

Memorability of Innovative Items

J. CHRISTINE HARMES, PhD

Christine Harmes is an assistant professor of graduate psychology at James Madison University

ANNE WENDT, PhD, RN, CAE

Anne Wendt, Ph.D., RN, CAE, is director of NCLEX Examinations Department of the National Council of State Boards of Nursing, Chicago, IL

Innovative items have the potential for expanding an examination's construct representation by providing opportunities to measure constructs or dimensions that cannot be measured in traditional multiple-choice items. It has been suggested that innovative items may be more memorable than traditional items, thus potentially posing security concerns. However, this claim has little or no empirical evidence to support or refute it. Given that the production of innovative items tends to be expensive and time-consuming, it is especially important that exam programs carefully consider strategies and rationale for production and operation of these item types.

A promising strategy for helping mitigate the security concerns related to item memorability and theft is the production of item variants. This approach can be used for both text-based and innovative items. Investigating the levels of memorability of various item types, as well as the specific kinds of elements that tend to be remembered by examinees, can be informative for guiding both initial item and item variant development.

The purpose of this study was to investigate the degree to which examinees are able to remember various types of items, both text and innovative. The findings from this study will inform future directions for item development, including strategies for item variation, and production and administration of innovative items.

Background

Given the great amount of time and money required for creating most innovative items, test developers are particularly concerned about the security of these types of items. The Internet provides opportunities for examinees to share items through the use of tools such as e-mails or "brain dump" Web sites. In some cases, examinees may take a test for the purpose of remembering and sharing as many items as possible. Security breaches such as these cause concern for the validity of the inferences made from scores on such tests (Davey & Stone, 2007; Impara & Foster, 2006).

It has been suggested that innovative test items may be easier for examinees to remember and then share (e.g., Sireci & Zenisky, 2006). However, there have not yet been any studies reported in the literature to test this notion. The cognitive psychology literature may provide some insight into the potential for certain types of items or tests to be more memorable than others. Of particular interest are findings related to novelty and memory and multimedia and memory.

Researchers have found that material that is unfamiliar, or novel, tends to be more easily remembered than familiar material (McDaniel, 1988; Tulvig & Kroll, 1995; Waddill & McDaniel, 1998). They have hypothesized that these differences in memorability are due to the ways in which these different categories of information are processed and encoded. If test items are presented in an unfamiliar way, contain unfamiliar types of information, or require unfamiliar actions to complete an item, these items may be processed and encoded such that they may be more accurately remembered by examinees.

In an instructional context, multimedia components are often added in the hope that they will increase a learner's understanding and retention of the material (see for example, Mayer, 2005 or Ricci & Beal, 2002). These theories are related to the idea that visual and verbal information are encoded differently (dual-coding; Paivio, 1995). By presenting material both visually and verbally, learners should have a greater potential for remembering the desired material. In the testing context, items that contain both types of content (visual and verbal) thus may be more easily remembered.

Procedures

An ongoing program of research at the National Council of State Boards of Nursing (NCSBN) has focused on the feasibility of including innovative items in nursing licensure exams (e.g., Wendt, 2004; Wendt, 2008). A recent study by Harmes, Wise, and Wendt (2007) involved the creation of innovative item templates for the following formats: graphics inclusion, audio inclusion, video inclusion, animation inclusion, and decision task item sets. These types are described as follows:

- *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.
- *Audio Inclusion*
An audio clip is included as the stimulus material for an item.
- *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 three or four items related to a common set of stimulus material. These items are presented in a set sequence, and an examinee's response to one of the items does not depend on the response to the others. Many of these item sets contain other innovative elements as well (graphics, video, nursing diagnosis lists, client chart information, etc.)
- *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 within the graphic.

Item development was focused on content areas in the NCLEX test plan for which items were needed and that had a promising match for the set of innovative item types. Additional consideration was given to direction from the NCSBN member boards regarding key skills to assess for entry level nurses (e.g., critical thinking and clinical reasoning, ethical and legal issues, psychomotor skills, and interpersonal skills). Item templates and prototype items were constructed 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 concepts are tested.

The current study was designed to further the efforts of the initial project (Harmes, Wise, & Wendt, 2007; Wendt, Harmes, Wise, & Jones, 2008) by creating a larger set of items from the templates and testing various aspects of their cognitive properties. As part of this project, item writers were asked to create at least one item variant for each item they developed from a template.

Pilot Testing

Instruments

Test. Two fixed test forms were created, using a total of 119 unique items. This included innovative items and their variants, along with a set of existing text-based multiple-choice items (non-operational). Forms were built 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 while the remaining 21 items were common across both forms. Text-based versions of the innovative items were created such that each item pair measured identical content. When forms were assembled, the text version of an item appeared on one form, and the innovative version on the other form. The item position was the same across the two forms. Item position was arranged so that the innovative items were interspersed throughout the test form, and an equal number of innovative items appeared on each form.

Memorability. At the beginning of the testing session, participants were instructed that they would first take a test, and that they should try to remember as many items and in as much detail as possible, as they would later be asked to record their remembrances. After completing the test, participants were presented with a large text box and asked to type out everything they could remember from the test. Participant reports were coded for the type of information remembered and the degree to which each item was remembered.

Participants

Six testing sessions were conducted at five schools, representing both Bachelor's and Associate's degree granting programs. A total of 89 senior-level nursing students participated. Of this group, 94% were female and 6% male. The majority of the participants were White-Not of Hispanic Origin (83%). Other demographic groups represented were African American (3%), Asian Other (5%), Hispanic (5%), Pacific Islander (1%), and Other (4%). Non-native speakers of English represented 10% of the participants. Participants were asked to rate their level of computer experience. Eighty-six percent identified themselves as being experienced or very experienced with computers and 97% reported being at least somewhat experienced with taking computer-based tests.

Data Analysis

Three content staff members from NCSBN mapped participant memory files to test items and rated the degree to which each participant remembered each item (using the rating scale provided in the Appendix). The rating scale ranged from 0 to 4, with 0 used when the participant remembered nothing about the item and 4 used when the participant remembered the main point of the item and most of the item details. Staff completed ratings individually and rater agreement was computed. If two out of the

three raters agreed, their rating became the final rating. Ratings were flagged for follow-up for cases in which all three raters disagreed. Raters then held discussions to come to a resolution for these cases, thus resulting in a single, final rating for each item for each participant. As they were going through the rating process, the raters were "blind" to which form the examinee was taking.

Results

The range of average ratings was 0 to 1.38 on Form A and 0 to 1.49 on Form B. Nearly all items across both forms received an average remembrance rating of less than 1. On Form A, 14 items had an average rating of more than 1, and on Form B, 18 items had an average greater than 1. The highest average was 1.38 on Form A. On Form B, the highest average was 1.49. It is important to note that these "highest" average remembrance ratings are actually quite low, corresponding to the participant having a general impression of the item (a rating of 1) and having a general description plus a few details. As depicted in Figure 1, the vast majority of item remembrances were rated as zero. Each bar in the graph represents the number of items receiving a particular rating, summed across all items of that type (innovative or text) for that test form. For example, there were a total of 611 ratings of zero for the innovative items on Form A and a total of 658 ratings of zero for the innovative items on Form B. A few items had a rating of 4 for some participants. On Form A, two text and three innovative items had *one* person for which the remembrance was at level 4; three text and three innovative items had *two* participants with a remembrance rating of 4; and three text items had *three* participants whose remembrances were rated at level 4. On Form B, seven text and ten innovative items had *one* person for which the remembrance was at level 4; one text and six innovative items had *two* participants with a remembrance rating of 4; and one text item had *three* participants whose remembrances were rated at level 4.

An item that was remembered to a high degree in its innovative format was generally also remembered to a high degree in its text form. For example, one item had an average rating of 1.21 in its innovative form and 1.07 in text form. Of the items for which there was a difference in the average remembrance rating between its innovative and text versions, most differences were negligible. However, seven items had differences greater than 0.5. These items were two decision task item sets. For these

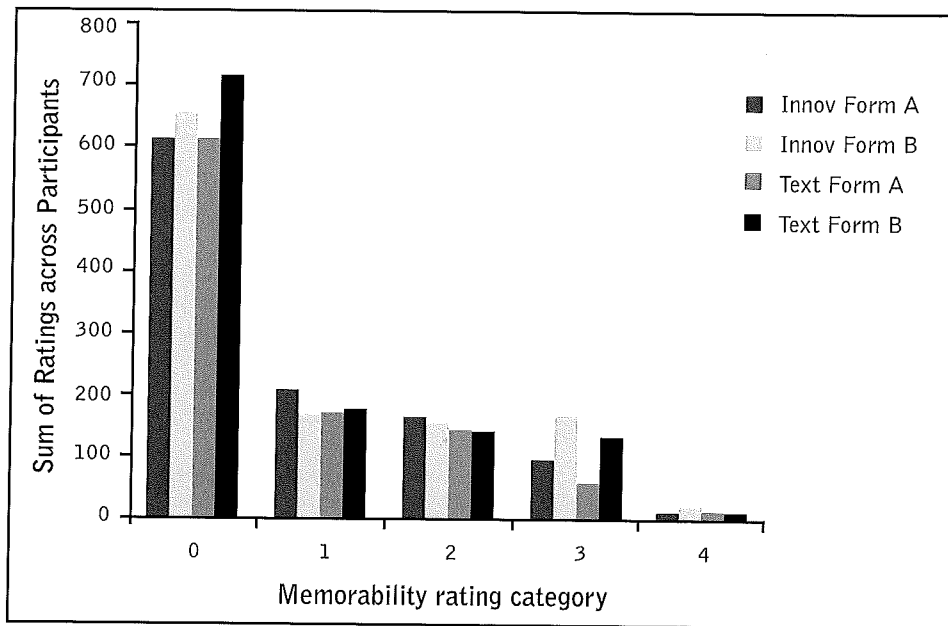


Figure 1. Sum of remembrance ratings across items and participants by category.

items, the innovative versions were remembered more than the text versions. One of these item sets required that examinees choose a needle, syringe, and injection site to administer a prescribed injection. The other item set was a variant of the first set, with a different client and prescribed injection. While the average remembrance ratings were higher for the innovative versions, it should be noted that the average values are still at the lower end of the scale. It is also interesting to note that these two item sets had higher average remembrance ratings in the innovative form, but actually had the same, or greater, number of participants whose remembrances were rated at level 4 in the text version.

Limitations

As with any research, results of this study are subject to limitations. First is sample size. As participation was voluntary, the sample size garnered was smaller than desirable, with approximately 45 participants taking each test form. The second concern is the motivation of the participants. Due to logistical constraints, we were only able to offer one level of incentive for participants in the study. It is possible that their level of motivation to remember items may have been higher if participants had been differentially rewarded based upon how many items

they remembered and how well they remembered them. This presents an interesting approach for future study.

Conclusions

The innovative items developed and tested in this study extend the range of nursing knowledge and competencies that can be assessed with the NCLEX. These innovative items often provide more direct measurement than their text-based counterparts. Moreover, the increased authenticity of these new items allows for the assessment of different areas of nursing practice that have been difficult to test.

The results of this study generally aligned with the theory from cognitive psychology suggesting that novel things are more readily remembered. However, when participants mentioned a novel item, they generally did so by describing the general interaction with the item format (e.g., “where we had to put a flag in the video”) instead of information that was specific enough to compromise an item.

In general, the innovative items were somewhat more memorable than the text-based items, which was likely due to the novelty of the items. Most of the participants remembered item formats followed by general item content but nothing specific that would, on the surface, compromise the item. For example, they would remember “listening” to breath sounds or watching videos on conflicts.

Overall our results suggest that participants do not remember many specific item components or keys. This suggests that item variants can be created successfully by using different distracters and keys. We also noted that examinees did not tend to remember specific information about clients or about individual actors.

Thus, results from this study indicate that the surface characteristics of the innovative items may be more memorable than a multiple-choice item but the specific details were not. In the future, strategies may be developed to

assist in the recall of innovative items and eventually lead to item compromise. Based on this study, it will take an organized effort to develop an innovative NCLEX item complete with an accurate stem, distracters and key. Single candidates will most likely not remember accurately an entire item. Moreover, based on NCLEX research presented to NCSBN's Joint Research Committee, it may take as many as 30 "remembered" items to affect a candidate's results (Reckase, 2008).

Based on this research, remembrances of innovative items are not more likely to lead to item compromise than remembrances from a multiple-choice item. There is a wealth of information in the memorability data that could be analyzed in order to understand the implications for item development and to mitigate security risks.

References

- Chuah, S. C., & Do, B. R. (2006). *Test administration as a means for improving test security: CAT versus paper and pencil*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Davey, T., & Stone, E. (2007, April). *Improving security under continuous testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Harmes, J. C., Wise, S. L., & Wendt, A. (2007). *JRC innovative items development*. Final report to the Joint Research Council of the National Council of State Boards of Nursing and Pearson VUE.
- Harmes, J. C., Barry, C. L., & Kaliski, P. K. (2007, November). *Are they really more memorable? Implications of innovative items for test security*. Paper presented at the annual meeting of the Florida Educational Research Association, Tampa, FL.
- Impara, J. C., & Foster, D. (2006). Item development strategies to minimize test fraud. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning*. New York, NY: Cambridge University Press.
- McDaniel, M. A., Dunay, P. K., Lyman, B. J., & Kerwin, M. L. E., (1988). Effects of elaboration and relational distinctiveness on sentence memory. *American Journal of Psychology*, 101, 357-369.
- Mead, A. D. (2006). *A comparison of traditional and simulation assessment memorability and security*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Paivio, A. (1995). Imagery and memory. *The cognitive neurosciences* (pp. 977-986). Cambridge, MA: MIT Press.
- Parshall, C. G., & Harmes, J. C. (2008, Summer). The design of innovative item types: targeting constructs, selecting innovations, and refining prototypes. *CLEAR Exam Review*, 19(2).
- Parshall, C. G., Harmes, J. C., Davey, T., & Pashley, P. J. (in press). Innovative item types for computerized testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice*. (2nd ed.). New York: Springer.
- Reckase, M. (2008). *Impact of item disclosure*. Paper presented to the March 2008 Joint Research Committee. Chicago, IL.
- Ricci, C. M., & Beal, C. R. (2002). The effect of interactive media on children's story memory. *Journal of Educational Psychology*, 94(1), 138-144.
- Sireci, S. G., & Zenisky, A. L. (2006). Innovative item formats in computer-based testing: In pursuit of improved construct representation. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain and long-term memory encoding. *Psychonomic Bulletin & Review*, 2, 387-390.
- Waddill, P. J., & McDaniel, M. A. (1998). Distinctiveness effects in recall: Differential processing or privileged retrieval? *Memory & Cognition*, 26, 108-120.
- Wendt, A. (2004, April). *Continuing the quest for authentic testing: Using innovative items*. Paper presented at the Annual Meeting of the American Education Research Association, San Diego, CA.
- Wendt, A. (2008). Investigation of the item characteristics of innovative item formats. *CLEAR Exam Review*, 19(1), 22-28.
- Wendt, A., Harmes, C., Wise, S. L., & Jones, A. T. (2008, March). *Development and evaluation of innovative test items for a computerized nursing licensure test*. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.

Appendix

Rating Scale for Participant Memories

Please use the following rating scale when evaluating individual examinee memory files. You will need to come up with a rating for each examinee for each item on the test.

To what degree does the participant remember the item?

- | | |
|---|--|
| 0 | None: Participant remembers nothing about the item. |
| 1 | Few: Participant has a vague or general impression of the item |
| 2 | Some: Participant has a general description of the item and includes some additional details |
| 3 | Many: Participant has a general description of the item and includes a lot of detail. |
| 4 | Most: Participant remembers the main content and focus of the item and most of the details about its components and few are missing. |