

## **The Incidence of Clueing in Multiple Choice Testbank Questions in Accounting: Some Evidence from Australia**

**Nicole L. Ibbett #**

School of Business

Western Sydney University, Sydney, Australia

Email: [n.ibbett@westernsydney.edu.au](mailto:n.ibbett@westernsydney.edu.au)

(#Corresponding Author)

**Brett J. Wheldon**

School of Business

Western Sydney University, Sydney, Australia

### **ABSTRACT**

*In 2014 Central Queensland University (CQU) in Australia banned the use of multiple choice questions (MCQs) as an assessment tool. One of the reasons given for this decision was that MCQs provide an opportunity for students to 'pass' by merely guessing their answers. The mathematical likelihood of a student passing by guessing alone can be removed, or at least minimised, by the way the test is administered. Consequently, the real risk associated with guessing rests in the quality of the questions asked; they must not give 'clues' that exacerbate any guessing risk. With the widespread use of multiple choice testing, having access to testbanks containing high-quality questions that don't, in themselves, add to the risk of student guessing is important for time-poor academics who are increasingly likely to rely on them as a source of test items. This study analysed the questions used to assess the same topic from the multiple choice testbanks provided by the publishers of six commonly used Australian financial accounting textbooks. A significant majority of the questions (almost two-thirds) showed evidence of at least one 'clueing' flaw that could actually increase the chances of a student guessing the correct answer. The findings provide a degree of support for CQU's decision and demonstrate some of the risks of adopting MCQ testbanks without adequate scrutiny prior to their use. Given the results of previous studies into the poor quality of MCQ testbanks, and the proliferation of guidelines for high-quality MCQ writing, it is both surprising and disappointing that such a large number of flaws are still being found in testbanks produced by what should be a well-informed academy. Clearly, care needs to be taken but multiple-choice tests that utilize 'quality' questions and that are appropriately administered can still be a reliable form of assessment. Perhaps CQU's decision to ban them entirely is too extreme a position to take in the current climate in higher education.*

**Keywords:** multiple-choice assessment; clueing signals; accounting education.

**JEL Classification:** M40

**PsycINFO Classification:** 3560

**FoR Code:** 1303; 1501

**ERA Journal ID#:** 35696

## Introduction

Multiple choice questions (MCQs) requiring selection of the correct answer from a set of alternatives have long been used as an assessment method in accounting and in a wide range of other disciplines, especially those that are professionally oriented. Such questions are widely used as assessment tools both within courses of study and in the entry examinations of many professional associations after graduation. It is now standard practice in many disciplines for publishers to provide MCQ testbanks to academics along with their textbooks, often with the ability to administer them online using the publisher's own website. Testbank questions can provide a comprehensive coverage of the text material and ensure consistency between terms and methods used in the relevant textbook with those presented to students (Clute et al., 1988).

Additionally, there is mounting pressure on academics to achieve greater efficiencies in their teaching through means such as increased use of MCQs, certainly in the accounting discipline where there can be large student cohorts (Frakes & Lathen, 1985; Nicol, 2007; Arthur & Everaert, 2012). All of these factors have combined to make the use of MCQ testbanks provided by publishers very attractive to time-poor academics who are increasingly likely to rely on them as a source of test items.

It has commonly been argued that multiple choice (MC) assessment can be both efficient and effective if properly administered: efficient, in that it can deliver the dual benefit of saving time for the academic while providing quicker feedback for students; and effective, in that it can produce reliable, valid results. (Bacon, 2003; Nicol, 2007; Douglas et al., 2012; Glass & Sinha, 2013)

There is little to dispute the efficiency argument but there is a large and growing body of literature questioning the effectiveness argument, especially around the issue of answer guessing. Despite the recent claim by Bush (2015) that the potential benefit of guessing can be significantly reduced or even eliminated through using more sophisticated MC test delivery formats, this is really only possible when well-written questions are used. Much prior research of MCQ results assumes that only well-written MCQs are used by examiners, however there is the potential for drafting flaws in questions that can confuse students and for clueing signals to assist student guessing that may lead to reduced reliability of MCQs as an assessment tool. Since results of examinations using MCQs affect student grades, determine academic awards, and so on, the reliability of the banks of questions used is important. As Clute et al. (1988) point out in their analysis of potential choice bias in MCQ testbanks in organizational behaviour, 'Clearly, the presence of any inherent bias or unreliability in these testbanks is a matter of critical importance to students and instructors alike'. (p. 124)

This issue of the reliability of MCQs became such a concern to the academic board of Central Queensland University (CQU) in Australia that in late 2014 the future use of MCQs as an assessment method at that institution was forbidden. In making the announcement CQU's Pro Vice-Chancellor for Teaching and Learning pointed to four reasons for this decision: (1) by their nature MCQs provide the mathematical potential for students to achieve a typical passing mark of 50% through mere guessing; (2) that MCQs do not present problems in the way students will face them in typical real-world workplaces; (3) there is a potential ethical problem in academics presenting students with partly-correct alternative answers to try to 'lead students astray'; and (4) some MCQs can be too complex for students to correctly comprehend, for example negative questions.

Support for CQU's second concern, which is not the focus of this study, can be found in Paxton (2000) who questions whether students are able to transfer material learnt for a MCQ test to complex real life problems. Her experience of the repeated use of MC tests in economics lead her to believe that such questions 'inevitably elicit answers

to isolated and decontextualised problems and real life problem solving situations do not exist in isolation' (p.113). Further, students may merely develop skills at MC test taking (i.e. become 'test-wise') but not develop essential problem-solving or personal judgement skills.

CQU's other three concerns can be linked to construction flaws in the MCQs themselves. For example, Haladyna (2004) warns against using 'trick items' designed to lead students astray. Likewise, phrasing options negatively is commonly included amongst construction flaws to avoid (Haladyna & Downing, 1989; Haladyna, 2004) while, conversely, stating the question in positive form is recommended (Haladyna & Downing, 1989; Gronlund & Waugh, 2009).

It should be acknowledged that the potential of success through guessing can be due to more than a mere mathematical phenomenon; significantly, it can be compounded by construction flaws in the questions themselves. Given the extensive literature that has developed around how to construct good MCQs, and the educational skills of textbook authors and publishers, it is reasonable to assume that question testbanks should only contain well-written questions that conform to best practice. MCQs that are poorly written may actually give students one or more 'clues' as to the correct answer either within the question itself, through the alternative answers provided, or both.

MCQs are widely used as an assessment tool in a number of business disciplines, including tests based on questions taken from testbanks. For example, Parker & Clow (2010) found that 27.8% of the respondents to their survey of American academics in the economics discipline used publisher supplied testbanks as their source of MCQ test questions. Importantly, none of those academics reported that they modified the questions provided, so reliance was placed entirely on the quality of the testbank questions without review.

The purpose of this paper is to consider CQU's first concern of potential student guessing in MCQs and whether testbank questions unintentionally increase student guessing success rates beyond the mere mathematical opportunity. This specific aspect of MC testing is yet to be studied comprehensively and is a first step in answering the call of Moncada & Moncada (2010) for more empirical research into the quality of textbook testbanks and the effect that guessing may have on exam results, especially for business students. We examine the MCQs for the same topic from a number of testbanks provided by publishers of commonly used financial accounting textbooks in Australia to determine if the questions are written or presented in such a way as to make guessing the correct answer easier.

## Literature Review

A significant body of literature has developed providing guidelines for the construction of high quality MCQs, including Gronlund (1988), Haladyna & Downing (1989), Haladyna et al. (2002), Haladyna (2004), Frey et al. (2005), and Haladyna & Rodriguez (2013). Although these guidelines can generally be applied to any context, a number of guidelines have been developed within particular disciplines, including Ellsworth et al. (1990) for nursing and Baldwin (1984) for accounting. In 2014, the International Accounting Education Standards Board (IAESB) released guidelines for the development of written examinations, which included a guide to writing MCQs.

While many question writing principles appear to be universally accepted, such as ensuring that there is only one correct answer (Gronlund, 1988; Haladyna & Downing, 1989; IAESB, 2014), others remain areas of contention. For example, Haladyna et al. (2002) report that some guidelines recommend using 'none of the above' as an alternative while other guidelines recommend never using 'none of the above'.

Consequently, there are no comprehensive rules for MCQ construction but there are only guidelines (Gronlund, cited in Hansen and Dexter, 1997).

What does appear to be well accepted though is that poorly written MCQs limit the effectiveness of this assessment tool. Various criticisms have been levelled against MCQs including inhibiting creativity in student responses and reducing knowledge to superficial facts (Osterlind, 1997), as well as limitations in assessing critical thinking skills. While Merino (2006) acknowledges that MCQs can foster critical thinking he suspects, however, that such questions taken from textbooks, testbanks or CPA professional exams are not designed in this way. Despite this, if designed well it is believed that MCQs can assess higher order thinking skills in addition to reliably measuring ability (Osterlind, 1997).

Studies comparing student performance in constructed response questions (CRQs) - that is short-answer, essay or problem style questions - and MCQs, are inconclusive. For example, Chan & Kennedy (2002) found economics students performed better in MCQs than in similar CRQs. Krieg & Uyar (2001) found that performance among business and economic statistics students in MCQs was better for some students, but worse for others, depending on their particular personal characteristics. CRQs and MCQs were found to produce similar results by Frakes & Lathen (1985) in introductory accounting level exams and by Bible et al. (2008) for intermediate accounting level exams. Following his study of tests undertaken by marketing students Bacon (2003, p.35) was of the view that 'MC questions offer comparable reliability and validity to SA [short-answer] questions'.

If there are no significant differences between results from MCQs and CRQs, assuming the questions assess at similar cognitive levels, it could reasonably be argued that the impact of guessing answers to MCQs is minimal. Simkin & Kuechler (2005) are of the view that many of the prior studies comparing MCQs and CRQs were flawed in that they incorrectly treated MCQs as 'homogenous entities' rather than considering their relative cognitive levels. They noted that it is difficult to write suitable MCQs to match CRQs at the higher cognitive levels. Thus, considering the mixed empirical results of the comparison studies, the role of guessing in determining performance in MCQs cannot be discounted entirely.

Osterlind (1997) points out that while a student may be able to prepare an answer to an essay type question in such a way that their writing style obfuscates their actual lack of knowledge of the content, they would be unable to do so when faced with a MCQ. This view supports the use of MCQs but presumes, of course, no bias achieved through guessing. Poorly constructed MCQs have the potential to increase the chance of guessing correctly, particularly for a test-wise student. An examiner may inadvertently clue the student to the correct answer by the very nature of the question asked and/or the alternative answers presented, thus increasing the student's chance of guessing correctly with limited or even no knowledge of the material being examined. This has the potential to produce an upwardly skewed result and a 'type-II error' – accepting that a student's MC passing test result is a valid assessment of their knowledge while it is not. Moncada & Moncada (2010) provide various example questions demonstrating some clueing issues and suggestions for re-writing them in a way that eliminates the clue.

MCQs have been criticised for their potential for guessing for many years (e.g. Biggs, 1999). A student knowing the correct answers to fewer than half of the set questions could actually succeed in the assessment by correctly guessing enough of the remaining answers to accumulate a passing result of, for example, 50% overall. When considering a MCQ examination of 100 questions with four answer options, Reed (2014) points out that 'once you factor in the randomness of guessing and the influence of binomial probabilities, we can only be 95% confident that the student

knows, and hasn't guessed, around one-quarter of their answers'. Mathematically, this risk can be reduced by raising the minimum pass mark above the 50% level or through the use of 'formula scoring' where a mark or part-mark is deducted for an incorrect answer. Both approaches have been suggested as ways to reduce the guessing concern, however Bush (2015) is of the view that many students believe negative marking is fundamentally unfair, with zero being a sufficient penalty for an incorrect answer. Students would need to understand the reasons for, and be convinced of, the appropriateness of the marking scheme if such an approach is used.

Another method of potentially reducing the guessing effect is to increase the number of answer options for each MCQ beyond the traditional four. This seems intuitively appealing, but in his meta-analysis of 80 years of research into the ideal number of answer options that should be used Rodriguez (2005) actually found that adopting a three-option approach was optimal in most situations, so long as the two distractors were 'plausible'. Moving from four to five options provides little additional reliability. From a practical viewpoint, increasing the number of options requires additional test time be given to students to read each question and, where there is a time constraint on the test, this limits the range of material that can be examined. Further, question writers would need more time to devise the extra distractors required, many of which may be implausible and easily discarded by students anyway. For example, Clute & McGrail (1989) discovered that option 'E' was only correct 5% of the time in the one cost accounting testbank they analysed that had answers with five-options. This supports the view that the fifth option is merely added to make up the numbers. The four-option approach appears to be the most common approach used.

If test validity is an issue then the method by which the test is administered also needs to be considered. Traditionally, paper-based MC tests have been conducted in-class and under supervision. In recent years the availability of technology has allowed them to easily be produced and delivered electronically. For such tests students are typically allocated a set number of questions at random from a testbank that has been uploaded to the system. For various reasons, such as limitations on computer facilities and costs, it is often difficult to administer these under supervision in computer labs on campus so they are, perhaps increasingly, being undertaken in an unsupervised fashion outside the classroom (Ladyshevsky, 2015). In their comparison of student performance in a supervised, paper-based midterm exam vs. an off-campus online final exam, Schmidt, et al. (2009) concluded that the online exam was preferred by both students and academics. Various reasons for this preference were cited, but the key concern expressed was the potential for cheating and the need for the academic to ultimately rely on student honesty. In his study of postgraduate business students Ladyshevsky (2015) found that the mean scores for unsupervised online tests were not significantly higher than for supervised in-class tests. His results add to a growing body of evidence that increased cheating in unsupervised online tests is not supported where only mean scores are considered (e.g. Peng, 2007; Anakwe, 2008; Yates & Beaudrie, 2009). Even with this evidence, cheating cannot be discounted entirely whenever unsupervised online assessments are conducted. The potential for students gaining assistance and/or sharing answers during MC exams that are not conducted under supervision makes the need for well-written questions, especially ones that do not provide answer 'clues', even more imperative.

In the accounting discipline, a number of studies have scrutinised MCQs in testbanks accompanying cost accounting textbooks (Clute & McGrail, 1989), auditing textbooks (Hansen & Dexter, 1997) and accounting textbooks in general (Moncada & Harmon, 2004), as well as textbook testbanks compared to professional certification exam questions (Bailey et al., 1998). Evidence of numerous flaws in MCQ construction was found to exist in all cases. These previous studies examined the overall quality of the MCQs used but not specifically whether the flaws identified could increase the chances of students guessing correctly. Some flaws increase the difficulty of the question,

suggesting the correct answer could actually be harder to guess. For example, the most common flaw reported by Hansen & Dexter (1997) was stating the question negatively. This type of question is less likely to help even the most test-wise student guess the correct answer as negatively worded questions, and those containing 'double-negatives', may cause the item to be more difficult to understand (Haladyna & Rodriguez, 2013).

Given this study's focus on correct MCQ guessing, the flaws of concern here are those related to positive clueing signals in the question (the 'stem'), in the correct answer, or in the incorrect alternatives (the 'distractors'). Analysis of the literature provides support for ten question flaws that can result in positive clueing signals to students:

### ***Clueing Signal 1: All of the above, none of the above, and complex format***

This study combines the flaws of using 'all of the above', 'none of the above' and complex format MCQs into one item for analysis. While it is acknowledged that each is a separate flaw, and in the case of 'none of the above' even debatable as to whether it is a flaw, all three are less discriminating and do appear to provide similar clueing signals (Haladyna et al., 2002). For example, when faced with a complex question, usually in the form of a list of options such as I, II and III and alternative answers based on various combinations of these options, a student only needs to be able to identify if one of the options as correct or incorrect, in order to immediately eliminate one or more alternatives. Similarly, if a student identifies two alternatives as correct they may confidently select 'all of the above' with no knowledge of other options (Hansen & Dexter, 1997). Conversely, if even one alternative is identified as incorrect then 'all of the above' automatically cannot be the correct answer. Harasym et al. (1998) specifically caution against using 'all of the above' as an alternative, supporting its clueing effect and the resulting significant enhancement in student performance. Evidence is not as conclusive regarding 'none of the above', with Haladyna (2004) summarising arguments both for and against its use and suggesting it can be suitable for complex quantitative questions but not for questions with lower cognitive demand. Although it has been identified that use of 'none of the above' can increase the difficulty of a question, it can also result in students with knowledge deficiencies receiving full marks (DiBattista et al., 2014) and hence will be included in this study as a clueing flaw.

### ***Clueing Signal 2: Specific determiners***

Specific determiners, such as 'never', 'always', 'completely' and so on, can help provide clues to the correct answer as they create alternatives so extreme they are seldom the correct answer (Haladyna & Rodriguez, 2013) or are more commonly associated with alternatives which are false (Gronlund & Waugh, 2009). Moncada & Harmon (2004) categorised 'none of the above' and 'all of the above' alternatives as forms of specific determiners when testing for clueing signals in MCQs. This study will follow the approach of Hansen & Dexter (1997) and Tarrant et al., (2006) of including this as a separate flaw.

### ***Clueing Signal 3: Grammatically incorrect stem and distractors***

Grammatical inconsistencies between the stem and distractors can assist in eliminating alternatives as possible options or can provide clues to the correct answer (Haladyna, 2004). The IAESB (2014) specifically identifies grammatical inconsistencies when warning against providing unintentional clues in questions. For example, if a student is required to select the correct alternative phrase to complete a statement (stem) then all of the alternatives need to be phrases that 'fit' grammatically with the

first part of the sentence. Miller et al. (2013) point out that attention is frequently directed to ensuring the stem and the correct answer are grammatically correct, but not the stem and the distractors. This may lead to clues to the correct answer being obvious when the grammatical structure of one alternative varies from the others (Gronlund & Waugh, 2009).

#### ***Clueing Signal 4: Similarity of wording in stem and distractors***

Similarity of wording in the question stem and distractor can provide a clue to the correct answer (Ellsworth et al., 1990; Hansen & Dexter, 1997). A key word from the stem that only appears in the correct alternative - a 'clang' association according to Haladyna & Rodriguez (2013) - is likely to lead a guessing student to the correct answer. Likewise, a key word appearing in all distractors, but not the correct alternative, provides a signal to the correct answer. Moncada & Harmon (2004) make the point that unintentional repeating of a key word can provide a clue to an uninformed test taker.

#### ***Clueing Signal 5: Inconsistency in length of distractors***

Evans (1984) identified the length of the correct option as having the potential to provide a clueing signal. Due to the tendency to include additional information in the correct option, test-wise students may pick up clues as to the right answer when alternatives vary in length (Gronlund & Waugh, 2009). Miller et al. (2013) suggest alternatives should be of approximately equal length, while Ellsworth et al. (1990) identified that length only becomes a significant issue when the correct option is at least one and a half times the length of the shortest alternative. For the purposes of this study, we followed Ellsworth et al.'s approach.

#### ***Clueing Signal 6: Pairs of options***

Options which are paired in such a way that they become highly related can provide a clue that the correct answer is one of the pair (Haladyna & Rodriguez, 2013). This appears to particularly be a problem where two options are effectively the opposite of each other, meaning one statement is a 'true' response to the stem with the other statement a 'false' response. As Hansen & Dexter (1997, p. 95) point out, 'This structure allows students to eliminate other alternatives because the inclusive pair covers all possibilities. An uninformed student would have a 50% chance of guessing the correct answer [rather than a 25% chance in a four-option question]'.

#### ***Clueing Signal 7: Implausible distractors***

Using distractors which are highly implausible become clues to the correct answer since they are likely to be easily eliminated as options, even by someone with no knowledge of the content being tested. Easily eliminated distractors can lead to questions losing measurement value (IAESB, 2014). Plausible distractors should be common errors (Haladyna & Downing, 1989; Haladyna & Rodriguez, 2013) and, when dealing with numerical questions, should be incorrect combinations of data from the question (Baldwin, 1984). This approach has the added benefit that selection of an incorrect distractor has feedback value because it will highlight the specific calculation error, or misconception, made.

#### ***Clueing Signal 8: Unfortunate coincidence***

Alternatives which are not necessarily implausible but can be arrived at via an 'unfortunate coincidence' lack feedback value, particularly in numerical questions. The 'unfortunate coincidence' (Davidson et al., cited in Baldwin, 1984) occurs when the

correct answer is able to be selected for the wrong reason. Baldwin (1984) gives the 'unfortunate coincidence' example of using a tax rate of 50%, resulting in equivalent tax amounts and after tax values. Assuming the question is asking for the after tax value, a student calculating the tax amount would still select the correct alternative. This flaw is not limited to MCQs, as it could occur in any calculation question, but has been included as a clueing flaw in this study as its occurrence within MCQs will improve the chance of guessing the correct alternative.

### ***Clueing Signal 9: Numerical order***

Strong support exists for ordering options logically or numerically (Haladyna et al., 2002; Haladyna, 2004). Most evidence supports this guideline on the basis that it eliminates unnecessary searching, making the question quicker to answer (Haladyna, 2004; IAESB, 2014). Gronlund & Waugh (2009) offer an additional justification when dealing with numerical questions in that ordering ascendingly will eliminate any possibility of clueing through the correct alternative being the only one out of order. The exception to this would be if the question requires identification of the smallest or largest item (Osterlind, 1997), necessitating random ordering of alternatives to avoid clueing.

### **Clueing Signal 10: Option bias**

Evans (1984) reported that the position of the correct option in the sequence of alternatives could influence the performance of a test-wise student. To avoid providing clues as to the correct answer, MCQ guidelines recommend balancing the spread of correct alternatives approximately evenly across the available options (Haladyna, 2004; Gronlund & Waugh, 2009; Haladyna & Rodriguez, 2013). Suggestions are that students choose option C if in doubt (Baldwin, 1984) or believe it is more common for either C or D to be the correct answer (Clute & McGrail, 1989). In contrast, edge aversion is the tendency to avoid the extremes and select options in the middle of the range (Attali & Bar-Hillel, 2003). All of this supports the oft-quoted adage of 'if in doubt pick C'!

In this study, the ten potential signals identified above were used to analyse a selection of MCQs from a number of testbanks accompanying financial accounting textbooks to determine the extent to which positive clueing signals were inherent in the questions provided.

## **Method**

Publisher supplied MCQ testbanks were sourced for six commonly used Australian textbooks, with three being at an introductory level (Birt et al., 2012; Hoggett et al., 2014; Atrill et al., 2016) and three at an advanced level (Henderson et al., 2011; Deegan, 2012; Leo et al., 2015). Details of these texts are listed in Appendix 1. All of the textbooks are well established titles, ranging from a fourth edition through to a fourteenth edition. A survey of accounting subjects/units offered during 2015 across 37 Australian universities shows all six books commonly appear as prescribed texts.

From the six testbanks, the questions relating to the Statement of Cash Flows were selected for analysis. Though isolating the sample to only one topic significantly limits the number of MCQs available for analysis, it does have a number of advantages. Firstly, preparation of the Statement of Cash Flows is traditionally a topic students struggle with but the ability to distinguish between cash flows and accounting profit is a fundamental accounting concept that needs to be mastered (Lucas, 2002). Secondly, the topic is usually taught at two levels in Australia. At the introductory level, students become familiar with the concept and the layout of the statement,



while at the advanced level students are generally required to prepare complex statements along with supporting disclosure notes. If students have passed the cash flow topic in introductory level units on the basis of guessing, this raises a serious issue as to their level of preparedness for advanced level units. Hence, it was deemed important to not only investigate the level of clueing signals provided across all six testbanks, but also to determine if there was any significant difference between the testbanks from the introductory level compared to the advanced level.

The six testbanks provided 270 MCQs, with 97 of these being numeric or calculation style questions and the remainder conceptual in nature. All MCQs in the testbanks had four-option answers, shown as 'A' to 'D'. Since item writing guidelines indicate an essential requirement for a MCQ is to provide only one correct answer (Haladyna & Downing, 1989) the selected questions were first checked for correctness. Of the 270 questions, seven were found to either have no correct alternative or multiple correct alternatives. These were eliminated from the analysis, leaving 263 questions (91 numeric and 172 conceptual) to be investigated for evidence of the ten clueing signals discussed above.

## Results and Discussion

Barring option bias, a total of 206 clueing signals were found across the 263 questions. Surprisingly, almost two-thirds of questions demonstrated at least one clueing signal, with 148 questions containing one signal, 27 demonstrating more than one clueing signal, and only 88 questions exhibiting no evidence of clueing signals.

Table 1 presents a summary of the incidences of nine clueing signals (signal 10 will be analysed separately): (1) all or none of the above or complex, (2) use of specific determiners; (3) grammatically incorrect stem and distractors, (4) similarity of wording in stem and distractors, (5) inconsistencies in length of distractors, (6) pairs of options, (7) implausible distractors, (8) unfortunate coincidences, (9) options not in numeric order.

**Table 1:**  
*Frequency of Clueing Signals, by Type (1) to (9)*

Clueing Signal	1	2	3	4	5	6	7	8	9	Total
Number of occurrences	64	2	5	10	12	23	17	10	63	206
% of total clueing signals	31%	1%	2%	5%	6%	11%	8%	5%	31%	100%

Of the 206 clueing signals found across the six testbanks, the most commonly occurring signal was the use of either all of the above, none of the above or a complex type MCQ. This result is not dissimilar to the finding of Hansen and Dexter (1997) that 25% of flaws in sampled auditing testbanks resulted from use of all of the above or none of the above. In the current sample, 21 of the questions used all of the above as an alternative and it was the correct option 67% of the time. In contrast, 23 of the questions used none of the above as an alternative, but it was *never* the correct option. This suggests that test-wise students who automatically eliminate this as a possible alternative would definitely be increasing their chances of guessing correctly from the remaining options. It is acknowledged that the use of none of the above as an option is controversial, with some viewing it as a viable alternative for numeric questions (Haladyna & Rodriguez, 2013). In this sample only 4 of the uses of 'none of the above' appeared in numeric questions.

Numeric questions, overall, were highly represented in the MCQs that had clueing signals. With 63 occurrences, displaying options out of numerical order was the second most common clueing flaw in the sampled testbanks. Since 91 of the MCQs sampled were numeric questions, this means 69% of all numeric MCQs displayed options out of numeric order. Not only does this increase the chances of guessing correctly, particularly when the correct answer is the only alternative out of order, it is also a less efficient method of displaying alternatives for students who have calculated the correct answer but have been forced to search the among the alternatives to find it (Haladyna & Rodriguez, 2013).

Overall, little difference was found in the incidence of clueing signals between the introductory level testbanks and the advanced level testbanks, with the exception of displaying numeric options out of order. The introductory level testbanks contained more numeric questions, with 39% of questions being numeric compared to only 26% for the advanced level questions. The introductory level testbanks also contained a higher proportion of numeric questions that were not displayed in numeric order, at 71% of all introductory level numeric questions compared to only 62% of all advanced level numeric questions.

Numeric MCQs represented all of the 17 implausible distractor incidences found. In these cases, the questions presented distractors which could not be calculated from any combination of the values provided. A student with no knowledge of the material being examined could easily eliminate these distractors as feasible alternatives, thus increasing their chances of guessing correctly. A further issue with this flaw, not related to guessing, is that the use of implausible distractors in questions limits the opportunity for useful feedback. Should a student select the implausible alternative there is no ability to specifically identify the source of the student's misunderstanding, or lack of knowledge. In contrast, when distractors are based on common errors, as recommended by Haladyna (2004), the misunderstanding can be quickly identified and useful feedback provided.

Ten numeric MCQs were identified as containing at least one alternative flawed by an 'unfortunate coincidence'. Examples included the calculation of a cash outflow paid being equal to the opening balance of the related payable account and the calculation of cash paid for inventory being equal to the unadjusted cost of sales amount. This type of flaw not only increases the chances of guessing correctly but has the added problem that it is impossible to identify situations when the correct answer has been selected for the wrong reason. In these circumstances, a student's misunderstanding is actually reinforced through their selection of the 'correct' alternative.

Of the conceptual questions the highest occurrence of clueing signals, after use of all of the above and none of the above, was found in the pairing of options. In contrast, Hansen & Dexter (1997) reported minimal occurrence of this flaw in auditing testbanks. In the current sample, these signals generally consisted of alternatives that resulted in the same outcome, for example, 'add an increase in a current liability' and 'deduct a decrease in a current liability'. A student with limited knowledge could identify that these alternatives are either both correct or both incorrect and eliminate them as feasible alternatives. In some cases the pairing of options gave opposite alternatives, for example, 'increase in current liabilities means a decrease in cash' and 'decrease in current liabilities means a decrease in cash'. A guessing student would quickly identify the correct alternative is likely to be one of these two options.

Representing approximately 6% of the clueing signals identified was the flaw of the correct answer being the longest alternative. This result is similar to the 7% reported by both Hansen & Dexter (1997) and Moncada & Harmon (2004). Interestingly, the latter authors found this flaw to be more common in advanced level testbanks than in

introductory level testbanks, whereas the current study showed little difference in occurrence between the two levels.

The remaining clueing signals each represented less than 5% of the total flaws found. This suggests these flaws are slightly less prevalent here than in previous studies. Incidences of the similarity of wording in the stem and distractors and grammatical inconsistencies between the stem and distractors appear marginally less in the current study than reported by Hansen & Dexter (1997). The least occurring clueing signal in the current study, use of specific determiners, at 1% also occurs slightly less than the 4% reported by Hansen & Dexter (1997). It is noted that the incidence of this flaw in both of these studies is well below the 25% occurrence rate as reported by Moncada & Harmon (2004).

In relation to the results for clueing signal 10, option bias, each of the six testbanks contained multiple choice questions with four options. Table 2 summarises the results from analysing the spread of correct options from all six testbanks.

**Table 2:**

*Frequency of Correct Answer, by Option (Clueing signal 10)*

Testbank	Option A	Option B	Option C	Option D	Total Items
1	27%	23%	26%	24%	66
2	18%	26%	32%	24%	50
3	17%	28%	35%	20%	65
4	17%	19%	38%	26%	42
5	29%	19%	19%	33%	21
6	26%	21%	21%	32%	19
Mean %	21%	24%	30%	25%	<i>n</i> =263

Except for Testbank 1 which had close to an even spread, none of the testbanks analysed had a near-even spread of correct options across the four possibilities. Testbanks 2 to 6 all had at least one option appearing as the correct alternative more than 30% of the time, instead of around 25% that would be expected for a four option MCQ.

Option C was the most commonly occurring correct alternative overall, and in three of the six testbanks. This presents a significant guessing opportunity to students defaulting to option C when in doubt (Baldwin, 1984; Moncada & Moncada, 2010). Further, the results support Clute & McGrail's (1989) claim that it is either options C or D that are more likely to be the correct answer, with 55% of the overall correct options being found as either C or D in this study. In five of the six testbanks, C or D appeared as the correct answer more than 50% of the time, with the sixth testbank having C or D correct 50% of the time. In contrast, if edge aversion (Attali & Bar-Hillel, 2003) is correct, with students more likely to avoid A or D when guessing, the results of this study are mixed, with three testbanks more likely to have B or C as the correct answer and the remaining three more likely to have A or D correct.

## Conclusion

It was surprising to find that almost two-thirds of the MCQs relating to the cash flows topic in the six testbanks analysed contained flaws that provide students with clues as to the correct answer. The findings demonstrate the risks of academics using testbanks without adequate scrutiny of the questions provided. Failure to review and amend the questions, where necessary, has the potential to invalidate the reliability of tests prepared using those testbanks.

A key finding was the extent of option bias (clueing signal 10) in a number of the testbanks, but this may become less of an issue with the electronic delivery of MCQ exams and the ability to 'shuffle' the correct alternative. This would be likely, nevertheless, to exacerbate the ordering problem. In particular, shuffling options for numeric questions would be likely to eliminate the ascending or descending order recommended.

The significant level of clueing flaws found in the MCQs examined in this study appears to provide support for CQU's decision to cease utilising MCQs on the basis that students may pass assessments by guessing. However, for the majority of the questions found to contain clueing flaws in this study it was evident that a rewording of the question could have easily reduced or removed the clueing effect. Accepting that well-constructed MCQs limit the chance of guessing correctly, this would suggest that prohibiting all use of MCQs means losing the potential advantages of MCQs. Along with the efficiency benefits of using MCQs, a well-constructed MCQ can have the potential of providing rapid feedback to the student regarding the error that has been made, as well as giving the examiner valuable feedback on common student errors or misconceptions. Rather than eliminate MCQ use entirely it would appear beneficial to include quality MCQs alongside other assessment tools. In short, to follow the call made by Douglas et al. (2012) to strike a balance between using MCQs and other assessment methods which engage with high cognitive levels.

### ***Limitations and Future Research Directions***

As part of our method we selected those types of question flaws we believed could provide positive 'clueing' to students. We also had to review questions and their answers in light of those categories of flaws. Thus, author subjectivity may have come into play. The authors are experienced academics and accounting practitioners, so any potential bias in this regard is considered to be insignificant.

It needs to be acknowledged that this study was limited in scope with only the questions for one accounting topic being examined. Further research which widens the scope to other topics in financial accounting, and beyond that into other areas of accounting, or to other disciplines, would make the results more generalizable. Extending the study to survey academics about the extent to which they create and use MCQs, what training they have undertaken in question writing techniques, how many answer choices they use, whether they review publisher provided testbanks before use, and how they administer and invigilate such tests would provide a fuller picture of the 'state of play'. Further, analysis of the level and type of feedback provided with MCQ results, and the use of this feedback by students, would also help contribute to an understanding of effective use of MCQs.

It must also be recognized that this study was limited to only investigating the incidence of clueing flaws. A worthwhile future study would be to investigate testbanks in light of CQU's other three reasons for eliminating MCQs, being lack of real world application, distractors leading students astray and level of question complexity.

### ***Concluding Remarks***

Given the documented history of clueing problems found in MCQs, it is reasonable to suggest that informed textbook authors and publishers would have taken steps by now to ensure such deficiencies were overcome in current testbanks. There are numerous guides available to assist in question writing so it is remarkable that these do not appear to have been used by the preparers of these testbanks. Osterlind (1997) hoped that poorly-prepared test items would become increasingly rare, but almost two decades later this does not appear to be the case. It is both surprising and

disappointing that such a high number of flaws are still being found in testbanks produced by what should be a well-informed academy.

Certainly, MCQs should not be the only method used to assess student outcomes, and care needs to be taken in their writing, but MC tests that utilize 'quality' questions and that are appropriately administered can still be a reliable and efficient form of assessment. Perhaps CQU's decision to ban them entirely is too extreme a position to take, especially in the current climate in higher education where financial stringency, the need to produce 'objective' evidence of student outcomes, and expectations on academics to increase the use of technology in teaching and assessment are the order of the day.

## Acknowledgements

We gratefully acknowledge the helpful suggestions made by two anonymous reviewers on an earlier version of the manuscript.

## References

- Anakwe, B. (2008). Comparison of student performance in paper-based versus computer-based testing, *Journal of Education for Business*, 84(1), 13-17.
- Arthur, N. & Everaert, P. (2012). Gender performance in accounting examinations: exploring the impact of examination format, *Accounting Education: An International Journal*, 21(5), 471-487.
- Attali, Y. & Bar-Hillel, M. (2003). Guess where: The position of correct answers in multiple-choice test items as a psychometric variable, *Journal of Educational Measurement*, 40(2), 109-128.
- Bacon, D. R. (2003). Assessing learning outcomes: A comparison of multiple-choice and short-answer questions in a marketing context, *Journal of Marketing Education*, 25(1), 31-36.
- Bailey, C. D., Karcher, J. N. & Clevenger, B. (1998). A comparison of the quality of multiple-choice questions from CPA exams and textbook test banks, *Accounting Educator's Journal*, 10(2), 12-29.
- Baldwin, B. A. (1984). The role of difficulty and discrimination in constructing multiple-choice examinations: with guidelines for practical applications, *Journal of Accounting Education*, 2(1), 19-28.
- Bible, L., Simkin, M. G. & Kuechler, W. L. (2008). Using multiple-choice tests to evaluate students' understanding of accounting, *Accounting Education: An International Journal*, 17(Supplement), S55-S68.
- Biggs, J. (1999). *Teaching for quality learning at university*. Buckingham, UK: Society for Research Into Higher Education and Open University Press.
- Bush, M. (2015). Reducing the need for guesswork in multiple-choice tests, *Assessment & Evaluation in Higher Education*, 40(2), 218-231.
- Chan, N. & Kennedy, P. E. (2002). Are multiple-choice exams easier for economics students? A comparison of multiple-choice and "equivalent" constructed-response exam questions, *Southern Economic Journal*, 68(4), 957-971.
- Clute, R. C., Jackson, J. H., & McGrail, G. R. (1988). Choice bias in multiple choice questions from organizational behavior test banks, *Journal of Management Education*, 12(4), 124-127.

- Clute, R. C. & McGrail, G. R. (1989). Bias in examination test banks that accompany cost accounting texts, *Journal of Education for Business*, 64(6), 245-247.
- DiBattista, D., Sinnige-Egger, J., & Fortuna, G. (2014). The "none of the above" option in multiple-choice testing: An experimental study, *The Journal of Experimental Education*, 82(2), 168-183.
- Douglas, M., Wilson, J. & Ennis, S. (2012). Multiple-choice questions tests: a convenient and effective learning tool? A case study, *Innovations in Education and Teaching International*, 49(2), 111-121.
- Ellsworth, R. A., Dunnell, P. & Duell, O. K. (1990). Multiple-choice test items: what are textbook authors telling teachers?, *Journal of Educational Research*, 83(5), 289-293.
- Evans, W. (1984). Test wiseness: An examination of cue-using strategies, *Journal of Experimental Education*, 52(3), 141-144.
- Frakes, A. H. & Lathen, W. C. (1985). A comparison of multiple-choice and problem examinations in introductory financial accounting, *Journal of Accounting Education*, 3(1), 82-89.
- Frey, B. B., Petersen, S., Edwards, L. M., Teramoto Pedrotti, J. & Peyton, V. (2005). Item-writing rules: Collective wisdom, *Teaching and Teacher Education*, 21(4), 357-364.
- Glass, A. & Sinha, N. (2013). Multiple-choice questioning is an efficient instructional methodology that may be widely implemented in academic courses to improve exam performance, *Current Directions in Psychological Science*, 22(6), 471-477.
- Gronlund, N. E. (1988). *How to construct achievement tests*, 4th Ed, New Jersey: Prentice-Hall.
- Gronlund, N. E. & Waugh, C. K. (2009). *Assessment of student achievement*, 9th Ed, Pearson, New Jersey.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*, 3rd Ed, New Jersey: Lawrence Erlbaum Associates.
- Haladyna, T. M. & Downing S. M. (1989). A taxonomy of multiple-choice item-writing rules, *Applied Measurement in Education*, 2(1), 37-50.
- Haladyna, T. M., Downing, S. M. & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment, *Applied Measurement in Education*, 15(3), 309-333.
- Haladyna, T. M. & Rodriguez M. C. (2013). *Developing and validating test items*, New York: Routledge.
- Hansen, J. D. & Dexter, L. (1997). Quality multiple-choice test questions: Item-writing guidelines and an analysis of auditing testbanks, *Journal of Education for Business*, 73(2), 94-97.
- Harasym, P. H., Leong, E. J., Violato, C., Brant, R. & Lorscheider, F. L. (1998). Cuing effect of "all of the above" on the reliability and validity of multiple-choice test items, *Evaluation and the Health Professions*, 21(1), 120-133.
- International Accounting Education Standards Board (2014). *Development and management of written examinations*, New York: International Federation of Accountants.
- Krieg R. & Uyar, B. (2001). Student performance in business and economics statistics: does exam structure matter?, *Journal of Economics and Finance*, 25(2), 229-241.

- Ladyshevsky, R. K. (2015). Post-graduate student performance in 'supervised in-class' vs. 'unsupervised online' multiple choice tests: implications for cheating and test security, *Assessment & Evaluation in Higher Education*, 40(7), 883-897.
- Lucas, U. (2002). Contradictions and uncertainties: Lecturers' conceptions of teaching introductory accounting, *British Accounting Review*, 34(2), 183-203.
- Merino, B. D. (2006). Financial scandals: Another clarion call for educational reform – A historical perspective, *Issues in Accounting Education*, 21(4), 363-381.
- Miller, M. D., Linn, R. L. & Gronlund, N. (2013). *Measurement and assessment in teaching*, New Jersey: Pearson.
- Moncada, S. M. & Harmon, M. (2004). Test item quality: An assessment of accounting test banks, *Journal of Accounting and Finance Research*, Fall, 28-39.
- Moncada, S. M. & Moncada, T. P. (2010). Assessing Student Learning with Conventional Multiple-Choice Exams: Design and Implementation Consideration for Business Faculty, *International Journal of Education Research*, 5(2), 15-29.
- Nicol, D. (2007). E-assessment by design: using multiple-choice tests to good effect, *Journal of Further and Higher Education*, 31(1), 53-64.
- Osterlind, S. J. (1997). *Constructing test items: Multiple-choice, constructed-response, performance and other formats*, 2nd Ed, Dordrecht: Kluwer Academic Publishers.
- Parker, T. & Clow K. E. (2010). Use of powerpoint slides and quizzes by economics faculty, *Journal of Economics and Economic Education Research*, 11(2), 75-81.
- Paxton, M. (2000). A linguistic perspective on multiple choice questioning, *Assessment & Evaluation in Higher Education*, 25(2), 109-119.
- Peng, Z. (2007). Giving online quizzes in corporate finance and investments for a better use of seat time, *The Journal of Educators Online*, 4(2), 1-18.
- Reed, R. (2014). Does the student a) know the answer, or are they b) guessing?, *The Conversation*, 19 September, viewed 4 June 2015, <<https://theconversation.com/does-the-student-a-know-the-answer-or-are-they-b-guessing-31893>>
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A Meta-analysis of 80 years of research, *Educational Measurement: Issues and Practice*, 24(2), 3-13.
- Schmidt, S. M. P., Ralph, D. L. & Buskirk, B. (2009). Utilizing online exams: A case study, *Journal of College Teaching & Learning*, 6(8), 1-8.
- Simkin, M. G. & Kuechler, W. L. (2005). Multiple-choice tests and student understanding: What is the connection?, *Decision Sciences Journal of Innovative Education*, 3(1), 73-97.
- Tarrant, M., Knierim, A., Hayes, S. K. & Ware, J. (2006). The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments, *Nurse Education Today*, 26(8), 662-671.
- Yates, R. & Beaudrie, B. (2009). The impact of online assessment on grades in community college distance education mathematics courses, *American Journal of Distance Education*, 23(2), 62-70.

## **Appendix 1: Textbook Testbanks Reviewed (Alphabetical order)**

- Atrill, P., McLaney, E. & Harvey, D. (2014). *Accounting: An Introduction*, 6th Ed, Frenchs Forest: Pearson Australia.
- Birt, J., Chalmers, K., Byrne, S., Brooks, A. & Oliver, J. (2012). *Accounting: Business Reporting for Decision Making*, 4th Ed, Milton: John Wiley & Sons Australia Ltd.
- Deegan, C. (2012). *Australian Financial Accounting*, 7th Ed, North Ryde: McGraw-Hill.
- Henderson, S., Peirson, G., & Herbohn, K. (2011). *Issues in Financial Accounting*, 14th Ed, Frenchs Forest: Pearson Australia.
- Hoggett, L. E., Medlin, J., Chalmers, K., Hellmann, A. & Beattie, C. (2015). *Financial Accounting*, 9th Ed, Milton: John Wiley & Sons Australia Ltd.
- Leo, K., Knapp, J., McGown, S. & Sweeting, J. (2015). *Company Accounting*, 10th Ed, Milton: John Wiley & Sons Australia Ltd.