawed questions had a tendency to confuse higher achieving students, leading to incorrectly answered questions. On the other hand, these flawed questions may have also provided unintended cues to lower achieving students, who are then able to “guess” correct responses.

Overall, when created and presented carefully to students, multiple-choice assessment is as reliable, and frequently more reliable a testing instrument, than constructed-response tests (e.g., essay examinations; Jensen et al., 2006; Parmenter, 2009; Ventouras, Triantis, Tsiakas, & Stergiopoulos, 2009; Wainer & Thissen, 1993).
Modifications to Conventional Multiple-Choice Assessment

Multiple-choice questions can be an inexpensive and efficient manner of assessing student knowledge if the appropriate levels of effort, care, and experience are incorporated into developing the questions (Bush, 2006; Holtzman, 2008; Jensen et al., 2006; Nicol, 2007; Tarrant & Ware, 2008; Williams & Clark, 2004). In order to avoid superficial assessment within this format, many test writers have attempted to create variations to the standard multiple-choice format (e.g., having one correct response along with multiple incorrect responses), while others have attempted to modify the test-creation process by utilizing an adapted marking scheme quite different from the commonly employed preset mark-per-correct grading scheme.

Haladyna et al. (2002) confirmed that many textbooks and standard tests employ six different multiple-choice formats. These formats include the following:

1. Conventional multiple-choice (MC)—typically containing three to five options;
2. Alternate-choice (AC)—essentially a two-option question;
3. True-false (TF)—a statement and declaration of whether the statement is true or false;
4. Multiple true-false (MTF)—multiple (i.e., 3–30) true or false statements held in one question;
5. Complex multiple-choice (CMC)—also referred to as “Type K,” typically consisting of groupings of answers. For example, i) 1 and 2 only, ii) 2 and 3 only, iii) 1 and 3 only;
6. Context-dependent item set (CDIS)—also referred to as “testlets,” containing a stimulus (e.g., pictorial representation, reading comprehension text) followed by a set of multiple-choice questions relevant to the stimulus.

Haladyna et al. also discovered that the MC and TF formats were used in 100% of the textbooks reviewed at that time, while the CDIS format was used in 70% of these texts, the MTF in 37%, the CMC in 31%, and the AC in 11%. Out of these six formats, four formats (MC, AC, CMC, CDIS) made use of the traditional or standard multiple-choice form of assessment. The questions that made up each of these types of tests varied in terms of the number of response options available (e.g., AC and MC), whether each response contained only one or multiple grouped responses (e.g., CMC), and whether there were multiple sets of response options (e.g., CDIS).

There is much discussion surrounding the optimal number of response options for a multiple-choice question as well as the abilities of the CMC and CDIS formats to achieve desirable levels of difficulty, reliability, and discriminatory ability.

**Optimal Number of Response Options**

Researchers have shown (e.g., Bruno & Dirkzwager, 1995; Landrum, Cashin, & Theis, 1993; Rodriguez, 2005; Rogers & Harley, 1999) through information theory modelling, empirical research psychometric research, and statistical/mathematical research, that three-option questions appear to be more effective than four- or five-option questions. Rogers and Harley (1999) suggest that three-option questions consisting of two equally plausible distractors are well-suited to optimize a test writer’s resources and time. While administrative work is lessened, the three-option format still manages to achieve desired levels of internal consistency, scoring reliability, and question difficulty.
Test validity can also be increased through the development of additional content-related questions, and by not having to create more than two distractors, educators are also able to save time (Trevisan, Sax, & Michael, 1994). Haladyna et al. (2002) have also suggested that having fewer item-response options does not necessarily exhibit as many benefits as having a larger number of items. However, considering the time and resource limitations experienced by educators, most teachers would likely support the use of two-distractor multiple-choice questions. This is supported by various claims that multiple-choice questions consisting of three options are capable of maintaining question integrity and discriminating power (Crehan et al., 1993; Haladyna et al., 2002). As a result, Haladyna et al. conclude that when implementing conventional multiple-choice assessments, three-option multiple-choice questions can effectively assess student learning.

**Response Option Modifications**

According to Haladyna (1992) and Haladyna et al. (2002), while CMC questions feature a level of difficulty greater than conventional MC questions, CMC questions are, at best, only equal to conventional MC questions in terms of their discriminating power and their reliability. In addition, CMC questions are seen as less efficient because they are time-consuming for both test-creator and test-taker. Haladyna et al. criticized the CMC format further, suggesting that this type of test may provide clues to the test-taker, enabling *test-wiseness* and test-taking strategies that play a role in determining student responses.

*Multitrak* items are less often seen than CMC questions and are, in essence, the inverse of the conventional MC response option format. When employing *multitrak*
items, test-takers are required to select the sole incorrect answer rather than the only correct response (Gergely, 2007). *Multitrak* items were used by Gergely (2007) to assess whether native Hungarians met a certain level of English grammar proficiency. What he was then able to confirm anecdotaly was that these items required a “different kind of thinking,” causing students to read over each response option carefully; an obviously desirable trait. Students at varying skill levels were impacted in different ways by the *multitrak* format. The results of the *multitrak* test were successful in providing more meaningful information in regard to higher ability students, because test content was assessed at a greater level of difficulty. It was further suggested that a well-constructed test would also include other types of assessment items, which might more appropriately assess students at varying levels of achievement.

Another way in which response options of multiple-choice tests have been modified is through the introduction of several *sets* of responses pertaining to the content of a question. The CDIS and two-tiered type questions are great examples of this. CDIS questions exhibit greater potential in measuring higher level thinking (Haladyna, 1992; Haladyna et al., 2002), while two-tiered questions typically use a second set of selected responses, affording students the opportunity to substantiate their initial answers, thereby providing a more sensitive and effective way of assessing meaningful learning (Chandrasegaran, Treagust, & Mocerino, 2007).

While the various multiple-choice formats have shown promise as effective assessment tools, there is a lack of related research regarding certain aspects of these assessments. Overall, these modified response formats would benefit from further research.
Modified Marking Schemes

Multiple-choice assessment is already considered to be one of the most commonly used methods of assessing student work, not only at postsecondary levels but across all levels of education (Heyborne et al., 2011), particularly when one is seeking an assessment format that is both relatively simple and objective (Kuechler & Simkin, 2010; Marsh et al., 2007). Educators, however, have often struggled with how to prevent or lessen the “success” of test-takers who possess little content knowledge or skills and who are often capable of obtaining inflated marks due to accurate guesswork (Betts et al., 2009; Bush, 2006). Paxton (2000) argued that guessing an answer correctly, when one possesses limited content knowledge, can be problematic. This can lead to false assurances that students actually understand a concept, when they have, in reality, only learned to recognize the correct answer. As a result, educators have attempted to combat this guesswork by manipulating the way in which marks are assigned to tests. Two of the most common modifications to the conventional multiple-choice marking scheme (where there is no penalty for an incorrect answer) are negative and confidence-based marking (Burton, 2006; Bush, 2006).

In an attempt to discourage test-takers from using arbitrary guesswork, negative marking can be used, though feasibility issues may detract a teacher from using it within a typical class. This form of marking is very similar to conventional marking, with the exception that incorrectly answered questions are generally penalized by a whole mark (Betts et al., 2009; Burton, 2006; Bush, 2006). In a study done by Scharf and Baldwin (2007), three different forms of negative marking—zero penalty (conventional multiple-choice), intermediate penalty (based on questions attempted), and maximum penalty
(penalties for incorrect or unanswered questions)—were analyzed and compared. Results indicated that the zero penalty scenario would likely be too lenient and would yield inflated marks due to guesswork. The opposing maximum penalty scenario would likely be too harsh, given that a large sum total of marks would, in essence, be penalized doubly for mistakes and questions simply not known by students. The intermediate option appears to be the most justified, yielding balanced results that fall between the zero and maximum penalty scenarios. In response to whether students would benefit from guessing in these scenarios, the results indicated a resounding yes in the zero penalty scenario. However, in the intermediate scenario, the opposite was true in that students, on average, do not benefit at all from guesswork. In the same vein, Burton (2006) suggested that if guesswork were eliminated, then a teacher would be able to assess an individual’s knowledge as well as his or her pool of misinformation, thereby helping to maintain overall levels of reliability (Burton, 2005). Nonetheless, there remain concerns that students’ overall levels of confidence and their willingness to risk guessing could play a non-content-related role in determining students’ scores (Burton, 2006).

Researchers have attempted to counter these concerns with a confidence-based marking scheme (Burton, 2006). While this more recent innovation has yet to make its mark in large, high-stakes standardized testing formats, it is currently used in areas of both formative and summative assessments and requires students to reflect upon the reasons behind their selected responses (Nicol, 2007). By having students assign a numerical confidence value to their responses and by using a mathematically based theory of information algorithm that shows how many points to reward or deduct accordingly, educators have been somewhat able to increase the viability of the process
through which they assign grades that are supposed to reflect an individual’s level of knowledge and understanding. Answers guessed correctly as a result of a lack of confidence or due to a lucky guess are no longer rewarded, because a confident and justified belief in a response would garner a much higher mark (Gardner-Medwin & Gahan, 2003).

Gardner-Medwin and Gahan (2003) have reiterated that the purpose of confidence-based marking is not to reward or discourage individuals but to encourage the use of practical and essential intrinsic learning tools such as reflection and self-awareness. Through the use of confidence-based marking on true/false questions, Gardner-Medwin and Gahan demonstrated that inherent confidence levels do not appear to influence the manner in which students respond to test items. Additionally, when testing the same students on the same content for both the bottom and top thirds of the class (as compared through mean values of \( r^2 \) for correlations between scores found in this marking method and conventionally marked tests), results indicated that the ability to discriminate among students was significantly higher in confidence-based marking as compared to the conventionally marked true/false questions of a test.

It is important to note that confidence-based marking is not intended to simply reward or discourage self-confidence. This marking method instead aims to not only encourage student reflection and self-awareness but also to serve as a formative learning experience in which students develop a greater sense of value and attention to the content area (Gardner-Medwin & Gahan, 2003). Interestingly, when students express a higher degree of confidence during test completion, they are likely more invested in their work and can therefore more effectively justify their responses. As part of this process,
students are also more likely to have solidified existing content knowledge and few misconceptions about the content of the test. This formative benefit is particularly evident when responses are accompanied by a high level of confidence and are marked *wrong* (Gardner-Medwin & Gahan, 2003). In these instances, students are often prompted to explore the reasons for their responses. In so doing, students wrestle with their own understandings and, occasionally, identify ambiguities or flaws within the questions themselves, making this a reflective practice for both test-takers and test-creators (Gardner-Medwin & Gahan, 2003).

However, one must remain cautious when employing unfamiliar testing formats, as these frequently leave test-takers perplexed or overly focused on developing alternative test tactics (Bush, 2006). Burton (2005) found that educators are concerned that such forms of evaluation as the negative marking scheme might reduce student attention to the actual test material and might subsequently measure risk-taking behaviours rather than targeted subject knowledge and understanding (Betts et al., 2009).

The value of certain modified marking schemes, such as confidence-based assessment, has provided educators with an increased richness in feedback generated from tests, while at the same time offering improved test validity (Swartz, 2006). On the other hand, researchers investigating the use of negative marking (e.g., Betts et al., 2009) have suggested that there is generally no significant advantage in utilizing this marking scheme.

In this researcher’s opinion, *all* modified marking schemes should be subjected to further investigation. Furthermore, it is important to note that these marking schemes are not necessarily beneficial for either classroom or large-scale assessment environments. If
these schemes are deemed worthwhile by educators and implemented for both environments, they could very well end up looking and operating differently—depending on the number of students receiving the assessment along with the context in which the assessments are being delivered. These marking modifications have the potential to improve teachers’ pedagogy and assessment efficacy as well as the student learning experience when these forms of multiple-choice assessment are appropriately designed and implemented.

**Feedback**

Constructive feedback is an important but often forgotten learning and assessment tool for all students and teachers (Karp & Woods, 2008; Paxton, 2000). Within multiple-choice assessment, feedback is used to inform educators about what students have learned or understood (Marsh et al., 2007) and can thus be an effective way of assuring assessment accuracy and quality of learning (Bush, 2006; Marsh et al., 2009). There are, however, noted limitations and/or a lack of productive feedback experienced by students following the completion of a multiple-choice assessment task (Paxton, 2000). These difficulties occur despite the fact that multiple-choice tests can be quickly returned and taken up in the classroom (Manning & Dix, 2008; Nicol, 2007).

However, the idea that feedback cannot be integrated into multiple-choice assessment is unjustified. When test results are supplemented with teacher feedback or immediate computer responses, multiple-choice questions have the ability to place student answers and marks into a learning context in which students are better able to draw accurate, pertinent, and critically timed insights from their assessment experience (Scharf & Baldwin, 2007). This affords students the opportunity to make updates and
corrections to their work, to counteract various negative testing effects (Fazio, Agarwal, Marsh, & Roediger, 2010), and to constructively apply the assessment experience towards future tests and learning (Scharf & Baldwin, 2007). Consequently, rapid and timely feedback can be an excellent tool towards the validation and success of the multiple-choice format.

**Bloom’s Revised Taxonomy**

Part 1 of Bloom’s original taxonomy (OT), entitled *Taxonomy of Educational Objectives, Handbook 1: The Cognitive Domain*, was initially written by a group of educators, psychologists, and school examiners, in an attempt to develop educational goals and measurements (Booker, 2008). This taxonomy was later revised in 2001 by a group of cognitive psychologists, curriculum and instructional researchers, and assessment specialists (Anderson et al., 2001; Bumen, 2007) in order to reflect the changes and advances in the fields of both cognitive psychology and educational research (Hanna, 2007). Both the OT and the newer Revised taxonomy (RT) are still widely used worldwide and especially so within the United States as conventional standards for assessing and evaluating student levels of cognitive ability and development (Booker, 2008).

Bloom’s OT, along with the RT, discuss student performance evaluations in which classification systems are established in hopes of assessing and evaluating students with greater precision (Booker, 2008; Bumen, 2007; Hanna, 2007). Many educators attempting to assess students’ cognitive levels and abilities are familiar with Bloom’s taxonomy (Kuhn, 2008). In the RT, which is considered to be the most commonly accepted educational objectives model, Ralph Tyler (as cited in Anderson et al., 2001,
Chapter 1) suggested that “the most useful form for stating objectives is to express them in terms that identify both the kind of behavior to be developed in the student and the content . . . in which this behavior is to operate” (p. 12). Agreeably, what is most useful in ascertaining teaching effectiveness is determined by a student’s level of learning (Bumen, 2007).

Not all researchers, however, consider Bloom’s work to be necessarily beneficial to the education system. Booker (2008) indicated that, for the most part, Bloom’s taxonomy is in use across all levels of education, even though it was originally intended only for postsecondary-level education and not the K-12 grades that it more predominantly influences. In actuality, there is little direct use of the taxonomy within the postsecondary context, which leaves the taxonomy seemingly misplaced and, perhaps, inappropriately used at the secondary school level and below. Reacting to Bloom’s RT focus on higher order learning, Booker hypothesized that many postsecondary educators feel that incoming students are often ill prepared for higher level education. Booker believes this is due to inadequacies in the development of basic knowledge, stemming from an overemphasis on higher order learning within elementary and secondary school education. As a result, students are left lacking the basic essentials when they begin postsecondary education.

Furthermore, Bloom’s OT and RT were both created with the U.S. education system in mind, where student achievement and school improvement are becoming increasingly standards driven (Davidson, 2009; Linn, 2005). A standards-driven environment often requires that terms, concepts, and methods be well defined in order to ensure that educators are held accountable for test results (Davidson, 2009). In order to
ensure the success of a standards driven environment across multiple levels, educators must be able to accurately identify age- or developmentally appropriate levels of achievement of their students through the use of a consistent framework similar to Bloom’s RT.

Based on the work of Anderson and Krathwohl (2001), Bloom’s RT boasts a new, two-dimensional framework that includes the components of knowledge and cognitive processes. This framework establishes a generic and working standard that educators can implement within their teaching practice.

Knowledge

Anderson and Krathwohl (2001) define the term knowledge within four key knowledge domains: factual, conceptual, procedural, and metacognitive.

**Factual** knowledge is defined as discrete or specific units, or bits of information (Anderson et al., 2001). These bits are vital to knowledge acquisition and have inherent value. They are used by experts, as well as educators in general, to communicate information concerning a specific academic discipline, such as subject-related terminology or facts pertaining to a given content area.

**Conceptual** knowledge covers a more complex and organized form of knowledge that typically involves relationships among and within categories and classifications (Anderson et al., 2001). This type of knowledge typically takes the form of general or specific laws, theories, principles, models, structures, or generalizations.

**Procedural** knowledge tackles the understanding of how something works or is carried out (Anderson et al., 2001). This knowledge form often involves a series of steps explicating when, where, or how to accomplish something. An understanding of the
process, as opposed to the product or outcome, is integral to procedural knowledge development.

**Metacognitive** knowledge is the knowledge of or about cognition (Anderson et al., 2001). Specifically, the focus of metacognitive knowledge lies in how to use, recognize, and understand one’s cognitive abilities. Flavell (1979) divided metacognition into three subcategories: strategic knowledge, task-oriented knowledge, and person or self-knowledge. Strategic knowledge refers to an understanding of general strategies for learning, thinking, and problem solving in order to memorize or comprehend material. Task-oriented knowledge allows one to understand how to complete a task and how a given task applies to and interacts with various cognitive learning tools. Finally, person, or self-knowledge refers to understanding oneself. This entails being able to accurately identify one’s strengths, weaknesses, breadth, and depth of knowledge, goals, and beliefs as well as how each of these personal qualities relates to oneself.

**Cognitive Processes**

Anderson and Krathwohl (2001) have written that two of the most important goals of education are promoting the retention of material and being able to transfer or apply that material to new problems and subject matter. According to Anderson and Krathwohl, if these two goals are accomplished, then meaningful learning has taken place. Six cognitive processes (i.e., remembering, understanding, applying, analyzing, evaluating, and creating) describe and elaborate in more detail how these educational goals are accomplished within different stages of cognitive processing.
**Remembering** information is likely the most self-explanatory and least confusing of the six processes (Anderson et al., 2001). It is simply the recognition or recall of information from long-term memory.

**Understanding** forms arguably the largest portion of learning necessary for the promotion of the transfer process (Anderson et al., 2001). It is exemplified through seven subprocesses: interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining. Together, these processes build connections between new and prior knowledge as they are integrated into existing cognitive schemes.

**Applying** is the actual execution and implementation involved in systematically solving a problem, which draws from the contents of procedural knowledge (Anderson et al., 2001). Execution is typically exemplified by the routine procedure of carrying out a task in an almost formulaic way or fixed order. Implementation relates to the application of conceptual knowledge where either no procedure has been developed or the procedural knowledge is in an undiscovered context. This subprocess is seen more as using conceptual knowledge to arrive at a solution in a free-form manner. It works in concert with the understanding and creating processes.

**Analyzing** is the breakdown and relationship of pieces and parts to an overall structure (Anderson et al., 2001). While difficult to define precisely, the analyzing process in fact consists of three subprocesses: differentiating, organizing, and attributing. In essence, these three subprocesses combine to establish relevance and importance, to determine how elements fit into a coherent structure, and to attempt to understand the meaning of the learning material by uncovering the underlying reason or point of view that supports this material. This process is often associated with the discernment of
different facets of an idea while also looking to uncover and understand relationships. Unfortunately, it is still somewhat vague and serves more as an extrapolation of understanding and a precursor to evaluating or creating.

**Evaluating** in essence takes on two self-explanatory forms in checking and critiquing the quality and quantity of work (Anderson et al., 2001). These include testing, detecting, monitoring, coordinating, and judging internal consistencies and are typically compared to external criteria or standards. Overall, this is effective in determining whether something is working properly and how to make improvements if necessary.

Finally, **creating** is the process of putting elements together to form a coherent and functional whole (Anderson et al., 2001). This process involves the generation of possible solutions from existing material either through the development of a workable plan or through the production of something previously planned or designed. The creating process employs many of the other five processes, and although it is seemingly linked to previous learning experiences, it still produces something new and unique.

The two dimensions of **knowledge** and **cognitive processes** form a taxonomy table (Figure 1) which can be used as an analytical tool. This tool is capable of providing educators with a framework through which they can model and examine their pedagogy (Bumen, 2007).

Another key term from Bloom’s RT pertaining to successful learning in the classroom is **alignment**. Alignment deals with the matching of three key elements: instruction and materials, objectives and standards, and tests. According to Anderson and Krathwohl (2001), when these elements are found in the same or proximal cells within the taxonomy table, a strong alignment is considered present. The supposition is that a
greater alignment of these elements will lead to more effective student learning (Bumen, 2007). Results of this study done by Bumen (2007) indicate that many teacher candidates wish to have their teaching practice strongly align with items within Bloom’s taxonomy table. It was suggested that through the alignment of instruction and materials, objectives and standards, and tests, candidates have an opportunity to further develop their critical and reflective thinking skills. At the same time, candidates are able to identify and follow up on missed teaching and learning opportunities that could lead to an improved and more well-rounded practice.

What Bloom has managed to produce is a useful body of work that binds theories of knowledge and thinking processes with practice, in a manner that is practical for educators to use. In fact, Stephen (2010) noted that Bloom’s OT and his foray into the cognitive domain were also designed with the thought of sharing a collective testbank of multiple-choice questions amongst various universities to measure specific objectives as indicated by a standard vocabulary in mind. The usefulness of Bloom’s work is supported by the research of Larkin and Burton (2008) for example, who found that long-lasting improvements to nursing practices were produced with the support and use of Bloom’s RT. Overall, Bloom’s work can be applied in a purposeful, logical, and scientific manner towards successful learning as well as to varying forms and levels of assessment.

Summary

This review of literature was undertaken in order to map out the historical and academic context within which multiple-choice assessment practices reside. Throughout this review, three important aspects of multiple-choice assessment were identified: (a) its
past, present, and future role within educational assessment; (b) its strengths and weaknesses; and (c) how it has been and can continue to be improved and modified in order to achieve targeted and beneficial learning. This review of literature also summarized relevant components of educational psychology (Bloom’s taxonomy), providing the tools and context from which assessment operates.
<table>
<thead>
<tr>
<th>Knowledge Type</th>
<th>Remembering</th>
<th>Understanding</th>
<th>Applying</th>
<th>Analyzing</th>
<th>Evaluating</th>
<th>Creating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factual knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conceptual knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Procedural knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacognitive knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1. Bloom’s revised taxonomy.*
CHAPTER THREE: RESEARCH METHODOLOGY

This chapter provides an explanation of the research design, selection of site and participants, instrumentation, data collection, data analysis, limitations, and ethical considerations. In this study, teachers interested in improving their ability to assess students through the use of a multiple-choice format were asked how they both approached and used multiple-choice assessment. The teacher-participants were involved in focus group interviews and a workshop which was delivered in order to provide them with an improved understanding of and potential applications for multiple-choice assessment. Pre- and post- interviews were conducted in order to gain as in-depth an understanding as possible of teachers’ multiple-choice perceptions, practices, and receptiveness to training. Analysis of the transcripts from the focus group interviews and survey responses revealed how educators perceive and approach the use of multiple-choice tasks in achieving their assessment goals.

Research Design

This research was undertaken to explore and understand the current realities of teaching and the perceptions and uses of multiple-choice assessment within secondary schools. Current teaching constraints surrounding multiple-choice testing were examined as well as the ways in which multiple-choice assessment training can benefit teachers in a practical way. Furthermore, the manner in which teachers respond to and view multiple-choice training and its potential uses were investigated.

Methodological Framework

This research was conducted using a qualitative-descriptive methodology that is frequently employed within the field of education and is emerging in its use among
nursing researchers (Caelli, Ray, & Mill, 2003). A qualitative methodology is considered to be somewhat amorphous in its naming across varied studies and disciplines. It has been referred to as “naturalistic inquiry,” “noncategorical qualitative research,” the “fundamental qualitative method,” and/or the “generic qualitative method,” to name a few (Sandelowski, 2000; Thorne, Kirkham, & MacDonald-Emes, 1997), but will be referred to in this study as the generic qualitative method or generic qualitative research.

Traditionally, academic views regarding this descriptive genre of qualitative research have often claimed that this methodology is an unreliable, disrespected, and weak form of research (Sandelowski, 2000). Opposing this, researchers within the field of nursing have argued that, in order to most effectively and accurately address an area of interest, a distinct form of inquiry that better aligns with a particular field’s philosophical and disciplinary objectives outside of the restrictions of traditional methodologies is sometimes needed (Thorne et al., 1997). Furthermore, Thorne et al. (1997) postulated that from a nursing point of view, this descriptive form of research is one of the best ways in which nurses could most effectively discover and implement knowledge that was directly applicable to their practice. Similarly, my study also benefited from being able to work outside of the traditional methodological confines of empirically based research by applying contextualized and inductive reasoning to a combined accumulation of theory (i.e., existing multiple-choice assessment research) and experiential knowledge (i.e., teachers’ experiences), which both logically and realistically resonates with educators today.

While generic qualitative studies conducted in the field of education have characteristically drawn from various theories within the areas of psychology and
sociology (Caelli et al., 2003), the methodology and design in my study more closely align with the reasoning behind emerging nursing studies, wherein methodologies and ideologies within similar and well-known qualitative designs (e.g., ethnography, action research, phenomenology, grounded theory, case study, and narrative inquiry) have formed their methodological framework (Caelli et al., 2003; Sandelowski, 2000; Thorne et al., 1997; Thorne, Kirkham, & O’Flynn Magee, 2004). According to Sandelowski (2000), generic qualitative methods have generally been seen as flexible and reasonable combinations of samples, group interviews, analyses, and data collections. In addition, the precise outcomes of analytical rigour found in traditional quantitative and qualitative studies may have also become unrecognizable or absent. The generic qualitative method, however, applies instead its analytical focus in a manner that attempts to fully engage the processes of inductive reasoning. It seeks to test and challenge existing interpretations and knowledge presented in order to ultimately shed new light on areas of interest in both a coherent and a meaningful manner (Thorne et al., 2004).

This study was undertaken in a qualitative manner for two main reasons, both of which are noted in Caelli et al. (2003):

1. Time limitations tied to simultaneous coursework along with limited access to research subjects played a large role by limiting the type and manner of research that I was able to conduct.

2. The exploration of research questions designed to pursue intriguing, salient, and perhaps unseen and unexpected branches of thought which ultimately lead to the benefit of both participants and researcher require, the freedom and flexibility inherent in this method.
Methodological Implementation

In order to interact meaningfully with and advise teachers in regards to the current reality and perceptions of multiple-choice assessment, this study gathered existing empirical and research literature surrounding the uses of the multiple-choice format across multiple levels of education. With this research, a framework was formed that would be more easily accessible to and understood by secondary school teachers. Of particular relevance were guidelines drawn from thorough and noted works done by Thomas Haladyna on how to reduce flaws in multiple-choice assessment and improve the efficacy of this form of assessment (i.e., Haladyna, 1992, 2004; Haladyna et al., 2002).

Prior to the presentation of findings from the research literature, a group interview was held with each of the groups of teachers in order to assess teachers’ experiences with and perceptions of multiple-choice assessment. The focus of the first round of interviews was to generate opinions and discussion pertaining to teachers’ current perceptions, use, and past training experiences in relation to multiple-choice assessment. In particular, these interviews attempted to distill the perceived strengths and weaknesses of the multiple-choice format and how teachers had or had not been using this format to its potential. In addition, a brief survey was administered to participants in order to collect profile information and brief descriptive statistics regarding participants’ perspectives, descriptions, and development of various aspects of multiple-choice assessment.

Upon completion of the interviews, points of interest as raised by teachers from these interviews (e.g., EQAO literacy testing) were further researched and combined with the research literature to be presented in a workshop to show the various ways in which teachers could improve their use of multiple-choice questions or implement these types of
questions in new and effective ways. This workshop was intended to provide practical and relevant information to teacher-participants and, was for many of them, the first time they had been involved in multiple-choice assessment training. Consequently, much of the focus of this training session provided teachers with a basic introduction to how the multiple-choice format could be implemented while addressing the questions that teachers had on multiple-choice assessment. Had there been more time, it would have been ideal to provide teachers with hands-on instruction on how to effectively construct multiple-choice questions.

Six to 12 weeks following the workshops, a follow-up set of interviews were conducted in order to assess the impact that the workshop had on teachers’ practices and perceptions of multiple-choice assessment and how teachers might approach future opportunities for training in relation to multiple-choice assessment.

In essence, this study attempted to explore some of the current perceptions held by a few secondary teachers towards multiple-choice assessment. This study also attempted to determine the potential utility of a multiple-choice assessment workshop through the data collected from the pre- and postworkshop interviews and the short survey. Interviews were designed to garner insight about teachers’ perceptions of multiple-choice assessment. Surveys were also used to provide concise data, which helped to supplement findings gathered during interview discussions. However, it is important to note that in no way were any findings of this study intended to be exhaustive.
Selection of Site and Participants

Before any potential schools or participants were contacted, ethical clearance was requested and granted from the southern Ontario school board where the study was to be conducted. Following this, clearance for conducting this research was obtained from the Brock University Research Ethics Board (10-131-ENGEMANN). Letters were then sent to all of the secondary school principals of this school board providing a brief explanation of the study and an offer to meet and discuss the details of this research project. Three principals responded with interest, and after meeting with them or an appointed contact in person to go over a more detailed overview of their involvement, each of these schools committed a group of teachers to take part in this study.

Teachers were then contacted either by their principals or by me, via email, and invited to participate in the study. During the preworkshop interviews, all participating teachers were given an informed consent form and were asked to read and sign this form. At this point, five teachers from the first school, seven teachers from the second school, and 16 teachers from the third school confirmed their participation, for a total of 28 participants. Participants had a wide range of teaching experience, taught various subjects across different levels, and were looking to improve their assessment practice in relation to their use of the multiple-choice format. Interviews, workshops, and any other in-person engagements were conducted on site at the participating teachers’ home schools.

Instrumentation

Data for this study were collected through the use of focus group interviews and a survey. The following section describes the instruments used within this study.
The focus group interviews provided the majority of the data and the richest data used in this study. Two sets of questions (Appendix A), one for teachers to respond to prior to the multiple-choice workshop and one to follow the workshop, were designed to elicit discussion and to probe more deeply into the processes, opinions, and perceptions that teachers held in relation to their use of multiple-choice questions. The profile of each participant was also gathered through a survey (Appendix B). This survey collected basic information to support interview findings and to provide a brief overview of each participant’s subject areas and grade levels taught as well as the number of years of teaching experience each participant possessed. Table 1 features a summary of participants’ information. The survey further asked teachers to identify their perceptions, experiences, and attitudes towards multiple-choice questioning.

The survey questions sought to ascertain teachers’ prior assessment training experience, their feelings towards multiple-choice questions, their perceived accuracy in assessing students, and the frequency with which they used multiple-choice questions in their classrooms. Together, these provided information on the background and intent of educators with respect to their use of multiple-choice questions.

**Data Collection**

Transcriptions of focus group interviews from audio recordings of these sessions along with the aforementioned survey instrument served as the primary sources of data for this study. Initially, meeting requests were sent out to participating teachers via email regarding the first set of interviews. Interview times were established, and all participants from a single school were interviewed at the same time. Each focus group interview
Table 1

*Participants*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Sex</th>
<th>Grades taught</th>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>M</td>
<td>KG–13</td>
<td>English, Family Studies</td>
</tr>
<tr>
<td>A2</td>
<td>F</td>
<td>6, 7, 9, 10, 11, 12</td>
<td>Religion</td>
</tr>
<tr>
<td>A3</td>
<td>F</td>
<td>10, 11</td>
<td>Religion</td>
</tr>
<tr>
<td>A4</td>
<td>F</td>
<td>9, 11, 12</td>
<td>Religion, Social Science</td>
</tr>
<tr>
<td>A5</td>
<td>F</td>
<td>9, 10, 11, 12</td>
<td>Visual Arts</td>
</tr>
<tr>
<td>B1</td>
<td>M</td>
<td>9, 10, 11, 12</td>
<td>Math, Special Education</td>
</tr>
<tr>
<td>B2</td>
<td>F</td>
<td>9, 10, 11, 12</td>
<td>Chemistry</td>
</tr>
<tr>
<td>B3</td>
<td>M</td>
<td>9, 10, 11</td>
<td>Math</td>
</tr>
<tr>
<td>B4</td>
<td>F</td>
<td>9, 10, 11, 12</td>
<td>Math</td>
</tr>
<tr>
<td>B5</td>
<td>F</td>
<td>9, 10, 11, 12</td>
<td>Biology, Science</td>
</tr>
<tr>
<td>B6</td>
<td>F</td>
<td>9, 10, 11, 12</td>
<td>Science</td>
</tr>
<tr>
<td>B7</td>
<td>F</td>
<td>9, 10, 11, 12</td>
<td>English</td>
</tr>
<tr>
<td>C1</td>
<td>F</td>
<td>9, 10, 11, 12</td>
<td>Math</td>
</tr>
<tr>
<td>C2</td>
<td>M</td>
<td>9, 10, 11, 12</td>
<td>English</td>
</tr>
<tr>
<td>C3</td>
<td>F</td>
<td>9, 12</td>
<td>Geography, Social Sciences</td>
</tr>
<tr>
<td>C4</td>
<td>M</td>
<td>11, 12</td>
<td>Social Science</td>
</tr>
<tr>
<td>C5</td>
<td>M</td>
<td>3, 6–12</td>
<td>Special Education</td>
</tr>
<tr>
<td>C6</td>
<td>F</td>
<td>9, 10</td>
<td>Social Sciences</td>
</tr>
<tr>
<td>C7</td>
<td>M</td>
<td>9, 10, 11, 12</td>
<td>Physical Education</td>
</tr>
<tr>
<td>C8</td>
<td>F</td>
<td>SK, 2, 3 8–12</td>
<td>Special Education</td>
</tr>
<tr>
<td>C9</td>
<td>F</td>
<td>9, 10, 11, 12</td>
<td>English, Religion</td>
</tr>
<tr>
<td>C10</td>
<td>F</td>
<td>10, 11, 12</td>
<td>Art</td>
</tr>
<tr>
<td>C11</td>
<td>F</td>
<td>11, 12</td>
<td>Science</td>
</tr>
<tr>
<td>C12</td>
<td>F</td>
<td>9, 10, 11</td>
<td>English</td>
</tr>
<tr>
<td>C13</td>
<td>F</td>
<td>10, 11, 12</td>
<td>Religion</td>
</tr>
<tr>
<td>C14</td>
<td>F</td>
<td>10, 11</td>
<td>Science</td>
</tr>
<tr>
<td>C15</td>
<td>M</td>
<td>9, 10, 11, 12</td>
<td>Technology</td>
</tr>
<tr>
<td>C16</td>
<td>F</td>
<td>9, 10, 11, 12</td>
<td>Math</td>
</tr>
</tbody>
</table>
lasted approximately one hour and consisted of teachers being asked to openly discuss their opinions and experiences in response to questions prepared as presented in Appendix A. Teachers were encouraged to engage the researchers and each other as freely as they would like in order to foster a genuine and in-depth discussion. Interviews were recorded and transcribed, thus forming the foundation of this study’s analysis. Surveys were also distributed during the interviews, and participants took approximately 15 minutes to complete and return them.

Interviews were conducted both before and after the workshops, with a follow-up round of interviews taking place at least 6 weeks following the workshop. One of the goals of the postworkshop data collection was to observe whether teachers’ perceptions, attitudes, and/or pedagogy towards multiple-choice assessment had changed as a result of the workshop training received and, if so, to what extent.

**Data Analysis**

As mentioned above, the primary source of data for this study were the recorded interviews. The interview transcripts were analyzed, and themes were identified and organized into relevant topics related to teachers’ perceptions of multiple-choice assessment as well as how this form of assessment is implemented in their classrooms. The focus of the postworkshop interviews was also to determine teachers’ own assessments of the workshops on the multiple-choice format. Through the use of QSR International’s NVivo 9 software, these transcripts were tagged and coded into nodes based on inductive analysis, bringing forward variances, commonalities, and salient data in relation to multiple-choice assessment. Creswell (2005) noted that thematic data analysis can help distill pertinent features of the data, allowing for greater insight and
detail into participants’ thoughts and activities. Subsequently, major points and themes were drawn from the analysis of the transcripts in order to organize and summarize responses. Survey data were analyzed and descriptive statistics compiled from survey questions related to the themes present in the analysis. Findings from this were used to supplement and support points drawn from the interviews. The overall process of analysis was kept as simple as possible in order to allow opportunities for themes to emerge during the data analysis process.

**Limitations**

While this study was straightforward in its conception and execution, several methodological limitations nonetheless presented themselves.

The first limitation involved participants’ responses and interpretations of multiple-choice assessment. Responses were taken at face value without other corroborating data, thereby limiting this study in terms of the verifiable nature of the data collected from participants.

Furthermore, the sample size and its representative teacher population were also limited to three schools within a single southern Ontario school board. This led to inconsistencies in the nature and culture of each school, including the number of volunteer participants from each of the three schools (5, 7, and 16), participants’ profiles (i.e., age, experience, subjects taught), and the reasons for which these schools chose to participate in this study.

In addition, certain school administrators were much more involved with this study, demonstrating a genuine and invested interest in this research, while other administrators were minimally involved with this study and its results. Most participants
were also teachers who were willing to learn about multiple-choice assessment, because participants were selected on a volunteer basis. As a result, the findings could have been more conclusive, had a larger and more consistent sample from numerous schools across several school boards been obtained.

Another limitation that should be taken into consideration is researcher bias. This type of investigative study involved the integration of many qualitative aspects over which I had control, particularly when conducting the surveys, leading the interviews, and interpreting the results. My perspectives and experiences as a lifelong student, new teacher, and researcher passionate about effective assessment (and multiple-choice assessment in particular) have likely shaped my approach to many areas of this study. To minimize researcher bias, researchers can self-reflect upon their actions and decisions, become aware of their own assumptions, and become more transparent in their presumptions (Osborne, 1990). While this study is far from perfect, I spent much time reworking my interview questions, attempting to read a diverse collection of multiple-choice literature, and bouncing ideas off my thesis supervisor in an attempt to minimize my own bias. My thesis supervisor was on hand to guide and support me and to provide feedback throughout the entire study. Educators should ensure that questions, tools, and results are as objectively and accurately used as possible.

**Ethical Considerations**

The research for this study was conducted in a manner that was in no way harmful or deceptive towards participating teachers. Participants were selected on a voluntary basis, were informed of the details of the study, and were only subject to a short survey, two 60-minute group interviews, and a single day, 2–3 hour workshop held at their
schools. Interview questions did not probe into personal matters and were always pertinent to the topic of multiple-choice assessment. All teachers read and completed the informed consent form and were thus aware of their right to withdraw from this study at any time. Local secondary school administrators and officials were also made aware of this study. Teachers were also informed that their participation would be kept confidential, that the data reported would remain anonymous, and that the raw data would be destroyed following the completion of the study.

As previously mentioned, the study was also reviewed and given clearance from both the school board’s research advisory committee and the Brock University Research Ethics Board (10-131-ENGEMANN).

**Summary**

The purpose of this study was to explore the current perceptions and teaching practices involving the use of multiple-choice questions within a secondary school context. In addition, this study investigated the effects of a professional development workshop on multiple-choice assessment on teachers’ perceptions of this form of assessment. The workshop was based on Haladyna’s development of multiple-choice guidelines in combination with Bloom’s work on the identification of targetable levels of thinking and learning. To gain an understanding of the current levels of multiple-choice assessment, data were collected and compiled from focus group interviews and surveys over a period of approximately four months.

The results, analyses, and implications of this study are presented throughout the remaining chapters.
CHAPTER FOUR: PRESENTATION OF RESULTS

Multiple-choice assessment has played a prominent role in the culture of North American secondary and postsecondary education. The manner in which this assessment format is used by educators influences the way in which students prepare for and respond to assessments (Parmenter, 2009). The main thrust of this study was to explore teachers’ perceptions of multiple-choice assessment through the use of two in-depth focus group interviews and a workshop on this form of assessment. At the same time, this study hoped to further inform educators of the function and potential of utilizing multiple-choice assessment in order to maximize the benefits of this format both for students and for themselves.

Participants in this study were secondary school teachers from three different schools within the same southern Ontario school board. Participants were interviewed in regard to their perspectives surrounding various aspects of the multiple-choice format. Teachers were interviewed twice, once prior to and once following a workshop that provided them with information on how to effectively approach the development and use of multiple-choice questions within their assessment practice. The research presented was based primarily on work undertaken by Thomas Haladyna (e.g., Haladyna, 1992, 2004; Haladyna et al., 2002). Focus group interviews from each participating school were transcribed, analyzed, and compared in order to investigate thematic similarities amongst teacher responses. Interviews were conducted using predetermined questions and were constructed in a manner in which participants’ responses reflected the strengths, weaknesses, and experiences of educators in relation to different aspects of the multiple-
choice format. In addition, each participating teacher completed a survey that was used to build his/her profile and to help supplement interview data.

Although this study gathered data for the benefit of future educators, it was also conducted for the immediate benefit of the participating teachers.

Findings

The participants in this study represented a diverse group of educators who ranged in age, teaching experience, and subject areas taught. Their responses during the interviews are reported in this chapter. These findings have been organized under three main headings that assist in answering each research question respectively.

Secondary School Teachers’ Experiences and Perspectives on Assessment Training

This section of the findings outlines the experiences and perspectives of the participants in terms of the training that they had or had not received on multiple-choice assessment and the training that they wished to receive.

Training received. In order to gauge the current and realistic depth of knowledge and understanding that teachers possessed, teacher-participants were asked to comment on their former experience and training in relation to the multiple-choice format and to assessment in general. While a number of survey respondents (68%) indicated that they had received some sort of general assessment training prior to this study, only 18% of respondents indicated that they felt at ease in creating and implementing multiple-choice questions much of the time. Most respondents (82%) indicated that they felt multiple-choice assessments were easy to create only some of the time. In general, interview and survey responses indicated that teacher-participants had a lack of
confidence in both the educational training they received surrounding multiple-choice assessment and in how to create good multiple-choice questions.

In response to general assessment training queries, there was a wide range in teachers’ responses. Teacher A5 stated, “I don’t think I learned any of this in teacher’s college . . . honestly!” On the other hand, teacher A3 commented, “I feel from my teacher’s college program there was huge emphasis on assessment and evaluation. So I felt very confident leaving the program, being very well-trained in that area.” Furthermore, teacher A2 exclaimed, “I don’t know if anyone else did, but I did have an inservice at least once, possibly twice, on how to write multiple-choice questions . . . and now I’m back! And I still don’t know how to write them!”

For the most part, responses indicated that teachers had received limited training, prior to the workshop that was given as part of this study in relation to assessment in general. Teacher C6 flatly stated, “We were never shown how to write a test.” Teacher B1 also asserted, “I’ve received no formal or informal training of multiple-choice question writing. Most of it is just from experience in being a teacher, but like I said, I couldn’t write a good question to save my life.” Teacher C6 similarly stated, “But they never actually—until I was on my first in-school, was I actually asked to—ok, put together a test however you want. So it’s like, Alright!” These comments support the belief of some teachers that almost everything useful learned in the field of teaching is gained through practical, hands-on experience rather than during formal learning opportunities (e.g., lectures, coursework).

While most teachers agreed that they were not sufficiently prepared to successfully implement any form of multiple-choice assessment, survey responses
indicated that every teacher did in fact use multiple-choice questions as assessment between one and six times per semester, with an average of 3.5 times per semester in at least one of their subject areas taught. When asked about their experience with multiple-choice training, teachers’ responses indicated they had received little explicit training. As teacher A4 stated,

For me, multiple-choice was not a topic that we covered in teacher’s college. And I still have not had much training in it. So I do not feel comfortable creating them on my own. And I’m sure I’m making big mistakes.

Teacher C4 further iterated,

I don’t think – I know I wasn’t at all prepared. I was never shown. I don’t know about anyone else, but I was never shown how to write a multiple-choice question. So that’s something. . . . I’ve been teaching 11 years now, and never been shown.

Similarly, teacher A1 indicated, “I don’t recall much training. I got some of it but I found most of it on the job using previous exams and going through them and constructing them following a template that was set out.” Teacher A1 further echoed the importance of on-the-job training, given that testbanks, previous exams, and the highly contextualized training, such as is occasionally involved in the preparation of the Ontario Secondary School Literacy Test (OSSLT), seem to have served as the only forms of support for teachers in creating their multiple-choice questions. As teacher A1 pointed out, while having undergone courses and/or training related to assessment,

I’ve also learned a great deal from my peers and one of our methods is to use previous exams. That’s where I’ve mostly learned my use of multiple-choice. . . .
Yeah, [that gives me] a template to follow. So my colleagues have been terrific in just finding last year’s exam or a few years’ exams . . . and then I go over them and say “Okay, how can I change it for this year?”

The limited amount of available resources addressing this specific matter suggests that testbanks have played an invaluable role serving as a template for the creation of multiple-choice questions. As teacher C6 stated,

Now we have testbanks, and they are pretty good . . . because you can go from a government test bank. . . . But I do find depending on the level you’re teaching, you really have to sometimes alter the questions, or the answers, or both.

In response to whether teachers felt prepared to use multiple-choice assessment, teacher B5 stated,

I think the most prep I’ve ever gotten is when we had to teach them [students] about the literacy test. . . . All [of] us Grade 10 teachers were brought out of class and [had] gone through workshops where we had to – where we learned what proper multiple-choice questions are supposed to look like . . . and [then] create them ourselves and submit them. But before that . . . [there was nothing].

Similarly, teacher A7 responded,

Sufficiently prepared? No. What did prepare me was the literacy review. . . . We actually sat down as a staff and took the test – the Grade 10 teachers. And then we talked about the answers . . . and I didn’t do very well.

Teacher A3 further agreed, “The after school literacy program this past couple weeks has taught me a lot about multiple-choice, so I feel a little bit prepared. I would say 20% [to] 25%.” This indicated that teachers did not feel adequately prepared, and that even when
presented with a small opportunity for training, they participated, and felt like they had
gained much knowledge and in turn felt better prepared.

Finally, the comment that may best summarize teachers’ opinions and the reality
of their experiences surrounding multiple-choice assessment was made by teacher C14,
who stated,

Half of us have never really learned like . . . that there’s [even] certain parts [that
go into creating multiple-choice questions]. [So] to say that I’ve actually been
instructed on how to make one properly, I’ve got to say, “No.” With so much at
stake for kids, especially knowing where they’re heading—in any kind of
postsecondary [education], I think it’s a big, big thing that we need to take steps
for.

This statement seemingly reflects that teachers recognized the importance of
multiple-choice assessment, yet many knew little about how to develop or implement it
appropriately. In general, teachers’ statements appear to suggest that this lack of
knowledge was because they had received a limited and less than ideal amount and level
of assessment training prior to and during their teaching experience. A subsequent lack
of confidence and preparedness in approaching the multiple-choice format was shared
among most teachers as well as a pervasive foreignness to multiple-choice assessment
training. In spite of these obstacles, teachers stated that they continued to consistently
use multiple-choice questions, indicating that they had come to rely on other self-taught
methods (i.e., testbanks, informal test sharing) in order to help them figure out how to
implement multiple-choice questions. As a result, teachers candidly indicated
disappointment in their lack of multiple-choice training.
Training desired. In light of this lack of training, teachers frequently expressed their desire for further preparation within different areas of multiple-choice assessment. For example, the desire to learn how to create good multiple-choice questions by capitalizing on the varying question formats, phrasings, and structures of this type of assessment was expressed by many teachers. Some were interested in learning how to control the level of difficulty of a multiple-choice test. Teacher B2 asked, “I would like to know how tricky to make them? Like what’s appropriate? And what’s too easy?” Other teacher-participants commonly requested training revolving around the development of questions targeting higher levels of cognition.

I’d also like to use higher level thinking in my multiple-choice questions. (A2)

Yea me too. That’s the main thing. (A5)

So we use them for knowledge. How can we use them for thinking? (A1)

Teacher B5 similarly stated,

. . . I know that I haven’t really changed the type of questions . . . so they’re still pretty much knowledge, [and] understanding. That’s something I still really struggle with. So if there is going to be any more sort of professional development, it would be trying to make those sorts of questions.

Teacher C1 further affirmed these sentiments: “So I still don’t have a good concept of what a good math question is that addresses more cognitive level(s) than just simple understanding.”

Teachers were also interested in creating questions that might fit into the appropriate learning categories of knowledge, thinking/inquiry, application, and communication as defined by the Ministry of Education of Ontario. As teacher A2 points
I use short answers for application and communication and usually thinking I put into essay form. . . . But usually multiple-choice is just for knowledge. But I would love to use multiple-choice questions for others—like for application [and] communication.

Alongside these cognitive and categorical targets, subject-specific queries were also raised. Teacher B5 stated, “I wouldn’t mind seeing ways to incorporate the different types of categories into [multiple-choice questions]. Again, just in science, more of the thinking and application types of questions for multiple-choice.” Teacher B3 similarly maintained,

you know what I’d like to do is, since we’re all different subjects, we can’t do it, but if we just had a couple math teachers like four or five, stick to like a Grade 10 curriculum and be like, “Hey, let’s create some multiple-choice.” Like, I want some hands-on, let’s do it, let’s talk about it.

Along the same lines, teacher C15 also claimed,

Based on the diversity of the subject areas in our department I would love the opportunity to sit down with other transportation teachers and create a bank of questions. . . . So I mean, anytime we could get the opportunity to . . . which would be subject-specific questions development-based, and coordinate things with other similar-subject teachers . . . I would love to do that.

Teacher C13 further iterated,

or even just more examples from you guys of what you sort of meant. Because we saw a few in terms of how you can sort of up the cognitive level of questions but
they weren’t necessarily applicable to all subjects. So maybe seeing one or two for each different subject so you [would] have more of an idea [of] what you can sort of do with your course content would be awesome.

According to teacher A1, “We’re [teachers are] always talking it through with one another, how did you do with this student . . . and it’s always something we wonder about – am I doing the best job evaluating?” Based on this statement, the need for teacher collaboration seems not only warranted, but sought after by teachers. In the past, teachers may have relied upon informal collaborative efforts to generate certain assessments, but now they indicated a desire for productive and purposeful subject-specific, collaborative work opportunities. This was particularly evident in discussing the development of multiple-choice assessment.

A slightly unexpected, though pleasant, request came from teacher B4, who commented,

I’d like to learn . . . what type of multiple-choice question is appropriate for any given level, and maybe the types of students that you’re dealing with. So if you have a classroom full of IEP students, what kind of multiple-choice questions can they be successful at answering?”

Further, upon considering potential ideas that could be incorporated into all of the differentiated instruction training teachers had received, teacher C2 stated,

And also, for further professional development, I would like to know how we can use this a little bit—like, that’s just one idea how we can incorporate multiple-choice into differentiated instruction. I’d like to learn a little bit more about that . . . maybe there are other ways we can implement multiple-choice into
differentiated instruction and if it can work in any way because that’s a big push right now—is DI.

Overall, teachers were honest with their questions and responses and frequently questioned the validity of varying multiple-choice formats. “But what [do] they do? Like you know, a, b, c, d, b and c, a and c, all of the above, none of the above . . . are those valid?” asked teacher A5. Similarly, teacher C1 asked,

What I’d like to know . . . can it be a question? “Which one is the most popular . . . ?” or whatever it is. Or [can it] have a question mark in the stem? Is that acceptable in multiple-choice questions? Or should it be a statement where they [the students] finish the sentence with one of the choices?

Overall, teachers indicated that while being insufficiently prepared to use multiple-choice assessment, their concern for students and their desire to learn about the components of multiple-choice questions that would be most or least effective in assessing students remained evident.

Secondary School Teachers’ Perceptions of Multiple-Choice Assessment

To help understand the current and potential use of the multiple-choice format in secondary school classrooms, determining the manner in which teachers considered and approached this form of assessment was crucial. Teachers’ perceived strengths and weaknesses of the multiple-choice format were diverse and well represented.

Several questions from the survey addressed teachers’ perceptions of multiple-choice assessment. When asked to describe their opinions and feelings towards multiple-choice assessment using at least three adjectives, teachers most frequently used adjectives that connoted some sort of difficulty (e.g., difficult, challenging, tricky, confusing, wordy)
During interviews, teachers indicated that these difficulties were most frequent when designing multiple-choice questions. Interestingly, the second most frequently used adjective from the survey was “easy” (~12%). Since the use of the adjectives “easy” and “difficult” referred to the same question on the survey, it is apparent that teachers were viewing the use of multiple-choice assessment from different vantage points. Furthermore, the use of the multiple-choice format was also supported by a number of assessment-positive descriptors (e.g., reasonable, fair, valid, useful, informative), which altogether made up roughly 8% of the reported words. Overall, interview discussions paralleled and supported the results of the survey, as the majority of teachers were typically in agreement on both the strengths and weaknesses of the multiple-choice method. Virtually all participants were in favour of, and interested in, the potential benefits that the multiple-choice format could offer to their assessment practice.

**Strengths.** Results of this study indicated that the multiple-choice format presents several distinct pedagogical advantages. Teachers frequently mentioned the fact that multiple-choice tests are easy to mark. Unsurprisingly, teachers also referenced the benefit of using Scantron cards: “I like multiple-choice because the Scantron does all the marking for me,” exclaimed teacher B1. In response to the various ways in which the multiple-choice format was seen as a strength, teacher A1 stated, “Well the first way is that they [can] mark it with the Scantron cards. They’re much easier to mark.”

Another frequently cited strength of the multiple-choice format was its intrinsically objective grading nature. As teacher A1 indicated, “I love it [multiple-choice] for English because it’s so objective.” Though seldom disputed, at the very least,
multiple-choice questions remove a considerable amount of subjectivity from the grading process, rendering it a much safer and reliable format through which teachers are able to compare student grades. Teacher A2 did her best to explain a teacher’s thought process:

And maybe there’s—maybe the multiple-choice is more objective. There’s less room for subjective evaluation. Because maybe . . . I understand what they’re [students are] saying even if they don’t say it, because you know what they’re trying to explain. But really they probably shouldn’t get the marks for that [something they did not say]. But because you [as a teacher] know what they’re talking about or [are] trying to express, you might give it [the marks] to them. So, multiple-choice is definitely more objective.

As teacher B7 conclusively pointed out, “There’s no way you can’t be consistent!”

Similarly, teacher B5 explained the objective benefits of multiple-choice as they relate to standardized testing:

I think for large-scale, like if . . . you’re using it mostly for an exam, it’s a lot harder for that sort of subjective marking so the kids . . . you know some teachers will mark a lot harder for a long answer, but another teacher [might not]. But if it’s mostly multiple-choice and you’re assessing whether they know it, it’s impossible for the hard teacher to mark harder compared to the other one, so it’s . . . it’s a little more of an even playing field I think.

One teacher also alluded to the fact that alongside the format’s objective nature in marking is the simplicity, clarity, and reliability of multiple-choice assessment, particularly in determining students’ prior knowledge of a topic. As teacher A2 stated,
I do like multiple-choice. I think it is valid, and I think it’s a valid indicator, and I think it is very reliable if the students have studied information. Because my questions are knowledge based, if they know the information, then they’re gonna know the answer. They should know the answer . . . but I think it is a pretty good indicator if they know the information.

In interviews, several teachers were quick to point out how efficient this form of assessment can actually be. Survey results similarly indicated that at least six teachers included the words “quick” or “efficient” when first prompted to describe their thoughts on multiple-choice assessment. Several teachers also indicated that they had already taken advantage of these “traits” and were thus able to capture student learning and progress at opportune times. “I can tell straight away if they don’t get the right answer, [and] whether or not they’ve read up until the point I need them to read,” stated teacher B7. Teacher B5 similarly explained,

for quizzes, I think they’re great! Just because there’s such a quick turnaround that you can get maybe not the best understanding of whether the kids get it, but you [can] get a pretty good idea . . . sometimes even by the end of the period [if] it’s a quick eight-question multiple-choice. If they’re doing seat work, quickly mark it . . . and pass them back and that way you know if . . . you need to readdress what you just taught them.

Along these same lines, teachers also commented on the fact that compared to more open-ended response type questions, multiple-choice was simply more effective in identifying whether or not students understood particular concepts. Teacher B2 observed the following:
I’ve noticed in math especially if you gave the same five questions as short answers that . . . were awarded part marks, students would do better on the short answer than they would on the multiple-choice . . . [from] showing some of the steps. [They would receive] part marks as opposed to just [being] right or wrong. Likewise, teacher B5 also indicated that “some multiple-choice questions narrow in on specifics, whereas [for] the short answers, they can give some broader terms and the main ideas and [still] get part marks for [them].” In addition, teacher A2 explained her reasoning in the following manner:

I like it because every area is covered. Because when I make my multiple-choice questions, I have four questions from this unit, four questions from this unit, and I make sure it’s balanced. So everything is covered there, and all the knowledge part is covered. So that’s kind of why I kind of like it. Because it gives me structure . . . and I know all the important topics in each unit are covered in the multiple-choice questions.”

Teachers also indicated several other multiple-choice characteristics that they favoured or found useful. For instance, teacher A4 stated, “with the open courses, I like the structure of multiple-choice. I think that that’s something that the students appreciate about it.” Additionally, when asked whether the multiple-choice format was valid, teacher A7 emphatically stated that the multiple-choice format had been an excellent indicator of overall student performance:

The same students that have really read the novel, they’ve done other things leading up to that. They have been working on questions and they have been working on essays. . . . So when it comes [to the multiple-choice section], the
ones who really succeed are the ones who are [already] doing very well in all the other areas. So when they succeed in the content of multiple-choice, they have also succeeded in essay writing. . . . Okay yes, that confirms that you’re a very bright student.

In the end, most teachers were in agreement in regard to the strengths of multiple-choice assessment and indicated that they would either begin or continue to take advantage of these strengths in their own practice. Altogether, the multiple-choice format was considered capable of providing good value for the amount of time and resources spent in creating and grading questions. As teacher C14 pointed out,

That is the nice thing . . . for a lot of us, for sure, for final exams and that sort of thing too. It’s a lot of bang for your buck. You put a lot [of work] into them, but to get through things without having tedious marking to do—as far as that goes, [it’s great]!

**Weaknesses.** As many of the strengths of the multiple-choice format were easily agreed upon and obvious to teachers, the weaknesses of this assessment format were just as clearly evident. While survey results indicated that positive language dominated the majority of responses in relation to teachers’ feelings and perspectives about multiple-choice questions, negative descriptive words such as “confusing,” “wordy,” and “unfair” were also present and frequently mentioned during interview discussions.

There were some small problems, such as the cost incurred in printing many pages, over which teachers expressed concern with regards to the multiple-choice format. “They take up a lot of paper, and we’re charged per page in our department, . . . and on exams they take up a lot of room. You can only fit like five on a page!” stated teacher
A5. However, the two major issues that dominated teachers’ opinions were the initial time consumption that is required to create multiple-choice questions and the constraints that are part of using a selected-response structure.

As teachers became increasingly informed throughout the workshop and interview discussions in regard to the “dos and don’ts” of designing multiple-choice questions, it became apparent to teachers that creating good multiple-choice questions from scratch requires time and effort. As teacher C1 explained,

The time consumption . . . I mean, we talked already about the flaws in the testbanks. So if you’re making them up on your own and you have so much formatting, with superscript on top of superscript and the subscript, it just becomes—for the value of a one-mark question? It just, . . . it doesn’t seem worth it.

Furthermore, teachers had begun to realize that questions were considered time-consuming because they were difficult to create. When developing new tests, or even when going back to adjust old ones, teachers indicated that this was already considered a hefty price to pay. Teacher A4 described her situation like this:

I felt it’s a little more difficult for me, but easier for the students. Difficult for me because I guess I’ve gotten so used to it [my way of creating questions] that now, having to go back and think about another possible answer, you know, is a little bit more time-consuming.

The second, and perhaps even greater concern voiced by many teachers during discussions was expressed in detail by teacher A2:
But the weakness, I would think it’s a weakness because there’s no ability to express an opinion, or to be able to actually write a response, or give a response to a statement, because it’s already been selected for them. Sometimes, it’s more than one answer . . . and of course we’re only going to put one answer there that’s going to be the correct answer. But maybe they can [express different answers]. If it was a short answer, they could actually write more about it. It’s not just [isolated] to that one closed response because there could be more [answers] to that question. So, maybe having the variety [or the open end] is still very important.

As teacher A4 responded, “So what you’re saying is . . . that it doesn’t allow them to think outside the box?” In many respects, this was a concern held by many teachers. The very strength of specificity that was heavily touted also appeared to be viewed as one of the primary weaknesses of multiple-choice assessment. Teacher B7 opined,

The weakness, I find [is] that it’s too specific. So it’s hard to come up with multiple-choice questions of characters and things [from English literature], ‘cause you only have certain characters to work with, certain themes to work with that they would have already known. So it’s . . . I find it’s too specific, and it’s, it may not be checking what I need [the question] to check.

Similarly, teacher C1 explained,

They can’t justify their work. And that’s what a lot of them complain about. “I’ve read it this way, so . . .” Just even how they answer the questions, or how they read them or what their literacy skills are like . . . or they’re overanalyzing it. Or,
“But I read it this way . . .,” but now they can’t support—they can’t justify it.

And they might be partially right to show that [justification or answer].

In fact, teachers suggested that some students would almost always need alternative ways in which to express their responses. “Those students need different ways of expression. Maybe not just in that [multiple-choice] structured way . . . [but] they need to express their answer to certain questions differently using different methods,” stated teacher A4.

Further extending upon this difficulty, teacher B4 suggested,

It’s difficult to gauge accurately how much they know. So they may know one part of the concept, and it’s [a] multiple-choice [question] . . . the answer could be “a and b.” So they know the answer is “a,” [but] they’re not too sure about the second part of the concept, so, they get the question wrong, and there’s no way to gauge that they know some . . . and [just] not all [of it].

Several teachers also felt that a student’s lack of exposure to the multiple-choice format could also be seen as a weakness. Teachers believed this to be a potential hindrance towards students performing at their actual level of understanding during an assessment.

As teacher A2 explained,

I think maybe a weakness is that when they have questions, homework questions or something—they’re always short answers. But then when they write a test, it’s multiple-choice. They don’t get enough practice. So that’s kind of a weakness.

So they only get multiple-choice on a test . . . on a summative evaluation.

Interestingly, according to survey results, 68% of teachers still responded that they felt they were able to accurately assess students much of the time, with the remaining 32% believing they were only able to accurately assess students some of the
time. Regardless, several teachers observed that taking multiple-choice tests embodied more than a written test. Teacher B4 commented,

I think [writing] multiple-choice questions is a skill—to learn to answer them . . . so if they [students] haven’t seen them on any other test or quiz within the year, it could be unreliable as a source of assessment.

The relevant and practical ramifications of a lack of multiple-choice test-taking experience and skill were concerns for several teachers. Teacher C11 stated,

I tend to increase it as the grades get older, because I’m thinking these 4U bio kids, their first and second years of university, it’s gonna be all just multiple-choice. So it would not be fair if I don’t let them practise those skills . . . or else they’re going to get burned in the following years.

Similarly, teacher C3 echoed,

It is a life skill though. It’s a life skill that they need . . . and if they go on in their career, or they go into the trades, or they want to get their boater’s license when they’re 45, it’s a life skill that they should know how to do. It’s never going to go away. So why not teach them how to do it? Why not train them?

It is strange, however, that teachers would view this issue as a weakness for two reasons. First, the lack of multiple-choice use is not really an attribute of the format itself. Secondly, one would think that it is more often than not that students do have much exposure to multiple-choice questions throughout their education. It is however possible that teachers who brought up this issue believe that, regardless of exposure, answering multiple-choice questions is difficult enough of a skill that not all students will acquire it at the same time or pace. Or perhaps, teachers find their assessments that contain
multiple-choice questions do not appropriately prompt students for their knowledge in a manner consistent with the way course content was taught throughout the year.

Most teachers shared concerns surrounding their inability to accurately assess student learning through the use of multiple-choice questions. Teachers also mentioned that, even when they did incorporate multiple-choice questions into a test, they felt most comfortable yielding only a small portion of their test to these types of questions. “We don’t use multiple-choice that much. I would probably say we use it maybe in 5% of formative assessments only. [For] most of the summative assessments, we don’t use multiple-choice at all,” stated teacher B7. “In the math courses I teach, I would say about 5% of the questions that are used on tests and in formative assessment [are multiple-choice],” agreed teacher B2. When asked if they would feel comfortable using only multiple-choice questions on a test, teachers all responded similarly. Teacher A3 stated, “I don’t think I would be comfortable relying on that alone. I think it’s way too absolute. I don’t like that.”

Teachers often felt that students performed worse on multiple-choice questions than expected, thereby further reducing teachers’ perceived validity of this testing format. Some teachers also pigeonholed multiple-choice questions, stating that they typically thought of using them only when assessing lower levels of cognition. Teacher A2 explained,

But I usually, right now, it’s strictly for knowledge . . . especially on the exam. And then I leave the rest of the exam where they can apply or communicate or express certain opinions or perspectives that they may have learned throughout the year.
Opinions were noted from teachers of various subject areas. As English teacher B7 stated, “I think for English . . . we just use it to check knowledge.” Math teacher B2 similarly commented, “They’re mostly just used for knowledge and to gauge specific skills when assessing during class.” Science teacher B4 also confirmed, “Yeah, for us it—multiple-choice—would typically [be] knowledge based. In some cases it could fall into another category. But they’re usually grouped together as knowledge questions.”

According to these interview statements, it appeared that teachers found it easier to place their multiple-choice questions under the knowledge category. This was embodied by teacher C11: “For me, I just chuck them under knowledge, because it’s easy.”

Overall, there were many interesting points brought forth in the data concerning teachers’ perceived weaknesses of the multiple-choice format. Numerous weaknesses were brought up; however only a handful were universally agreed upon. At the root of those agreed-upon weaknesses appeared to be a simple lack of time for teachers to develop good test questions, which might otherwise have solved those issues. In addition, general misinformed and ill-preconceived notions appeared to tarnish and hamper the perspectives that teachers had on the potential and usefulness of the multiple-choice format. It seems possible that if provided additional training that would serve not only to correct misinformation, but to demonstrate perhaps new ways in which multiple-choice could and should be used, teachers would readily adapt this into their assessment and teaching practices. Therefore what appears to be hampering the use of multiple-choice is less so the intrinsic attributes of the format itself, but rather, the opportunity it is given to be properly understood, and therefore to excel, within a teacher’s mind and curriculum.
Teachers’ perspectives of student responses. Although teachers were able to identify several challenges and several advantages in developing multiple-choice assessments based on their experiences, they also consistently indicated that they needed to take cues from students in determining the effectiveness of their assessments. According to the teacher-participants, students and teachers alike shared overlapping views of their likes and dislikes surrounding the multiple-choice format. To many teachers, it appeared that students enjoyed the process, speed, and simplicity of the format itself and, conversely, more often than not students were not at all pleased with the test results they received.

Teacher responses generally indicated that senior students, along with those in the academic streams, appeared to more often resent multiple-choice questions. As teacher A4 indicated,

I do get a lot of the students . . . academic students, that come to me and say, “Well Miss, we don’t get part marks here. You know it’s a lot more difficult. I can’t express myself in this type of assessment as I could, let’s say, with the long answer [questions].”

On the other hand, junior students and those from the applied streams appeared to look forward to the multiple-choice format. As teacher B5 recounted,

It’s interesting to see when you get to the upper levels, when you start to tell them [the] breakdown of tests, you hear more so the senior students will be the ones where they’re like, “Oh, I hate multiple-choice!” Whereas the Grade 9s and 10s when you tell them the number of multiple-choice they’re like, “Yes!”
In response to how students have reacted to multiple-choice questions and to whether instructions for students on how to approach multiple-choice questions would be beneficial, teacher C14 responded, citing two perspectives,

It depends on the level that they’re at too, whether they’re in an academic class or applied class. Because some of them think that it’s nice that the information’s there, but sometimes the applied kids will say, “I don’t have to study because the answers are there.” So there’s this difference—[they think] it doesn’t matter since I just have to choose [the answer]. So it depends on where they’re coming from. There are students who know they’re going to struggle. They’re lacking that confidence. They’re like, “These stress me out because I can’t explain [myself].”

Teachers seemed to suggest that it was a difference in perception, rather than a noticeable difference in test outcomes, that was responsible for the mindset and approach that different students took towards assessment in general. Therefore, even though it appeared that younger students or those within less academic streams enjoyed the multiple-choice format more, teacher A2 stated, “They [applied students], like the structure of it . . . [but] they still . . . struggle to find the answer.” The results that the younger and applied streamed students achieved on tests were not indicative of their attitudes towards this assessment format. Teacher A2 suggested that in general, “the student perception is that they are going to be tricky . . . and a lot of them are very nervous about doing multiple-choice.” In another response to how students have reacted to their multiple-choice questions, teacher C14 observed the following:

I think they pick up . . . fairly quickly [and] recognize the difference [in the level of difficulty of the test] whether it’s just knowledge [or something else]. They’ve
[students] come back a couple of times on our last test saying that “they’re really tough, they’re really a bit tricky and I had to think it over.” So at that level, it’s hard [for them], because it’s almost not what they’re expecting, except the knowledge level.

Teacher A2 gave this account of her firsthand experience observing a drastic change in attitude from her students towards multiple-choice:

They were so excited. They finished quickly. They were so happy about doing this [a multiple-choice test]. . . . There were no short answers, no essay. [They were] thrilled! However, most students did a lot [worse]. Their results were not as good as when they do have short answers and even the essay . . . so they weren’t so excited when they got the results, because they just didn’t do as well [as they thought they would].

Teacher A2 also explained that, due to its specificity and the number of concepts covered, students could no longer get away with studying only bits and pieces of the information that might be covered on the test. According to teacher A2, “they [students] have to know everything about everything. There’s no choice. . . . They have to focus in [on all subject area details].” Again, it appeared that in many instances students have a preconceived notion that equates less writing on a test with less of a need to study, which might appear to incorrectly suggest a lower level of test difficulty.

Overall, teachers’ perspectives on how their students felt about multiple-choice questions seemed to suggest a range in which most students were initially dismissive and unconcerned towards multiple-choice assessments but then later grew to see the importance, value, and difficulty that this form of assessment actually presented. Survey
results seemed to align similarly with this range of possibilities, as approximately 64% of teachers indicated that they felt their students perceived the level of difficulty of multiple-choice questions to be “somewhat difficult,” while an even split of the remaining 36% of teachers indicated that students perceived their questions to be either a little difficult (18%) or often difficult (18%).

Though not likely to have been a main consideration of the teachers when responding to survey questions, it is worthwhile to mention that teachers indicated the importance and the existence of a third group of students who do not necessarily fall neatly under the categories of academic or applied, such as those students with special needs, ESL (English as a Second Language) students, and/or ELL (English Language Learners) students. For reasons considered to be “out of their control,” some teachers indicated that the multiple-choice format may not only have been perceived as being more difficult, but may in actuality have been more problematic for certain students. In response to potential difficulties experienced by these students, teacher B4 stated,

I think these are also difficult questions for those who are learning English as a second language. A lot of the questions are a play on words, and the incorrect answers could be hard to decipher if someone doesn’t understand the language. Especially in science, it’s a completely new language for them, just learning the science words. So . . . it could be difficult.

Teacher B7 continued,

So not only do they have to translate the word into their language, they have to translate all the wrong words as well, and then choose the right word. So it takes
double the time. Whereas if it was just a fill in the blank, or like a short-answer question, they could find the word that they needed to use.

Teachers noted that these difficulties were also quite apparent even for certain students whose first language is English. Teacher C3 offered some insight into this area,

I found that a lot of students don’t like multiple-choice because they have to do too much reading. So if I just give them a short answer question, there’s a lot less reading in that short answer question than there is in multiple-choice because I now have to read all the responses to [the] multiple-choice [questions]. If I’m a weak reader, I’m not going to read the whole test . . . [if] I can’t read all the multiple-choice ones . . . . So for the weak readers, it is a problem.

In response to difficulties perceived by students, teacher B7 extended these oft ignored circumstances to those with Individual Education Plans (IEPs) and raised some important considerations:

I wonder how autistic kids feel about multiple-choice? Because in the autistic kid world it’s black or white, so if you have four different choices there, a kid would have a meltdown. Like if they didn’t see the answer they perceived to be the right answer, . . . I don’t know. With the growing number of IEP’d students at this school coming in every single year, I just don’t know how that particular form of assessment is gonna do well for these kinds of kids.

While these points were not often raised by teachers during the focus group interviews, teachers who did raise these points were uncertain about how exactly to approach multiple-choice assessment in light of these concerns. Fortunately, participants demonstrated an understanding and a willingness to further tackle these issues.
Overall, teachers believed they were aware of how their students viewed multiple-choice questions. As several teachers indicated, students *will* let you know if you have created terrible multiple-choice questions. Teacher A5, and many other teachers, also pointed out, “They’ll put their hand up and [say], ‘Yea . . . I don’t understand, it’s confusing, what do you mean? . . . This [other answer] is right too!’ You don’t want to be in that situation.”

**Secondary Teachers’ Perspectives of the Value of Multiple-Choice Assessment and Training After Participating in a Workshop on the Multiple-Choice Format**

All teachers took part in a short, one-day workshop held at each of this study’s three participating schools. At this workshop, teachers were presented with (a) information and examples detailing the makeup of multiple-choice questions, (b) the various forms in which multiple-choice assessment can be used, (c) guidelines for question creation, (d) an explanation of Bloom’s revised taxonomy detailing the four kinds of knowledge and six levels of cognition, (e) how to analyze multiple-choice questions, and (f) how to approach multiple-choice questions found within the Ontario Secondary School Literacy Test (OSSLT).

In the series of interviews following each workshop, teachers were asked to comment on how they felt their perceptions towards multiple-choice assessment may or may not have changed as a result of participation in the workshop. Following each of the workshops, most teachers responded with gratitude and reported having gained new and valuable assessment knowledge. Many teachers voiced their desire for more workshops and discussions in light of their realization that they lacked understanding and training concerning multiple-choice assessment. As teacher C1 poignantly stated, “I think people
realize they have a lack of training in this area.” However, those who mentioned the workshop to other teachers found it difficult to express the awareness they had gained and felt that they had to leave it up to teachers themselves to choose to further investigate this topic. Teacher C4 pointed out that in the midst of a hectic teaching schedule, teachers often asked, “How [does] this change what we are supposed to do?”

In response to further training opportunities, many teachers agreed that they were in favour of participating in future workshops and, in fact, requested training in specific areas concerning multiple-choice assessment. One of the more prominent requests was to be provided with dedicated time in which to apply knowledge that was presented during the workshops and to gain greater hands-on experience developing multiple-choice questions. As teacher C9 suggested,

Something that you could add to the workshop would be maybe making us do work. Making us like write questions with the answers to questions while you’re actually here . . . and making us put it into practice right here and assessing it.

Teacher B3 similarly suggested,

You know what I’d like to do is, since we’re all different subjects, we can’t do it, but if we just had a couple math teachers, like four or five, stick to like a Grade 10 curriculum and be like, “Hey, let’s create some multiple-choice.” Like, I want some hands-on; let’s do it, let’s talk about it.

Following the multiple-choice workshop, the most prominent response from teachers across all three schools was that they now held a much improved understanding of this assessment format. This enabled teachers to approach multiple-choice questions in new and challenging ways while making improvements to the quality of their
assessment practice and, consequently, becoming mindful of potential errors that accompany this form of assessment. For some teachers, the workshop resulted in a changed outlook on multiple-choice assessment. Teacher A1 attested,

I just loved it because it’s given me a whole new look at it [multiple-choice]. It’s given me a whole new understanding of the importance of them [multiple-choice questions] and how to use them in the best way to help the students.

For others, it shed light upon a new area of training that they may not have previously considered undertaking. “But, yeah honestly . . . when I’m making tests, I don’t even have . . . any multiple-choice questions on my math tests. But I really like what a lot of the material says, and I’m always thinking about it,” stated teacher B3. Finally, there were teachers who felt that the workshop had changed the way in which they approached the design and structure of their assessment practice and preparation. As teacher B4 explained,

Right now, it’s in the back of my mind, and I’ve been conscious when I’m creating the latest tests and quizzes, and looking through those [guidelines]. Even in the textbook when you assign practice questions to the tests, you see the errors in some of them. So I try to avoid those. And I’ve also been conscious of making sure that they’re [students are] prepared. That they don’t just see multiple-choice on the exam or on the tests . . . so the quizzes all have them on them and with the review from the textbook as well.

Similarly, while reviewing their own multiple-choice tests, teachers explained specific items that they had begun to take into consideration in light of the training session. Teacher C13 stated,
I just think for me it’s definitely great advice in terms of my practice. So things like using “all of the above,” which I before thought was totally acceptable, I’ll take second looks now. When I was going through my exam I [was] kind of like, “Oh, I should probably fix that one,” you know at least eliminating some of them . . . and also doing the whole negative thing using “this is not.” I’ve definitely become more aware of that [using negation]. So it’s good in terms of helping me review which questions weren’t necessarily great.

Teacher B5 also commented,

I think I know for me I’ve definitely changed the negatives, not putting the incorrect sort of thing. And I rarely now do the all of the above, none of the above, a and c only, sort of thing. . . . Now I know [this] when I’m creating the tests, I’ll make an effort to not do it.

Many of the other teachers echoed similar sentiments in terms of making changes to their assessment practice and their views on future training.

When asked whether they felt that creating multiple-choice questions had become more challenging after receiving training in this area, most teachers agreed that it had indeed become a more time-consuming process. Teacher B5 responded in this way:

Oh, much more difficult definitely . . . you know. But in terms of discrepancy, ‘cause you know how sometimes the kids will come back and argue for a point? But now you have the solid, “No! That is the most correct answer!” sort of thing. And so on that end it was easier, but in terms of creating, it was more difficult.

Overall, teachers appeared to be most willing and able to implement portions of the multiple-choice creation guideline in order to make immediate and practical changes
to their multiple-choice test structure and design. This was the case, despite the fact that most teachers found creating multiple-choice questions to be even more difficult following their participation in the workshop. Nonetheless, it appeared that being equipped with “multiple-choice aids” had increased teachers’ willingness and confidence to make improvements to their assessment practice. “It will take time to do [write the multiple-choice questions], but in the end, it will give the kids a little [more] variety and [further] opportunities [to succeed],” stated teacher C14.

Teachers were anxious to apply the multiple-choice form of assessment within their practice, to watch it succeed and benefit their students, and to continue to learn and improve their assessment practice. One suggestion for potentially further improving a teacher’s assessment practice would be to implement an integrative assessment approach. This could be accomplished by saddling multiple-choice assessment alongside the Ministry of Education of Ontario’s ongoing push for ubiquitous differentiated instruction throughout schools. Teacher A2 saw this potential, suggesting,

I mean, this could work even with this whole DI movement. When you want to do a formative [assessment] . . . they have these exit trees, or when you come in, . . . maybe they can fill out, like do a multiple-choice-based [assessment] on that lesson, and then you know what they’ve actually learned on that lesson. Then the next day, you can say, “This is what we’ve learned, this is what we need to review because we didn’t learn this yesterday and I would like you to learn this.” And so you could use it for assessment for [learning]. . . . Because a lot of [the] time, you just use it for assessment of—like a final test. Like you could use it as part of the exit cards, you could give a few multiple-choice exit cards for that specific lesson
. . . just two or three. So it can even fit in with the whole DI. . . . So we can use that [multiple-choice for formative learning] and also for further professional development.

In summary, teachers’ perceptions of multiple-choice assessment and of training aimed at improving and extending the use of the format itself had begun to change. For the most part, teachers had been presented with a new way of looking at something that they had grown a narrow view towards, and what was presented was exciting to them. They could see the potential for success in effectively applying multiple-choice assessment in their classrooms. Many teachers expressed the opinion that this was the initiation of a shift in their understanding and appreciation for the multiple-choice format. Teachers claimed they would no longer take their use of this format lightly and that they would look to take advantage of future trainings and collaborative opportunities with fellow colleagues. Ultimately, teachers believed that multiple-choice assessment was worthwhile, as it can improve the efficacy of their pedagogy and benefit student learning in both the present and the future.
CHAPTER FIVE: DISCUSSION

The multiple-choice form of assessment has become not only a commonly used assessment tool in our education system but an essential component of our society as a tool for making assessments and gathering information in the workplace. While multiple-choice assessment is considered well known and often used by many, particularly in relation to high-stakes assessment (Gergely, 2007), conclusive and summarizing research regarding preparation, training, and perceptions in regard to the implementation of the multiple-choice format have been relatively scarce. Compounding the circumstances and difficulties of the use of this format is the lack of teaching and advice being given to students on how to approach multiple-choice tests.

The purpose of this study was to explore the current perceptions and realities of multiple-choice assessment use within the Ontario secondary school education system as well as changes in teachers' views after participating in a professional development workshop dedicated to the effective design of multiple-choice assessments. This study attempted to ascertain the current landscape of multiple-choice use through teachers’ divulgence of their perceptions and understandings of the multiple-choice format. This research, validated by teaching experience, attempted to highlight both the strengths and weaknesses of multiple-choice assessment in light of existing studies and to provide teachers with a way to assess students effectively using this format. Analyses of the interview transcripts recorded from each school were the primary source of data for this study. To consolidate the different aspects of the research that was conducted, a summary, discussion of the findings, and an investigation of the implications of these research findings are presented in this chapter.
Summary of the Study

In this study, two focus group interviews were conducted at three southern Ontario secondary schools. Interviews were conducted separately, preceding and following a multiple-choice professional development workshop, and were approximately three to four months apart. During these interviews, participating teachers, who possessed various levels of experiences and teaching subject backgrounds, were asked to respond to questions designed to probe both their views on multiple-choice assessment and the role that multiple-choice assessment had played in their own teaching practice. A short survey was also distributed to supplement the data collected from the interviews. The interval of time between the first and second interviews was planned deliberately in an attempt to capture the teachers’ changing perceptions towards multiple-choice assessment across a period of several months. Overall, teachers revealed a wide range of beliefs about their approach to multiple-choice assessment. While all teachers acknowledged having used multiple-choice questions in their classrooms, most teachers acknowledged their lack of training and, unsurprisingly, a lack of understanding or interest in developing their multiple-choice questions. The findings revealed that there was a general sense of unpreparedness among teachers toward this form of assessment. For many participants, teacher education programs and a lack of ongoing professional development had not adequately prepared them to develop assessment materials appropriately. In particular, many teachers were left with little knowledge of how to develop multiple-choice questions and, as a result, were unaware of the important role that the multiple-choice form of assessment plays in students’ current and future academic success. The findings
further revealed that teachers felt unprepared in using the multiple-choice format for assessing higher levels of cognition among students. 

There was a convergence of teachers’ views, both positive and negative, towards the multiple-choice format. On the positive side, teachers indicated that the potential for objective grading, the ability to efficiently cover a wide range of material, and the marking efficiency of the format were of particular interest and value. On the negative side, teachers voiced several concerns regarding the use of multiple-choice assessment, including (a) the time required to create effective multiple-choice questions, (b) the perceived limitation of this tool to assess only lower level thinking, and (c) the inability of the multiple-choice format to wholly evaluate students’ understandings and abilities.

In response to suggested improvements that could be made to the multiple-choice format, findings revealed that teachers were willing to implement simple guidelines that would primarily involve fine-tuning the language, structure, and presentation of their multiple-choice questions. Toward the end of the study, several teachers indicated that they had already begun to alter the way in which they developed multiple-choice questions and had begun to apply some of the multiple-choice question writing guidelines in their practice. Following the multiple-choice assessment workshop, teachers stated that, either consciously or subconsciously, they had begun to regard multiple-choice questions with greater importance and to scrutinize their own questions through a wider and more critical lens.

In addition, the findings revealed that most teachers wished to incorporate multiple-choice questions of greater cognitive difficulty into their assessments. The usefulness of having opportunities for further professional development in the creation of
multiple-choice questions, alongside year-long teacher collaboration to create and share test materials (from within the same subject area), were also mentioned during interviews and converged with literature suggesting the need for more professional learning communities (Popham, 2009; Volante, 2009). At the same time, however, teachers were aware that opportunities for additional training and teacher collaboration would be quite rare. As such, the limited time and experience provided to teachers through the multiple-choice workshop was generally viewed in a positive light, indicating that continued third-party research efforts may be necessary to serve as a catalyst for action so that teachers might continue to participate in ongoing professional development in this area.

**Discussion**

In this section, the results from the study are examined in light of three research themes: (a) secondary school teachers’ assessment training, (b) teachers’ perceptions of multiple-choice assessment, and (c) teachers’ changing perceptions of the value of multiple-choice assessment and training following a one-day workshop. The order in which these research questions are presented is important, as they progress logically from one to another.

It was important to first gather insight into teachers’ prior training in the area of assessment for several reasons. First of all, teachers’ prior training in multiple-choice assessment, or lack thereof, helped to lay the foundation for this research, while also providing valuable insight into the general landscape of teachers’ knowledge and skill in using the multiple-choice format. As such, teachers’ knowledge, understanding, and perceptions of the multiple-choice format were analyzed. Painting a broad landscape was necessary to the formation of this study, as there appears to be a lack of current research
surrounding the knowledge and perspectives that secondary school teachers possess in relation to the benefits, hindrances, and uses of the multiple-choice assessment format. In general, research into teachers’ knowledge and understanding in regard to their pedagogy and teaching practice is already considered rare (Voss, Kunter, & Baumert, 2011). When considering the types of assessment tools used by teachers, it became clear that very few teachers had ever been exposed to training in the area of multiple-choice assessment, nor had they ever seriously considered the value of training in the multiple-choice format. In general, researchers and measurement specialists believe that additional training is needed for educators in all areas of measurement (Peyton, 2010). Through interviews with teacher-participants, this study sought to more accurately identify ways in which existing multiple-choice research (e.g., Haladyna et al., 2002) might benefit teachers in clarifying misconceptions and helping them to use their time and resources effectively in order to benefit students.

Finally, the impact that the interviews and workshops had on teachers’ perceptions of the importance and value of multiple-choice assessment and its training were reviewed. My intention was to explore these perceptions and to suggest future areas of study that might aim to continue improving the efficacy of multiple-choice assessment in teaching. Furthermore, I hoped to also lend reason and support for the setup of future professional development sessions of this nature within schools.

**Teachers’ Experiences and Need for Assessment Training**

In attempting to understand how multiple-choice testing fits in to the assessment practices used in education today and to gain insight into how and why teachers might respond differently to this form of assessment, it was important to first understand
teachers’ prior knowledge and background on the multiple-choice format. Responses to the following questions were used to help guide the findings and implications of this study: (a) Do you feel that you have been sufficiently prepared to develop and use multiple-choice questions as part of your assessment? (b) How comfortable are you with using these types of questions? Teacher responses to these questions helped to shed light on the following research question: How well are secondary school teachers trained in multiple-choice assessment?

Interview results revealed that while a fair number of teachers had previously been through some form of assessment training, most of them felt uneasy in creating assessment questions of the multiple-choice variety. This makes sense given the level of difficulty encountered in attempting to construct good multiple-choice questions (Shuhidan, Hamilton et al., 2010). It appears that little discussion, if any, on the topic of multiple-choice assessment had been included in previous assessment training or workshop sessions. Teachers were leaving these sessions feeling uninformed as to how they might improve their understanding and practice involving multiple-choice assessment. According to an aggregate result of both discussion findings and survey results, the majority of teacher-participants claimed to have received some prior form of assessment training (not necessarily specific to multiple-choice assessment). One would hope that assessment training topics covered the basic and most prominent areas of assessment or, at least, the areas in which teachers are experiencing difficulties. Unsurprisingly, however, survey results indicated that teachers frequently thought of the word difficult when describing their use of multiple-choice questions. It is therefore interesting that findings also indicated that teachers implemented multiple-choice
assessment at least once at some point in their classes and, on average, 3.5 times per semester, suggesting that despite their difficulties with multiple-choice assessment, most teachers still regularly employed multiple-choice questions.

Consequently, a common thread throughout interview responses was that a heavily emulated approach in constructing multiple-choice questions was taken on by necessity rather than by choice. To borrow materials from other teachers, to rely upon often suboptimal commercially developed testbanks (Simkin & Kuechler, 2005), and to review previous exams, all the while hoping for the best, does not help teachers cultivate a growing understanding of multiple-choice assessment or develop teachers’ assessment practices. Making matters worse is the danger of falling victim to a plethora of poorly constructed guidelines, which are often made available through untrained item-writing and test-construction authors (Irish, 2006). However, based on the teachers’ responses in this study, simple and practical guidelines remain the preferred and most relatable method for teachers who are willing to implement and attempt future changes to their assessments. Teachers also commented that when comparing their own multiple-choice questions to the item-writing guidelines provided, they were quite often confronted with their own “bad” habits and errors to which they had never given much thought.

Therefore, in concert with the fact that teachers admitted to having put little thought or time into how they constructed their multiple-choice questions, it seems likely that teachers have been conducting multiple-choice assessment with limited direction or regard for the efficacy of their assessments.

Questions regarding how to increase the level of difficulty of multiple-choice items, what makes a question too difficult, or how to shift the level of cognition tested to
a higher level were frequent questions. These points were often seen as limitations and barriers hindering teachers from utilizing multiple-choice questions as an effective means of assessment. Of particular relevance and concern to teachers appeared to be how they might create their multiple-choice questions to assess the wide range in levels of cognition or curriculum mandated assessment categories. The majority of suggestions indicated that teachers were also interested in further focus group-oriented meetings organized by subject area, so that teachers might be able to field questions and receive support from other same-subject teachers. Other teacher requests were for concrete or hands-on workshop times that would be directed by the researchers of this study. This comes as no surprise as several studies confirm that hands-on learning in workshops and training are almost universally preferred by learners (e.g., Lee, 2004—2005; Shriner et al., 2009; Struyven, Dochy, & Janssens, 2008).

Teachers indicated that the aforementioned suggestions were often informally discussed within daily conversation but that it would have been necessary to receive administrative approval in order to formally create assessment training opportunities. Haviland et al. (2010) indicated their support for workshops tied closely to teachers’ own classroom practice and professional development opportunities, claiming that teachers’ attitudes, understanding, and confidence levels can benefit greatly from this type of endeavour. It would therefore appear that if any individuals were to implement assessment training opportunities, many teachers would be interested in this beneficial path leading to effective assessment and improved practice.

One of the more intriguing training suggestions from teachers, however, revolved around whether multiple-choice questions can be beneficially integrated into a teacher’s
method of differentiating instruction. This is relevant as Ontario continues to promote equitable and accessible education for all of its students, even in the face of an ever-diversifying classroom and alongside a growing number of students with special needs (Daniel, 2005; Graham & Jahnukainen, 2011; Ontario Ministry of Education, 2010). This greatly increases the importance and relevance of the multiple-choice format if teachers are able to use these questions not only to effectively assess students, but to also adhere to Ministry-level efforts. While there is still little in the way of research commending the inclusive approach of the multiple-choice format, and although the research that exists suggests there is the potential for cultural bias for ESL students in large-scale multiple-choice assessments (Lampe & Tsaouse, 2010), Gatfield and Larmar (2006) have suggested that, at the very least, this form of assessment is a nonbiased method of evaluation spanning various nationalities.

Overall, teachers appeared to acknowledge an overall lack of and definite need for a consistent use of multiple-choice assessment training. While the participants desired future training in relation to multiple-choice questions, their responses indicated that they were most ready and willing to implement suggestions that were included within the multiple-choice writing guidelines provided. This was likely due to the fact that the guidelines presented were relatively easy to understand and would present an improved value-imbued assessment when taking into consideration the additional time required to implement these changes.

**Teachers’ Perspectives of the Multiple-Choice Format**

Just as a one-size-fits-all approach to teaching is unlikely to be championed as the most effective way to teach (Skwarchuk, 2004), a one-size-fits-all approach to assessment
is equally unlikely to be effective, particularly given an ever-diversifying population of students found within secondary schools (Garcia & Floyd, 2002). When it comes to creating and implementing multiple-choice questions, likewise there is no “best” or “one-and-only” way of constructing questions. Many researchers still consider the construction of multiple-choice questions more of an art form than a science (Crehan et al., 1993; Downing, 2005; Droegemueller et al., 2005; Haladyna et al., 2002). It is up to teachers themselves to discern the value and effectiveness of the assessments that they choose to implement. Their perspectives on the multiple-choice format—having taken into consideration the realities of the teaching world—will ultimately shape their desire, willingness, and confidence to use this form of assessment. If teachers are to be equipped with the best possible tools, to be given solutions to their everyday challenges, and to be shown how multiple-choice can deliver improvements to the quality of their practice, then it is all too important to understand the current landscape of how “active” teachers view the multiple-choice form of assessment.

The following subsections will reveal teachers’ perspectives towards the multiple-choice format through two key areas in the hopes of delivering a comprehensive and salient understanding of the following research question: What are secondary school teachers’ perceptions of multiple-choice assessment?

**The strengths and weaknesses of multiple-choice assessment.** Identifying and realizing how to take advantage of the strengths of an assessment format are crucial to the successful implementation of any form of assessment. Understanding how teachers go about identifying the strengths and weaknesses of a potential assessment tool depends on how they view and understand its strengths (or weaknesses) in light of how it fits into
their teaching realities. While this may vary among specific individuals and the contexts within which they reside, teacher consensus and agreed-upon realities establish a basic foundation for the strengths of the multiple-choice format as an exceptionally valuable tool for assessment.

Findings from all three participating schools consistently revealed that the perceived strengths of the multiple-choice format most often stemmed from the method in which questions were evaluated. One of the easiest and unanimously cited strengths of the multiple-choice format was its objectivity and marking ease. Having cited the removal for the potential of subjective bias, in addition to the added marking ease provided by facilitative technology found in using Scantron cards, participants were in agreement in identifying these format strengths. The grading-oriented strength of the multiple-choice format is also made clear throughout much of the existing literature on multiple-choice assessment (Jensen et al., 2006; Marsh et al., 2007; Moreno et al., 2006; Odegard & Koen, 2007). Additionally, teachers noted the value of being able to rely on an objective grading method, particularly when dealing with large-scale assessments that might span more than a single class or institution. With the use of multiple-choice questions, teachers indicated that it becomes impossible for any one teacher to grade “harder” than another teacher. Studies have similarly suggested that subjectively graded assessments, such as short answer or essay questions, fall victim to varying forms of bias unrelated to the content answer and that the objectively graded nature of the multiple-choice format promotes reliability (Rogers & Harley, 1999; Parmenter, 2009).

Another strength of the multiple-choice format as described by several teachers is its speed and efficiency. When taking into consideration the time required for scoring,
immediate feedback, and precision of feedback, “multiple-choice exams rate higher in efficiency than essay exams” (Hautau et al., 2006). Though these traits were popular responses to a survey question asking participants to describe the multiple-choice format using three adjectives, participants who did not list adjectives pertaining to the speed in which one can score multiple-choice questions were perhaps mindful of the amount of time it takes to create effective multiple-choice questions. In regard to the areas of test-taking and feedback, interview findings and previous research (e.g., Farrell & Leung, 2004) mention the ability to have a quick turnaround in providing feedback and grades to students while at the same time covering a relatively large amount of material in a brief period of time. This may be a way to immediately address subject material with which students might be struggling. This would appear to also lend its support to the increasingly popular philosophy of “assessment for learning” (So & Lee, 2011). This concept directly contributes to student learning in a more transient and less stagnant manner, where knowledge and instruction can be passed back and forth between student and teacher, cultivating a more interactive learning environment. During interviews, several participants were quick to point out that students’ performances on multiple-choice questions were often found to be accurate indicators of a student’s overall academic performance and understanding of subject material. Participants were thus provided with a quick and often accurate assessment of students’ overall academic achievement in their class. In short, the sheer amount of time saved during the marking process is noteworthy.

However, along with the reported strengths of multiple-choice assessment stemmed an equal number of weaknesses, according to the participants. Responses
seemed to suggest that participants realized that the multiple-choice format held considerable potential, yet they were unable to gain a firm understanding of this form of assessment due to their own ill-conceived notions and perceived weaknesses of multiple-choice. A number of commonly perceived weaknesses were mentioned throughout discussions.

Participant responses, however, indicated that the time required to create quality, multicognitive-level questions was of genuine concern. While participants were willing to make small adjustments to improve their tests, any changes attempting to target new levels of cognition were seen as too challenging by most participants. It is, without a doubt, difficult and time-consuming to create effective multiple-choice questions (Slem, 1981) and, as a result, teachers have simply become “used to it.” That is, teachers have become accustomed to the way in which they have always created questions, and if they were to now change the manner in which they use multiple-choice assessment, this would be time-consuming in and of itself, let alone the time considerations for the corrections that would need to be made. Slem (1981) suggested that it is the difficulty involved in creating alternative “distractor” response options of quality that consumes much of a teacher’s time. Unsurprisingly, participants admittedly appeared to be using textbook questions, not because the presumed quality of these questions was superior to their own, but simply because questions and their respective response options had already been developed and were ready for use. As former research has suggested, it may be more difficult to create questions of a higher cognitive level, but it is still very much possible (Fellenz, 2004; Hampton, 1993; Torres et al., 2011). For many teachers, the simple thought of the time required to create a large bank of quality questions seems
unfathomable in light of already numerous teaching obligations. This is an important area that teachers will likely need to reconcile before implementing this format. Research that has in part remained mindful of this weakness (e.g., Bruno & Dirkwager, 1995; Landrum, Cashin & Theis, 1993; Delgado & Prieto, 1998; Rodriguez, 2005; Rogers & Harley, 1999) suggests that three-option multiple-choice questions are typically more effective than four- or five-option questions.

Findings from this study have also revealed that another major concern, according to most participants, was that the multiple-choice format hindered the ability of students to freely express their unique responses to each question. Some teachers believed that this format would not allow students to justify their own work or think outside the box, hindering students’ abilities to correctly convey their response. However, I would submit that this opinion is narrow in its view. Provided that the desired function and purpose of multiple-choice questions are appropriately identified, multiple-choice assessment is simply doing what it was intended to do. Slem (1981) suggested that multiple-choice questions are successful in highlighting the skills of discriminatory thinking, comprehension, application, synthesis, and evaluation. Participants who were in favour of the specificity generated by multiple-choice questions noted that a major benefit of the targeted approach was its ability to test many specific content areas of knowledge and understanding, thereby increasing test reliability. This would not be possible to undertake in any reasonable amount of time using an open-ended constructed form of assessment (Parmenter, 2009; Slem, 1981). Some participants also indicated that explanations given by students for the opportunity to justify and lend support to their responses were more often than not an error in question recognition or interpretation and an attempt by students
to creatively manipulate their way into undeserved marks. Therefore, it would appear that the perceived weakness of indulgent specificity would be valid in some instances but that, overall, the specificity in this context is viewed as a strength that teachers can and should employ to potentially bolster the reliability and scope of their assessment practice.

Other weaknesses suggested by teachers revolved around the potential cost of administering large sets of multiple-choice questions. However, it would be just as costly to administer an equally large number of constructed-response questions. Participants also observed a noticeable presence of confusing and often ineffective uses of language in presenting a question (i.e., questions deemed as “wordy,” “unfair,” or “confusing”). However, this could likely be corrected through the simple use of research-proven question writing guidelines or principles (Downing, 2005; Haladyna et al., 2002; Tarrant & Ware, 2008). Findings further reveal that, given the infrequent use of multiple-choice questions as a formative and daily assessment tool, the times when it is used could be seen as lacking in validity, particularly when used as a substantive portion of a class’ summative assessment. Consequently, participants indicated that these concerns might mix in extraneous grade-influencing factors that would end up damaging the integrity of the assessments delivered.

**Teacher’s perspectives of student responses.** Prior to stating a conclusion regarding the perceived value and use of the multiple-choice format, there is an additional factor that teachers must consider when striving for successful assessment implementation. Studies indicate that the manner in which students perceive and approach their education and perhaps, more important, the way in which students perceive upcoming assessments, is directly related to the way in which they prepare for
them (Dogan, Atmaca, & Aslan Yolcu, 2012; Parmenter, 2009; Yonker, 2011). Given that students oftentimes feel that multiple-choice tests require less preparation than other types of assessment formats (Struyven et al., 2008), it would be prudent for teachers to take into account how students view multiple-choice assessment. In fact, findings indicate that the participants believed students should study content material even more thoroughly when having to respond to multiple-choice questions because, in essence, they need to know everything about everything.

Participants, however, noted that not all students approached the multiple-choice format with the same mindset. They noted that both upper-year and academic-streamed students more frequently exhibited resentment towards the multiple-choice format, whereas lower-year and applied-streamed students seemed to relish the opportunity to be assessed via multiple-choice questions. Participants further noted that there appeared to be a misconception among students that, because they were not required to synthesize a given response onto a piece of paper, there was no need to synthesize a complete response in their head. Students, therefore, might incorrectly infer that the requisite amount of studying time and effort can be lessened because all they need to do is choose the correct answer rather than apply any cognitive fortitude in generating a response. Simply put, multiple-choice questions strike a positive chord with weaker and/or younger students because these students feel that identifying the correct response among a series of possible answers improves their abilities and, therefore, exempts them from studying (Parmenter, 2009). A study by Watering, Gijbels, Dochy, and Rijt (2008) similarly indicated that younger and more inexperienced students are more likely to incorrectly perceive the true level of difficulty and/or the nature of the concept being assessed. This
may lead to an ineffectual approach towards multiple-choice test-taking and other comparable types of assessment.

Perhaps what is sustaining this misconception stems from elementary school where multiple-choice questions may have been initially presented as the “easy” portion of a test and were likely taken lightly. However, within a secondary school setting, there is an increased presence and level of difficulty that surrounds multiple-choice assessments (Nicol, 2007). It is therefore possible that students’ early successes with the multiple-choice format might actually have given them a false sense of security in regard to their ability to excel at this format. Teachers similarly suggested that by the time students are given the opportunity to experience several multiple-choice assessments at the secondary level, their perspectives on the format have already gradually become inverted. At this point, students might realize that the grades they receive on multiple-choice tests are simply not as good as they envisioned them to be. In fact, participants indicated that students demonstrated a lack of confidence in their abilities, coupled with elevated levels of stress, as they began to perceive the multiple-choice format as the vehicle through which many high-stakes assessments are driven (i.e., EQAO standardized tests).

Findings from this study further indicated that, according to the participants, common complaints and excuses from students tended to revolve around their inability to acquire at least partial marks for responses that were incorrect. According to the participants, students think that if they are given the chance to completely express themselves, they will be able to compose a valid response. This is one of the reasons why an investigation into varying modifications of the multiple-choice format might be valuable in allowing evaluators to better gauge the extent of a student’s understanding of
a specific concept. Not only might this type of investigation grant students the ability to provide responses in greater depth, but this might also help to reduce the perceived “guessing game” weakness of the multiple-choice format. In turn, students with no existing knowledge of a concept would be prevented from correctly guessing the answer and receiving the maximum allotted marks for that question.

Educators need to sometimes be reminded that some of their “tried and tested” questions and assessments may no longer assess intended course content for students at all levels of learning. As participants indicated, insight from a student’s perspective can impart valuable knowledge to teachers on how to construct and deliver purposeful, equitable, and successful assessments. Students’ differences in reading ability, comprehension of concise texts, and spelling and grammar skills can impact their eventual understanding of course content. These differences are overlooked by some teachers, and educators can find it difficult to maintain an awareness of the unique learning abilities of each student. The manner in which all students perceive multiple-choice questions, however, can never completely and accurately be accounted for, though this statement most likely holds true for any form of assessment. Participants additionally revealed that students are typically not reserved in making their opinions of test questions known. As a result, participants in this study often stated that they were aware of how students perceived their multiple-choice questions and were subsequently willing to consider students’ feedback and opinions towards the creation of future multiple-choice assessments.

Different forms of assessment need to be viewed as tools that are useful only if used appropriately. Teachers should capitalize on the strengths of an assessment format,
while at the same time attempting to minimize its weaknesses whenever possible.

Findings from this study suggest that one of the issues that presented itself when applying the multiple-choice format was that participants were utilizing this assessment method with little understanding of how it works. As a result, the participants’ assessments were not always aligned with the methods of instruction and materials that were being used, nor with the overarching goals and curriculum expectations of the class. In spite of this all, what was apparent in the findings was that the strengths of the multiple-choice format lay within its grading ease and objectivity as well as its ability to assess large amounts of material within a relatively short period of time. The overall amount of time saved, particularly in large-scale assessment environments, is very important to teachers (Parmenter, 2009), and this was reflected in the participants’ interview responses. The primary perceived weaknesses of the multiple-choice format, according to the participants, were that students have perhaps not been given opportunities to fully or appropriately express their responses, that the initial creation of the multiple-choice format requires too much time and effort, and that this format is incapable of assessing levels of cognition beyond the lowest levels of Bloom’s RT. Findings from this study further suggest that these perceived weaknesses lose much of their momentum when placed alongside the strengths of the multiple-choice format.

Overall, findings from this study indicated that multiple-choice assessment was viewed as providing a lot of “bang for its buck.” This was particularly true for participants utilizing this assessment format on a continual basis throughout the school year rather than as a “one time only” year-end assessment. Participants expressed the desire to have a tool that could be built up and refined repeatedly over time through
teacher collaboration. Furthermore, taking into account the perspectives of a diverse student population while constructing and refining assessments on an ongoing basis will help teachers to develop their level of proficiency with multiple-choice assessment. In short, if teachers’ perspectives were to grow and coalesce around a proper understanding of the multiple-choice format, they would be presented with an innovative and effective assessment tool that would quickly improve their assessment practice while also effectively preparing students for elevated academic success.

**Teachers’ perspectives of multiple-choice assessment and its training following a workshop on the multiple-choice format.** Having looked at the participants’ current perceptions, along with the theoretical and practical implications of multiple-choice usage within the secondary environment, it is evident that teachers would benefit from further learning opportunities in this area. While this study was not designed as an exhaustive training tool in developing multiple-choice writing skills, sufficient new and relevant information was presented that impacted participants’ perceptions of the multiple-choice format and further multiple-choice training.

Following the group interviews and the workshop, participants were asked how their opinions in regard to the multiple-choice format had changed. Though there exists some research in relation to the varying results and impact of the multiple-choice format, there is limited research that investigates and summarizes secondary teachers’ general opinions toward multiple-choice assessment. One possible reason for this could be that there are few teachers who would find interest or value in research that targets the improvement and success of the multiple-choice format. It was during the final set of interviews that participants indicated that they could now see that they actually had a
relatively insufficient understanding of the multiple-choice format and possessed a lack of training in this area. This would suggest that the workshop, at the very least, had acted much like a litmus test towards teachers’ assessment training, revealing to them that their level of assessment knowledge, especially in relation to multiple-choice assessment, is lacking.

Though I would not expect sudden changes in teachers’ attitudes or assessment practices, over the course of this study, findings revealed that participants possessed a new outlook, not only toward utilizing the multiple-choice format itself but toward future training in this area. Some participants indicated that they had gained a new respect for the format, demonstrating an openness towards participating in future training sessions on multiple-choice assessment. For other participants, compelling changes were evidenced through an explicit desire to not only further their knowledge and understanding of the multiple-choice format but to put forth greater effort to develop multiple-choice questions that are valid, reliable, discriminatory, cognitively challenging, preparatory for postsecondary education, and reusable over time.

For the most part, findings revealed that participants had gained an understanding of theoretical concepts and advice pertaining to the multiple-choice format. Participants expressed a strong desire for a more experiential and hands-on learning approach to training so that they might be able to apply the knowledge they had been given. Studies (e.g., Alkhawaldeh, 2011; Shawer, 2010) suggest that a hands-on approach to learning is important in pursuing effective and lasting understandings. A brief tour of the Ontario Ministry of Education and the Council of Ontario Directors of Education websites reveals documents (e.g., The Journal of Literacy and Numeracy for Ontario, CODE Advisories)
demonstrating a strong commitment to increased and more purposeful professional development for Ontario teachers.

Findings from the study similarly suggest that future multiple-choice affiliated training that is less theoretical and more practical in nature may legitimize the limited amount of time that teachers would end up spending in training. Tang (2010) similarly affirms this, indicating that expert knowledge is developed through the process of transforming theory into practical knowledge. Additionally, while the time and money saved in administering multiple-choice assessments could be viewed as a secondary benefit, the amount of material that can be covered on a multiple-choice test increases the value of even small gains made through this format, rendering it much more feasible and relevant.

Participants indicated that the 2- to 3- hour workshop was informative, but insufficient and undesirable as a standard format for ongoing training. Still, this workshop would undoubtedly benefit those teachers who have not yet received any form of multiple-choice training, as is suggested by Guskey (2002), Persellin and Goodrick (2010), and Shriner et al. (2009).

Overall, findings from this study suggest that participants appeared to be acutely aware of the fact that more training would be necessary if one were intent on acquiring greater depth of knowledge and skills on multiple-choice assessment. Furthermore, training that may prove to be most beneficial would need to apply teachers’ existing theoretical knowledge within a subject-specific context amongst colleagues. This would be accomplished not only to maximize resources but to capitalize on the rarity of having a concentrated and collaborative teacher effort to create appropriate and effective questions
for students. Though many participants realized that this would initially require a substantial investment of their time and effort, to many, the value of the projected long-term payoff would far exceed initial costs. In the meantime, teachers seemed to suggest that this process has led to improvements to both their current assessment practices and overall pedagogy.

**Implications**

In this section, the implications of the findings with regard to teachers’ assessment practices and overall pedagogical approaches are examined. As well, implications for further research are discussed. Considerable insight can be drawn from these areas, and these insights will provide support in establishing a base from which continued growth and an improved and effective use of the multiple-choice format can take place.

**Implications for Teachers’ Assessment and Teaching Practices**

The findings of this study indicated that many participants held a more favourable and supportive stance on multiple-choice assessment following their participation in this study and they could also see this assessment format being applied in practical, innovative, and ultimately successful ways.

One of the central findings of this study was that, while most participants had administered multiple-choice assessments, few used these tasks for formative assessment purposes. In fact, there appeared to be an overwhelming sense that multiple-choice questions have become relegated solely to evaluative purposes and that participants had not considered using these questions as an assessment for learning. Participants, for the most part, reported that they had rarely considered using or even seen multiple-choice questions outside of a unit-ending summative or large-scale assessment environment, let
alone as a purposed formative assessment tool. In order to fully capitalize on the strengths of this format, teachers needed to be aware of exactly when, where, and how they could and should rely upon these strengths. The ability to implement a format such as the immediate feedback assessment technique (IFAT; DiBattista, 2005) or brief activities and quizzes would enable teachers to better grasp and diagnose the abilities of their students. At the same time, teachers would be able to establish formative and more comprehensive learning experiences that may, in turn, help to tailor their specific methods of instruction to students while also delivering accurate future assessments. Teachers could benefit greatly from the versatility and formative potential of the multiple-choice format. Examples of this versatility include the implementation of interactive student response systems (i.e., Clickers; Gupta, 2010) to having students themselves create multiple-choice questions in order to engage with course content, apply metacognitive skills, and take ownership of their learning (Pittenger & Lounsbery, 2011). Multiple-choice assessment still has room for innovation and creativity and needs to be established as more than just a form of summative assessment.

Other findings indicate that the teachers in this study now understood that results gleaned from multiple-choice questions need not be relegated to Bloom’s lowest levels of knowledge and recall. Participants had not previously considered how they might use the multiple-choice portions of their assessments to engage students in varying levels of cognition, as studies (e.g., Hancock, 1994; Torres et al., 2011) have shown. While further training would be required in order for them to develop the necessary understandings and skills to accomplish this, participants demonstrated a genuine interest in participating in such an activity. This shift in perception supports the integrity of the multiple-choice
format in its ability to reliably assess varying levels and manners of thought and should pique teachers’ interest and warrant meaningful use from their assessment tool belts. Some participants even suggested that the multiple-choice format could perhaps benefit unique student thought processes that had previously presented challenges or difficulties in the classroom. Some participants also felt that multiple-choice questions could represent potential solutions in developing effective assessments in the areas of ESL support, student success, and differentiated instruction. The multiple-choice format might be able to not only enhance the average student’s learning experience, but also tackle specialized situations and requirements that had been previously challenging for teachers.

One of the most fascinating and perhaps the most visible way in which the multiple-choice format could impact education today, would be in its role in relation to the emergent push to have learner-centered technology—essentially any form of technology that allows the student to take ownership and feel responsible for her or his learning (An & Reigeluth, 2011)—integrated into the classroom. Gupta (2010) found the multiple-choice format when coupled with this kind of technology to be an effective formative assessment tool. It is therefore encouraging to see—due in large part to its simple and automated grading abilities (Ventouras et al., 2009)—that this affinity between the technology oriented classroom, education’s increasingly online presence, and the multiple-choice format, is gradually being recognized by educators. The ability of the multiple-choice format to effectively integrate technology opens up many possibilities for teachers in actual classrooms. In particular, this potential remains wide open for those teachers who are able and who choose to creatively take advantage of the capabilities of
technology to efficiently deliver high volumes of material and provide immediate feedback to students.

Overall, the implications of this research are far-reaching and require further thought and elaboration. It has been shown that the multiple-choice format is adaptable in both its use in a variety of contexts and the numerous ways in which its marking scheme can be modified. Multiple-choice assessment is also openly embraced for its efficacy. This form of assessment could eventually impact the way in which teachers approach their preparation and delivery of all assessments and could eventually improve the way in which teachers plan for and approach the development of their course curricula.

This study has shown that many teachers are lacking multiple-choice assessment training. Therefore, all teachers should be afforded opportunities to participate in multiple-choice workshops and training. Furthermore, if the desire is to have teachers embrace and capitalize on the strengths presented by the multiple-choice format in order to improve the quality of their questions, an accountability system and a more structured approach to teachers’ curricula and assessments will be needed. It is worthwhile noting that one school did have a heavily involved administrator in the entire process and “coincidentally” had the greatest number of teacher participants and the most professional development experience in the area of multiple-choice assessment. Research seems to support this, as Volante (2009) suggests that the manner in which administrators lead their schools can impact the school’s direction and success and that administrators might consider helping teachers by initiating the development of cohesive professional learning communities rather than by coaching teachers individually, something that Popham (2009) agrees can be quite effective.
Implications for Further Research

The findings of this study have sparked interest regarding the potential of the multiple-choice format. After taking into account some of the widely agreed upon strengths of the multiple-choice format, such as grading efficiency, course content coverage, and delivery of feedback, participants acknowledged that there is still much to learn from this form of assessment. The findings below suggest several areas that could profit greatly from extended research.

In this study, some of the weaknesses of the multiple-choice format expressed by participants were somewhat unclear. Results from previous research (e.g., Haladyna et al., 2002, Hancock, 1994; Nicol, 2007; Tarrant & Ware, 2008) as well as data gathered from within this study suggest that there are various misconceptions attributed to multiple-choice assessment, including (a) its perceived rigidity as a selected response form of assessment and (b) its perceived inability to assess beyond the lowest levels of cognition (i.e., remembering and understanding from Bloom’s RT) and students’ subsequent lack of exposure to multiple-choice questions that assess higher order thinking. The fact that many participants believed these weaknesses or misconceptions to be entirely legitimate is worthy of further investigation.

While some research exists in relation to the potential for multiple-choice assessment to reveal students’ cognitive abilities, there is little that provides a comprehensive analysis of the ability of multiple-choice questions to assess higher levels of thinking and learning. The closest studies matching this ideal (e.g., Fellenz, 2004; Jensen et al., 2006; Pittenger & Lounsbery, 2011) focused primarily on the formative side of assessment. These studies involved having students create their own multiple-choice
questions related to course content, resulting in higher ordered and varied ways of thinking as questions were constructed. While modifying the multiple-choice format to more effectively target higher level thinking has been tried and has generated some success (Briggs et al., 2006; Haladyna et al., 2002), there remains a paucity of studies documenting the ability of multiple-choice questions to successfully assess the various levels of cognition across most disciplines using both formative and summative forms of assessment.

Participants often stated that they were not the only ones lacking exposure and training in relation to the multiple-choice format. They believed that their students lacked experience or knowledge of effective approaches in reading and responding to multiple-choice questions. Participants indicated that they knew students were often confused or were making “poor” non-content-related decisions in their responses, suggesting that students could benefit not only from further exposure to multiple-choice questions but from dedicated instruction on how to logically approach multiple-choice questions. Further research investigating the benefits of consistent and more frequently delivered multiple-choice assessment throughout the school year accompanied by dedicated lessons on how to appropriately process multiple-choice questions is needed.

The findings of this study further indicate that multiple-choice assessment was, and still remains, very popular among teachers. Interestingly, it was also found that due to an increased affinity between emerging technologies, large-scale assessment formats, and value-laden time – cost efficiencies, multiple-choice assessment appears primed for an increase in future usage. Though not all participants recognized the potential of this facet of multiple-choice assessment, further research into the practical integration of
technology into the classroom (An & Reigeluth, 2011) and, in particular, the ability to capitalize on the affinity between technology and the multiple-choice format as both a formative (i.e., Farrell & Leung, 2004) and summative assessment tool is required.

In the end, while this study presented qualitative findings, the appointed length of this study and the accompanying workshop along with the number of teacher-participants did not easily allow the collection of quantitative measures. In addition, though participants from three distinct schools within the same school board took part in this study, this research would likely benefit teachers across different schools, boards, and geographical regions. Moreover, a longitudinal study could be conducted in order to determine the long-term impact that a multiple-choice workshop could have on teachers’ assessment practices. Research extending upon and solidifying the findings of this study is needed to further verify and fully maximize the potential of multiple-choice assessment.

All of this potential research will most importantly continue to improve the effectiveness of both teachers’ overall pedagogical approaches and students’ capacities for learning and success in both their present and future academic careers.

**Final Thoughts**

Assessment is undeniably an essential part of the teaching and learning process in education. With effective assessment comes a better understanding of student learning and, subsequently, an ever-evolving and accurate blueprint on how we can support students to their full potential. Teachers’ time and resource constraints have brought a seemingly unending struggle to efforts for well-developed and effectively administered assessment. We cannot simply concede. The underlying motivation for this study
alongside existing research efforts suggests that perhaps multiple-choice assessment can begin to make headway against these constraints. While it is far from a perfect form of assessment—arguably, there is no such holy grail—when multiple-choice testing is used appropriately, research shows that it can truly assess students at a level of efficacy that is comparable to many other common forms of assessment. The findings of this study show that teachers are familiar with the multiple-choice assessment format, yet many misunderstand its intent and abilities as a successful piece of the assessment puzzle. Teachers may not be capitalizing on the strengths of the multiple-choice format and may instead be focusing on the “traditional” and misconstrued views about this form of assessment. This study has further shown that this, however, has not marked the demise of multiple-choice. Teachers are ready and willing to embrace this format for what it is. It is important for all educators and administrators to continue supporting one another in effectively utilizing this form of assessment; especially given the fact that most teachers already use this method, albeit often suboptimally. Teachers can be engaged through thoughtful, hands-on experiential training. This research provides some insight into how teachers relate to multiple-choice assessment and some motivation for frequent and effective multiple-choice training in the future. It is hoped that this study also helps to demonstrate that multiple-choice assessment can be used as an effective teaching tool in the hands of many secondary school teachers.
References


doi:10.1080/02602930701773091


doi:10.1007/s12129-007-9031-9


doi:10.1111/1467-9620.00242


Cappel, J. (2007). And the answer is…*Education Canada, 47*(2), 41.


doi:10.1177/0017896907080136


doi:10.1177/014662.1608.320523


Odegard, T. N., & Koen, J. D. (2007). “None of the above” as a correct and incorrect alternative on a multiple-choice test: Implications for the testing effect. *Taylor and Francis Group. 15*(8), 873–885. doi:10.1080/09658210701746621


doi:10.1046/j.1365-2923.2001.00003.x


Appendix A

Interview Questions

PRE

1. (a) What is the role of assessment in your class?
   (b) What do you perceive to be the role of multiple-choice assessments as they relate to your curriculum outcomes?

2. What do you believe are the limits to multiple-choice assessment? What makes it easy? Hard?

3. What are student reactions like typically following a multiple-choice test?

4. Describe the process you go through when constructing multiple-choice tests. What are you trying to accomplish when you create distractors?

5. What else would you like to see happen in your multiple-choice assessment practice, and in what areas would you like to learn about the use of this format?

POST

1. How has your perception towards multiple-choice assessment changed, if it has changed?

2. Overall, how has your level of confidence changed, if it has, in accurately assessing students?

3. Have your students been responding differently to your tests? To anything else?

4. How has your process in creating and implementing multiple-choice tests changed?
Appendix B

Survey

Name: ____________________________________  Years taught: __________

Grades taught: _________________________________

How many years at each grade level: _________________________________________

Which faculty of education did you attend: ____________________________________

1. What subjects do you currently teach and at what grade levels?
   ________________________________________________________________

2. Out of these subjects, for which do you use multiple-choice questions? ALL
   If not all subjects, please specify which subjects:
   ________________________________________________________________

3. Have you ever attended or led an assessment-related course or workshop?
   YES   NO
   If YES, please specify the nature and duration of course/workshop and provide a
   brief description:
   ________________________________________________________________
   ________________________________________________________________
   ________________________________________________________________

For all of the following questions except #7, please circle the answer that best applies to
you.

4. Do you prefer using
   (a) Selected-response questions (i.e., multiple-choice, true-false, matching)
   or
(b) Constructed-response questions (i.e., short answer, essay)

5. List at least 3 adjectives you feel best describe your perceptions/feelings towards multiple-choice questions (ex. fun, difficult, easy, boring, invalid, boring, etc.)

__________________________________________________________________

6. Altogether, approximately how many times during a semester do you give quizzes/tests/exams that employ multiple-choice questions?

< 3  3 < 5  5 < 7  7 < 9  10+

7. In your opinion, are multiple-choice tests easy to create and implement?

Not at all  Sometimes  Most of the time  All of the time

8. On average, how long does it take you to come up with 20 multiple-choice questions?

______ minutes

9. In terms of the level of difficulty, how do you think students perceive your multiple-choice questions?

Very difficult  Somewhat difficult  Not too difficult  Not difficult at all

10. When implementing multiple-choice questioning, do you feel you are able to accurately assess student learning?

Not at all  Sometimes  Most of the time  All of the time

11. On average, how long do you expect each student to spend on each question?

< 30 sec  30 > 60 sec  1 > 2 minutes  2+ minutes

12. On average, approximately what percentage of your tests is composed of multiple-choice questions?
13. Which types/levels of cognitive processes do you assess through multiple-choice questions? (Circle all that apply)

Remembering  Understanding  Applying  Analyzing  Evaluating  Creating

14. What type/level of knowledge do you typically assess? (Circle all that apply)

Factual  Conceptual  Procedural  Metacognitive

15. How many response options do you typically have?

2  3  4  5+

16. How many of your distractors (incorrect response items) are typically plausible?

1  2  3  4+

17. Do you try to ensure that your correct answer stands out from the rest of the distractors?

Not at all  Sometimes  Most of the time  All of the time

18. How often do you use all of the above or none of the above?

Not at all  Sometimes  Most of the time  All of the time

19. How often do you negatively phrase either questions or response items? (e.g., using the words NOT or EXCEPT)

Not at all  Sometimes  Most of the time  All of the time

20. Do you feel your questions align closely with the manner in which you have been teaching and presenting curricular content?

Not at all  Sometimes  Most of the time  All of the time
Appendix C

Constructing Valid Multiple-Choice Questions

Joe Engemann, PhD
Derek Oei, MEd Graduate Student
Brock University, Faculty of Education

Contact Information
engemann@brocku.ca (905) 688.5550 ext. 3943
derek.oei@gmail.com
**Glossary of Assessment Terminology**

**Assessment** is the gathering of reliable information pertaining to students’ knowledge and understanding of critical concepts or achievement in relation to the specific curriculum expectations. Phases of assessment include diagnostic, formative, and summative.

**Evaluation** is the process of making judgments based on assessment data from a variety of sources in order to quantify achievement or progress.

**Traditional Assessments** tend to assess content knowledge. Examples of traditional assessments include selected-response and constructed-response tests.

- **Selected-Response Item Tests** - students attempt to choose a pre-determined correct answer (True/False, Multiple-Choice, Matching Items).
- **Constructed-Response Item Tests** (Short-Answer, Essays).

**Taxonomy** is a system for naming and organizing things into groups that share similar characteristics.

**Reliability** is the extent to which a test produces consistent, reproducible measures of performance.

**Validity** is the degree to which a test measures what it is intended to measure and is aligned with instructional practice.

**Item difficulty** is the proportion of examinees that answered the question correctly.

**Item discrimination** is a calculation of how a test item discriminates between individuals who have scored high on the overall test and those who have scored low on the overall test.

**Grading** is translating evaluation information into letters, numbers, or marks to indicate the quality of students’ performances and to communicate meaningful information about students’ learning and achievements.

**Marking** is assigning a number or letter score to a particular piece of student work or performance.
Anatomy of a Multiple-Choice Question

The following is an example of a typical multiple-choice question. The question stem is the lead-in to the question. Among the four answer options, there are three distractors (incorrect yet plausible answers) and one correct response (the best answer to the question stem).

As an educator, Heather believes that individuals learn best when they actively build knowledge and understanding in light of their own experiences. Heather is a

A) Behaviourist  
B) Constructivist  
C) Evolutionist  
D) Socialist

Stem

The stem should provide a complete idea of the knowledge to be indicated in selecting the right answer.

Correct Response

The correct response is the one and only right choice. In a question format, the correct choice can be a word, phrase, or sentence. In some cases, it can be a paragraph, drawing, or photograph (provided the distractors are also paragraphs, drawings, or photographs). With the incomplete stem format, the second part of the sentence is the option, and one of these is the right answer. With the best-answer format, all options are correct, but only one is unarguably the best.

Distractors

Each distractor must be plausible to the test-takers who have not yet learned the knowledge or skill that the test item is supposed to measure. To those who possess the knowledge, the distractors are clearly wrong choices. Each distractor should resemble the correct response in grammatical form, style, and length. Subtle or blatant clues that give away the correct choice should always be avoided.
Multiple-Choice Formats

Conventional MC Format

There are three variations of this format:

Question Format

Who is John Galt?
A. rock star
B. movie actor
C. character in a book

Completion Format

John Galt is a character in an Ayn Rand novel who is remembered for his
A. integrity
B. romantic tendencies
C. courage

Best Answer Format

Which is the most effective safety feature in your car?
A. seat belt
B. front air bag
C. anti-lock braking system

Best answer format items are best followed by a constructed response part that asks the test-taker to justify his/her response. Test-takers with limited English proficiency may have more difficulty with the completion format since they have to retain the stem in short-term memory while completing the stem with each option.

Completion format items that include a blank space in the middle or at the beginning of the stem are problematic because they require the retention of stem information in short-term memory. This is particularly problematic with the so-called blanket-blank format as shown below.

Child abuse is an example of __________ violence, whereas sexism is an example of __________ violence.

A. aggressive; structural
B. emotional; psychological
C. structural; emotional
**Matching Format**

This format uses a set of options that seems useful for two or more items. The matching format begins with a set of options at the top followed by a set of stems below. The instructions that precede the options and stems tell the test-taker how to respond and where to mark answers. The example below has five options and six statements. The list of statements could be expanded, which makes the set of items more comprehensive in testing student learning.

Mark your answer on the answer sheet.

For each item select the correct answer from the options provided below.

A. Minnesota  
B. Illinois  
C. Wisconsin  
D. Nebraska  
E. Iowa

1. Home state of the Hawkeyes  
2. Known for its cheese heads  
3. Land of many lakes  
4. Cornhuskers country  
5. The largest of these states  
6. Contains Cook County

The following are the best contexts for matching items:
- persons and achievements  
- dates and events  
- terms and definitions  
- rules and examples  
- symbols and concepts  
- authors and books  
- English and non-English equivalent words  
- machines and uses  
- plants or animals and classification  
- principles and illustrations  
- objects and names of objects  
- parts and functions
Extended-Matching Format

An extended-matching format has four components:

a. a lead-in statement
b. a question
c. a set of stems
d. a set of options

The lead-in statement might be a scenario or vignette. This puts the problem in a real-life context. The set of stems should be independently answered.

A 54-year-old man has a year’s history of steadily progressive personality changes. He has become increasingly apathetic and appears depressed. His main complaint is increasing frontal headaches. On examination, he has word-finding difficulties. EEG shows frontal slowing that is greater on the left.

Which test (Options A-J) would you consider using for each situation described below?

1. You are concerned that he may have an intracranial space-occupying lesion.
2. Test indicates that his current performance IQ is in the low average range.
3. The estimate of his premorbid IQ is 15 points higher than his current performance IQ. It is recommended that he has a full WAIS IQ assessment to measure both performance and verbal IQ. On the WAIS, his verbal IQ is found to be impaired over and above his performance IQ. Which test is part of the WAIS verbal subtests?
4. An MRI can show a large meningioma compressing dorsolateral prefrontal cortex on the left. Which test result is most likely to be impaired?

Options

A. Cognitive Estimates Test
B. Digit Span
C. Go-No Go Test
D. Mini Mental State Examination
E. National Adult Reading Test
F. Raven’s Progressive Matrices
G. Rivermead Behavioural Memory Test
H. Stroop Test
I. Wechsler Memory Scale
J. Wisconsin Card Sorting Test
**Multiple True-False Format**

The multiple true-false format offer a stem and a number of options that are either true or false.

Mark each of the options below T (true) or F (false).

The set of scores 4, 5, 5, 7, 8, 9, 9, 15 has

A. a median of 7.5.
B. a symmetric shape.
C. more than one mode.
D. a larger mean than median.

**Context-Dependent Items Sets**

The context-dependent item set has an introductory stimulus and usually 2 to 12 test items related to this stimulus. The stimulus for any item set might be a work of art, photograph, chart, graph, figure, table, written passage, poem, story, cartoon, problem, experiment, narrative, or reference to an event, person, or object. The item set is well suited to testing problem solving and higher-level thinking.

Types of context-dependent item sets include reading comprehension, problem solving, pictorial, and interlinear.
Reading Comprehension

The reading comprehension item set shown below presents a poem for elementary grade language arts students and asks questions to measure student understanding of the poem. Within a reading comprehension item set, some items might systematically ask for the meaning of words, phrases, or the entire passage. Some items might ask for prediction (e.g., what should happen next?). Other items might analyze characteristics or plot.

“The radiance was that of full, setting, and blood-red moon, which now shone vividly through that once barely discernible fissure of which I have before spoken as extending from the roof of the building, in a zigzag direction, to the base. While I gazed this fissure rapidly widened—there came a fierce breath of the whirlwind—the entire orb of the satellite burst at once upon my sight—my brain reeled as a saw the mighty walls rushing asunder—there was a long, tumultuous shouting sound like the voice of a thousand waters—and the deep and dank tarn at my feet closed sullenly and silently over the fragments of the House of Usher.”

1. What is Poe referring to when he speaks of “the entire orb of the satellite”?
   A. The sun
   B. The moon
   C. His eye

2. What is a “tarn”?
   A. A small pool
   B. A bridge
   C. A marsh

3. How did the house fall?
   A. It cracked into two pieces.
   B. It blew up.
   C. It just crumbled.

4. How did the speaker feel as he witnessed the fall of the House of Usher?
   A. Afraid
   B. Awestruck
   C. Pleased

5. What does the speaker mean when he said “his brain reeled”?
   A. He collected his thoughts.
   B. He felt dizzy.
   C. He was astounded.
Problem Solving

The example below contains an item set in science. The stimulus is a scientific experiment involving a thermos bottle and some yeast, sugar, and water. The questions involve the application of principles of science. Each item tests a different step in problem solving. Item 1 asks the student to apply a principle to predict what happens to the temperature of the water. Item 2 gives the reason for this result. All four options are judged to be plausible. Item 3 calls for a prediction based on the application of a principle. Item 4 addresses possible changes in sugar during a chemical reaction. Item 5 tests another prediction based on this chemical reaction.

A thermos bottle is filled with a mixture of yeast, sugar, and water at 15 degrees C and the contents are examined 24 hours later.

1. What happens to the temperature?
   A. Increases
   B. Stays the same
   C. Decreases

2. What is the reason for that result?
   A. Yeast plans respire.
   B. Yeast plants do not respire.
   C. Yeast plants absorb heat in order to live.
   D. Heat cannot be conducted into or out of the thermos bottle.

3. What has happened to the number of yeast plants?
   A. Increased
   B. Decreased
   C. Remained about the same

4. What has happened to the amount of sugar?
   A. Increased
   B. Decreased
   C. Remained the same

5. What has happened to the contents of the thermos bottle?
   A. Increased in oxygen
   B. Decreased in oxygen
   C. Increased in carbon dioxide
   D. Decreased in carbon dioxide
**Pictorial**

The example below provides an example of a table showing the number of participants and number of injuries for 10 sports. Test items can be written to test one’s understanding of the data and inferences that can be made from these data.

<table>
<thead>
<tr>
<th>SPORT</th>
<th>INJURIES</th>
<th>PARTICIPANTS&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Basketball</td>
<td>646,678</td>
<td>26.2</td>
</tr>
<tr>
<td>2. Bicycle riding</td>
<td>600,649</td>
<td>54.0</td>
</tr>
<tr>
<td>3. Baseball, softball</td>
<td>459,542</td>
<td>36.1</td>
</tr>
<tr>
<td>4. Football</td>
<td>453,648</td>
<td>13.3</td>
</tr>
<tr>
<td>5. Soccer</td>
<td>150,449</td>
<td>10.0</td>
</tr>
<tr>
<td>6. Swimming</td>
<td>130,362</td>
<td>66.2</td>
</tr>
<tr>
<td>7. Volleyball</td>
<td>129,839</td>
<td>22.6</td>
</tr>
<tr>
<td>8. Roller skating</td>
<td>113,150</td>
<td>26.5</td>
</tr>
<tr>
<td>9. Weightlifting</td>
<td>86,398</td>
<td>39.2</td>
</tr>
<tr>
<td>10. Fishing</td>
<td>84,115</td>
<td>47.0</td>
</tr>
</tbody>
</table>

<sup>1</sup>Reported in millions

1. Which sport has the greatest number of participants?
   A. Basketball
   B. Swimming
   C. Bicycle riding

2. Which sport has the least number of injuries?
   A. Football
   B. Weightlifting
   C. Fishing

3. Which of the following sports has the highest injury rate per number of participants?
   A. Basketball
   B. Bicycle riding
   C. Roller skating
Interlinear

This item set does not take up a great amount of space but cleverly gets the test-taker to choose between correct and incorrect grammar, spelling, capitalization, and punctuation.

For each numbered, underlined pair of choices, choose the letter next to the correct spelling of the word and fill in your answer sheet with that letter next to the number of the item.

There (1. A. our or B. are) many ways to invest money. You can earn (2. A. intrest or B. interest) by buying savings bonds. Or you can (3. A. bye or B. buy or C. by) corporate bonds. Or you can become a (4. A. part-owner or B. partowner) of a company by owning stock in a company. As a shareholder in a company, you can share in company (5. A. profits or B. prophets).
Guidelines for Writing Multiple-Choice Questions

The following guidelines are a synthesis of 30 years of assessment research.

**Content Guidelines**

1. Every item should reflect specific content and cognitive processes as called for in the table of specifications.
2. Base each item on important content to learn; avoid trivial content.
3. Use novel material to measure understanding and the application of knowledge and skills.
4. Keep the content of an item independent from content of other items on the test.
5. Avoid overspecific or overgeneral content.
6. Avoid unqualified opinion-based items.
7. Avoid trick items.

**Style and Format Concerns**

8. Format items vertically instead of horizontally.
9. Edit items for clarity.
10. Edit items for correct grammar, punctuation, capitalization, and spelling.
11. Simplify vocabulary so that reading comprehension does not interfere with testing the content intended.
13. Proofread each item.

**Writing the Stem**

14. Make directions as clear as possible.
15. Make the stem as brief as possible.
16. Place the main idea of the item in the stem, not in the choices.
17. Avoid irrelevant information.
18. Avoid negative words in the stem.

**Writing Options**

19. Develop as many effective options as you can.
20. Assign the position of the right answer randomly.
21. Keep options independent; choices should not be overlapping.
22. Keep the options homogenous in content and grammatical structure.
23. Keep the length of options about the same.
24. “None of the above” should be used sparingly.
25. Avoid using “All of the above.”
26. Avoid using negative words such as “not” or “except.”
27. Avoid options that give clues to the right answer.
28. Make distractors plausible.
29. Use typical errors of students when writing distractors.
## Table of Specifications for Test on Weather

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Knowledge/Understanding</th>
<th>Inquiry</th>
<th>Communication</th>
<th>Making Connections</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discriminate between <em>weather</em> and <em>climate</em></td>
<td>5%</td>
<td>5%</td>
<td>5%</td>
<td>15%</td>
<td></td>
</tr>
<tr>
<td>Predict weather patterns using data from weather reports</td>
<td></td>
<td>20%</td>
<td></td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>Describe the water cycle</td>
<td>5%</td>
<td></td>
<td>5%</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>Define terms such as <em>temperature</em>, <em>precipitation</em>, <em>humidity</em>, <em>wind chill factor</em>, <em>barometric pressure</em>, and <em>cloud cover</em></td>
<td>10%</td>
<td></td>
<td>5%</td>
<td>15%</td>
<td></td>
</tr>
<tr>
<td>Summarize how weather conditions affect the activities of humans and animals and differentiate the impact of these conditions among various human societies.</td>
<td>20%</td>
<td>20%</td>
<td></td>
<td>40%</td>
<td></td>
</tr>
</tbody>
</table>
Bloom’s Revised Taxonomy of Educational Objectives: The Knowledge and Cognitive Domain

The unidimensional nature of the Taxonomy was changed with its revision (Anderson et al., 2001) by separating the knowledge and cognitive process aspects into two dimensions. In the revised Taxonomy, the subject matter content is found within the Knowledge dimension and the description of what is to be done with that knowledge is located within the Cognitive Process dimension.

The Knowledge dimension contains four categories, which are described in the following table:

<table>
<thead>
<tr>
<th>Knowledge Category</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
</table>
| Factual Knowledge        | The basic elements that students must know to be acquainted with a discipline or to solve problems in it. | • Knowledge of terminology  
                          |                                                                                     |   • Knowledge of specific details and elements |
| Conceptual Knowledge     | The interrelationships among the basic elements within a larger structure that enable them to function together. | • Knowledge of classifications and categories  
                          |                                                                                     |   • Knowledge of principles and generalizations  
                          |                                                                                     |   • Knowledge of theories, models, and structures |
| Procedural Knowledge     | How to do something; methods of inquiry, and criteria for using skills, algorithms, techniques, and methods. | • Knowledge of subject-specific skills and algorithms  
                          |                                                                                     |   • Knowledge of subject-specific techniques and methods  
                          |                                                                                     |   • Knowledge of criteria for determining when to use appropriate procedures |
| Metacognitive Knowledge  | Knowledge of cognition in general as well as awareness of one’s own cognition.                 | • Strategic knowledge  
                          |                                                                                     |   • Knowledge about cognitive tasks, including appropriate contextual and conditional knowledge  
                          |                                                                                     |   • Self-knowledge |


The following table describes the structure of the *Cognitive Processes* dimension of the Revised Taxonomy:

| 1.0 Remember | Retrieving relevant knowledge from long-term memory. | 1.1 Recognizing  
1.2 Recalling |
|--------------|------------------------------------------------------|------------------|
| 2.0 Understand | Determining the meaning of instructional messages, including oral, written, and graphic communication. | 2.1 Interpreting  
2.2 Exemplifying  
2.3 Classifying  
2.4 Summarizing  
2.5 Inferring  
2.6 Comparing  
2.7 Explaining |
| 3.0 Apply | Carrying out or using a procedure in a given situation. | 3.1 Executing  
3.2 Implementing |
| 4.0 Analyze | Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose. | 4.1 Differentiating  
4.2 Organizing  
4.3 Attributing |
| 5.0 Evaluate | Making judgments based on criteria and standards. | 5.1 Checking  
5.2 Critiquing  
5.3 Justifying |
| 6.0 Create | Putting elements together to form a novel, coherent whole or make an original product. | 6.0 Generating  
6.1 Planning  
6.2 Producing |
Key Verbs for Use with Bloom’s Revised Taxonomy

**Level 1: Remember** – exhibits previously learned material by recalling facts, terms, basic concepts, and answers.

*Key verbs:* count, define, describe, draw, enumerate, find, identify, label, list, match, name, quote, read, recall, recite, record, reproduce, select, sequence, state, tell, view, write

**Level 2: Understand** – demonstrating understanding of facts and ideas by comparing, translating, interpreting, giving descriptions, and stating main ideas.

*Key verbs:* classify, cite, conclude, convert, describe, discuss, estimate, explain, generalize, give examples, illustrate, interpret, locate, make sense of, paraphrase, predict, report, restate, review, summarize, trace, understand

**Level 3: Apply** – solving problems by applying acquired knowledge, facts, techniques, and rules in a different way.

*Key verbs:* act, administer, apply, articulate, assess, change, chart, choose, collect, compute, construct, contribute, control, demonstrate, determine, develop, discover, dramatize, draw, establish, extend, imitate, implement, interview, include, inform, instruct, paint, participate, predict, prepare, produce, provide, relate, report, select, show, solve, transfer, use, utilize

**Level 4: Analyze** – examining and breaking information into parts by identifying motives or causes; making inferences and finding evidence to support generalizations.

*Key verbs:* analyze, break down, characterize, classify, compare, contrast, correlate, debate, deduce, diagram, differentiate, discriminate, distinguish, examine, focus, illustrate, infer, limit, outline, point out, prioritize, recognize, research, relate, separate, subdivide

**Level 5: Evaluate** – presenting and defending opinions by making judgments about information, validity of ideas, or quality of work based on a set of criteria.

*Key Words:* award, choose, conclude, criticize, decide, defend, determine, dispute, evaluate, judge, justify, measure, compare, mark, rate, recommend, rule on, select, agree, interpret, explain, appraise, prioritize, opinion, support, importance, criteria, prove, disprove, assess, influence, perceive, value, estimate, influence, deduct
**Level 6: Create** – compiling information together in a different way by combining elements in a new pattern or proposing alternative solutions.

*Key Words*: build, choose, combine, compile, compose, construct, create, design, develop, estimate, formulate, imagine, invent, make up, originate, plan, predict, propose, solve, solution, suppose, discuss, modify, change, original, improve, adapt, minimize, maximize, delete, theorize, elaborate, test, improve, happen, change
Cognitive Levelling of Multiple-Choice Questions

It is common practice to identify a category indicating the nature of cognitive challenge required to answer a multiple-choice question. Assessment authors (e.g., Masters et al., 2001) have postulated that only four of the six levels of Bloom’s Taxonomy are suitable for use with multiple-choice questions (i.e., remember, understand, apply, and analyze). These categories, as they apply to the cognitive levelling of multiple-choice questions, are described as follows:

Remember

This level is distinguished by the retrieval of knowledge in long-term memory in order to find the answer to the multiple-choice question that is consistent with the material presented in the textbook or lesson. Answering a remember-level question requires the recall of concepts, principles, and theories. The question stem and correct option for this level of question usually contains words and phrases identical to or synonymous with those found in the textbook or lesson. Occasionally, a remember-level question will be set in the context of a real-life example or contain extra information, but this addition is superfluous since the remaining text in the stem and/or correct option is such that only information retrieval is required. The following is an example of a remember-level multiple-choice question:

When referring to judgements people make about the causes of their own and other people’s behaviours, psychologists use the term:

A) attitudes  
B) stereotypes  
C) social norms  
D) attributions
Understand

This level is distinguished by selecting an answer in the multiple-choice question that represents

- the same answer as that in the textbook or lesson, but which is presented differently (e.g., presented numerically instead of verbally).
- a specific example or illustration of a concept or principle found in the textbook or lesson.
- an entity that belongs to a category presented in the textbook or lesson.
- a summarization of a general theme or a major point presented in the textbook or lesson.
- a logical conclusion drawn from information presented in the textbook or lesson.
- a comparison with a concept or principle found in the textbook or lesson.
- an explanation of a concept or principle found in the textbook or lesson.

Understand-level multiple-choice questions are written in a manner that requires more than recall of information from the textbook or lesson. Typically, an understand-level question is written in the form of a real-world example or illustration of a concept, principle, or theory where interpretation of the information in the question stem is required before the correct option can be selected. Understand-level questions also contain text that is paraphrased from the wording in the textbook. The following is an example of an understand-level multiple-choice question:

Juan is in the process of forming an opinion about someone when a friend who is taking a psychology class tells him to avoid making snap judgements and to carefully consider the evidence. The net result of this advice is that Juan feels more accountable for his opinions. The advice of Juan's friend should most likely decrease:

A) the recency effect
B) the primacy effect
C) the fundamental attribution error
D) the self-serving bias
Apply

This level is distinguished by selecting an answer in the multiple-choice question that represents the application of a procedure to a familiar or unfamiliar task (e.g., finding the solution to a word problem). Multiple-choice questions written for apply-level thinking are the most difficult of the four levels of question to construct and the most infrequently occurring within a test bank. This is due to the fact that such questions must demonstrate answer options that involve application of concepts, principles, and theories in a new situation or the demonstration of the correct use of a procedure. The following is an example of an apply-level multiple-choice question:

Lady Marion Seafood, Inc. sells 5-pound packages of Alaska salmon. Assume its variable cost per package is $30, and its fixed cost is $250,000. It wants a target profit of $38,000 on a volume of 16,000 packages. What should it charge for a five-pound package of salmon?

A) $25.00
B) $30.00
C) $40.00
D) $48.00
E) $55.00
Analyze

This level is distinguished by selecting an answer in the multiple-choice question that requires the test-taker to break down the concepts presented in the question stem, the multiple-choice answer options, or both, and relate how the parts relate to one another in order to determine the correct answer option. Analysis involves the breaking down of information so that the relationship among the parts may be understood and applied. It can also necessitate the recognition of fallacies in reasoning, the identification of cause-and-effect relationships, and the prediction of outcomes. Analyze-level multiple-choice questions typically involve higher-level reasoning in the selection of the correct answer option. The following is an example of an analyze-level multiple-choice question:

The manufacturer of a new kind of fat-free ice cream that has the consistency and taste of regular ice cream is thinking of using a skimming pricing strategy for its new product. Which of the following conditions would suggest using a skimming pricing strategy for the tasty fat-free ice cream?

A) The ice cream market is highly conservative.
B) A large portion of the market has inelastic demand for ice cream—over a fairly broad range of prices.
C) Economies of scale in production are substantial.
D) Retailers are willing to pay for new brands of premium ice cream in an extremely overcrowded category.
E) Once the initial price is set, it is nearly impossible to lower price because of the possibility of alienating early buyers.
Why do students struggle more with multiple-choice questions on the OSSLT than other formats?

Some possible reasons:

1. **Students are unfamiliar with the format and presentation, and therefore do not know how to approach the selection of their response.**
   
a. They are choosing answers that they think “sound” correct, rather than the answer they think carries the correct content.

   *i.e., students might be distracted by answers that perhaps sound more formal, comprehensive, or “intelligent”.*

b. They are having difficulty translating or matching their own thoughts to the answers being presented as multiple-choice response options. Compared to constructed-response situations, the manner in which students are asked to express their “unique” thoughts, is much more constrained.

c. They are not reading through all the options and are just choosing the first one they like, OR they might be following a “pattern” they think they think they have discovered in the questions.

d. They are unfamiliar with periodical literature (i.e., newspapers, magazines, reports, etc.), short stories, and graphic texts which are commonly used as the stimulus in order to test levels of reading comprehension.

2. **Two or more plausible response options are often listed. Selecting the “best” or “most correct” answer is difficult as students are not always accustomed to this format and/or distractors of this nature.**

   a. Students are sometimes easily swayed when presented with novel or interesting response options.

   b. When an explicit or direct answer is not available, students are unable to place plausible responses on a gradient from least correct to most correct.

   c. The uncertainty of not having *the* correct answer in front of them leads students to second guess themselves – which in turn often leaves them with an incorrect answer.
3. Most questions require rereading of the text in addition to added uncertainties (i.e., grammar, vocabulary, confusion, etc.) As a result, some students are simply running out of time and/or are rushing their decisions.

   a. For students of limited vocabulary and limited exposure to everyday English vernacular, uncommon English conventions found both in the text, and particularly in the questions, are the source of great difficulty.

   b. The combined effects of time constraints, external pressures, and difficulties with the questions as mentioned above, can lead to elevated levels of anxiety, which in general, lead to poor performances.
Looking at the OSSLT Multiple-Choice Questions

There are three main types of multiple-choice associated questions presented in the OSSLT. They occur at varying levels of cognition (Remember, Understand, Apply, Analyze according to Bloom’s Revised Taxonomy) and they are:

i)  **Reading Comprehension**


ii) **English Conventions and Grammar**

    *i.e., Punctuation, conventions, symbols, etc.*

iii) **Graphic Texts**

    *i.e., Navigating and understanding how graphics and text go together*

What distinguishes most of the questions found on the test from many English tests is not the question itself, but the *distractors* provided.

The *distractors* do not typically change the assessed level of cognition, but they often greatly increase or decrease the question’s level of difficulty.

The “most correct answer” provided as the correct response as it relates to the information provided, often determines the *level of cognition* invoked by the question.

*note that in the OSSLT, rarely is the cognitive domain ‘Remember’ assessed.*
Reading Comprehension

*Cognitive Level = CL
*Level of Difficulty = LD

1 Food scientists from Penn State University “ruined the jellybeans” when they tried to
   A add moisture.
   B improve the taste.
   C soften the candy’s centres.
   D perform tests and use chemicals.

   CL →
   LD →

8 Scoville makes dangerous dives because he wants
   F to show that he is a brave and skillful diver.
   G to discover valuable treasure on sunken ships.
   H the experience of seeing sunken ships first-hand.
   J additional practice using decompression techniques.

   CL →
   LD →

4 Which questions are answered in the first sentence of paragraph 3?
   F who and what
   G what and how
   H how and where
   J where and when

   CL →
   LD →

Researchers at the University of New Brunswick (UNB) have solved a mystery that has baffled candy-makers for more than 100 years. Confectioners wanted to understand why it takes up to two days for jellybeans to dry before they can be polished. They believed that jellybeans dried as moisture from the outer layers evaporated. Several years ago, candy-makers asked food scientists at Pennsylvania’s Penn State University to help find ways of speeding up the jellybean manufacturing process, but the scientists’ tests and chemicals ruined the jellybeans. Then, they heard about a lab in Canada that used magnetic resonance imaging technology, commonly known as MRI—often used to detect tumours in humans—to peer inside things like concrete, pharmaceuticals and wood for industrial applications. A Penn State scientist flew to Fredericton and made jellybeans in the lab. For almost three days, UNB researchers took MRI images of the insides of a jellybean, which showed waves of moisture moving toward the centre. Scientists at UNB had figured out why nothing had worked: the moisture in a jellybean travels in, not out. That is why a jellybean centre is moist and why the manufacturing process can’t be speeded up—a jellybean takes time to age to perfection.

You might call Don Scoville and Jim Kennard shipwreck detectives. Scoville, 33, a highly skilled technical diver and Kennard, 33, an electronics engineer and recreational diver, look for sunken historic ships. They had been looking for the remains of the Milan for years, after seeing a reference to the shipwreck in an old newspaper.

They searched for months using a sonar device and finally, in the summer of 2005, they saw a clear image of a ship with the masts still standing. But the ship was too deep to be explored with regular diving equipment.

A year later, in the fall of 2006, they returned to the wreck with an unmanned remote-operated vehicle (ROV). Pictures from the ROV were amazing—the masts, anchors, riggings and deck were all intact. But the ROV became tangled in a line, and they couldn’t bring it back up.

A diver would be dangerous. The wreck was so deep that the water pressure would be great. A diver would have to return to the surface very slowly, taking at least 45 minutes to decompress and avoid harm. To last 20 minutes at that depth in such extremely cold water, a diver would have to breathe a mixture of oxygen, nitrogen and helium. Only a specially trained and equipped diver like Don Scoville could reach the wreck safely.

To Scoville, the most exciting moment in his life was worth the risk. Searching for shipwrecks is “90 percent boredom and 10 percent total fun. When you find what you’re looking for—that’s your payoff,” says the brave diver.

For a few minutes he stood on the deck of the Milan, a ship that had sunk almost 160 years before. “It’s something no one’s seen since the 1840’s. To actually be there and seeing it with my own eyes… I can’t describe the feeling.” It’s one of the prettiest ships he’s ever seen on the bottom of the Great Lakes, Scoville claims. Modern technology could float it to the surface but it would cost tonnes of money and even more to preserve it once it was up. “Leave it where it is… keep it,” says Scoville.

Driven by the Sun: Solar Car Sets World Distance Record

On September 16, 2004, the Midnight Sun VII arrived home to the University of Waterloo (UW) to cheers from hundreds of supporters. After 40 days and 15,079 kilometres of travelling, the car had set the world record for the longest journey by a solar-powered vehicle. This distance broke the official Guinness record of over 7800 kilometres set by the Queen’s University solar car in July 2003 and the 2002 unofficial record from Australia of over 13,800 kilometres.

In 1987, the UW established a solar car project to provide renewable energy in transportation. The first UW solar vehicle circled the university campus. Seven rebuilds later, a student team made a score that could circle North America. The Midnight Sun VII rolled through seven Canadian provinces and 15 American states. “We drove through sun, torrential downpours, traffic congestion in the crowded streets of Los Angeles, the vast desert of Arizona, the Rocky Mountains and the open plains. It was an incredible journey,” wrote one team member.

The Midnight Sun VII is a low, sleek three-wheeled car that is propelled by electricity generated only by sunlight. Solar cells that cover much of the upper body of the car convert sunlight into electricity. When the car is moving, the power is sent directly to the motor. When the car is stopped, power is transferred to the batteries.

“We hope this tour has inspired people to recognize the potential of alternative forms of energy,” says Daniel Yun, the project team manager.
English Conventions and Grammar

5 Choose the best closing sentence for the paragraph.

When the Canadian government decided in 1878 to connect Montreal and Vancouver with a railway, it faced a difficult task. A railway would cost a great deal of money to build. Workers had to push through the rocks and swamps of northern Ontario. Tracks had to cross vast stretches of the prairie. It was uncertain whether suitable passes through British Columbia’s mountain ranges could be found.

A The loans weren’t repaid for many years.
B Many wondered if these obstacles could be overcome.
C British Columbia became a popular tourist destination.
D The United States had built many railways by this time.

2 Which of the following is written correctly?

a Samuel de Champlain a famous explorer lived in the 1600s.
b Samuel de Champlain a famous explorer, lived in the 1600s.
c Samuel de Champlain, a famous explorer lived in the 1600s.
d Samuel de Champlain, a famous explorer, lived in the 1600s.

1 Which is the best way to combine all the information in the following sentences?

Ontario is Canada’s largest province.
Ontario has many different physical regions.
Ontario’s wildlife is very diverse.

a Ontario, Canada’s largest province, has diverse wildlife and many different physical regions.
b Canada’s largest province is Ontario because it has different physical regions and diverse wildlife.
c Ontario has many physical regions because of the diverse wildlife and it is Canada’s largest province.
d Ontario is Canada’s largest province and it has many different physical regions but its wildlife is very diverse.
ii) Graphic Texts

SportsPlex Stadium Preliminary Design Proposal Submitted to City Council

This first drawing should give you an idea of the design for a multi-purpose sports facility. I will be happy to discuss hiring an engineer to develop a drawing to scale that will include the structure of the stadium’s frame. I am looking forward to future discussions with you about design details and cost.

Respectfully, N. Gilmour, Architect

LEGEND

Main seating area (20,000 seats)
Track area and extra seating (10,000 seats)
Main field
Seating area entrance
Sponsor logo (to be determined)
Megaview video screens
Nearby streets
Vaulehrooms
Wheelchair accessible
Concessions
Ticket booth and stadium entrance/exit
Parking
Bus transportation

COST CONSIDERATIONS

No Cost
- Land donation: value of $1.2 million
- Existing municipal parking lot

Cost of Current Design
- Current seating capacity (15,000): $4.6 million

Additional Cost Considerations
- Upper balcony seating: $800,000
- VIP box seats and lounge: $200,000
- Rooftop lights: approximately $100,000 (depending on roof materials)

2 In this selection, “Bird’s-Eye View” means
a a scale view.
b a detailed view.
c a structural view.
d an overhead view.

6 Which of the following details does N. Gilmour expect to be a topic for future discussions with city council?
   a the purchase of land
   b the availability of parking
   c the number of video screens
   d the cost of including VIP box seats
Multiple-Choice Questions and the OSSLT

*NB- Not all OSSLT questions are well-constructed. However, they do for the most part test above-average multiple-choice question levels of thinking, and present a greater level of difficulty (whether this is done fairly and/or equitably, is another question).

How to create questions of similar intent for your students (in no significant order):

1) Create your responses in a manner such that, what would be THE outright correct answer, is **not** one of the available options. Instead, phrase your correct answer in a manner in which students must infer, deduce, or make a connection from the information provided to the best/most correct answer you have provided.

   *i.e., You want to model your questions after the “choose the best or most correct answer” convention. For example, if the correct answer located in the text was ‘thief’, you could make the correct response ‘the person who was not the rightful owner’*

2) Create distractors that are actually indirectly, or partially correct, yet that would **not** be considered the “best” or “most correct” among the available options.

3) **All distractors** should attempt to reflect common student errors, misconceptions, and potential areas of confusion.

   *i.e., Create distractors which are perhaps plausible outside of the correct context or target information, but not plausible if the question is properly understood*

4) Test reading comprehension by tapping into students’ abilities to resolve and discern metacognitive knowledge, meaning, character traits, abstract thoughts, and higher level thinking taking advantage of textual implications and the surrounding context.

   *i.e., Why does this selection include a description of this? What was the intent of this project? Why would they do that first? Which of the following links these paragraphs together? What does this say about that?*

5) Test reading comprehension by getting students to use inductive reasoning to determine the meaning of uncommon vocabulary or phrases from context. Test also their understanding of the specific, or general, plot/storyline. Keep in mind when you are determining the order of the questions they do **not** need to be chronologically ordered nor presented in a predictable linear fashion.
6) Through reading comprehension as well, test the understanding and use of different English language structures and conventions.

i.e., format – font effects, paragraph, sentence, and thought sequencing, etc.
symbols – dashes, ellipsis dots, quotes, semi-colons, asterisks, etc.
grammar – use of pronouns, adjectives, verbs, etc.
literary devices – metaphors, similes, expressions, etc.

7) Test English conventions, syntax, and grammar in isolation. Attempt to test what could be seen as intuitive grammatical and topical conventions. Afterwards be sure to explain what makes this ‘intuition’ correct.

i.e., Which sentence is written correctly? (From a given paragraph) Which sentence doesn’t belong? Which sentence best summarizes these ideas?

8) Use graphic texts and ask questions which demonstrate an understanding of how the graphics content and structure enhance the text and information provided in the document.

i.e., Why are these groups of texts positioned in this manner? Why did they use a chart rather than a graph to represent this information?

9) In all areas, attempt to assess a student’s ability to extract and identify a concise and complete summary statement or idea. Check to see whether they understand the thematic and overarching point(s) of the text and/or graphic.

How to create questions of similar structure for your students:

1) Keep your question stems short and simple (~less than 15 words).

2) Try to have four response options (3 distractors + 1 correct answer).

3) Use texts/graphics that are from roughly 150, to 500 words in length.

4) Number your paragraphs in your text so that you can refer to specific sentences and paragraphs within the text from within your question.

   i.e., What is Bob’s intention in paragraph 5?

5) Mix in some ‘completion format’ type questions

   i.e., “The description of the way the men boarded the boat shows their…”

   Afterwards, be sure to explain to students the reasoning behind why the correct answers you have chosen are in fact the correct answers!
Multiple-Choice Question Item Analysis

The accuracy of a multiple-choice question may be judged by statistically analyzing the performance of students who have completed the item. Common item analyses include: item difficulty, item discrimination, and point-biserial correlations.

**Item Difficulty**

Item difficulty is the proportion of students giving the correct answer to an item. Item difficulty is important because it reveals whether an item is too easy or too hard. In either case, the item may add to the unreliability of the test because it does not aid in differentiating between those students who know the material and those who do not. For example, an item answered correctly by everyone does nothing to aid in the assignment of grades. The same is true for items that no one answers correctly.

An item difficulty score is reported as a decimal rather than a percentage (e.g., 0.43). It is recommended that questions with an item difficulty level of less than 0.30 or greater than 0.90 be chosen with caution by an instructor. For instance, an instructor should consider how appropriate the cognitive challenge of the question is for his/her students.

**Item Discrimination**

Item discrimination is the technique used to determine how useful a test item is in discriminating between individuals who have scored high on the overall test and those who have scored low on the overall test. If an item discriminates well between these two groups of students, then a greater percentage of those who score high on a test will get a given item correct as compared to those who scored low on a test. One way to ascertain the degree to which a test item discriminates between high and low scorers is by rank ordering the students according to their total score and then determining the upper quartile or $Q_3$ (i.e., the numerical value that separates the top 25% scores from the bottom 75% scores) and the lower quartile or $Q_1$ (i.e., the numerical value that separates the bottom 25% scores from the top 75% scores). Then, for each test item, the percentage of students in the upper and lower quartiles answering correctly is calculated. The resulting difference is a measure of item discrimination. Item discrimination is reported as a decimal (e.g., 0.57). The formula is:

$$\text{IDis} = (\text{Upper Quartile \% Correct}) - (\text{Lower Quartile \% Correct})$$

The maximum item discrimination difference is 100%. This would occur if all those in the upper group answered correctly and all those in the lower group answered incorrectly. Zero discrimination occurs when equal numbers in both groups answer correctly. Negative discrimination, a highly undesirable condition, occurs when more students in the lower group than the upper group answer correctly. It is recommended that questions with an item discrimination level of less than 0.20 be chosen with caution by an instructor and may need to be modified. Again, an instructor should consider how appropriate the cognitive challenge of the question is for his/her students.
There is an interaction between item difficulty and item discrimination:

a. very easy or very difficult test items have little discrimination;
b. items of moderate difficulty (60% to 80% answering correctly) generally are more discriminating.

**Point-Biserial Correlation**

Another way to ascertain the degree to which a test item discriminates between high and low scorers is to calculate a point-biserial correlation. A correlation is a statistic that quantifies the relationship between two variables. The point-biserial is a statistic that accurately estimates item discrimination. The formula for point-biserial is as follows:

\[
r_{pbi} = \frac{M_p - M_q}{S_t} \sqrt{pq}
\]

\(r_{pbi}\) = point-biserial correlation coefficient
\(M_p\) = whole-test mean for students answering item correctly
\(M_q\) = whole-test mean for students answering item incorrectly
\(S_t\) = standard deviation for whole test
\(p\) = proportion of students answering correctly
\(q\) = proportion of students answering incorrectly

It is recommended that questions with a point-biserial correlation coefficient score of less than 0.15 be carefully considered by an instructor (i.e., for the cognitive challenge of the question) before including it as an assessment item.
References


