Producing Better Quality MCQs at First Year Level: Are Guidelines and Templates Enough?

C. Sue Price
Mitchell Hughes

Abstract

Largely owing to high student numbers and the constraint of having to operate a common curriculum over two campuses (Westville and Pietermaritzburg), the School of Information Systems & Technology (IS&T) at the University of KwaZulu-Natal (UKZN) has adopted multiple choice questions (MCQs) as its primary assessment method at first year level. In this paper, the authors reflect on initiatives undertaken within the School to improve MCQ construction against a backdrop of considerable student and staff diversity, particularly in terms of language. These initiatives centre on the use of a set of MCQ guidelines and an MCQ template, together with a structured cycle of review and feedback. This reflection on current practice forms part of the first stage of an action research programme of MCQ construction improvement in the School. The use of the template and guidelines has been beneficial since they have made staff more aware of quality issues. In addition, however, issues such as staffing, the time needed to develop questions, the importance of assessment, and question review also play a significant role in producing quality MCQs. The paper concludes with suggestions for further work.

Keywords: Multiple choice questions (MCQs), MCQ construction, Bloom's taxonomy, student diversity, staff diversity, teacher-practitioner.

Background and Context

First Year Information Systems and Technology at UKZN

The School of Information Systems & Technology (IS&T) operates over two campuses (Westville and Pietermaritzburg). Despite this divide, the School strives for near complete commonality in its four first year modules. Module outlines, learning outcomes, prescribed textbooks, lecture plans, practicals, tutorials and, most importantly, assessments are therefore identical across both campuses. The language of instruction and assessment is English.

Each module is divided into multiple topics, each coordinated by a Topic Leader who is responsible for drawing up a topic outline, including lecture plans and learning outcomes for each lecture, and ensuring that the topic's tutorials and practicals (if applicable) are prepared on time. Each topic therefore has both a Topic Leader on one campus and another staff member to lecture on the topic on the other campus. Topic Leaders report to the overall Module Coordinator, who is usually situated on the Westville campus. In Pietermaritzburg, there is also a Module Coordinator who addresses issues that arise on that campus.

The modules and their roles in the IS&T curriculum are summarised in Table 1.

Module Code	Module Name	Role in Curriculum
ISTN100	End User Computing	Service module for non-IS&T majors
ISTN101	IS&T for Business	Compulsory for all B.Com students
ISTN102	IS&T Development Fundamentals	Compulsory for IS&T majors
ISTN103	Development & Applications Fundamentals	Compulsory exit course for Accounting majors; elective for other non-IS&T majors

Table 1: First year IS&T modules

In this paper, the authors concentrate on the two second semester modules,

ISTN102 and ISTN103, as both authors were intimately involved in these modules.

In an ideal world, with lower student numbers, one would assess student learning using various constructed response assessment methods. However, with larger student numbers, even though this situation is not ideal, the School has adopted MCQs as the preferred means of assessing large classes, primarily for logistical reasons (large classes can be assessed more quickly and faster feedback can be given). With the exception of IST102 (which has a short answer component, part of which was assessed in 2010 via a practical programming test), assessment in all modules occurs entirely via MCQs. It is therefore critical that the School focuses on constructing MCQs that are of acceptable quality and fair to a large and diverse student population. In attempting to do so, the School operates within several key environmental constraints, as outlined below.

Student Numbers and Diversity

The School is responsible for upwards of 2000 students per semester across its first year modules. The average student numbers for each module per campus between 2008 and 2010, are summarised in Table 2.

Module Code	**Semester 1**		**Semester 2**		**Total Examinees**
	West.	**Pmb**	**West.**	**Pmb**	
ISTN100	262	137	383	138	920
ISTN101	1187	416			1603
ISTN102			426	152	578
ISTN103			872	305	1177
Total students	**1449**	**553**	**1681**	**595**	**4278**
Total students/semester	**2002**		**2076**		

Table 2: Average first year IS&T student numbers (2008 – 2010)

First year students in the School come from a wide variety of cultural and educational backgrounds and speak a wide variety of languages. Students across all the modules in 2008, 2009 and 2010 spoke between 19 and 25 home languages (see Table 3). On average, 50.8% of these students had English as a home language, while 41.0% had Zulu. For ISTN102 and ISTN103, an average of 52.2% and 58.1%, respectively, had English as a home language.

	2008	2009	2010	Average
All first year IS&T students				
Total no. of home languages	19	25	19	21.0
Percentage with English as home language	51.6%	51.4%	49.6%	50.8%
Percentage with Zulu as home language	40.8%	41.0%	41.2%	41.0%
ISTN102 students				
Total no. of home languages	14	13	14	13.6
Percentage with English as home language	50.6%	50.7%	55.3%	52.2%
Percentage with Zulu as home language	41.0%	41.8%	34.9%	39.2%
ISTN103 students				
Total no. of home languages	10	18	16	14.6
Percentage with English as home language	61.6%	56.3%	56.4%	58.1%
Percentage with Zulu as home language	33.9%	37.5%	35.1%	35.5%

Table 3: Home language diversity amongst first year IS&T students (2008 – 2010)

The student body is not only diverse in terms of language. The modules draw students from the Faculties of Management Studies; Science and Agriculture; Law; and Humanities, Development and Social Sciences, which means that there is considerable diversity in terms of the skills that they bring into the modules.

Staff Numbers and Diversity

In addition to a diverse student body, the staff members who assess first year modules are also numerous and diverse. There can be upwards of 10 staff members involved in any of the modules during a semester and staff members assigned to first year modules tend to be newer and/or less-experienced. In both 2008 and 2009, there were eight staff members who each had less than two years lecturing and assessment experience contributing to the assessments. In 2010, as no new staff members joined the group of first year lecturers, there were only four lecturers in the group with less than two years experience (see Figure 1). Although there is a School

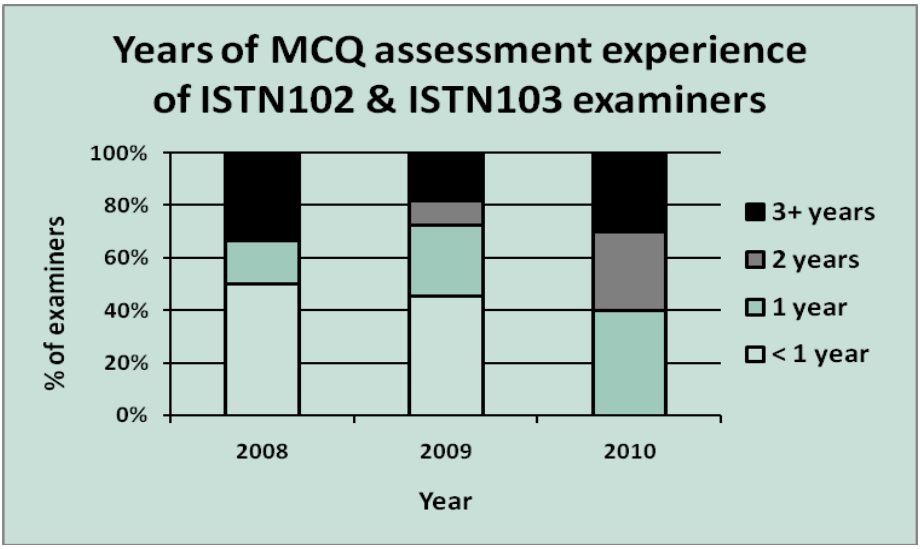

Figure 1: MCQ assessment experience of the ISTN102 & ISTN103 examiners (2008 – 2010)

policy that all lecturers should teach at first year within a two-year cycle, this has not been fully implemented as some lecturers' expertise is needed in other modules. The modules had 19 different examiners between 2008 and 2010, with only four being involved as assessors over all three years. With high rates of diversity and turnover amongst staff, it is naturally difficult to build, retain and pass on the capacity to construct MCQs of acceptable quality from year to year.

Similar to the student body, there is also language diversity amongst the staff. In the ISTN102 and ISTN103 modules, 45% to 75% of the lecturers who contributed examination questions had English as a home language (see Figure 2).

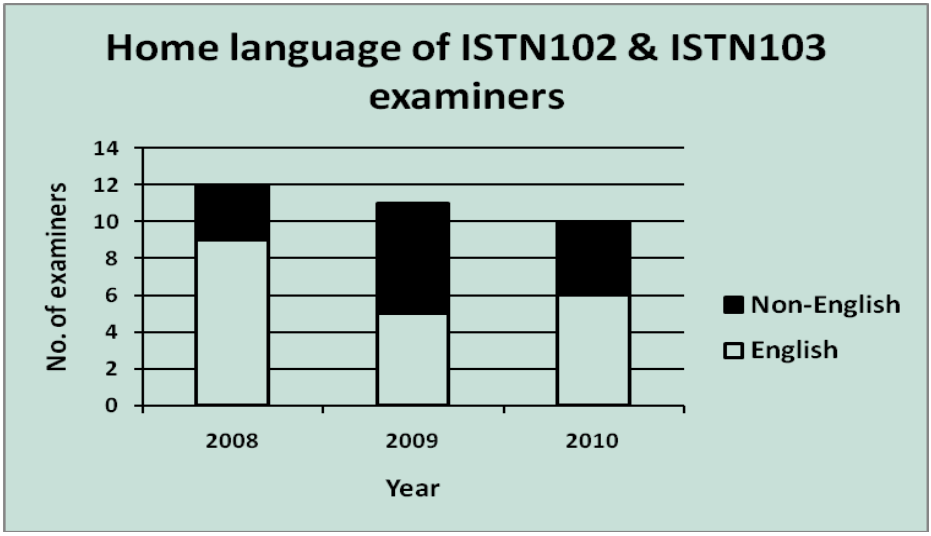

Figure 2: Home language of ISTN102 & ISTN103 examiners (2008 – 2010)

School Interventions: The MCQ Template (2005) and the MCQ Guidelines (2008)

In 2005 the ISTN101 moderator highlighted several problems with the MCQs. As a result a training course was held to improve the lecturers' MCQ

construction skills. Over 20 staff members attended, although by 2008 only four remained in the first year lecturing team.

Also in 2005, the MCQ template was instituted, which contained the MCQ itself, and a table showing its solution, the learning outcome, Bloom's taxonomy level (i.e. whether it is a 'Recall', 'Comprehension' or 'Application' type question), where the solution could be found, and the question developer's name. An example of this template is provided in the Appendix.

The MCQs favoured by the School are typically the best-answer variety. These have a stem which describes the problem to be solved, a single correct or 'best' answer, and several distractors (incorrect answers) (Clegg & Cashin 1986:1; Hansen & Dexter 1997:94; Tarrant *et al*. 2006:663). There are typically between two and four distractors for each question.

In 2008, as a result of regular question writing flaws still occurring, a set of MCQ guidelines was drawn up and circulated to staff. The guidelines cover several broad areas in MCQ construction, including academic content, formatting, writing the stem, and developing distractors. The School has also focused on the project management aspect of assessment compilation as part of its drive to construct better quality MCQ assessments.

Characteristics of Acceptable Quality MCQs

Based on the literature, a MCQ is deemed to be of acceptable quality when:

- questions are testing a learning outcome and not some obscure content (Haladyna *et al*. 2002:312; Jozefowics *et al*. 2002:158; McCoubrie 2004:709; McCoubrie & McKnight 2007:507).
- a range of learning outcomes are being tested (Clegg & Cashin 1986:2; McCoubrie 2004:709-710).
- students' content knowledge is being tested, and not their knowledge of the content *and* the English language (Farley 1989b:11; McCoubrie 2004:709-710; Stupans 2006:62).
- students' content knowledge is being tested, and not their ability to write MCQ tests (McCoubrie 2004:709-710).
- questions are phrased unambiguously and clearly (Thompson *et al*. 2002:7).

- there is one correct or best answer to each question (Farley 1989b:12; McCoubrie & McKnight 2007:507).
- there is a uniform format to the questions (Ware & Vik 2009:241), as well as a lack of grammatical errors (Clegg & Cashin 1986:3; Haladyna *et al*. 2002:312), so that the student is not distracted from answering the questions by oddities which might increase their test anxiety, or given clues as to the answer.

A survey of the literature relating to MCQs follows.

Literature Review

MCQ Usage in Higher Education

MCQs have been used as an assessment method in a variety of disciplines, including the medical professions, such as nursing (Masters *et al.* 2001:25; Clifton & Schriner 2010:12), medicine (Downing 2005:134; Ware & Vik 2009:238), pharmacology (Stupans 2006:59) and radiology (McCoubrie & McKnight 2007:506). MCQs are also used in the humanities (Riecker & Makhoba 2009), to assess business subjects (Gatfield & Lamar 2006:107), computer programming (Simkin & Keuchler 2005:74) and IT (Woodford & Bancroft 2004:948). MCQs are used to assess learning at a variety of levels of education, from first year (Riecker & Makhoba 2009) to final year 'high stakes' assessments, the outcome of which assesses students' readiness to move into a profession or not (McCoubrie & McKnight 2007:506; Tarrant & Ware 2008:199; Ware & Vik 2009:238).

The Disadvantages of MCQs

There are numerous disadvantages of MCQ assessments. Using MCQs to assess learning should only take place when it does not matter how well a student can put together their own response (Clegg & Cashin 1986:2; Schuwirth & van der Vleuten 2003:643). Not all learning outcomes can be tested via MCQs. For example, one cannot test whether a student can draw a particular diagram; one can only test whether a diagram can be interpreted. Because one MCQ option is correct, students could guess the answer (Clegg

& Cashin 1986:2), thereby circumventing the purpose of the test, which is to assess what knowledge the student has gained. Some students find MCQs disempowering, as they cannot state the response in their own words, but have to choose from the options given to them (Paxton 2000:114).

The Benefits of MCQs

There are several benefits of using MCQs to assess student learning. They can test a wide range of content in one assessment (Clegg & Cashin 1986:2; McCoubrie & McKnight 2007:506). They can be used to test all levels of learning (as defined by Bloom 1956 – see Figure 3), from 'Knowledge' or 'Recall' to 'Evaluation' (Clegg & Cashin 1986:1; McCoubrie & McKnight 2007:506).

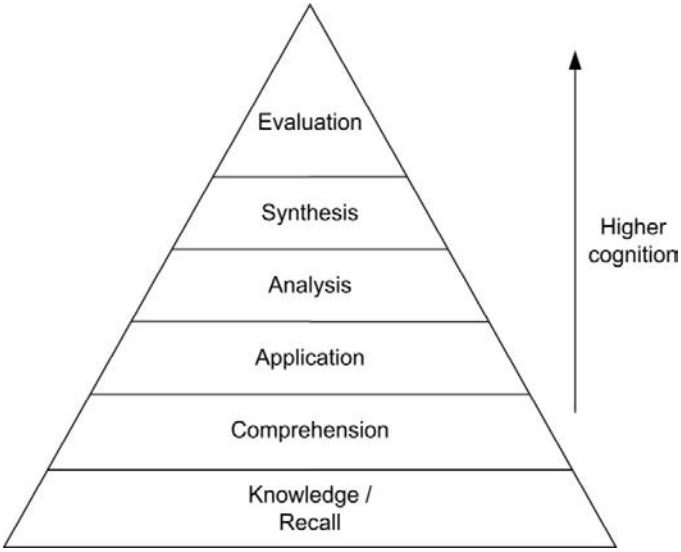

Figure 3: Bloom's levels of cognition (Woodford & Bancroft 2004:948)

MCQs take relatively less time for students to answer (Clegg & Cashin 1986:2). Since marking can be undertaken by a machine, marking is quicker and is more uniform (Clegg & Cashin 1986:2), especially when compared to the marking of short answer questions and essays (Schuwirth & van der

Vleuten 2003:644). These factors make it easier to assess more students at a time, and give the students their test scores back quicker (Simkin & Kuechler 2005:75). Some students, whose home language is not the language of assessment, prefer MCQs as they do not have to worry about the spelling and grammar of their response (Gatfield & Larmar 2006:109).

Issues in MCQ Construction

While it is fairly easy to construct an MCQ that assesses knowledge (or recall) (Clegg & Cashin 1986:1), it takes a long time to generate better questions of higher taxonomic levels (Clegg & Cashin 1986:1; Tarrant *et al*. 2006:669). The amount of time that it takes to create MCQs is generally underestimated (Clegg & Cashin 1986:2). This could explain why there is an abundance of recall MCQs asked in tests (Jozefowics *et al*. 2002:159; Stupans 2006:60; Tarrant & Ware 2008:200) and found in textbook question test banks (Masters *et al*. 2001:27). Van Hoozer *et al*. (1987:280) estimate that developing a single question could take a professional item writer over an hour. Schuwirth and van der Vleuten (2004:977) estimate that a novice question compiler could take up to three hours to develop a context-rich question.

In order to implement MCQs successfully there are several pre-conditions (Clegg & Cashin 1986:2). Firstly, the question developer should have mastered the subject matter thoroughly. This includes not only the course content, but common misconceptions and fallacies about the subject. Secondly, a set of educational objectives must have been developed and used for the course material. This enables one to assess the learning gained for specific outcomes, and also the level at which the learning has been gained. Thirdly, one needs to understand the examinees' educational backgrounds and intellectual abilities so that the difficulty or complexity of the questions can be adjusted accordingly. Finally, the question developer should be excellent at writing simply and concisely in language which the students can understand. This is particularly important for those students whose first language is not the language of assessment (Paxton 2000:115; Tarrant *et al*. 2006:668). Cassels and Johnstone (1984:613-615) found that students' marks improved when questions were stated more clearly or briefly, when using

common words, phrasing the question in a positive way, and using the active voice. In a workshop on assessment, Dempster (2010) recounted how learners from African schools in South Africa ignored options which contained words they did not understand (even if the option was correct). In an article on the same study, Dempster and Reddy (2007:919) reported that higher sentence complexity negatively affected the learners' ability to answer questions correctly. From an assessor's point-of-view, Tarrant *et al.* (2006:668) comment on the large number of unclear or ambiguous questions in the test bank they were reviewing: they surmise that this might have been due to the large number of contributing teachers who do not speak English as a first language.

'If tests are not well-constructed, assessments of student performance may be invalid' (Tarrant *et al*. 2006:664). This means that questions must be free of flaws. Multiple choice question flaws include issues like giving word clues, grammatical clues or logical clues in the stem or options, implausible options, making the correct option markedly shorter or longer than the incorrect options, and having no correct answer, or several correct answers. When a question includes a flaw, it either makes the questions easier or harder than they should be (Downing 2002:S104; Downing 2005:141; Tarrant & Ware 2008:198-199). This sometimes affects whether borderline students pass or not (Downing 2005:141); it could also affect whether a student gets a distinction or not (Tarrant & Ware 2008:203).

There are many guidelines available which describe well-formed questions and how to avoid common flaws, e.g., books such as Case and Swanson (2001) and Haladyna (2004); also articles, such as Clegg and Cashin (1986); Hansen and Dexter (1997); Haladyna *et al.* (2002); Masters *et al*. (2001); Tarrant *et al*. (2006); McCoubrie and McKnight (2007); Tarrant and Ware (2008); Ware and Vik (2009). In spite of these guidelines being easily available, flawed questions are still being included in tests and examinations (Downing 2002:S103; Downing 2005:140; Tarrant *et al.* 2006:667; Tarrant & Ware 2008:202; Riecker & Makhoba 2009). In many cases, the question's flaw could be removed by re-wording the question (Tarrant *et al*. 2006:667).

The content of an assessment steers students to what they should learn (Swanson & Case 1997:74). If students are only tested on recall questions, they will spend their time cramming facts into their heads (which

will probably soon be forgotten). In fact, Boud (2009) asserts that no recall questions should be included in summative assessments if students are to be encouraged to gain habits of life-long learning. If students are to have learned skills of content recall, comprehension and application by the end of the module, these skills need to be assessed (Tarrant *et al.* 2006:663). There is no research that shows appropriate proportions of recall / comprehension / application questions to include in tests for different years of study (Masters *et al*., 2001:26). However, it is thought that for higher years of study, there should be fewer lower level questions (Tarrant *et al*. 2006:669). A quality MCQ assessment, therefore, is made up of questions which do not have flaws, which assess the student's knowledge of the range of learning outcomes and assesses the student's knowledge at different cognitive levels.

A possible reason MCQs in tests and examinations have so many flaws and are often asked at the lower levels identified by Bloom is the relative importance that academics place on lecturing compared to assessment (Jozefowics *et al*. 2002:157; Tarrant *et al*. 2006:669). Failing to pay enough attention to assessment means that not enough time and effort is given to the development of questions, and the process of developing questions is not begun early enough (Clegg & Cashin 1986:3). This attitude towards assessment fails to take account of the fact that if the examination questions are flawed, then the examination may not be valid, and all the effort put into lecture development and delivery has been a waste of time.

The following consolidated list outlines several key principles which, if followed, should yield better quality MCQs:

- Seek training in the writing of MCQs. This has been shown to help improve developers' questions (Hansen & Dexter 1997:96; Jozefowicz *et al*. 2002:159; Schuwirth & van der Vleuten 2003:644).
- Make sure that the person responsible for the assessment is an expert in the subject (Clegg & Cashin 1986:2).
- Use learning outcomes for the course, and then assess the students based on those learning outcomes (Clegg & Cashin 1986:2; McCoubrie 2004:710). No questions should test obscure facts or knowledge (McCoubrie 2004:710).

- Use a test blueprint to outline which of Bloom's levels each topic should include in the test (Farley 1989a:4).
- Start early with question generation (Clegg & Cashin 1986:3). Time is needed to think of questions, and time is also needed to review the questions (both self-assessment and peer review). If necessary, time may also be needed to improve the question or to replace it. Time is also needed to practice writing good questions (Farley 1989b:39).
- Concentrate on developing higher order questions (on the Bloom's hierarchy), as lower level questions are easier to develop. Avoid using textbooks' test banks as a source of possible questions, as they tend to contain a high proportion of recall questions (Masters *et al*. 2001:27; Tarrant *et al*. 2006:668).
- Write using simple English (Tarrant *et al*. 2006:668). Do not include unnecessary text in the question (Haladyna *et al*. 2002:312). Make sure that knowledge is being tested, not knowledge *and* English skills (Stupans 2006:62).
- Note the Bloom's taxonomy level and other data about the question (e.g. the developer's name and the learning outcome it is testing), so that these can be included in the test blueprint (Tarrant *et al*. 2006:665).
- Assess your own questions against the guidelines. Give your questions to others to review (Jozefowics *et al*. 2002:160). Accept feedback and improve or replace questions as necessary (Tarrant *et al*. 2006:669; Clifton & Schriner 2010:16).
- Assess the test for uniformity of format (Ware & Vik 2009:241). List MCQ options vertically (not horizontally), as this is easier for the examinee to read. Use capital letters for the options. Check that the correct answer is distributed over the options (Clegg & Cashin 1986:4).
- After the students have written the test, perform distractor analysis on the questions to see if the distractors need improving or replacing (Farley 1990:9; Downing 2002:S104; Tarrant *et al*. 2009).

Methodology

This paper forms the first (i.e. diagnostic) phase of an action research programme (Baskerville 1999:6) of improving MCQ construction in a first year IS&T context against a backdrop of considerable student and staff diversity. This phase involves reflecting on current practice before designing, introducing and studying future interventions (Baskerville 1999:6). The authors approach the subject as Teacher-Practitioners (Pollard *et al*. 2005:9) using an interpretivist lens.

MCQ Construction at UKZN IS&T

This section describes the School's existing MCQ construction strategy (implemented in 2008) as envisaged by the Coordinator(s). Actual practice will be reflected upon in the next section.

At the start of each module, the Module Coordinator circulates the MCQ guidelines and MCQ template to all staff members (see Figure 3). The Coordinator also circulates a schedule indicating when questions must be sent to Topic Leaders, the number of questions required from each staff member for each topic, the nature of questions required (MCQ or short answer) and what percentage of questions should be 'Knowledge' ('Recall'), 'Comprehension' or 'Application' type questions (Bloom 1956). It is expected that every staff member use the template and guidelines to prepare their allotted number of questions before the deadlines.

The MCQ guidelines outline principles which staff members can use to assess the quality of the MCQs they and their peers construct. The guidelines cover issues such as academic content, formatting, stem construction and developing distractors.

The MCQ template is primarily an administrative intervention to ensure that all questions are received in a similar format to aid in the compilation of the paper. It is also useful for creating the model answer and for examining Bloom's levels and whether solutions are well distributed among the options.

The first level of quality assurance is self-reflection, i.e. carried out by the staff member him/herself. The questions are then forwarded to the Topic Leader (and any other staff members involved in that particular topic)

for the second round of review. Questions which need attention are edited or replaced. When the Topic Leader is satisfied that all questions for the topic are of a suitable quality, they are sent to the Module Coordinator, who is responsible for the compilation of the full paper. After a final quality check by the Coordinator, the full assessment is circulated to the entire lecturing team for comment.

Once a draft version of the assessment has been finalised, all staff members attend a review meeting where every question is re-checked. Only once the question paper has passed through the review process and the issues flagged are addressed is it considered ready for moderation (both tests and examinations are moderated). The School takes the position that all staff members involved in the module share the responsibility for the quality of the assessment.

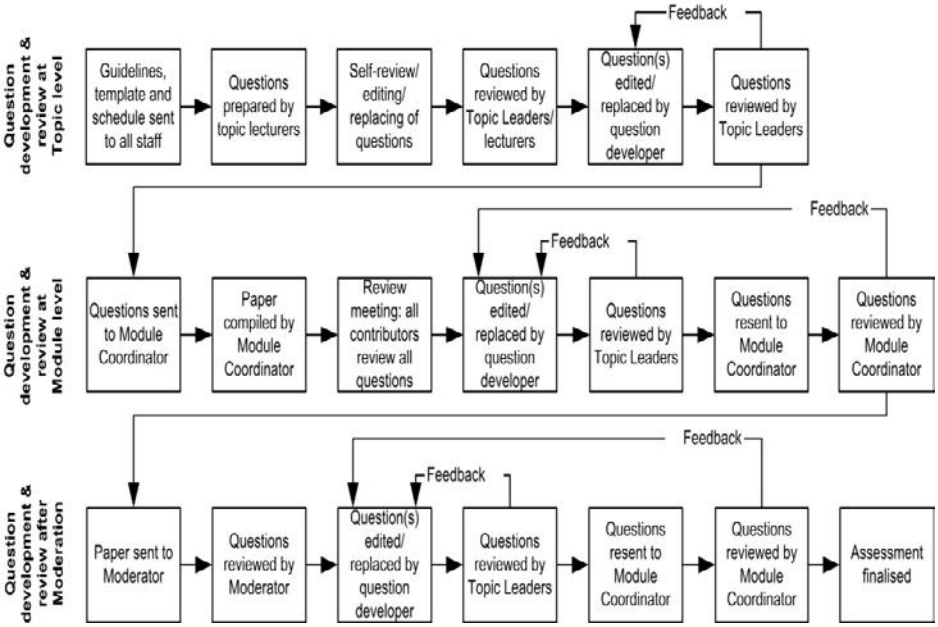

Figure 4: The MCQ construction process of UKZN IS&T

In summary, the template and guidelines have several desired outcomes. The template ensures a uniform 'look and feel' for every assessment that students encounter during a module. The template also allows the Module Coordinator to review the paper to ensure that an appropriate mix of Bloom's taxonomy levels is used in the assessment. The guidelines allow staff members both to self-assess their own work, as well as assessing that of their peers. Using the guidelines helps to ensure that all MCQs are understandable and unambiguous for a large and diverse student group. The template and guidelines also serve as a developmental tool, particularly for newer members of staff.

The Authors' Reflections on MCQ Construction in Two First Year Modules

In this section, the authors reflect on their experiences and observations of the construction of MCQs in two first year modules between 2008 and 2010. This reflection concentrates on the second semester courses, ISTN102 and ISTN103, after the template and guidelines were introduced. Both authors are content specialists (and not educational experts). The first author was a Topic Leader in both modules in 2008 and 2009, and lectured in both modules in 2010; the second author was the Pietermaritzburg Module Coordinator for both modules for 2008 and 2009 and lectured in both modules in 2008; he also moderated the tests and examinations in the modules in 2010. These reflections do not necessarily represent the opinions of other first year lecturers.

Layout and Formatting

The impact of the template and guidelines has largely been positive and drastically reduced paper compilation time. As the Module Coordinator sometimes delegates the compilation of a test or examination to other academic staff members, more people are appreciative of how much time the template saves.

Topic Leaders have been encouraged to return questions not conforming to the template or guidelines to the question developer

immediately. This does not always happen, particularly when the Topic Leader is junior to his/her colleague. However, formatting issues are usually addressed at the review meeting. The grammar, spelling and formatting of the final papers have generally improved. The template and guidelines are a constant reminder to staff to think about the quality and format of their questions. They have also proved useful for newer staff members, who can quickly learn the expected question layout.

Evolving Module Content and Recurriculation

Since IS&T is a highly dynamic discipline, the School is frequently faced with the need to recurriculate and update its modules in line with the latest developments in technology. This means that the School has limited capacity to build up a catalogue of acceptable quality MCQs that can be used from year to year. It also means that staff members do not become 'experts' in certain areas. This again impacts negatively on the quality of the MCQs, as staff tend to fall back upon more superficial or 'Recall' questions.

Timing and Deadlines

Staff members, especially those who are new to lecturing and/or to a particular topic, tend to underestimate the amount of time needed to construct MCQs of acceptable quality. Construction is often started too late, which means that initial question submission deadlines are missed. This, in turn, reduces the time available for effective review, moderation, and modifying/replacing questions, which has implications for the quality of the final product. Often staff members have reduced time to respond to the Moderator's comments. As new staff members become more experienced, their delivery on the initial deadlines tends to improve.

The timing of the MCQ construction process is further exacerbated by the number of staff members involved in these modules. When more people are involved there is greater potential for delays. This implies that deadlines for initial submission of questions should be set earlier; however, this might mean that lecturers have to construct MCQs long before they have actually started to lecture on their topic. While this may be easy for a seasoned lecturer, a novice lecturer may find this quite difficult.

Review

Having reviewed the paper beforehand, all staff members involved in the module attend the review meeting. During the meeting each MCQ is reviewed, any necessary clarification is sought, and suggestions for improvements are made. It is the question developer's responsibility to make the changes, and the Topic Leader sees that all questions for his/her topic are suitably edited or replaced before they are re-sent to the Module Coordinator (see the middle row of Figure 4). It is at these meetings that issues which make a quality MCQ are addressed.

Issues such as clarity, ambiguity, conciseness, poor grammar, formatting, the Bloom's taxonomy level and question difficulty are raised. If one person on the lecturing team does not understand the question, it is likely that at least one student will not understand it either. A student writing a test is stressed, and has no time to ponder ambiguities or meaning. These issues therefore need to be addressed. It is usually at the review meeting that the grammar and language of questions are addressed (if not addressed before the meeting). In these meetings the Bloom's taxonomy level is also reviewed and adjusted. It has been noted that the more a question has been reviewed, edited or changed before the review meeting and moderation, the less likely it is to have change requests at the review meeting or moderation stage of the question construction process. It is often the questions about which there was uncertainty or differing opinions prior to the review meeting that are flagged by the lecturing team or Moderator for attention.

It is difficult to accept critique of one's questions, especially in the first review meeting one attends. Also, junior or less experienced staff members are sometimes loathe to offer critique on their senior colleague's questions. However, the review meetings present a learning opportunity for staff members on how to improve their questions. It has been encouraging to see increased engagement by the first year lecturing team at recent review meetings.

In the review meeting, if there is an issue on which there is more than one opinion, this is left to the Moderator to resolve. The less experienced the lecturing team is at constructing MCQs, the more work the Moderator has, to suggest changes or request replacements, which makes this task more onerous. It is the authors' opinion that another review meeting and/ or another moderation needs to take place after each question developer has

responded to the Moderator's comments. (Some lecturers find themselves wondering if their replacement phrases or questions are worse than the original.) However, this would require additional time, which means that the delivery of initial questions needs to be made earlier.

It should be noted that the reviewing process is very time consuming. Questions have to be reviewed, and reviewed again each time a change is made (at topic level, module level and paper readiness level). The amount of time required for this is generally underestimated.

Quality of Questions

With the increasing experience of the staff members on the two modules, and the increasing engagement with the guidelines and template, the quality of the questions in the assessments has generally improved. The more a question has been reviewed, the better it is likely to become. This means that staff need to allow enough time for question development, reflection, feedback and improvement of their MCQs.

Discussion and Conclusions

All first year IS&T courses at UKZN are primarily assessed via MCQs. The language of instruction and assessment is English, which can prove challenging for 49% of students, whose home language is not English. This paper does not debate the merits of conducting assessment through MCQs. Rather, its aim is to reflect on how MCQs of acceptable quality can be constructed, so that the diverse student body is not disadvantaged.

In order to construct MCQs of acceptable quality, the School has introduced an MCQ template and a set of MCQ guidelines, together with a cycle of review and feedback. The template and guidelines assist staff members to construct MCQs which are more easily understood by students without trivialising the learning outcome being tested. In addition, they provide a starting point for lecturers to engage with understanding the fundamentals of constructing MCQs of acceptable quality. The template and guidelines also assist staff members who construct MCQs to create a uniform-looking assessment. A uniform style is a criterion of quality that Ware and Vik (2009:241) also strive to achieve.

Although the template and guidelines do contribute to the construct-

tion of better quality MCQs, these alone will not necessarily yield a high quality overall assessment. It is believed that staffing issues, MCQ development time, stressing the importance of assessment (relative to lecturing), and MCQ review also need to be addressed in order to improve the overall quality of MCQs in assessments.

As mentioned, the staff complement at first year is made up of new or less experienced lecturers. However, as Clegg and Cashin (1986:2) state, it is important for the MCQ developer to be an expert on the topic in which they are developing questions. This could be addressed in several ways. Firstly, MCQ construction training needs to be given to staff (Jozefowics *et al.* 2002:159; Downing 2005:142; Tarrant *et al.* 2006:669; Tarrant & Ware 2008:202; Clifton & Schriner 2010:16). Even though staff members believe they know how to construct MCQs, they still need training (Clifton & Schriner 2010:15). Training has been found to improve the quality of MCQs in assessments (Hansen & Dexter 1997:96; Jozefowics *et al.* 2002:157). Related to training is the issue of the grammar and style of questions, i.e., the question developer's skill in the language of assessment (Tarrant *et al.* 2006:669). Thirdly, a more even mix of experienced and inexperienced lecturers needs to be teamed up to teach first year modules on the two campuses. If a new topic is being introduced, or an existing topic reworked, at least one of the lecturers allocated to such a topic needs to be someone who has the ability to develop appropriate learning outcomes and good MCQs. More experienced MCQ writers also need to be available to help those who are less experienced (Clifton & Schriner 2010:16). Finally, defining, and then assessing learning outcomes needs to be stressed, so that assessment questions do not concentrate on trivia, but on the main issues which need to be learned (Farley 1989b:10; Jozefowics *et al.* 2002:158; McCoubrie 2004:710; McCoubrie & McKnight 2007:508).

Echoing the findings of Clegg and Cashin (1986:3), the amount of time it takes to develop high quality MCQs is underestimated by most staff members. Whilst it is fairly easy to develop MCQs that test 'Knowledge' (or 'Recall') (Clegg & Cashin 1986:2), it takes much longer to construct high quality questions at high taxonomic levels (Clegg & Cashin 1986:1; Clifton & Schriner 2010:12). This can, in part, be addressed by ensuring that there is an expert in the topic area team, who is able to start and guide the commencement of the MCQ construction process earlier. Good project

management is also an important factor, as a late start often means a delay in the delivery date. An early start, together with frequent peer feedback or mentoring, would also help each staff member to develop higher-order questions timeously. Time also needs to be allocated to self-assessment of questions and peer review, as well as question improvement or replacement (Clegg & Cashin 1986:3; Tarrant *et al.* 2006:669; Clifton & Schriner 2010:16). In addition, examiners whose home language is not English may need to start the MCQ construction process earlier than their English first language counterparts, so there is more time for review.

The importance of assessment, compared to lecturing, is the third area which needs to be stressed. If the assessment is flawed, much of the lecture preparation and delivery efforts (which form a large part of a lecturer's life) could be wasted. Perhaps lecturers think of themselves as lecture preparers and deliverers because their title is lecturer, not assessor? However, in the allocation of workload and in the process of the semester's work, assessment needs to be highlighted as an integral part of the lecturer's duties (Jozefowics *et al.* 2002:157). The importance of assessment could be emphasised by a discussion of the MCQ tests' difficulty and distractor analysis (Farley 1990:8-9; Downing 2002:S104; Tarrant *et al.* 2009). Mentorship in good practice would also be beneficial.

Lastly, the review of questions which have been developed needs to be emphasised. Numerous authors have noted the importance of the review process in producing MCQ assessments which do not contain question flaws, and which are valid assessments of the course's learning outcomes (Jozefowics *et al.* 2002:158; Tarrant *et al.* 2006:669). The more times that questions have been self-assessed or peer-assessed and improved by the time they are submitted to the Module Coordinator, the more likely it is that a higher quality test or examination will be produced, which in turn will mean a fairer assessment for students. This approach of improving the quality at the source and reducing time wastage is what is advocated in quality improvement initiatives (Christopher & Rutherford 2004:26). In the case of IS&T, when poorer quality questions were submitted to the Module Coordinator, the participating reviewers and the Moderator had to work harder to ensure that the assessment was of an acceptable quality. It is pleasing to note, however, that the School does plan time for review, unlike instances reported in the literature (Jozefowicz *et al.* 2002:158).

In this review process, it needs to be remembered that the purpose of the review is to detect if there are ambiguities in the questions (which would confuse students in the examination), or confusing instructions or options. The discovery (and improvement) of these before the examination greatly enhances the validity of the examination (Tarrant *et al.* 2006:668). As relatively few staff members participate in this review, this increases the workload (in terms of reviewing and giving feedback) of those who do participate. Appointing an MCQ expert, who is not involved with the module, as an external question reviewer and mentor to those who are developing questions would be beneficial. This would ease the workload of the current participators, as well as providing a learning opportunity for those whose questions need improving. It would also help more lecturers who are reluctant to comment on the work of their colleagues, especially when these colleagues are considered experts (Jozefowicz *et al.* 2002:158).

While assessment through MCQs is arguably not ideal, MCQs have become the preferred assessment method in the School and in other disciplines with large student numbers. At UKZN, MCQs may also be preferred because of the quick turn-around time between the final examinations and supplementary examinations. If constructed properly, MCQs can test a wide range of content in a single assessment (Clegg & Cashin 1986:2; McCoubrie & McKnight 2007:506). Well-prepared MCQs can also be used to test all levels of learning, as defined by Bloom (1956), from 'Knowledge' (or 'Recall') through to 'Evaluation' (Clegg & Cashin 1986:1; McCoubrie & McKnight 2007:506). While it is imperative to have guidelines for good quality MCQ construction and a template to aid formatting consistency, they alone are not sufficient to ensure the development of a quality assessment.

Although this paper has been written in the context of IS&T, there is no reason why its principles cannot be applied in all modules that deal with large and diverse student populations. The paper's principles arguably become less generalisable for more senior years with fewer students, fewer staff members and the scope for different types of assessment.

Recommendations for Further Investigation

This paper expresses the opinions of the authors, based on observation, experience and anecdotal evidence. This forms part of the first phase (i.e.

diagnostic) of an action research programme to improve MCQ construction within the context of a diverse student body and staff complement.

There are several opportunities for further investigation. Even though assessments have been developed for the past three years using the MCQ template and guidelines, MCQ writing flaws are still present in these assessments. One could assess the extent to which IS&T MCQs contain writing flaws. One could also investigate the extent to which writing flaws are present in IS&T textbooks' test banks, and the proportion of questions on the lower levels of Bloom's taxonomy. One could also research the recommended percentages of Bloom's taxonomy levels for questions in tests at various years of study, as there are currently no guidelines for this (Masters *et al.* 2001:26). One could also examine past IS&T MCQs to compare question categorisation using the old and revised Bloom's taxonomy (Anderson *et al.* 2001). An exemplar for a range of IS&T topics at different levels of Bloom's taxonomy could be developed. This is important, as some authors believe that MCQs cannot be developed for higher levels of Bloom's taxonomy (Masters *et al.* 2001:26; Simkin & Kuechler 2005:91). A comparative 'before' and 'after' question analysis showing how questions can be improved after intervention would also be useful as a pedagogical guide. This could include the type of analysis undertaken by Dempster and Reddy (2007), to see if students are guessing the answers, based on which words in the options are understood, or which words appear in both the stem and the options.

The effect that training in MCQ writing has on reducing MCQ construction flaws could also be investigated. Finally, insights from MCQ developers who have used the template and guidelines could be explored.

Acknowledgements

The MCQ template used evolved through many staff contributions towards first year level student assessment, commencing in mid-2005. The MCQ guidelines were originally developed by Mr. M. Hughes and Mrs. M. Hughes (no relation) of the School of IS&T, Pietermaritzburg campus. Additional comments and suggestions were made by members of the lecturing teams on both campuses during the years 2008 to 2010.

The authors gratefully acknowledge colleagues' comments on previous versions of this paper. They also thank the reviewers whose comments and suggestions improved the paper.

Appendix – The UKZN IS&T MCQ Template

1. Gigabyte Technologies is developing a telecommunication system. They want to know the maximum number of users the system can handle before it starts giving network failure problems (like dropping calls). Which test should they use?

[A] Integration testing

[B] Acceptance testing

[C] Unit testing

[D] Volume testing

[E] Beta testing

Q	Solution	Lecture & Learning Outcome	Source of Content: Chapter/Notes & Page	Bloom's Taxonomy: R or C or A	Examiner
1	D	SAD, Lec 2, LO 4	Stair *et al.* (2008: 450)	A	JJ

Note: Bloom's taxonomy key: R = Recall (or Knowledge); C = Comprehension; A = Application.

References

Anderson, LW, DR Krathwohl, PW Airasian, KA Cruikshank, RE Mayer, PR Pintrich, J Raths & MC Wittrock 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Baskerville, RL 1999. Investigating Information Systems with Action Research. *Communications of the AIS* 2,3: 2-32.

Bloom, BS 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Handbook 1: Cognitive Domain. New York: McKay.

Boud, D 2009. Reconstructing Assessment for Longer Term Learning. Presented at the University of KwaZulu-Natal 3rd Annual Teaching & Learning Conference. Durban, 21-23 September 2009.

Case, SM & DB Swanson 2001. *Constructing Written Test Questions for the Basic and Clinical Sciences.* 3rd Edition. Philadelphia: National Board of Medical Examiners.

Cassels, JRT & AH Johnstone 1984. The Effect of Language on Student Performance on Multiple Choice Tests in Chemistry. *Journal of Chemical Education* 61,7: 613-615.

Christopher, M & C Rutherford 2004. Creating Supply Chain Resilience through Agile Six Sigma. *Critical Eye* June-August: 24-28.

Clegg, VL & WE Cashin 1986. *Improving Multiple-choice Tests.* Manhattan, Kansas: Center for Faculty Evaluation and Development, Kansas State University. (Idea Paper 16.)

Clifton, SL & CL Schriner 2010. Assessing the Quality of Multiple-choice Test Items. *Nurse Educator* 35,1: 12-16.

Dempster, E & V Reddy 2007. Item Readability and Science Achievement in TIMSS 2003 in South Africa. *Science Education* 91,6: 906-925.

Dempster, E 2010. What can we Learn from TIMSS 2003? Presentation Handout: Workshop on Using the New Bloom's Taxonomy in Assessment, University of KwaZulu-Natal. Pietermaritzburg: 15 June 2010.

Downing, SM 2002. Construct-irrelevant Variance and Flawed Test Questions: Do Multiple-choice Item-writing Principles Make any Difference? *Academic Medicine* 77,10: S103-S104.

Downing, SM 2005. The Effects of Violating Standard Item Writing Principles on Tests and Students: The Consequences of Using Flawed

Test Items on Achievement Examinations in Medical Education. *Advances in Health Sciences Education* 10,2: 133-143.

Farley, JK 1989a. The Multiple-choice Test: Developing the Test Blueprint. *Nurse Educator* 14,5: 3-5.

Farley, JK 1989b. The Multiple-choice Test: Writing the Questions. *Nurse Educator* 14,6: 10-12,39.

Farley, JK 1990. Item Analysis. *Nurse Educator* 11,1: 8-9.

Gatfield, T & SA Larmar 2006. Multiple Choice Testing Methods: Is it a Biased Method of Testing for Asian International Students? *International Journal of Learning* 13,1: 103-111.

Haladyna, TM 2004. *Developing and Validating Multiple-Choice Test Items*. 3rd Edition. New Jersey: Lawrence Earlbaum Associates, Inc.

Haladyna, TM, SM Downing & MC Rodriguez 2002. A Review of Multiple-choice Item-writing Guidelines for Classroom Assessment. *Applied Measurement in Education* 15,3: 309-334.

Hansen, JD & Dexter, L 1997. Quality Multiple-choice Test Questions: Item-writing Guidelines and an Analysis of Auditing Testbanks. *The Journal of Education for Business* 73,2: 94-97.

Jozefowicz, RF, BM Koeppen, S Case, R Galbraith, D Swanson & RH Glew 2002. The Quality of In-house Medical School Examinations. *Academic Medicine* 77,2: 156-161.

Masters, JC, BS Hulsmeyer, ME Pike, K Leichty, MT Miller & AL Verst 2001. Assessment of Multiple-choice Questions in Selected Test Banks Accompanying Text Books Used in Nursing Education. *Journal of Nursing Education* 40,1: 25-32.

McCoubrie, P 2004. Improving the Fairness of Multiple-choice Questions: A Literature Review. *Medical Teacher* 26,8: 709-712.

McCoubrie, P & L McKnight 2007. Single Best Answer MCQs: A New Format for the FRCR Part 2a Exam. *Clinical Radiology* 63,5:506-510.

Paxton, M 2000. A Linguistic Perspective on Multiple Choice Questioning. *Assessment & Evaluation in Higher Education* 25,2: 109-119.

Pollard, A, J Collins, M Maddock, N Simco, S Swaffield, J Warin & P Warwick 2005. *Reflective Teaching: Evidence-informed Professional Practice*. London: Continuum.

Rieker, M & M Makhoba 2009. Decoding Multiple Choice Assessments: Lessons from the Extended Curriculum Programme in the Faculty of

Humanities, Development and Social Sciences. Presented at the University of KwaZulu-Natal 3rd Annual Teaching & Learning Conference. Durban: 21-23 September.

Schuwirth, LWT & CP van der Vleuten 2003. ABC of Learning and Teaching in Medicine: Written Assessment. *British Medical Journal* 326,7390: 643-645.

Schuwirth, LWT & CP van der Vleuten 2004. Different Written Assessment Methods: What can be said about their Strengths and Weaknesses? *Medical Education* 38,9: 974-979.

Simkin, MG & WL Kuechler 2005. Multiple-choice Tests and Student Understanding: What is the Connection? *Decision Sciences Journal of Innovative Education* 3,1: 73-97.

Stupans, I 2006. Multiple Choice Questions: Can they Examine Application of Knowledge? *Pharmacy Education* 6,1: 59-63.

Swanson, DB & SM Case 1997. Assessment in Basic Science Instruction: Directions for Practice and Research. *Advances in Health Sciences Education* 2,1: 71-84.

Tarrant, M, A Knierim, SK Hayes & J Ware 2006. The Frequency of Item Writing Flaws in Multiple-choice Questions Used in High Stakes Nursing Assessments. *Nurse Education Today* 26,8: 662-671.

Tarrant, M & J Ware 2008. Impact of Item-writing Flaws in Multiple-choice Questions on Student Achievement in High-stakes Nursing Assessments. *Medical Education* 42,2: 198-206.

Tarrant, M, J Ware & AM Mohammed 2009. An Assessment of Functioning and Non-functioning Distractors in Multiple-choice Questions: A Descriptive Analysis. *BMC Medical Education* 9,40, doi:10.1186/1472-6920-9-40.

Thompson, SJ, CJ Johnstone & ML Thurlow 2002. *Universal Design Applied to Large Scale Assessments (Synthesis Report 44).* Minneapolis: University of Minnesota, National Center on Educational Outcomes. Available at: http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html. (Accessed on December 17, 2010.)

Van Hoozer, HL, BD Bratton, PM Ostmoe, D Weinholtz, MJ Craft, CL Gjerde & MA Albanese 1987. *The Teaching Process: Theory and Practice in Nursing*. Norwalk: Appleton-Century-Crofts.

Ware J & T Vik 2009. Quality Assurance of Item Writing: During the Intro-

duction of Multiple Choice Questions in Medicine for High Stakes Examinations. *Medical Teacher* 31,3: 238-243.

Woodford, K & P Bancroft 2004. Using Multiple Choice Questions Effectively in Information Technology Education. Presented at the 21st ASCILITE Conference: 'Beyond the Comfort Zone'. Perth: 5-8 December 2004.

C. Sue Price
Information Systems & Technology
University of KwaZulu-Natal
pricec@ukzn.ac.za

Mitchell Hughes
Information Systems & Technology
University of KwaZulu-Natal
hughesm1@ukzn.ac.za