Nurse Educator

Nurse Educator
Vol. 34, No. 3, pp. 109-113


# Developing and Evaluating Innovative Items for the NCLEX

## *Part 2, Item Characteristics and Cognitive Processing*

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

*This article is a continuation of the research on the development and evaluation of innovative item formats for the NCLEX examinations that was published in the March/April 2009 edition of Nurse Educator. The authors discuss the innovative item templates and evaluate the statistical characteristics and level of cognitive processing required to answer the examination items.*

Innovative items contain content or functionality that is not possible in a text-based, multiple-choice question (item).[1] Thus, these item types have the potential for expanding an examination's construct representation by providing opportunities to measure constructs or dimensions that cannot be measured, or cannot be measured well, using traditional multiple-choice items.[2] Innovative types of items are also considered to have the capacity to tap higher levels of cognitive processing as compared with traditional text-based, multiple-choice items. Prior research on the statistical characteristics and cognitive processing of items that use alternate formats (innovative items) provided evidence to support the initial development of these item types for the NCLEX examinations.[3-5] Because the development of innovative items is expensive and time-consuming, examination programs have to carefully consider strategies and rationale for production and operation of these item types. The primary purpose of this project was to inform future directions for item development by investigating the levels of cognitive processing required to answer various types of innovative items and the statistical characteristics of the different item types.

## Cognitive Processing

Many methods of evaluating an individual's knowledge, skills, and abilities in a content domain such as nursing practice involve assessing that individual's cognition or cognitive processing ability. However, there is a paucity of research on (*a*) a taxonomy for categorizing cognitive processing of items and (*b*) a methodology for assessing the cognitive processing required to respond to items.

**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).

### Taxonomy for Categorizing

Various taxonomies attempt to categorize the different levels of cognitive processes.[6-8] However, there is no consensus about which taxonomy should be used to categorize the cognitive processing required to respond to items.[9] Some researchers note that the format of the item (multiple-choice versus constructed response) may not influence higher cognitive processing. Other researchers note that the novelty of the innovative item formats may interfere with examinees' ability to articulate their cognitive processing.[10] Because this study focused on item format and the cognitive processing needed, it was important to select a taxonomy that was familiar to the NCLEX stakeholders. Thus, this study used a variation of Bloom's taxonomy, which has been used for more than 6 years to categorize NCLEX items: remember, understand, apply, analyze, evaluate, and create.[6]

### Methodology for Assessing

The ability to assess how someone thinks has challenged many researchers. The think-aloud protocol, or verbal report,[11,12] has been used successfully with other research projects seeking to identify cognitive processing or cognitive strategies (for example, see Refs. 5,13-18). In this method, participants are generally asked to verbalize their thought processes as they proceed through completion of a task. This verbalization can take place either during completion of the task (concurrent) or after the task has been completed (retrospective). Concurrent verbal reports generally garner more data and are not subject to participants' inability to remember detail or the possibility that participants may alter their description of their thought processes. This study used a concurrent think-aloud protocol.

While there is confusion on how to best assess a domain of knowledge and the varying levels of cognitive processing required for the domain, there is consensus that the ability to critically think and reason is essential for the assessment of professionals.[17,19] These issues were addressed in this study

by investigating the thought processes used by nursing students when responding to various types of items, some of which were specifically created to require a higher degree of critical thinking and reasoning.

## Procedures

Item templates for various types of innovative items were developed and tested in part 1 of this multistage research project.[20-22] In this second stage (part 2), the item template formats were refined (Figure 1). The content for the initial item development was directed at creating items with the purpose of expanding the domain coverage of the NCLEX, either by testing skills and processes that could not be tested with text-based, multiple-choice items or by improving the ways in which certain concepts are tested.

### Item Writing, Refinement, and Production

The first step in this project was refining the item templates and producing innovative item variations. A group of subject matter experts (SMEs) revised items and templates from part 1 and developed new items. The group was asked to develop variations of the items that would enable the researchers to determine how nursing students process the items and to gather statistical information about the items.

Once the innovative versions of items were completed, text-based versions of the same items were created and refined as much as possible to have "parallel" test forms. There were some innovative items for which it was not possible to create a text-based item with any fidelity, so the items appeared in the innovative format on both test forms.

## Pilot Testing

### Participants

A total of 103 senior-level nursing students participated in this study across 6 testing occasions. Participants represented both baccalaureate and associate degree nursing programs. Ninety-four percent of the participants were female, and 6% were male. Eighty-three percent were white-not of Hispanic origin. Other demographic groups represented were African American (3%), Asian other (5%), Hispanic (5%), Pacific Islander (1%), and other (4%). Ten percent were nonnative English speakers. When rating their level of computer experience, 86% identified themselves as being *experienced* or *very experienced* with computers. Regarding their experience with computer-based tests, 97% were at least *somewhat experienced*, with 36% experienced and 38% very experienced. Of the 103 participants, 89 took the test under normal conditions in a computer laboratory, and 14 were tested in individual think-aloud sessions.

### Instruments

#### Test

Once the innovative items were produced and refined, a set of existing, nonoperational multiple-choice items was selected so that a representative number of items could be administered. In terms of item content, whenever possible, all the unique items were developed in pair, with 1 item in the traditional text format and the other in an innovative format. Each item pair measured identical content. Two fixed forms were constructed to include a combination of text-based and innovative items. Each resulting test form contained 70 items, 49 of which were unique to a test form, whereas the remaining 21 items were common across both forms. The item position, regardless if text-based or innovative, was the same across both forms. Item position was arranged so that the innovative items were interspersed throughout the test.

### Think-Aloud

The research question regarding the levels of cognitive processing required for answering various types of items was addressed by administering the test to a smaller set of participants. In this phase, 14 participants were individually tested using a think-aloud protocol. Participants were asked to verbalize their thought processes as they completed each item.

- **Graphics Inclusion**
  A photograph or drawing is included as the stimulus material for an item and may also be included as the item's response options
- **Graphics Interaction**
  A photograph or picture is the primary element in the item. The examinee interacts with the graphic by selecting one or more areas in the graphic
- **Audio Inclusion**
  An audio clip is included as the stimulus material for an item and may also be included as the item's response options
- **Video Inclusion**
  A video is included as the stimulus material for an item and may also be included as the item's response options
- **Video Interaction**
  A video is the primary element in the item. The examinee interacts with the video by playing the video and then marking the video at one or more points
- **Animation and Audio Inclusion**
  An animation with accompanying audio is included as the stimulus material for an item
- **Decision Tasks Item Sets**
  Each decision task contains 3 or 4 items related to a common set of stimulus material. These items are presented in a set sequence with an examinee's response to one of the items does *not* depend on the response to the others.

Figure 1. Item template formats.

They were encouraged to explain their reasoning for selecting the answer to an item before moving on to the next item.

## Analysis and Results

### Item Performance

The examination was delivered on computer through a Web-based interface. Once participants logged in, the software randomly assigned them to either form A or B. As participants progressed through the examination, statistical information was gathered; both classic item statistics and Rasch calibrations were computed.

Examinee responses from the 89 participants, those who were not included in the think-aloud sessions, were used to complete item analyses. A total of 42 participants completed form A, and 47 participants completed form B. Difficulty values for items presented in both innovative and text-only format were generally similar. For cases in which the difference in difficulty was noticeable, the innovative format was usually more difficult (approximately 10 items were more difficult in the innovative format, and 2 items were more difficult in the text-only format). Item total correlation values were similar across item formats. When differences were noticeable, the innovative items tended to have better discrimination (this was the case for 13 items). However, there were 5 items for which the text-based items had slightly better discrimination values. The video interaction items (marking a point in the video) were generally more difficult in the innovative format. This is not surprising as there were as many possible response options as there were video frames in the innovative versions as compared with a much smaller set of options in the text format.

Rasch item difficulties were calibrated,[23,24] and results indicated that, overall, item difficulties (b-parameters) of the same-content item pairs were comparable across the 2 item formats. As seen in Figure 2, most item calibrations fell close to the reference line. There was a slight trend, however, that some innovative items were more difficult than their counterparts in text format. When interpreting these findings, one should be mindful of the sample size limitation in the current study. Because of the small sample sizes, 7 of the 119 items calibrated contained no variability in candidate responses. That is, either all respondents answered an item correctly or all answered incorrectly. This lack of response variability rendered the resulting item difficulty estimates unstable. Figure 2 also contains calibrations of the 21 common items, ranging from a difficulty of −3.34 to 3.15 logits.

In general, the innovative items were more difficult and had better discrimination than the paired text-based items. This finding is consistent with previous research on innovative items and is quite important for item development. Understanding the statistical characteristics of innovative items can assist the NCLEX program to develop items to targeted difficulty levels.

## Cognitive Processing Ratings

Using the modified version of Bloom's taxonomy[6] as a rating framework, 3 SMEs rated each participant's interaction with each item. If the participants stated that they did not know the answer and were choosing their response by guessing, raters coded that as G (guess). Similarly, if the participants did not provide enough material to allow the raters to choose a cognitive processing level, the item was coded as NB (no basis to judge). Rating focused on the cognitive process the examinees used when interacting with the item. Raters were specifically instructed to focus on the examinee's verbal report and not on the item content and quantity of respondent's words (verbalizations). Finally, raters were cautioned to carefully focus on the thought process being verbalized, not the correctness of the rationale or final response.

Subject matter expert raters worked independently and were blind to item content and format. For ratings on which

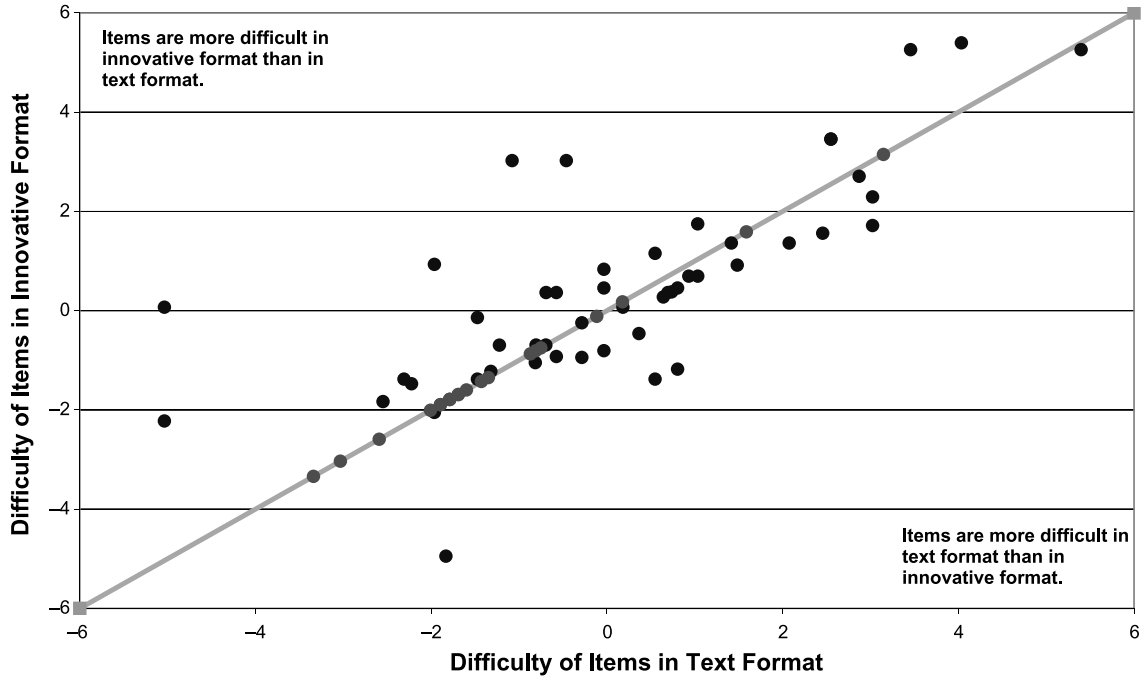


Figure 2. Item difficulty calibrations.

the SMEs disagreed (no 2 raters gave the same rating), a fourth content expert evaluated the transcript and served as the arbiter to determine the final cognitive process rating. To make comparisons between the innovative and paired text-based items, the cognitive ratings were converted to numbers using the following scale: *remember* = 1, *understand* = 2, *apply* = 3, *analyze* = 4, *evaluate* = 5, and *create* = 6. Items coded as G and NB were considered missing.

In general, only slightly more cognitive processing was used by participants to answer the innovative items as compared with the paired text-based items. There were 26 innovative items that had slightly higher cognitive ratings and 8 innovative items with substantially higher cognitive ratings (0.5 or greater). There were 16 innovative items developed specifically to determine if the format required more cognitive processing than a paired text-based, multiple-choice item. Of these 16 items, 10 required more cognitive processing. Three of these 10 had a cognitive processing rating that was 0.5 or greater than their text-based counterparts.

There were 3 innovative items (3 parts of a decision task item set) that required less cognitive processing. In this 3-part decision tasks item addressing breath sounds, the text descriptor of the breath sounds in the item stem was unfamiliar to the participants. This lack of familiarity caused the students to process the descriptor and then process the problem at length, whereas for the innovative item, the participants selected a breath sound without verbalizing very much processing. It could be that participants were not able to verbalize how they process sound. Another likely explanation is that the SMEs rated this greater amount of verbalization as being indicative of a higher degree of cognitive processing.

Another example of an unexpectedly difficult text-based item was one in which a participant had to identify the area where an injection should be given. The innovative item included a *picture* of potential sites, whereas the text item included a *description* of sites. The participants in the think-aloud verbalized that they did not understand the text description "inner surface of the forearm," which seemed to be quite clear to the SMEs who developed the item. However, when multiple textbooks were consulted, the term *ventral* (or *dorsal*) aspect of the forearm may have been a more correct and familiar term, thus underscoring the importance of careful item review.

In summary, most of the innovative items that were written specifically to assess the cognitive processing of item formats were rated higher based on participants' think-aloud. The few instances in which the text-based version of an item was more difficult could be related to examinees' lack of familiarity with the content, content errors, or the ability of the participants to speculate and reason about the problem without having response options to cue them.[13] Also noteworthy was the observed tendency of the SME raters to rate a participant's cognitive processing higher if the participant was verbose. This indicates that additional SME rater training activities may be needed. In general, many of the innovative items expected to require higher levels of cognitive processing did indeed require more processing. It may be difficult for some examinees to assess their thinking about sounds as this process seems to be automated. Thus, it may be difficult to find any differences in cognitive processing between innovative items and text-based audio items. Yet, clearly the innovative audio items are more directly assessing the ability to assess breath sounds and in a much more realistic manner. Previous research on alternate items supports these findings.[3-5]

Similarly, the video interaction items, for the most part, required more cognitive processing and were more authentic than a paired multiple-choice item. And yet, the researchers expected a greater degree of cognitive processing differences. As mentioned previously, it could be that the use of answer options in the multiple-choice item allowed the participants to verbalize their reasoning more completely as compared with being faced with no answer options from which to generate thoughts and rationale.

## Limitations

Limitations to be considered when viewing the results of this study are, first, sample size and its impact on the statistical properties of the items, and the degree to which these properties can be compared across test forms. Second is the categorization schema used for rating cognitive processing. Although the framework used has a strong foundation in the literature, its ability to fully address the cognitive processes pertinent to nursing practice may be limited. A final limitation is the variation among the think-aloud participants in their ability to verbalize their thought processes. Some participants may have engaged in higher levels of cognitive processing but simply were unable to adequately verbalize this.

## Recommendations for Further Research

As with many research studies, the answer to some questions lead to further questions. As noted previously, results from this study provide important information to help make decisions regarding further pursuit of innovative items for the NCLEX programs.

Future studies should include LPN/VN participants to ensure similar findings regarding cognitive processing and item statistical characteristics. Additional studies should evaluate the use of Bloom's taxonomy to categorize cognitive processing. There may be other taxonomies that would be more sensitive to critical thinking skills and higher-order thinking skills and would thus allow for differences in cognitive processing to surface. Moreover, future studies should consider examinees with various cultural, ethnic, and educational backgrounds.

## Conclusion

Effective clinical decision making is an essential skill for the newly licensed nurse. A large component of the effective clinical decision making is the ability to think critically and to understand complex issues. The introduction of innovative items that use sound and video should extend the domain of nursing practice that is being assessed as well as to assess some areas more authentically. Additionally, some of the item formats (such as video) are designed to assess the nursing skill and critical thinking in a qualitatively different way. The statistical properties of these innovative items are comparable to multiple-choice items. There seems to be evidence to support the development of innovative items for the NCLEX program. However, it is important to note that the investigation of various types of innovative item formats for the NCLEX is still at the research level. Additional

research and policy discussions will be needed to determine whether any of these innovative item formats will be incorporated into the NCLEX.

## References


1. Parshall CG, Harmes JC, Davey T, Pashley PJ. Innovative item types for computerized testing. In: van der Linden WJ, Glas CAW, eds. *Computerized Adaptive Testing: Theory and Practice*. 2nd ed. New York: Springer; In press.
2. Sireci SG, Zenisky AL. Innovative item formats in computer-based testing: in pursuit of improved construct representation. In: Downing SM, Haladyna TM, eds. *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates; 2006.
3. Wendt A. Continuing the quest for authentic testing: using innovative items. Paper presented at: Annual Meeting of the American Education Research Association; 2004; San Diego, CA.
4. Wendt A. Investigation of the item characteristics of innovative item formats. *CLEAR Exam Rev*. 2008;19(1):22-28.
5. Wendt A, Kenny L, & Marks C. Assessing critical thinking using a talk-aloud protocol. *CLEAR Exam Rev*. 2007;18(1):18-27.
6. Anderson LW, Krathwohl DR, eds. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. London: Longman; 2001.
7. Bloom B. *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook 1*. New York: McKay; 1956.
8. Sternberg RJ, Grigorenko EL, Jarvin L. Improving reading instruction: the triarchic model. *Educ Leadersh*. 2001;58(6):48-52.
9. Martinez M. Cognition and the question of testing item format. *Educ Psychol*. 1999;34(4):218-241.
10. Ryan K, Ryan A, Arbuthnot K, Samuels M. Student's motivation for standardized math exams. *Educ Res*. 2007;36(1):5-13.
11. Ericsson KA, Simon HA. *Protocol Analysis: Verbal Reports as Data*. Boston, MA: MIT Press; 1984.
12. van Someren MW, Barnard YF, Sandberg JAC. *The Think Aloud Method: A Practical Guide to Modeling Cognitive Processes*. London: Academic Press; 1994.
13. Ankenmann R, Moore J. Students' reasoning on multiple-choice and constructed-response versions of eighth-grade mathematics assessment tasks. Paper presented at: Annual Meeting of the American Educational Research Association; 2004; San Diego, CA.
14. Christie L. The relationship between experience and information processing in a clinical judgment task: analysis of think-aloud protocols for a group of professional nurses. Paper presented at: Annual Meeting of the American Educational Research Association; 1997; Chicago, IL.
15. Duran RP, Enright MK, Peirce LP. *GRE Verbal Analogy Items: Examinee Reasoning on Items. GRE Board Professional Report No. 82-20P*. Princeton, NJ: Educational Testing Service; 1987.
16. Fonteyn ME, Kuipers B, Grobe SJ. A description of the think aloud method and protocol analysis. *Qual Health Res*. 1993;3(4):430-441.
17. Leighton J, Gierl M. Defining and evaluation models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educ Meas Issues Pract*. 2007;26(2):3-16.
18. Norris SP. Effect of eliciting verbal reports of thinking on critical thinking test performance. *J Educ Meas*. 1990;27(1):41-58.
19. Renaud R, Murray H. The effect of higher-order questions on critical thinking skills. Paper presented at: Annual Meeting of the American Educational Research Association; 2003; Chicago, IL.
20. Wendt A, Harmes, JC. Evaluating innovative items for the NCLEX, Part 1: usability and pilot testing. *Nurse Educ*. 2009;34(2):56-59.
21. Harmes JC, Wise SL, Wendt A. JRC innovative items development. Final report to the Joint Research Council of the National Council of State Boards of Nursing and Pearson VUE. Chicago, IL: National Council of State Boards of Nursing; 2007.
22. Wendt A, Harmes C, Wise SL, Jones AT. Development and evaluation of innovative test items for a computerized nursing licensure test. Paper presented at: Annual Meeting of the American Educational Research Association; 2008; New York, NY.
23. Linacre JM. *A User's Guide to WINSTEPS: Rasch Measurement Computer Program*. Chicago, IL: MESA Press; 2003.
24. Linacre JM. *WINSTEPS Rasch Measurement. Version 3.20*. Chicago, IL: Institute of Measurement; 2004.


### Copper Surfaces Effective in Controlling MRSA

Results of international laboratory tests and clinical trails indicate that copper and copper alloys (such as brass and bronze) can help control the growth of Methicillin-resistant *Staphylococcus aureus* (MRSA). An international conference focusing on this issue was held in Athens, Greece in November of 2008. Scientists from the U.S., the U.K., Germany and Greece presented evidence in support of incorporating copper surfaces into healthcare environments to help to reduce infection risk as a method to protect public health.

The U.S. Copper Development Association (CDA) is leading international efforts in this area of research. CDA has registered copper and copper alloys as antimicrobial agents with the Environmental Protection Agency. Independent laboratory tests demonstrated that copper, brass, and bronze were 99.9 percent effective in killing certain disease-causing bacteria, including. CDA has also initiated clinical trials to compare the amount of bacteria on stainless steel, plastic and aluminum surfaces in intensive care units with that the amount of bacteria found on the same surfaces made with antimicrobial copper alloys. CDA proposes that copper alloys can lessen both cross-contamination and infection rates. The clinical trials are funded by the U.S. Department of Defense under the Telemedicine and Advanced technologies Research Center.

*Source: Medical NewsTODAY. January 6, 2009. International Copper Industry Defines Role In The Fight Against Hospital Infections. Available at http://www.medicalnewstoday.com/articles/134436.php. Accessed on January 22, 2009.*