Nurse Education in Practice (2005) **5**, 238–242

ELSEVIER

Nurse Education in Practice

www.elsevierhealth.com/journals/nepr

FOCUS

Assessment of learning with multiple-choice questions

Anne-Marie Brady *

School of Nursing and Midwifery Studies, Trinity College Dublin, TCD Faculty of Health Sciences Building, St. James' Hospital, Dublin, Ireland

Accepted 15 December 2004

KEYWORDS
Multiple-choice questions;
Assessment

Summary Nurse educators aspire to learning outcomes that will produce competent practitioners. Assessment strategies should be consistent with desired learning outcomes [Manogue, M., Kelly, M., Bartakova Masaryk, S., Brown, G., Catalanotto, F., Choo-Soo, T., Delap, E., Godoroja, P., Murio, I., Rotgans, J., Saag, M., 2002. Evolving methods of assessment. European Journal of Dental Education 6 (3) 53–66] and there is a requirement to diversify approaches. Objective testing, specifically multiple-choice questions (MCQ's) are one of the approaches that may diversify the assessment approach in Irish undergraduate nursing education. In this paper, Quinn's [Quinn, F.M., 2000. The Principles and Practice of Nurse Education, fourth ed., Stanley Thorne (Publishers) Ltd, Cheltenham] cardinal criteria of assessment, practicality, reliability, validity and discrimination provide a useful framework to discuss the issues associated with the implementation of this type of assessment strategy. MCQ examinations that are consistent with educational outcomes can be used to effectively assess aspects of student performance and can facilitate timely feedback and contribute to the process of self-learning. Significant commitment is required to prepare MCQ test items and examination formats that are reliable and consistent with curriculum objectives. Appropriately constructed MCQ examinations are efficient, objective, capable of discrimination and can be combined with other assessment strategies to contribute to a comprehensive student assessment strategy for use in nursing education.
© 2005 Elsevier Ltd. All rights reserved.

Introduction

A diverse range of formative and summative assessment strategies are at the disposal of nurse educators (Milligan, 1996; Quinn, 2000; Race and Brown, 2001; Manogue et al., 2002). Assessment is critical to the success of any educational program and is essentially a gate keeping mechanism through which people do or do not make progress in their chosen profession or occupation (Jarvis, 1985; Manogue et al., 2002). Critical thinking is one of the

* Tel.: +353 1 6083004; fax: +353 1 473 2984.
E-mail address: abrady4@tcd.ie.

1471-5953/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.nepr.2004.12.005

hallmarks of higher education and aspects of performance such as analysis, reasoning, and decision making are just some of the critical indicators by which we can evaluate graduating nurses (Vaughan-Wrobel et al., 1997). The measurement of a student's ability to think critically is a consistent challenge for nurse educators and there is considerable debate on the best assessment methods to assess critical thinking skills (Morrission and Free, 2001; Staib, 2003; Manogue et al., 2002). ''Assessment can take many forms and it can be argued that the greater the diversity in the methods of assessment, the fairer the assessment is to the students'' (Race and Brown, 2001, p. 41). Over use of one assessment strategy can favor particular students and result in one-dimensional or limited evaluation of student performance. Current approaches to class assessment of undergraduate nursing students in the Irish Republic appear to focus on a variety of seen and unseen examination essay formats. Essays items are well suited to assessing learning outcomes in nursing as they require students to develop a rational to support their thinking and position, to consider multiple perspectives and present ideas logically (Oermann, 1999), but as Holsgrove (1992) highlights can be limited in terms of reliability and objectivity. At present, there exists a will in the Irish Nursing education circles to develop some diversity in the current approach. This paper will explore objective testing, specifically multiple-choice questions (MCQs), and analyze its effectiveness as an assessment strategy. Quinn (2000) outlines the following, as the cardinal criteria of assessment, practicality, reliability, validity, and discrimination, and these will be used as a framework to facilitate analysis.

Practicality/usability

The forthcoming changes in nursing education in the Irish republic will result in increasing numbers of students, with diverse needs, causing greater workload in terms of assessment for nursing faculty. Students require feedback in a timely fashion so that assessment is actively incorporated into the learning process. Pamplett and Farnill (1995), state that the importance of MCQs lies is in their ability for testing large numbers of students in a short time and for testing different aspects of a course. Gibbs et al. (1997) explain that large classes do not have a negative impact on student assessment performance in MCQs, but did have a negative impact on performance in essay questions. MCQs lend themselves to standardization and can expedite the scoring of exams as they can be machine graded (Hammond et al., 1998; Morrission and Free, 2001). The MCQ provides objective evaluation of performance as it has the capacity to overcome the subjectivity that may exist in the assessment of essays and oral examinations (Hammond et al., 1998). They can motivate students positively and can assist students in monitoring and affirming their own learning (Flannelly, 2001). Multiple-choice question's have been criticized for testing only superficial learning and regurgitation of facts and have been considered by some to be inappropriate for courses using problem solving, and self-directed approaches (Pamplett and Farnill, 1995). Controversy persists as to whether this method assesses anything other than recall. Masters et al. (2001), explain that while the synthesis and evaluation levels in the cognitive domain of Blooms Taxonomy (1956) cannot be assessed with MCQ, they can competently evaluate knowledge, comprehension, application, and analysis levels. Questions that require the student to recognize problems or discrepancies and infer their causes and devise solutions are examples of MCQ formats capable of challenging the analytical skills of students (Diog, 2000). In their study of undergraduate nursing students' use of on-line MCQ to support learning, Honey and Marshall (2003) endorsed that use of MCQs in formative assessment as they can facilitate inquiry based learning and are a vehicle to actively encourage nursing students to take responsibility for their own learning. This will only be successfully achieved with effectively and accurately constructed questions.

Reliability and validity

Educators need to take steps to ensure their assessment practices and instruments are well-designed and valid (Race and Brown, 2001). The challenge for examiners, in designing objective tests, is to produce assessments that are suitable and consistent in measuring, what they are supposed to measure. Masters et al. (2001) suggest that questions should be written to reflect the level of sophistication that students are expected to perform in practice. MCQs appear to lend themselves to objective scrutiny of their design by several people. It is recommended that lecturers seek critical comment from their colleagues, with teams of lecturers examining new test items to determine suitability for inclusion (Quinn, 2000; Race and Brown, 2001).

Although they appear to be a simple and straightforward style of examination, MCQs are difficult to write (Holsgrove, 1992). All multiple-choice

question formats are prone to construction errors. Holsgrove and Elzubeir (1998), found a high proportion of imprecise terms, in their study of questions included in various medical examinations in the United Kingdom. This affects the reliability and validity of the examination tool. They warn that candidates taking high stakes examinations have a right to expect them to be fair, and accurate, while exam boards have a moral and ethical duty to ensure reliability and validity in examination.

It is important to avoid providing clues through the construction of questions. Schuwirth et al. (1996), highlight the challenge of two types of cueing in question construction. They determined that cueing had a bi-directional effect depending on the level of expertise of the candidates, and the difficulty and content of the questions. Students could be cued to answer correctly or incorrectly. If tests are biased because of poorly written questions, faculty evaluations of student competency will be distorted. One concern evident in the literature was that multiple-choice questions might be obtained from test banks that are questionable in terms of statistical analysis. In their study, Masters et al. (2001) examined multiple-choice questions in test banks accompanying textbooks used in nursing education. The study evaluated the following:

- The adherence to generally accepted guidelines for multiple-choice questions.
- Cognitive level as defined by Blooms Taxonomy of educational (1956).
- The distribution of correct answers.

The most common violations were inadequate spacing, uneven length of options, negative questions, more than one correct answer, implausible options, and grammar errors. The study revealed that 47% of questions were at knowledge level, 25% at comprehension, 22% at application, with only 6.5% at analysis level. On the positive side, the study found the correct answer was fairly distributed among the four options. The implication for faculty is that test banks need to be evaluated and used only to generate ideas. Questions in test banks may not as rigorously assessed for reliability and validity as would be necessary in an examination.

It is a commonly held belief that changing answers in multiple-choice questions is detrimental to performance. Gaskins et al. (1996) discount this theory in their study of student's beliefs and benefits gained from changing answers. They found that students change answers most often to their advantage despite their belief that changing answers was detrimental. They concluded that the inherent reliability produced in good question design, was a means to overcome this potential problem. Questions that are clear in construction will require minimal interpretation or rethinking. They do warn that this is difficult when assessing higher levels of learning such as application, and analysis (Gaskins et al., 1996).

When considering alternatives to essays, one needs to ask, is that what is measured, in 3 h, more reliable than that measured in 1 h? Furthermore, do we intent to measure learning achieved or are we focused on measuring how students communicate learning? Essays are undoubtedly easier and less time consuming to construct, can test higher order learning but can limit the quantity of curriculum sampled. Farley (1989) estimates that it requires 1 h to write a good multiple-choice question, so considerable time need to be diverted to the pre assessment phase of design and construction.

A clear relationship should exist between the educational objectives and quality of questions on a MCQ examination (Downing, 2003). The MCQ is a reliable method to sample a wide spectrum of knowledge and with careful design, are capable of assessing higher cognitive functions such as application and analysis (Morrission and Free, 2001). Some argue that multiple-choice questions are challenged in terms of validity, especially their ability to consistently measure the desired higher order learning outcomes (Blooms, 1956). Designing questions that require knowledge of multi logical thinking and using plausible alternatives to the correct answer to challenge the student's discriminating judgment will increase the ability of MCQs to measure critically thinking skills (Morrission and Free, 2001). Holsgrove (1992) provides five rules for writing good questions

- Use correct grammar.
- Avoid negative stems.
- Avoid over complicated questions.
- Avoid ambiguity.
- Do not use trick questions.

Pamplett and Farnill (1995) place importance on the ability of examiners to develop banks of improved question items based on statistical analysis of answers. With the assistance of scoring and item-analysis computer software, post-test analysis of facility and discrimination indexes can assist in improving the quality of MCQ's (Quinn, 2000; Morrission and Free, 2001). Facility index informs us on the percentage of students that answer a question correctly while the discrimination index attempts to differentiate the performance of one student from another (Quinn, 2000).

Discrimation

The multiple-choice examination has been criticized for being artificial and not reflective of real life clinical situations. The potential for cueing effect in multiple-choice question instruments is highlighted by Veloski et al. (1993). The undesirable effect of cueing and the short list of potential options, furnishes hints to examinees that could not be there in clinical practice. They argue that the cueing effect of five choices makes it impossible for examiners to find out about specific clinical abilities of the student. In a later work, Veloski et al. (1999), argues that test items followed by one correct and four incorrect answers are artificial. The practitioner has to perform competently in a setting that does not offer short lists of five choices. They further argue, that while MCQ's may have content validity they are doubtful of their predictive value. However the same argument could be constructed against any classroom assessment strategy. The ability of any classroom assessment method to discriminate or predict performance in the clinical real life situation is doubtful. This is reflected in the requirement of any healthcare education program to balance any classroom assessment with effective assessment strategies in the clinical area.

Dixon (1994) recommends careful scrutiny of exam questions for relevance, potential ambiguities, correctness and balance with course objectives. In his study of the quality of multiple-choice questions, Dixon (1994) found 60% of questions were insufficiently discriminating between the overall best and worst student performance. He did conclude that this weakness could be overcome by careful question design.

Another strategy to improve the discrimination ability of the MCQ format is negative marking. Pamplett and Farnill (1995) explain the advantage of negative marking is that a larger spread of marks is obtained, which is important for ranking with a more normal distribution and therefore the absolute levels of marks reflects the true performance of the class better. Some examiners believe that negative marking deters guessing. Students are often discouraged from guessing in MCQs examinations that employ negative marking. Pamplett and Farnill (1995) explain that an individual who guesses blind would poorly serve the clinical area. This is an insufficient and unconvincing argument to penalize students in this way. Holsgrove (1992) believes negative marking can be used to the advantage of streetwise candidates who are good guessers or realize that −1 is not a sufficient deterrent. Hammond et al. (1998) studied the level of certainty in the answers of candidates undertaking MCQ's and established that students only answered correctly 90% of the 50% of the MCQs where they were certain of the answer and that educated guesses produce the correct answers 75% of the time. Therefore students could fail by not answering enough questions as opposed to losing marks through incorrect guesses. Negative marking will not turn out better practitioners by itself. If candidates are insufficiently prepared, the weakness is likely to be at some other juncture of the educational process.

The pro negative marking argument produces many questions regarding our intentions in health care education. Are our assessment methods geared to facilitate students to demonstrate their knowledge and to discover areas for improvement and indeed consolidate the knowledge gained? Some would feel that the function of assessment is to assess students on how they perform against each other in an effort to eliminate those who do not meet criteria. The former objectives may well be served with out resorting to negative marking which may well impact negatively on the student's motivation.

Anecdotally, it is suggested that negative marking may unfairly disadvantage anxious students. Pamplett and Farnill (1995) in their study of student anxiety, when using true/false questions exclusively, found negative marking gave a much better spread of marks, allowing more accurate ranking of the students. However, in this study the level of anxiety at the time of examination appears to have little effect on the final result obtained suggesting that concerns about anxious students being disadvantaged by negative marking are unwarranted. The level of anxiety studied here was based on a negative mark of 0.5, which may not have produced anxiety in students, as the consequences of guessing were not great. They did discover that women students have a higher level of anxiety than men. Holsgrove (1992) explains that if negative marking is to achieve its goal, to discriminate between students performance, there will have to be a great enough penalty to discourage guessing. Therefore, effective negative marking may indeed cause the significant stress for students as suspected.

Conclusion

Over the coming years in the Irish Republic, we will see the amalgamation of schools of nursing in the third level settings with a dramatic increase in

the size of groups involved in classroom teaching. It can appear that the most significant product of higher education is the qualifications that students gain, rather than the quality of the learning experience. A balance is necessary between enhancing the students learning experience and determining qualifications that are valid, and predictive of an individual's suitability to practice. Milligan (1996) describes assessment as an ethical activity where we need to reasonably certain of the accuracy of our judgments, essentially that we are right, and that we actively support the judgments that we have made. Multiple-choice questions as all assessment strategies have limitations and application to some content more than others. Good multiple-choice questions should be short, understandable and discriminating. MCQs can fulfill the criteria for effective assessment as suggested by Quinn (2000) as they can be an integral component of the teaching and learning process and can assess performance in relation to the aims of the curriculum. MCQs are efficient, objective, easy to grade and can be used to test a broad sampling of the curriculum and facilitate timely feedback and self-assessment. However, significant commitment is required in preparation to produce reliable and valid examination tools. Well designed MCQs can be combined with other assessment methods to provide an educational strategy to enhance the learning process and provide an accurate and comprehensive evaluation of student performance.

References

Blooms, B.S., 1956. Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain. Longmans, London.

Diog, K., 2000. Common formats of multiple choice questions testing analytical skills in clinical laboratory science. Clinical Laboratory Science 13 (1), 40–45.

Dixon, R.A., 1994. Evaluating and improving multiple-choice papers: true–false questions in public health medicine. Medical Education 28 (5), 400–428.

Downing, S.M., 2003. Validity: on the meaningful interpretation of assessment data. Medical Education 37, 830–837.

Farley, J., 1989. The multiple-choice test: writing the question. Nurse Educator 14 (6), 3–5.

Flannelly, L., 2001. Using feedback to reduce students' judgment bias on test questions. Journal of Nursing Education 40 (1), 10–16.

Gaskins, S., Dunn, L., Forte, L., Wood, F., Riley, P., 1996. Student perceptions of changing answers on multiple-choice examinations. Journal of Nursing Education 35 (2), 88–90.

Gibbs, G., Lucas, L., Spouse, J., 1997. The effects of class size and form of assessment on nursing student's performance, approaches to study and course perceptions. Nurse Education Today 17 (4), 311–318.

Hammond, E.J., McIndoe, A.K., Sansome, A.J., Spargo, P.M., 1998. Multiple-choice examination: adopting an evidence-based approach to exam technique. Anaesthesia 53, 1105–1108.

Holsgrove, G., Elzubeir, M., 1998. Imprecise terms in UK medical multiple-choice questions: what examinees think they mean. Medical Education 32 (4), 343–350.

Holsgrove, G., 1992. Guide to post graduate exams: multiple-choice questions British. Journal of Hospital Medicine 48 (11), 757–761.

Honey, M., Marshall, D., 2003. The impact of on-line multiple-choice questions on undergraduate student nurses' learning. In: Crisp, G., Thiele, D., Scholten, I., Barker, S., Baron, J. (Eds.), Interact, Integrate, Impact: Proceedings of the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education, vol. 1. Ascilite, Adelaide, pp. 236–243.

Jarvis, P., 1985. The Sociology of Adult and Continuing Education. Croom Helm, London.

Manogue, M., Kelly, M., Bartakova Masaryk, S., Brown, G., Catalanotto, F., Choo-Soo, T., Delap, E., Godoroja, P., Murio, I., Rotgans, J., Saag, M., 2002. Evolving methods of assessment. European Journal of Dental Education 6 (3), 53–66.

Masters, J., Hulsmeyer, B., Pike, M., Leichty, K., Miller, M., Verst, A., 2001. Assessment of multiple-choice questions in selected test banks accompanying test books used in nursing education. Journal of Nursing Education 40 (1), 25–32.

Milligan, F., 1996. The use of criteria-based grading portfolio in formative and summative assessment. Nurse Education Today 16 (6), 413–418.

Morrission, M., Free, K., 2001. Writing multiple choice test items that promote and measure critical thinking. The Journal of Nursing Education 40 (1), 17–24.

Oermann, M., 1999. Developing and scoring essay tests. Nurse Educator 24 (2), 29–32.

Pamplett, R., Farnill, D., 1995. Effect of anxiety on performance in multiple-choice examination. Medical Education 29, 298–302.

Quinn, F.M., 2000. The Principles and Practice of Nurse Education, fourth ed. Stanley Thorne (Publishers) Ltd, Cheltenham.

Race, P., Brown, S., 2001. The Lecturers Toolkit, second ed. Kogan Page, London.

Schuwirth, L., Van der Vleuten, C., Donkers, H., 1996. A closer look at cueing effects in multiple-choice questions. Medical Education 30 (1), 44–49.

Staib, S., 2003. Teaching and measuring critical thinking. Journal of Nursing Education 42 (11), 498–508.

Vaughan-Wrobel, B., O'Sullivan, P., Smith, L., 1997. Evaluating critical thinking skills of baccalaureate nursing students. Journal of Nursing Education 36 (10), 485–488.

Veloski, J., Rabinowitz, H., Robeson, M., 1993. A solution to the cueing effects of multiple-choice questions: the Un-Q format. Medical Education 27 (4), 371–375.

Veloski, J., Rabinowitz, H., Robeson, M., Young, P., 1999. Patients don't present with five choice: an alternative to multiple-choice tests in assessing physician' competence. Academic Medicine 74 (5), 539–546.

Available online at www.sciencedirect.com

SCIENCE DIRECT®