The Critical State of Measurement in Nursing Education Research

Over the past year, I’ve had the opportunity to review dozens of grant proposals seeking funding for evaluation of innovations in nursing education, as well as scores of manuscripts reporting studies of nursing education innovations. It is clear to me how little investment there has been in developing suitable measures for the assessment of learning outcomes, particularly those relevant for a practice discipline and those that might be sensitive to our educational interventions. As we see renewed opportunities for funding nursing education innovations in response to the Institute of Medicine’s (IOM) The Future of Nursing report (IOM, 2010), I hope that we can marshal the resources for developing and testing instruments critically needed for evaluating these innovations.

There are three broad categories of learning outcome assessments that have been used in nursing education research. Each of these categories have advantages and disadvantages with regard to their practicality, their relevance in a clinical practice discipline, their validity, and their logical relationship and sensitivity to particular educational interventions

Student Perceptions, Attitudes, or Opinions

In this grouping are students’ ratings of learning, their attitudes toward an instructional activity, or their opinions about the effectiveness of an activity. Each of these are most commonly measured using an investigator-developed instrument, although there are now commercially available instruments for benchmarking an individual program’s student or graduate ratings, such as the Educational Benchmarking, Inc., system. The self-reports or opinion surveys are relatively easy to develop and provide preliminary evidence for program effectiveness, but they are far from adequate for grounding instructional decisions or for the adoption of educational innovations. For example, research on simulation, to date, relies largely on student opinions about the effectiveness of this activity (Cant & Cooper, 2010; Ogilvie, Cragg, & Foulds, 2011); but there are many variations in how simulation may be conducted, and the relative effectiveness of these variations cannot be adequately assessed by student self-ratings.

Tests of Content Knowledge or Skills Relevant for Nursing Practice

In this grouping is a wide range of tests used to evaluate individual student knowledge, then aggregated for use in evaluation of an educational intervention. Tests are commonly teacher made, but investigators have also used commercially available standardized content knowledge tests (e.g., the Educational Resources, Inc.—part of Assessment Technologies Institute, LLC—or Elsevier’s Health Education Systems, Inc. [HESI] Exit Examination) or NCLEX pass rates as indicators of learning outcomes. Teacher-made tests frequently lack psychometric evaluation beyond internal consistency reliability and minimal content validation. Standardized tests have the advantage of psychometric testing, with some normative data, but both predictive tests such as the HESI and the NCLEX are developed to measure the minimal level of knowledge and may not be sensitive to variations produced by educational interventions. For example, interventions directed toward improving clinical reasoning, such as case-based instruction, may not affect content knowledge scores.

Other commonly used standardized tests are those that measure general critical thinking skills, such as the Watson-Glaser Critical Thinking Appraisal or the California Test of Critical Thinking Skills. These tests have been used as outcome measures for program evaluation, as well as for the assessment of the effectiveness of educational innovations such as simulation (Cant & Cooper, 2010). Like other standardized tests, normative data are available. It is still unclear the extent to which these tests are related to specific nursing performance, such as clinical reasoning (Fero et al., 2010), or are sensitive to nursing education interventions.

Tests of nursing skills, such as skill checklists, are often used for individual student assessment. In the past several years, there has been an increased use of objective structure clinical examinations, often focused on safe performance of technical skills (Jones, Pegram, & Fordham-Clarke, 2010; McWilliam & Botwinski, 2010; Walsh, Bailey, Mossey, & Koren, 2010), or health assessment skills in advanced practice students (Kurz, Mahoney, Martin-Plank, & Lidicker, 2009). Objective structure clinical examinations have been used to evaluate the effectiveness of high-fidelity simulation training (Alinier, Hunt, Gordon, & Harwood, 2006).

Tests of Performance Within a Particular Clinical Context

In a practice discipline, content knowledge is a necessary but insufficient condition for safe practice. Nurses must also have the ability to recall the knowledge that is relevant for a particular clinical situation, an ability referred to as “situational cognition” (Wooley & Jarvis, 2007), or “sense of salience” (Benner, Surphen, Leonard, & Day, 2009). High-fidelity simulation, coupled with rating scales or assessment rubrics (e.g., Lasater, 2007), holds great promise for testing students’ ability to use relevant knowledge in particular contexts and have been used in research to a limited degree (Dillard et al., 2009). Of course, such observational tools require costly rather than efficient and time-consuming development and validation.
maintain interrater reliability. In addition, the content specificity of clinical reasoning performance (Norman, 2005; Tanner, 2006) requires that multiple case scenarios be used to reliably assess students’ performance. Both of these requirements drive up costs of assessment. Research that explores the relationship between higher-cost, higher fidelity assessments, such as simulation performance, with lower cost, general measures of ability, such as critical thinking assessments, may be helpful (Fero et al., 2010).

Summary
Nursing education research lacks common metrics or standardized approaches for the evaluation of learning outcomes that are relevant for a practice discipline and that assess not only students knowledge gains, but also their ability to use it during the provision of patient care. To have high quality evaluation of educational innovations, we must have investment of resources—investigator expertise, time, and money—to develop measures that are appropriate for a clinical practice discipline and that will reflect variations in educational approaches. Only then can we hope to build the evidence essential for quality nursing education.

References


Christine A. Tanner, PhD, RN, FAAN
Editor
The author has no financial or proprietary interest in the materials presented herein.
doi:10.3928/01484834-20110819-01