

5-2011

Assessing The Use of High Quality Multiple Choice Exam Questions in Undergraduate Nursing Education: Are Educators Making the Grade?

Theresa Guentzel Reichert
St. Catherine University

Follow this and additional works at: http://sophia.stkate.edu/ma_nursing

Recommended Citation

Reichert, Theresa Guentzel, "Assessing The Use of High Quality Multiple Choice Exam Questions in Undergraduate Nursing Education: Are Educators Making the Grade?" (2011). *Master of Arts in Nursing Theses*. Paper 15.

This is brought to you for free and open access by the Nursing at SOPHIA. It has been accepted for inclusion in Master of Arts in Nursing Theses by an authorized administrator of SOPHIA. For more information, please contact ejasch@stkate.edu.

Assessing The Use of High Quality Multiple Choice Exam Questions in Undergraduate Nursing
Education: Are Educators Making the Grade?

Scholarly Project
Submitted in Partial Fulfillment
of the Requirements for the Degree of
Masters in Nursing Education

St. Catherine University
St. Paul, Minnesota

Theresa Guentzel Reichert

May, 2011

ST. CATHERINE UNIVERSITY
ST. PAUL, MINNESOTA

This is to certify that I have examined this
Master of Arts in Nursing
written by

Theresa Guentzel Reichert

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by
the final examining committee have been made.

Graduate Program Faculty

Name of Faculty Project Advisor

Date

DEPARTMENT OF NURSING

Author Note

Theresa A. Reichert, Collegeville, Minnesota

A special thanks to Roberta Hunt, Nursing Department, St. Catherine University; who assisted in the preparation of this manuscript.

Correspondences regarding this article should be addressed to Theresa A Reichert, PO Box 5755, Collegeville, MN 56321. E-mail: treichert@csbsju.edu

Abstract

Evidence based guidelines for the development and assessment of multiple choice test item quality and validity are well documented in the literature, however contemporary evidence indicates that multiple choice exams used to assess student competency in undergraduate nursing education are heavily flawed. These exams subject students to invalid, low quality measures of their academic performance. The results from the literature review are significant because multiple choice exams serve as a primary assessment of student success. Failure to adequately assess student performance can have dire consequences on student grades, career opportunities, and future scholarship. Inaccurately assessing student competency dramatically affects the reliability and legitimacy of academic institutions and can adversely impact the integrity of the nursing profession itself. The literature recommends recognizing the prevalence of flaws in multiple choice exams, increasing faculty education regarding multiple choice exam development and assessment, and developing peer review teams to assess the exams and offer recommendations prior to the exam's administration.

TABLE OF CONTENTS

1. **Abstract** 6
2. **Background** 8
3. **Cause for Concern** 10
4. **Recommendations** 13
5. **Conclusion** 15
6. **References** 16

Guidelines for properly written multiple choice test items have been widely documented in the literature. Despite the fact that these evidence based guideline for best practice in multiple choice test item construction have been developed, and guidelines for revision are prevalent in the literature, current evidence reveals that the use of evidence based multiple choice test item evaluation methods is severely deficient. The Code of Ethics for Nurses (American Nurses Association, 2001) as well as the Core Competencies of Nurse Educators (National League for Nursing, 2005) requires nurse educators to accurately assess and evaluate student competence; however this is not consistently being done. Students are being subjected to low quality, inadequate assessment measures of their competency and academic performance. This is detrimental to student success as well as institutional integrity. This paper seeks to reveal the scope of the problem, to expose the importance of evaluating students based on appropriately constructed and assessed multiple choice test items, and to recommend long term solutions for educators to improve their multiple choice test item writing and assessment of test bank items. Appropriately assessing student success informs and reflects the integrity of academia, educational institutions, and the nursing profession nationally and globally.

Background

Nurse educators carry a large responsibility on their shoulders. The National League of Nursing's (NLN) Core Competencies of Nurse Educators (2005) oblige educators to implement teaching strategies that are appropriate to learner needs based on evidence-based teaching practices. The competencies challenge all nurse educators to use a variety of evidence-based assessment and evaluations strategies that are timely, appropriate for the learner, and evaluate learning in all domains. Tarrant, Knierim, Hayes, and Ware (2006) charge nurse educators with the ethical and legal responsibility to ensure that the methods of evaluation used in their programs are valid. Masters et al. (2001) state that these evaluation methods should "reflect the level of sophistication at which students are expected to practice"(p. 26). Evaluation and assessment of students in nursing education should be effective, efficient, and valid. The quality of assessment and evaluation however, is often hindered due to shortages in nursing faculty, time, or other compromises.

In 2008, the mean age for nurse educators averaged 51.5 years of age. Allen (2008) warns that the rate of anticipated faculty retiring will increase, making it difficult for academic institutions to keep up with the demand for educating new nurses. On top of the impending shortage of nurse educators, Allen (2008) predicts a need for over 1 million registered nurses by the year 2020 due to growth and replacements. This statistic indicates that the shortage of prepared nursing faculty will be challenged by the demand of educating additional nursing students. Shortages in faculty and time, as well as an increase in students lead many faculty to use multiple choice questions (MCQ's) to quickly measure knowledge in nursing education.

Multiple choice test questions are advantageous for several reasons. They allow educators to test a large number of students objectively and efficiently and can be graded electronically. Several studies (Hansen & Dexter, 1997; Masters et al., 2001) document that MCQ's can measure a wide variety of content, and can prepare students for the format of the NCLEX-RN exam. Multiple choice test questions also assess student knowledge within four of the six levels of Bloom's Taxonomy. Bloom's taxonomy is used as a framework for much of the research regarding the use of MCQ's in education. Though it hasn't been validated, Bloom's Taxonomy is frequently referenced as a measure of appropriate assessment in nursing education (Clifton & Shriner, 2010). Learning objectives are often written based on the hierarchy of classifications within the taxonomy beginning with knowledge, the lowest level of cognitive assessment. Morrison and Walsh Free (2001) describe knowledge as memorizing, or habitual thinking. Comprehension, an understanding of the knowledge encompasses the second level of Bloom's taxonomy, followed by application, analysis, synthesis, and evaluation. To appropriately evaluate higher-level and critical thinking, higher level MCQ's should be written at a higher cognitive level, particularly in upper level courses (Morrison & Walsh Free, 2001). The NCLEX-RN examination test items are written within each level within Bloom's taxonomy, however more questions are at the application and analysis level (Masters et al, 2001).

Despite the advantages for using MCQ's, there are challenges. To be effective, MCQ's must accurately assess higher levels of cognitive thinking and uphold high standards of student achievement. Ensuring this isn't easy. MCQ's are time consuming and difficult to create, particularly at the higher cognitive assessment levels. Few faculty have formal education in MCQ test item construction (Tarrant et al., 2006), nor the time to create their own MCQ's. The research widely documents that it can take up to one hour to create one quality test item (Clifton

& Shriner, 2010; Masters et al., 2001; Tarrant et al., 2006). Poorly constructed MCQ's have the tendency to be more difficult, produce lower passing rates, and threaten the integrity of nursing education (Downing, 2002). According to Boscher (2008), many nursing schools have also recognized an increase in students within their programs who culturally and linguistically diverse. These students are particularly challenged with multiple choice exams.

Due to the difficulty in constructing multiple choice items, coupled with limited time to develop MCQ's, many faculty rely on text book derived test questions for their examinations. Clifton and Shriner argue that one of the difficulties with test bank MCQ items is that they assess lower cognitive levels and are often rife with bias. (2010).

Cause for Concern

Standards are well known in nursing and nursing education, as they communicate criteria for competence (Bandaranayake, 2008). The standard for best practice in MCQ test item construction has been developed and supported by research. In the last decade, new and ever-evolving research involving the format, design, and construction of multiple choice questions have been published (Considine, Botti, & Thomas, 2005). Much of the research and many guidelines for constructing MCQ's are based on the works of Thomas Haladyna and Steven Downing, who have developed an empirically validated set of MCQ test item writing guidelines based on a review of over 50 textbooks, studies and reviews (Haladyna, Downing, & Rodriguez, 2002) These research based guidelines form a best practice standard for the development of effectively constructed MCQ's. Nursing educational research continues to support Haladyna and Downing's work.

There is a potential for students to be drastically affected by the measures used to assess their competence. Research on MCQ tests administered within a 5-year period in one undergraduate nursing program indicated that 47.3% of all MCQ's were flawed. Over 85% of flawed test items contained frequently cited test item violations, violations which are well documented in literature as being flawed (Tarrant & Ware, 2008). A second study evaluated test bank questions and discovered that almost half of the MCQ's assessed contained flaws and bias; over 90% were written at cognitive levels below the application and analysis level (Tarrant et al., 2006). Research by Masters et al. (2001) reveals that 76% of studied test bank questions contained item writing flaws. They also discovered that 6% of the test items were written at the analysis level compared to 46% written at the knowledge level of Bloom's taxonomy. Test bank authors also have limited formal training in MCQ construction, therefore MCQ's taken from test banks are equally susceptible to item writing violations as teacher developed questions (Tarrant et al., 2006). These findings indicate that educators are poorly evaluating student achievement when using most test questions. When established item-writing guidelines are not incorporated in MCQ exam construction, the exams validity decreases, it may favor test-wise students, and may improperly represent student competency (Stagnaro-Green & Downing, 2006).

These findings indicate a major problem. Educators are responsible for the accurate assessment and evaluation of student competence and for valid reasons; the multiple-choice testing format is commonly used to do so. The standards of high quality MCQ design, construction, and evaluation are not being followed. Reported test item flaws in multiple choice test items of over 50% should be alarming and initiate a need for action, particularly because those test items are regularly used as a primary measure of student competency. This means that

less than half of items evaluated within the research passed the MCQ item writing guideline standards. By inaccurately evaluating student competency, nurse educators are failing their students, their educational institutions, and the profession of nursing itself.

Holding nursing education to high standards of evaluation and outcome assessment is vital. It is imperative that students are evaluated accurately and appropriately, as MCQ's are often primary assessments of final grade and content mastery in nursing programs. The results of these grades can have lasting consequences for students as the grades are often absolute, important, or irreversible. (Considine et al., 2005). Unreliable and inaccurate grades can impede career pathways for students and may inhibit accreditation or credentialing processes within the clinical domain as well (Considine et al., 2005).

As shortages of well-trained nursing faculty threaten proper assessment and evaluation of the student's capabilities, educators cannot forget that as nurses, the American Nurses Association's nursing standards mandate them with the obligation and responsibility to advance and expand the body of knowledge within the profession through "development, evaluation, dissemination and application of knowledge in practice" (American Nurses Association, 2001, Provision 7.3). The standards state that nurses and nurse educators are charged with the responsibility of maintaining and preserving the integrity of the profession and of ourselves. In the interest of preserving academic integrity, educators must recognize when assessment methods are ineffective, and that the consequences of ineffective assessment and evaluation, particularly for students, may be irreversible, lasting, or final.

Recommendations

Arguably one of the most prevalent recommendations from the literature was to implore educators to recognize the prevalence of test item flaws in test bank items and learn how to assess them for flaws. To do so, Tarrant et al. (2006) argues that it is the responsibility of academic institutions to hire experts to properly train and faculty. Proper training in multiple choice test item construction and assessment improves the quality of MCQ's developed or assessed by faculty (Tarrant et al., 2006). Institutions should also take note of faculty responsibilities, and allow adequate time and training for faculty to develop, properly assess, and administer high quality MCQ's. Institutions need to recognize that nursing faculty not only educate student nurses, they also have other responsibilities such as research, grant writing, and maintaining competence in the realm of nursing and nursing education (Allen, 2008). Appropriate time and appropriate faculty ratios must be appropriated to meet the objectives of both the nursing department and the academic institution.

The responsibility for preparation does not lie solely on the shoulders of the institution. Educators must also be motivated to increase their own education through the evaluation of the literature surrounding MCQ item writing, which is rich and abounding with suggestions for faculty and guidelines for item writing development. Educators spend an immense amount of time preparing lectures, slides, and preparing course materials; yet insufficient time is allotted towards test preparation (Tarrant et al., 2006). Adequate time to assess test bank MCQ's is pertinent because the research shows test bank multiple choice items to be highly flawed yet still highly utilized. Faculty must question their use of test bank items, use them cautiously, or allocate sufficient time to evaluate the quality of test bank MCQ's and MCQ revision. Masters et

al. (2001) recommend developing a test plan prior to writing an exam to map out the desired distribution of questions and appropriate allotment of Bloom's taxonomy cognitive levels within the exam. Unfortunately, there is not sufficient research suggesting guidelines as to the percentage of questions to be written within each level of Bloom's taxonomy, however Masters et al. (2001) suggests using higher level questions with higher level courses. When faculty create their own multiple choice test items, the literature recommends that prior to test administration, a review process should be implemented which should include peer review. Tarrant et al (2006) describe peer review as an examination of test items by a review team composed of members who are adequately trained in writing MCQ exam items. The review team would be responsible for evaluating and eliminating item writing flaws, for the appropriateness of cognitive domain within the exam, and for offering suggestions and guidelines for improvement of underperforming multiple choice test items.

Another recommendation is to train select faculty leaders to be test item reviewers for the NCLEX-RN exam, and to utilize their knowledge to educate other faculty and keep them versed to current changes within item writing format. Downing (2005) endorses faculty efforts to develop their own methods to reduce or even eradicate inappropriate multiple choice test items. Clifton & Shriner (2010) exemplify Downing's endorsement by recommending the implementation of easy to read guidelines for faculty to use as a reference for test item construction and test item review.

Finally, the compilation of research seeks to bring an increased awareness to the prevalence of poor MCQ quality and validity, as well as mindfulness to the dramatic consequences of poorly evaluated student competence. An increased attentiveness to poorly assessed student aptitude will increase the accountability of educators and institutions to take

action and to review their own MCQ assessment standards, format, and their use of evidence based guidelines.

Conclusion

The manner in which assessment and evaluation is approached needs to be rethought and redesigned to appropriately assess student success. Educators bear a heavy responsibility, as they are accountable for preparing proficient nurses by using appropriate and meaningful standards of evaluation and accomplishment. They are mandated to advance the profession through the standards and values to which the profession is held. They are held accountable to their students, to the public, to licensing programs, and to the nursing profession. Both the academic and clinical realms of nursing should be greatly alarmed with the current evaluation practices used in nursing education. Faculty must utilize their own realm of obligation to reflect the integrity of academia, educational institutions, and the nursing profession nationally and globally. It is imperative that nurse educators recognize the need to change the system, a system that lacks the preservation of integrity and one that is in dire need of reform.

References

- Allen, L. (2008). The nursing shortage continues as faculty shortage grows. *Nursing Economics*, 26 (1), 35-40.
- American Nurses Association. (2001). *Code of ethics for nurses with interpretive statements*. Washington, DC: Author
- Bandaranayake, R. (2008). Setting and maintaining standards in multiple choice examinations: AMEE guide No.37. *Medical Teacher*, 30, 836-845. doi:10.1080/01421590802402247
- Boscher, S. (2003). Barriers to creating a more culturally diverse nursing profession: Linguistic bias in multiple-choice nursing exams. *Nursing Education Perspectives*, 24 (1), 25-34.
- Considine, J., Botti, M. & Thomas, S. (2005). Design, format, validity, and reliability of multiple choice questions for use in nursing research and education. *Collegian*, 12, 19-24. doi:10.1016/S1322-7696(08)60478-3
- Clifton, S. & Schriener, C. (2010). Assessing the quality of multiple choice test items. *Nurse Educator*, 35 (1), 12-16. doi: 10.1097/NNE.0b013e3181c41fa3
- Downing, S. (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, 133-143.

doi: 10.1007/s10459004-4019-5

Haladyna, T., Downing, S., & Rodriguez, M. (2002). A review of multiple-choice item writing guidelines for classroom assessment. *Applied Measurement in Education, 15* (3), 309-334. doi: 10.1207/S15324818AME1503_5

Masters, J., Hulsmeyer, B., Pike, M., Leichty, K., Miller, M., & Verst, A. (2001). Assessment of multiple-choice questions in selected test banks accompanying text books used in nursing education. *Journal of Nursing Education, 40* (1), 25-31.

Morrison, S., & Walsh Free, K. (2001). Writing multiple choice test items that promote and measure critical thinking. *Journal of Nursing Education, 40* (1), 17-24.

National League for Nursing. (2005). *Core competencies of nurse educators*. Retrieved from <http://www.nln.org/profdev/corecompetencies.pdf>

Stagnaro-Green, A., & Downing, S. (2006). Use of multiple-choice items by the New England Journal of Medicine for continuing medical education. *Medical Teacher, 28* (6), 566-568. doi:10.1080/01421590600711153

Tarrant, M., Knierim, A., Hayes., & 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. doi:10.1016/j.nedt.2006.07.006

Tarrant, M., & Ware, J. (2008). Impact of item-writing flaws in multiple choice questions on

student achievement in high-stakes nursing assessments. *Medical Education*, 42, 198-

206. doi: 10.1111/j.1365-2923.2007.02957.x