Nurse Educator

Nurse Educator
Vol. 34, No. 2, pp. 56-59

Evaluating Innovative Items for the NCLEX, Part I

Usability and Pilot Testing

Anne Wendt, PhD, RN, CAE
J. Christine Harmes, PhD

National Council of State Boards of Nursing (NCSBN) has recently conducted preliminary research on the feasibility of including various types of innovative test questions (items) on the NCLEX. This article focuses on the participants' reactions to and their strategies for interacting with various types of innovative items. Part 2 in the May/June issue will focus on the innovative item templates and evaluation of the statistical characteristics and the level of cognitive processing required to answer the examination items.

Among the advantages of a computer-based test over a paper-and-pencil test is its capability to administer innovative types of test questions (items). Innovative items refer to test items that incorporate functionality and elements beyond those available in a traditional text-based, multiple-choice format. These innovative elements may be realized in the structure of the items or tasks or may refer to the addition of multimedia such as graphics, audio, animation, or video.[1] The allure of such items to testing programs is their potential for measuring constructs that are difficult to measure using the traditional multiple-choice item format. To the extent that such constructs comprise a particular measurement domain of interest, the addition of innovative items may ensure better coverage of the domain, thereby increasing the content validity of the inferences that can be made from the test scores.

The National Council of State Boards of Nursing (NCSBN) recently conducted preliminary research on the feasibility of including various types of innovative items in its National Council Licensure Examinations for nurses (NCLEX-RN and NCLEX-PN). Previous research by NCSBN yielded several alternate item types that have been incorporated into the NCLEX.[2-4] This article describes the most recent research that involved developing a trial set of innovative items designed to advance the assessment of nursing practice. The specific project goals were to (*a*) identify a set of nursing skills and processes whose measurement is likely to improve through the use of innovative item types or elements (eg, those that reflect higher order thinking, aspects of physical assessment, or communication skills), (*b*) develop innovative item templates and prototype items for assessing a subset of the identified skills and processes, (*c*) conduct a usability study for a set of items based on these templates, and (*d*) perform a pilot study using innovative items. The purpose of this article was to describe the development process and examine results from evaluative components (usability study and pilot study of a set of trial innovative items) of the research study.

Authors' Affiliations: Director (Dr Wendt), NCLEX Examinations Department, National Council of State Boards of Nursing, Chicago, Illinois; Director (Dr Harmes), PhD Program in Assessment and Measurement, James Madison University, Harrisonburg, Virginia.
Corresponding Author: Dr Wendt, National Council of State Boards of Nursing, 111 East Wacker Dr, Chicago, IL 60601 (awendt@ncsbn.org).

Test Development

The initial phase of test development involved defining a set of specific skills and processes that would be particularly suitable for new types of items and relevant to the practice of newly licensed registered nurses.

To define the target skills and processes, a group of subject matter experts was convened. Two sources were especially influential in focusing this process. The first source document was the Boards of Nursing's ranking of 5 nursing competency areas.[5] These areas, according to their importance ranking, were (*a*) application of knowledge to practice, (*b*) critical thinking, (*c*) ethical/moral standards of practice, (*d*) competence in performing clinical skills, and (*e*) effective communication skills. Second was a set of competencies identified in a report by the Institute of Medicine,[6] which included (*a*) provide patient-centered care, (*b*) work in interdisciplinary teams, (*c*) employ evidence-based practice, (*d*) apply quality improvement, and (*e*) use informatics. Based on the recommendations from these reports, the expert panel identified 4 of the 8 sections from the NCLEX-RN test plan that may benefit most from innovative item development: (*a*) management of care, (*b*) safety and infection control, (*c*) psychosocial integrity, and (*d*) basic care and comfort. Within each of these areas of the test plan, specific objectives were selected.

Once the target skills, processes, and objectives were identified, the second phase focused on item template development. The item templates and resulting items had to meet the following criteria:

- be able to assess a problem-solving or higher order skill needed by RNs

but that is difficult to effectively assess using traditional multiple-choice items;
- be targeted to a construct that may be better assessed via an innovative item;
- be suitable for specification in an *item template*, from which many similar items could be readily developed;
- be suitable for computer-adaptive administration;
- have a clearly correct answer; and
- be immediately scorable by the computer.

Templates were initially developed for the following item types: (*a*) graphics inclusion, (*b*) graphics interaction, (*c*) audio inclusion, (*d*) video inclusion, (*e*) video interaction, (*f*) animation inclusion, and (*g*) decision task item sets. Figure 1 presents a computer screen print of video interaction addressing communication skills. Figure 2 presents a computer screen print of a graphics interaction item addressing safety. A Flash-based application was developed to administer these item types and to collect detailed examinee action and response data. After discussion regarding the match of item types to target skills and processes, item writers created items within each of the templates. Item content and specifications were revised in subsequent sessions.

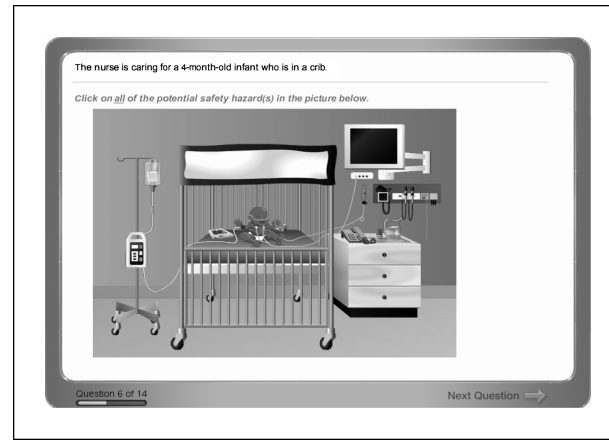

Figure 2. Graphics interaction item addressing safety.

Usability Testing

In the third project phase, 2 evaluative studies were conducted using the pilot innovative items. The first was a usability study, using the methods of expert review and user testing. As recommended by Harmes and Parshall,[7] expert review was conducted as part of a user-centered, iterative design strategy. At various points in the development of item templates and prototype items, experts from measurement, nursing, and interface design reviewed the prototypes. After each review, appropriate modifications were made. User testing was then conducted, in which the set of prototype items was administered to a small sample of examinees representative of the target examinee population.

User testing was conducted in 3 iterative stages. The first iteration included 2 students from a bachelor's degree program, the second included 1 student from a bachelor's program, and the third included 3 students from an associate's degree program. Four of the 6 students were women, and 2 were men. One student was a non-native speaker of English.

The think-aloud protocol was used as examinees proceeded through the test items.[8] These individual sessions were digitally recorded for transcription and analysis. After each participant's completion of the examination, the interviewer posed a set of structured interview questions related to impressions of the test items and suggestions for revision. Results from both data collection methods (observation and interview) were highly informative and were used to make substantial modifications to the items before pilot testing. Examples of changes included (*a*) changing the colors of buttons and tabs to allow for easier recognition, (*b*) reducing the amount of scrolling required, (*c*) increasing the window size for viewing videos, and (*d*) modifying item instructions. Based on the detailed results from the usability testing, the final set of 23 items was selected and refined for use in the pilot test.

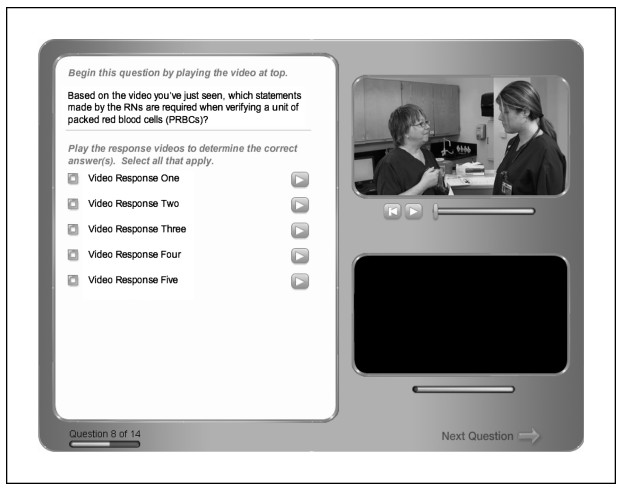

Figure 1. Video interaction addressing communication skills.

Pilot Test

Examination Results

The pilot test was administered to 224 students from 4 nursing programs. Two of the programs offer associate's degrees, and 2 offer bachelors' degrees. Most participants (84%) were women, and 79% were native English speakers. At the conclusion of each student's pilot test, a post-test survey was administered to gather demographic information and assess the students' perceptions of the pilot test.

The prototype items yielded item difficulty values ranging from very

difficult (2 items with *p*-values of approximately .2) to very easy (1 item with a *p*-value of approximately .9). However, item difficulty values for most items were in the middle range from .4 to .7. The students used the innovative components, playing and replaying the video and audio clips, as well as referencing the "exhibit" material. The participants did not seem to have problems using any of the innovative features for accessing item content or responding to items.

Survey Results

Data analysis for each of the perceptions questions was broken down by degree type and by whether English was the student's first language. A χ^2 test (at the 0.01 level of statistical significance) was used to determine whether either of these variables was associated with the responses to each perception question. In most cases, there was no significant association with degree type or whether English was the student's first language.

Computer Experience

Students tended to be experienced with using the computer in general, with 68% stating that they were *experienced* or *very experienced*. A similar pattern was found for students' experience using computers for testing. Sixty-seven percent said that they were experienced or very experienced.

Perception of Item Quality

Students strongly felt that the items assessed higher order thinking skills, with 93% responding either *somewhat* or *very effectively*. They also thought that the items were fairly difficult, with 89% responding either *somewhat* or *very difficult*. Compared with text-based multiple-choice items, the students generally viewed the items as more realistic (63% rated them as somewhat or much more realistic). Similar results were observed for the relative challenge of the items (53% rated them as somewhat or much more challenging than text-based multiple-choice items). Finally, 60% responded that the test items, relative to text-based multiple-choice items, allowed them to demonstrate their nursing competence *somewhat better* or *much better*. Overall, these responses indicated that the items were viewed very favorably by the students.

Impressions of Test Delivery Software

Generally, students felt that the items clearly indicated what they were to do to answer the items (64% responding *very clear*) and that the testing software was easy to use (66% responding *very easy*).

Inclusion on the NCLEX

Students were asked if they felt that the types of items on the pilot test should be part of the NCLEX. Although many students responded *yes*, *definitely* (47%), compared with those who responded *no*, *definitely not* (13%), a sizeable percentage responded *unsure*. It was unclear whether the frequent unsure ratings were due to generally uncertain feeling about the items or whether they were due to examinees wanting some (but not all) of the pilot test items to be on the NCLEX.

Quality of Specific Item Types

There were 3 survey questions that assessed realism, likeability, and difficulty of the items. Eighty-six percent of students felt that the video items were *somewhat* or *very realistic*, and 77% responded that they *liked* them or *liked* them *very much*. Ninety percent of students responded that the videos were *very clear*, 72% felt that the graphics were *very clear*, and 94% of students rated the quality of audio clips as *good* or *excellent*. Finally, regarding the supplemental item information in "exhibits," 88% reported looking at the exhibits and finding them *somewhat* or *very useful*. The final question addressed the groups of items all pertaining to the same nursing scenario (decision task item sets). A high percentage (94%) rated these items as somewhat or very effective.

The overall conclusion that can be drawn from these survey responses is that the students generally had positive or very positive reactions to the pilot test items. They tended to feel that the items were (*a*) clear, (*b*) of appropriate difficulty, (*c*) realistic, and (*d*) able to measure their higher order levels of nursing competence. The survey question with the least positive responses concerned whether students felt that the types of items on the pilot test should be part of the NCLEX. It is unclear how to interpret the finding that 40% of the students responded *unsure*, particularly in light of the positive responses to the other survey questions.

Educational Importance

It is essential that the licensure examination for nurses should assess the entry-level nurse's ability to practice safely and effectively to protect the public from unsafe practitioners. The introduction of innovative items, such as those developed in this study, may allow for more authentic assessment of some important nursing skills that, heretofore, have not been as directly assessed on the licensure examination. In addition, some higher order cognitive processing skills, such as *evaluate* and *create*, may be assessable using innovative items. Furthermore, these item types may allow some candidates more and perhaps better opportunities to demonstrate their competence than with the exclusive use of text-based multiple-choice items. Based on the results of this study, the nursing students who participated believed the items to be more representative of the work actually performed by nurses on the job. Additional research should focus on these issues as they are essential for most professional licensure examinations and will ultimately help protect the public.

References

1. Parshall CG, Harmes JC. Design of innovative item types: targeting constructs, selecting innovations, and refining prototypes. *CLEAR Exam Rev*. 2008;19(2):18-25.
2. Wendt A. Beyond multiple choice: the quest for authentic testing. Paper presented at the annual meeting of the American Educational Research Association; April 12-14, 2004; San Diego, CA.
3. Wendt A, Kenny L, Marks C. Assessing critical thinking using a talk-aloud protocol. *CLEAR Exam Rev*. 2007;18(1):18-27.

4. Wendt A. Investigation of the item characteristics of innovative item formats. *CLEAR Exam Rev*. 2008;18(1):22-28.
5. NCSBN. *2006 NCSBN Annual Business Book*. Chicago, IL: NCSBN; 2006.
6. Institute of Medicine. *Health Professions Education: A Bridge to Quality*. Washington, DC: The National Academies Press; 2003.
7. Harmes JC, Parshall CG. An iterative process for computerized test development: Integrating usability methods. Paper presented at the annual meeting of the Florida Educational Research Association; November 8-10, 2000; Tallahassee, FL.
8. Van Someren MW, Barnard YF, Sandberg JAC. *The Think Aloud Method: A Practical Guide to Modeling Cognitive Processes*. London, England: Academic Press; 1994.

MARTTI: Wireless Interpretation System Improves Communication

Teaching our students culturally appropriate communication skills can be a challenge. Many people seeking health care have limited proficiency with the English language. Safe, timely and effective diagnosis and treatment of presenting conditions depends upon adequate interpretation of a patient's symptoms, concerns and needs. Neither health care facilities nor nursing programs have the resources to hire qualified medical interpreters to fit every communication challenge that arises. The Language Access Network (LAN) has developed a HIPPA approved wireless, two-way video and audio wireless connection to a skilled medical interpreter known as *MARTTI* (My Accessible Real-Time Trusted Interpreter). This device is designed to assist with interpretation and language barriers to health care. *MARTTI* provides 24 hour a day, 365 days per year access to interpreters for over 150 languages including American Sign Language (ASL). *MARTTI* allows healthcare providers to use a small screen for face-to-face communication or a wide 19 inch screen for emergency situations. A document translation service allows patients to discuss forms in their native language and in a transparent environment. Interpreters employed by LAN take education programs that focus on various healthcare related topics including hospice, medical terminology and sexual health. LAN staff interpreters have degrees in ASL interpretation and national certification through the *Registry of Interpreters for the Deaf*.

Nurses at both Aultman Hospital in Canton, Ohio and Texas Children's Hospital in Houston report successful use of *MARTTI*, Ann Bykes RN, MSN a clinical specialist at Texas Children's Hospital says that using *MARTTI* is more accurate than trying to "wing it" on her own, and an improvement over standard phone interpretive services. Debby Taylor, RN, MSN, MBA of Aultman Hospital notes that the best thing about *MARTTI* is that the device can go anywhere that assistance with interpretation is needed. Go to the LAN website (http://www.languageaccessnetwork.com/) to evaluate *MARTTI* and also to determine if your students should be made aware of this technology.

Source: Wood D. (November 2008). Wireless interpretation system improves communication. NurseZone.com: Devices and Technology. Available at http://www.nursezone.com/Nursing-News-Events/devices-and-technology.aspx?ID=18461&Tab=1. Accessed November 13, 2008.