Testing Smarter with Technology?

Anne Wendt, Ph.D., R.N.
National Council of State Boards of Nursing, Inc.
Chicago, IL

Abstract

This article briefly discusses the factors affecting the National Council of State Boards of Nursing's decisions to use computerized adaptive testing to administer the National Nursing Licensure Examination (NCLEX(r) examination). Practical issues associated with implementing computerized adaptive testing for large-scale, high-stakes test such as the National Council's Licensure Examination (NCLEX(r)) are discussed.

Have you wondered why so many tests are now being administered via computer? Using the National Council of State Boards of Nursing licensing examination (the NCLEX examination) as a model, we will examine some of the issues surrounding computer based testing, particularly the advantages and potential problems that can occur with its implementation.

Reasons for Using Technology in Testing

Aside from the excitement of using up-to-date technology for testing, one of its major advantages is convenience to the candidates. Prior to April 1994, the NCLEX examinations were paper-and-pencil examinations which were administered to about 190,000 nursing candidates per year. Each examination was administered twice a year for one day (NCLEX-PN exam) or two days (NCLEX-RN exam). Now candidates call and schedule their examination at a time and place that is convenient for them. Candidates are tested via computer in one of approximately 250 Sylvan Technology Centers that are located in each state or U.S. territory. In addition, candidates are tested in a quiet, private, self-paced setting rather than in the large auditoriums or halls when everyone began and ended their examination at the same time.

Another major advantage for the National Council of using technology for testing is that a large volume of candidates can now be tested almost daily on a year-round basis. In addition, candidate results are reported to the respective Boards of Nursing within 48 hours. Thus, candidates who are competent are able to join the workforce much more quickly. And Boards of Nursing are able to quickly identify those individuals who are not competent to practice (Zara, 1996). Therefore, by taking advantage of computer technology, Boards of Nursing are better able to perform their role of public protection while licensing nurses in a timely manner.

A third advantage to the National Council of utilizing advances in technology is to improve the security of the examination. Major causes of security breaches for paper-and-pencil examinations are lost and/or stolen test booklets, copying answers from other test-takers during the examination, and hiring impersonators to take the examination (Scheuneman, 1997). The use of the computer adaptive testing has virtually eliminated the first two causes of security breaches. With computerized testing, examination materials can be encrypted. The adaptive nature of computerized adaptive testing (CAT), where each candidate is administered a unique set of test questions tailored to his/her competence level, eliminates the possibility of copying answers from the person taking the examination in the next testing cubicle. Technology is also used to enhance security measures. For example, the National Council uses strict security procedures which include a sign-in log, appropriate picture identification, a digital photograph taken at the site which can be forwarded with the examination results, thumbprinting, and video monitoring of the testing event. All of these measures help to ensure that the correct candidate takes the examination and that the testing event (as much as possible) is the same for all candidates.

Lastly, use of CAT for the NCLEX examination allowed a reduction in testing time without a loss in the precision or accuracy of the results. That is because candidates do not "waste time" trying to answer questions that are too hard or too easy for them. Basically, that is the goal of CAT— to determine competency based on the difficulty of the questions answered correctly, rather than the number of questions which are answered correctly as is the case with many paper-and-pencil or computer linear examinations (Wainer, 1990; National Council, 1995). While every candidate's examination meets the NCLEX Test Plan, only questions which contribute to the measurement of the individual candidate need to be administered. For the NCLEX-RN examination, the testing time decreased to five hours from two days. And, because NCLEX CAT is a variable-length examination, many of the candidates taking the examination finish in less than the allotted 5 hours. Thus, the increased efficiency of computerized adaptive testing for the NCLEX examination can be translated into savings to the candidate and organization.

In summary, there are a number of compelling reasons to take advantage of changes in technology to begin "testing smarter." By implementing computerized adaptive testing, the National Council has been able to increase test security, shorten testing time, enhance Member Boards' mission of public protection, and streamline the testing process.

Practical Issues

While the decision to take advantage of technological changes may seem to be an easy one, there are some practical issues that need to be addressed. One of the major ones is how to maintain the psychometric soundness and legal defensibility of a CAT administered examination. The National Council spent a great deal of time and effort to ensure that an NCLEX CAT examination would be legally defensible. Much

of this process is documented in the Collected Works on the Legal Aspects of Computerized Adaptive Testing (National Council, 1991). In addition, the National Council conducted field tests and a large-scale beta test to ensure the psychometric soundness of the NCLEX examination administered by CAT methodology (Way, 1994). The preparation for implementing CAT for the NCLEX examination took considerable time and effort. Do not underestimate the time, effort, and money that it will take to address these issues.

Now let us move on to the more "practical" aspects of testing smarter with technology. Two of the major issues that need to be addressed are the cost of setting up a system for computer delivery of the multiple-choice test questions and the quantity of questions that are available for converting to computer delivery. The former may be easily achieved if the questions are already formatted and available in electronic files. The latter may be quite costly. There are several reasons that many questions (items) are needed for CAT. First, large numbers of items are needed to ensure that there is an "acceptably low" level of item overlap between candidates. Limiting item exposure in this way prevents the items from becoming common knowledge among candidates. In addition, items are needed in all areas of a test plan at all difficulty levels to ensure precise estimates of candidates with varying competence levels. Because the field of measurement does not have a good way to perform "on-line item calibrations", the items will need to be tried out (given as unscored items to candidates in order to gather statistical information about the item) before administering the items as "operational" scored items. This extensive task is very time-consuming.

It should also be noted that computerized adaptive testing items may need to be classified more carefully using dimensions other than the test plan categories. When a paper-and-pencil test form is produced, a content expert selects the items and reviews the composition of the test form for such things as cueing and overlap. This is not possible when the computer automatically selects each candidate's items based on an algorithm. Therefore, items may need to be classified for additional dimensions. Additional and in-depth item classification is especially important in a health-care related examination where there are likely to be frequent changes in practice and/or terminology necessitating frequent reviews of the items in the pool.

More detailed item classification is also important when considering item pool needs. With large item pools, there may be no easy or effective way to determine voids in content areas without more detailed item coding. With smaller item pools it may be possible for one or two people to be the "gurus" of the item pool, i.e., know what specific concepts are and are not addressed in the items. But with any more than about 3,000 items, in-depth knowledge of the pools is impossible. Detailed item classification allows for a more in-depth knowledge of test content and an easier determination of voids in that content.

Now let us turn to some administration issues. As noted previously, one of the compelling reasons for computer-administered examinations is the reduction in security issues. However, it should be noted that continued vigilance in matters of security is absolutely necessary. Candidates for high stakes examination may try to find ways to circumvent the security measures. Any irregularities need to be investigated. The time and effort that this may take should not be underestimated. Candidates may complain about adverse environmental conditions such as excessive heat or noise. These complaints also need to be investigated and followed up. With multi-site, daily testing it is important to be diligent in tracking issues and trends. To effectively handle all of these administration issues, additional staffing may be required.

These practical issues are not intended to be an exhaustive list of what can be encountered when taking advantage of technological advances. It is up to the test-sponsoring organization to assess if the advantages to its stakeholders outweigh the disadvantages. In particular, some of the issues to consider are "Will computer-based testing save time or money for the program or candidates?" "Will security improve?" "Will better decisions about candidate competence be made?"

References

National Council. (1991). Collected Works on the Legal Aspects of Computerized Adaptive Testing. Chicago: National Council of State Boards of Nursing.

National Council. (1995). The NCLEX Process. Chicago: National Council of State Boards of Nursing.

Scheuneman, J. (1997). Testing and measurement issues: potholes on the road to computer-based testing. CLEAR Exam Review. Volume VIII, Number 1. Kentucky: Council of Licensure, Enforcement, and Regulation.

Wainer, H. (1990). Computerized adaptive testing: a primer. Hillsdale New Jersey: Erlbaum Associates, Publishing.

Way, W. (1994). Psychometric results of the NCLEX beta test. Paper presented at the Annual Meeting of the American Educational Research Association.

Zara, A. (1996). "Overview of a successful conversion." In Computer-based examinations for board certification. Evanston, Illinois: American Board of Medical Specialities.

Anne Wendt, Ph.D., R.N.

Anne Wendt is the NCLEX(Content Manager at the National Council of State Boards of Nursing, a not-for-profit organization responsible for the development of the National Council nursing licensure examination (NCLEX(Examination). She received her BSN from the University of Minnesota, her MSN from Loyola University, and her Ph.D. in Psychometrics from the University of Chicago.

Anne Wendt has a unique perspective of nursing licensure exams because she comes to her position as a nurse, a psychometrician, and as an educator. She was instrumental in the National Council's transition from a paper-and-pencil NCLEX examination to its current computerized adaptive testing (CAT) form. She has co-authored the NCLEX test plans and detailed test plans since March 1993. She has also been influential in the publication of such documents as The NCLEX™ Process, The NCLEX™ Manual and Assessment Strategies for Nursing Educators.

She is the author and co-author of numerous articles and has presented papers nationally. She continues to be actively involved in various research projects, particularly in the areas of item development and test construction.