Title:
Nurse Educators' Practices in the Measurement of Student Achievement Using Multiple-Choice Tests

Susan Birkhead, DNS, MPH
Samaritan Hospital School of Nursing, Troy, NY, USA
Linnea Jatulis, PhD
Department of Nursing, The Sage Colleges, Troy, NY, USA

Session Title:
Use of Testing to Foster Success in Students
Slot:
F 05: Saturday, April 9, 2016: 2:45 PM-4:00 PM
Scheduled Time:
2:45 PM

Keywords:
multiple-choice testing, nurse educators and testing

References:

Abstract Summary:
Examinations composed of multiple-choice questions (MCQs) are widely used to assess student achievement. This study explored New York State nurse educators' adherence to recommended MCQ test construction and test analysis practices. Adherence was not consistent. Mentoring and professional development conferences were identified as important sources of learning about MCQ testing.

Learning Activity:

<table>
<thead>
<tr>
<th>LEARNING OBJECTIVES</th>
<th>EXPANDED CONTENT OUTLINE</th>
</tr>
</thead>
<tbody>
<tr>
<td>The learner will be able to list seven best practices in item writing, test construction and analysis of test results</td>
<td>Review of the literature identifying best practices in item-writing, test construction and analysis of test results, and presentation of scores on the Best Practice Index. Selected practices include: 1) blueprinting; 2) writing rationales for question answers; 3) writing three total answer choices; 4) avoiding all-of-the-above questions; 5) avoiding none-of-the-above questions; 6) using readability analysis; 7) editing commercially prepared test bank questions prior to use; 8) vetting test questions [review by other nurse educators prior to administration to students]; 9) using an e-scoring system; 10) nullifying questions; 11) using the correct procedure when nullifying</td>
</tr>
</tbody>
</table>
The learner will be able to contrast and compare faculty practices in assessment of student achievement with respect to faculty educational preparation, teaching experience and program type (AD or BS).

Results from the study will be presented: The effect of formal preparation as a nurse educator on adherence to best practices (no significant difference); the effect of number of years teaching on adherence to best practices (no significant difference); whether faculty who reported teaching in baccalaureate programs were more likely to adhere to best practices than those who reported teaching in associate degree programs (no significant difference).

The learner will be able to identify the two most important sources cited by faculty for learning about how to write test questions, how to construct tests, and how to analyze test results.

Respondents were asked to identify their most important source of learning about how to write item, how to construct tests and how to analyze test results. In descending order, the responses were as follows: professional development conferences 31% (n = 60); being mentored 29% (n = 57); formal coursework 21% (n = 40); other 8% (n = 15); training from the NCSBN/NLN/AACN/ATI 8% (n = 15); professional literature 3% (n = 5).

Abstract Text:

Examinations composed of multiple-choice questions (MCQs) are widely used to assess student achievement in pre-licensure nursing education (Bevis & Watson, 2000; Clifton & Schriner, 2010; Masters et al., 2001; Morrison & Free, 2001; Smith, 2000; Su, Osisek, Montgomery, & Pellar, 2009; Tarrant & Ware, 2008). This study explored nurse educators’ practices in the measurement of student achievement using testing in pre-licensure nursing education in New York State. This research sought to describe three phenomena: (1) the prevalence of the use of multiple-choice testing to measure student achievement in pre-licensure nursing education in New York State; (2) the relationship between nurse educator characteristics (educational preparation, age, years teaching nursing) and self-reported practices in the use of testing to measure student achievement; (3) the relationship between nursing education program institutional characteristics (program type, program size, accreditation status, policies) and nurse educators’ self-reported practices in the use of testing to measure student achievement.

Background: Student progression in the theoretical component of nursing courses may, in large part, or indeed exclusively, be based on the grades based on their performance on multiple-choice exams, despite recommendations that a variety of assessment methods be used (Benner, Sutphen, Leonard, & Day, 2010; McDonald, 2014; National League for Nursing [NLN], Board of Governors, 2012). The few studies of nursing exams have shown that they may be poorly constructed (Bosher & Bowles, 2008; Clifton & Schriner, 2010; Tarrant, Knierim, Hayes, & Ware, 2006; Tarrant & Ware, 2008). When exams are poorly constructed, some knowledgeable students may be prevented from progressing based on results which are not reliable and not supported by strong validity evidence – ‘false negatives’, while other students may progress who should not – ‘false positives’ (Downing, 2005; Tarrant & Ware, 2008). The authors of these studies commonly suggest that faculty development, i.e. training faculty in the principles of test construction and analysis of test results, is imperative to prevent inaccurate measurement and the negative impact that this may have on students (Halstead, 2013). However, little is known about the
extent of nursing faculty preparation in pedagogical techniques for assessing student achievement (Considine & Thomas, 2005; Tarrant, Knierim, Hayes, & Ware, 2006). Graduate curricula for nurse educators should include content on measurement of student achievement. However, many nurse educators may have earned their graduate degrees in other areas of specialization such as nursing administration or as a nurse practitioner. The curricula for these degrees would likely not have included content on assessment and measurement in education. These educators may not possess the skills required to create multiple-choice exams and to analyze these exams for difficulty, reliability, and validity. A review of the literature did not find any studies examining these suppositions. Furthermore, there are no studies that explore the policies and practices of nursing education institutions that would promote the principles and standards for educational measurement (e.g. a policy requiring test 'blueprinting').

**Methods:** This was a descriptive, correlational, quantitative study. The data were obtained using an anonymous online survey of nurse educators in New York State. A 49-question online survey of demographics, testing practices and program characteristics was sent to 1559 nurse educators whose professional email addresses were available on the websites of nursing education programs in NYS in 2014. The response rate was 19 percent \((n = 297)\). Ninety-seven respondents who did not meet eligibility criteria were disqualified. The final study cohort consisted of 200 respondents. Independent variables included nurse educators’ demographics (age, years teaching nursing), data related to nurse educators’ formal education at the graduate level, and institutional factors (size, type of program, accreditation status, policies) describing the nursing education programs in which they taught. Dependent variables included self-reported practices used in the measurement of student achievement. From the survey, a Best Practices Index (BPI) was constructed that included twelve item-writing, test construction and results analysis practices recommended in the scholarly literature, such as blueprinting or vetting (review of a test by other educators prior to the administration of a test) (Bosher & Bowles, 2008; Haladyna, Downing & Rodriguez, 2002; Masters et al., 2001; McDonald, 2014; Morrison & Free, 2001; Tarrant et al, 2006). Each of the twelve practices was given equal weight (one); possible scores on the BPI ranged from zero to twelve. The researcher measured the extent to which respondents reported engaging in these best practices.

**Results:** This study found a high prevalence of use of multiple-choice questions on tests or exams in both associate and baccalaureate degree programs. Sixty-five percent of the respondents \((n = 130)\) indicated that at least 80% of a typical course grade was derived from testing. Respondents reported that MCQs comprised a mean of 81% of the questions on a typical test. Those who taught in associate degree programs were more likely than those in baccalaureate programs to report a high reliance on the use of multiple-choice tests to derive course grades \((p < .05)\). Overall, the mean score on the Best Practice Index was 7.3 out of a possible 12 recommended practices; the range of BPI scores was zero to eleven. Thirty-three percent reported engaging in nine or more practices, 10% of the respondents reported engaging in four or fewer recommended practices. This differs from the findings of Killingsworth, Kimble & Sudia (2015), who reported a high rate of adherence to recommended practices in classroom testing. There was no significant relationship between respondent age, respondent years teaching nursing, and respondent educational preparation and their reported use of best practices in item-writing, test construction and analyses of test results. In particular, those nurse educators who had formal coursework in the measurement of student achievement were no more likely engage in best practices (score higher on the BPI) than those who did not. With respect to program characteristics, there was no significant relationship between program size, program type and class size and respondents’ scores on the BPI. Respondents from associate degree programs were significantly \((p < .05)\) more likely than those in baccalaureate programs to report that their programs had a written testing policy. However, there was no significant relationship between the reported existence of a written testing policy and respondents’ scores on the BPI. Respondents from associate degree programs were significantly \((p < .05)\) more likely to report the existence of a blueprinting policy than those in baccalaureate programs. Those who reported that their program required blueprinting were more likely to engage in this practice \((p < .05)\). Respondents from associate degree programs were significantly \((p < .05)\) more likely to report that the program had a policy requiring vetting before administering it to students. Those who reported that their nursing education program required vetting were more likely to engage in this practice \((p < .05)\).
Mentoring and professional development activities were cited as the most important source of faculty learning about measurement of student achievement.

Limitations: This study is limited because it only surveyed nurse educators in New York State. According to the NLN (2013), in 2012 there were 99 basic nursing programs in New York State, or five percent of all 1839 basic nursing education programs in the United States. The number of nursing education programs in NYS was second only to California, which had 125 such programs. The mean number of basic nursing education programs per state was 37. The respondents in this study represented 61 different nursing education programs from urban, suburban and rural areas of the state. So, while the limitation is to be considered, it should also be noted that the study was conducted in a populous state with a large number of nursing education programs. An additional limitation is the low response rate among eligible participants. The final limitation is the fact that the websites of some nursing education programs do not provide professional email contact information for faculty members. Those nursing education programs may differ in some unknown way from those that do provide contact information on the internet.

Conclusions/recommendations: Evaluating student achievement is one of the most important responsibilities in teaching, and the educator has an ethical responsibility to be accurate and fair in this process. Since this study confirms that MCQ testing is widely used, nurse educators must ensure adherence to best practices in the use of MCQ tests to assess student achievement. This research found that the respondent nurse educators did not consistently adhere to recommended practices in the measurement of student achievement. Failure to adhere to best practices may at best be seen as incompetence, and at worst, outright disregard. However, nurse educators may be unaware of their faulty assessment practices. It is hoped that this research will increase awareness and will inspire improvements. Deans and directors of nursing education programs should ensure the existence of and adherence to written testing policies. They should also ensure that nurse educators are well prepared, through mentoring and professional development, to use MCQ tests to measure student achievement. Nurse educators should hold themselves accountable for adhering to recommended practices in test construction and analysis of test results.