Review of Computerized NMTCB Certification Examination

Bhaskar R. Dawadi

Nuclear Medicine Technology Certification Board, Atlanta, Georgia

Key Words: certification examination; IRT

J Nucl Med Technol 2001; 29:31–33

The Nuclear Medicine Technology Certification Board (NMTCB) started certifying technologists through examinations by offering its first paper-and-pencil (P&P) multiple-choice (MC) examination on September 15, 1978. Up until 1995, the Board regularly administered 2 P&P MC certification examinations per year. During 1996, the Board converted its P&P MC examination into computer-based testing (*1*). The NMTCB computerized examination has been one of the premier certification examinations not only in the field of nuclear medicine, but also in the field of computerized testing. The purpose of this article is to provide a general review of the existing computerized version of the NMTCB certification examination. In particular, we will review specifications of the computerized test, conversion of P&P into the computerized version, the reliability and validity of the computerized test, and the trend of computerized NMTCB testing in the future.

Computerized testing provides several advantages over P&P tests. The most frequently mentioned benefits are scheduling convenience for test-takers, short testing time, a secure testing environment, flexibility in the use of various item formats, a testing environment more realistic to the real world, and immediate score reporting. The NMTCB examination uses computerized adaptive testing (CAT) for classification. Under CAT, candidates are classified most accurately and efficiently into 2 categories, such as pass/fail, or minimally competent/not minimally competent. The computer administering the test stops when it determines that the ability of the examinee taking the test is clearly above or below the minimally competent level. However, the validity of the test score is questioned if a candidate below minimum competence passes the test, or if a candidate who is minimally competent to practice fails the test. These 2 errors are called false-positive and false-negative, respectively. A statistical technique called sequential probability ratio test (SPRT) is applied to control the false-positive and false-negative classification errors in the NMTCB computerized adaptive testing. The SPRT method is used widely to control such types of errors (*2*).

For correspondence or reprints contact: Bhaskar R. Dawadi, PhD, Executive Director, Nuclear Medicine Technology Certification Board, 2970 Clairmont Rd., Suite 935, Atlanta, GA 30329; Phone 404-315-1739; Fax 404-315-6502; brdawadi@nmtcb.org.

DETERMINING A PASSING SCORE

In a certification examination, determining a passing standard that divides the test scores into 2 categories of minimally competent/not minimally competent is an integral part of the examination development process. This determination is also an integral part of the CAT for classification. In the early days of the NMTCB test development process, the corrected Nedelsky method was used for determining the cutoff score for the examination (*3*). The examinations administered in September 1993 and June 1994 with the original 200-item tests were selected as the standard reference forms to determine the cutoff score for the computerized testing. The 200-item test refers to the P&P test with 200 multiple-choice questions used in the NMTCB examination. With these 2 administrations, 1322 candidates were tested. The passing score was 123 items (questions) correct out of 200 items (questions), or 61.5% (*4*). This passing score was determined using the aforementioned method. American College Testing, Inc. (ACT), a well-established testing company, helped the NMTCB to determine the passing score of its computerized adaptive testing by using 3 steps (*4*). First, all of the multiple-choice items included on the standard reference forms were calibrated using the 3-parameter item response theory (IRT) logistic model, a mathematic model that fit well with the NMTCB examination data. Calibration is the method by which item characteristics such as difficulty, discrimination, and guessing values of those items are determined for computerized testing. Second, the standard reference set test characteristic curve (TCC),[1] based on the

[1]TCC incorporates characteristics (such as difficulty, discrimination, and guessing) of individual test items and provides characteristics of the test as a whole. All individual test items (questions), when calibrated using the 3-parameter logistic IRT model, will have numerical values of difficulty, discrimination, and guessing based on the responses of examinees to those items. Item characteristic curves (ICC) are drawn for each individual test item (similar to the above TCC) using numerical values of discrimination, difficulty, and guessing. An ICC is used to determine the probability of an examinee answering an individual item correctly based on the level of his/her ability. In IRT, ability is denoted by the Greek letter theta (θ) and represented by the X-axis. The Y-axis represents the probability, $P(\theta)$, of correct response at a given ability level θ. In the case of a typical test item, the probability of answering the item correctly will be small for examinees of low ability and large for

calibrated data, was drawn using the IRT model (Fig. 1) (*4*). Finally, the cutoff score (θ_{Passing}) for the computerized testing that corresponded to the 61.5% on the P&P test was determined using the TCC curve. Candidates scoring above the cutoff are classified as minimally competent, and candidates scoring below the cutoff are classified as not minimally competent. The decision is based on the candidates' responses to the items presented during the test. Items are presented to candidates based on 3 prespecified constraints: testing time limit, content coverage, and item exposure control.

TESTING TIME LIMIT

Candidates who take the current NMTCB computerized adaptive test answer a total of 80–90 test items in 1 h and 45 min. Within this time limit, each candidate responds to a different number of test items because of its adaptive nature. However, each candidate responds to a total of 50–75 operational test items used to determine pass/fail status. The remaining 15–30 items to which candidates respond are pretest items, which have no impact on the pass/fail decision (*5*). These items (questions) are being pilot-tested for future operational use and are very important for maintaining the high quality of the item bank.

CONTENT COVERAGE

All of the items on the test are selected from 4 areas: radiation safety, instrumentation, clinical procedures and radiopharmacy. The 4 areas and the percentages of the total number of items from these areas were determined by the information received from task analysis. On each examination administered, it is very important to make sure that the number of items from the given 4 areas are selected based on 15%, 20%, 45% and 20% of the total number of test items, respectively (*5*).

ITEM EXPOSURE CONTROL

This is an important issue in computerized testing. If many candidates see the same item repeatedly in a relatively short span of time, the item loses its usefulness. Thus, it is critical to present each item in the item bank only to a certain number of candidates taking the test. For the NMTCB certification examination, the target exposure rate is set to 15%. It is reassuring to note that the actual average exposure rate of all operational items in the NMTCB item bank has been around 14% (*5*).

During the computer administration of the test, candidates can fail the test under certain conditions even if they are minimally competent. Candidates fail the test if they do not respond to the minimum number of 50 items, if they quit before answering all the items presented, or if they fail to complete the test in the allotted time. Candidates cannot skip any item presented, but they can go back to review and change their own responses if they have enough time. This feature is much appreciated by NMTCB test takers.

examinees of high ability. Similarly, the TCC provides the probability of an examinee's performance on that test at a given level of his/her ability.

RELIABILITY

A measurement instrument is considered reliable when the instrument provides consistent readings for multiple measures. For example, when a patient's body temperature was taken twice consecutively by the same thermometer, and if the 2 readings were not similar (e.g., 103° versus 98.1°), you would not use that thermometer again because you would consider the thermometer unreliable. The certification examination is used to ascertain the ability of examinees, and should provide consistent results when the testing is repeated on a group of individuals. The reliability of the NMTCB certification examination is evaluated by checking the consistency with which it would classify candidates if they were to take the same test (or one parallel to it) twice. By conducting CAT simulation studies, ACT was able to evaluate the reliability of the NMTCB certification examination by calculating 2 indices: proportion of agreement (P_0) and Cohen's kappa (*5*). The index P_0 is the proportion of candidates that is consistently classified as pass/pass and fail/fail. However, this index (P_0) is not sensitive to classifications that are correct by chance. Cohen's kappa is corrected for chance and is interpreted as the proportion of consistent classification after correcting for chance. These indices range from 0 to 1, with the higher values indicating higher reliability. The proportion of consistent classifications agreement and Cohen's kappa for the NMTCB CAT are estimated to be 0.95 and 0.76, respectively. Because the Cohen's kappa index is adjusted for chance, its value is lower than the value of the index P_0. However, the Cohen's kappa of 0.76 for the NMTCB examination indicates that this certification examination is performing well.

VALIDITY

Although reliability of an examination is very important, it is not a sufficient condition for a test score to be valid. In other words, a high level of reliability does not mean that

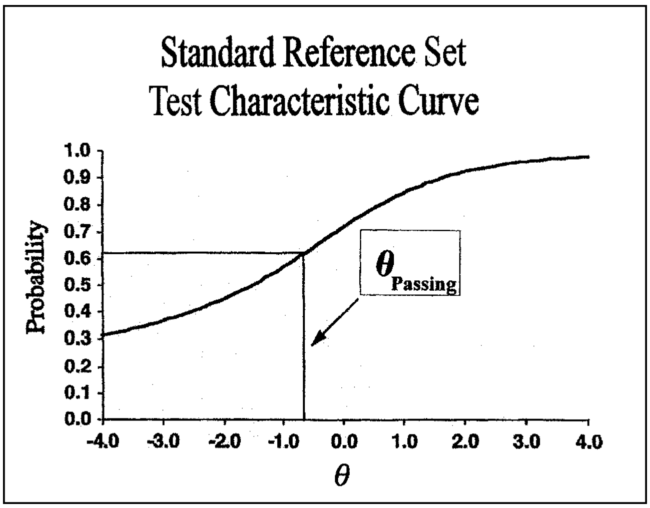

FIGURE 1. Test characteristic curve drawn using the data from the original 200-item multiple choice test administered in September 1993 and June 1994.

test scores obtained from the reliable test are valid automatically. The Standards (*6*), an authoritative source of professional technical guidance in the field of certification testing, states, “Validity is, therefore, the most fundamental consideration in developing and evaluating tests.” The validity of test scores needs to be established, and it can be established by “accumulating evidence to provide a sound scientific basis for the proposed score interpretations (*6*, p. 9).” It is important to note that it is not the test that is to be validated, but it is the score interpretation or the inferences made from test scores that have to be validated.

However, there is no simple way—such as calculating a coefficient—to show that inferences made from test scores are valid. The best way to validate score interpretation is by collecting various evidence that supports the intended interpretations and their relevance to the proposed use. Construct-related, content-related, and criterion-related evidence could be collected to enhance the intended interpretation of test scores for the proposed purpose (*6*). Out of these various sources, however, some types of evidence are more appropriate in a given case than others, and for certification examinations it is the content-related evidence that is more heavily relevant (*7*, *8*). Providing content-related evidence for the nuclear medicine technologist certification examination involves not only defining the domain of test content and making sure that the test adequately covers the domain, but also defining critical skills and abilities necessary to practice the profession safely.

To make sure that all of the content-based evidence is covered and updated for the NMTCB examination, the NMTCB has regularly conducted task analyses. The first task analysis was conducted in 1982 (*9*). Subsequent analyses were conducted on a regular basis, and the information was disseminated through *JNMT* (*10–15*). A new task analysis is planned for 2002. Updating the task analysis every 3–5 years is a standard practice in the field (*16*).

Criterion-related evidence is collected by relating the score of the examination to some outside criteria. In 1996, the Board required an additional 45 h of course work for alternate eligible candidates to qualify to take the NMTCB certification examination. This requirement was added to already existing requirements. During the period of 1992–1995, before the additional 45 h of course work was required, the passing rate of alternate eligible candidates was between 56–66%. However, during the period of 1997–1999, after the additional 45 h of educational course work was implemented, the passing rate of the alternate eligible candidates jumped to 87–100%. This additional requirement has had a positive impact on the passing rate of alternate eligible candidates.

Future of the Test

The rapid development of computer technology and easy Internet access holds many new promises and opportunities for the NMTCB certification examination. The ease and convenience of Internet access makes it likely that candidates will eventually take their certification examinations without leaving their campuses or the hospitals where they are trained. In addition, the availability of inexpensive but fast computers means that interactive, simulation-based testing could be implemented. Performance-based examinations are also considered more realistic and can test more skills critical to the day-to-day tasks of technologists than multiple-choice examinations.

Your Contribution

To ensure the continued success of the NMTCB certification examination, it is extremely important that you, as an expert practicing nuclear medicine, participate in all activities whenever you are approached by the Board. If contacted, you would have been selected through the random sampling statistical method; your response is necessary to enhance the power of the statistical analysis. Your responses to the survey questions about task analysis or your participation in item-writing workshops are very important to the Board. In general, your feedback to the Board on any examination issue is critical to the future examination development process.

REFERENCES

1. Nuclear Medicine Technology Certification Board. *Certification Examination Application Instructions*. Atlanta, GA;1999.
2. Spray JA and Reckase MD. Comparison of SPRT and sequential Bayes procedures for classifying examinees into two categories using a computerized test. *J of Educational and Behavioral Statistics*. 1996;21:405–414.
3. Determining passing score for an examination: the corrected Nedelsky Method. Internal Report, ACT Inc., March 1993.
4. Spray JA. A report on the estimation of latent abilities from simulated NMTCB CAT administrations. Internal Report, ACT Inc.; 1997.
5. NMTCB Technical Summary, Internal Report, ACT Inc.; May 2000.
6. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 1999.
7. Stoker H W and Impara J C. Basic Psychometric Issues in Licensure Testing. In: Impara James C, ed. *Licensure Testing: Purposes, Procedures, and Practice*. University of Nebraska-Lincoln, Buros Institute of Mental Measurements 1995:167–186.
8. Fabrey, L. Basic psychometric principles. In Browning AH, Bugbee Jr. AC and Mullins MA, ed. *Certification: A NOCA Handbook.* Washington, DC: National Organization for Competency Assurance. 1996;1–40.
9. Nuclear Medicine Technology Certification Board. NMTCB certification examination validation report. *J Nucl Med Technol.* 1982;10:210–22.
10. Nuclear Medicine Technology Certification Board. NMTCB critical task validation study: identification of entry level domain. *J Nucl Med Technol*. 1984;12:192–200.
11. Nuclear Medicine Technology Certification Board. Reexamination of NMTCB critical task survey: a response to changing entry level practice. *J Nucl Med Technol.* 1986;14:228–234.
12. Blosser N, Drew H and Steves A. NMTCB task analysis validation update. *J Nucl Med Technol.* 1988;16:216–221.
13. Cizek GJ and Bridges JA. Task analysis: foundation for a valid examination. *J Nucl Med Technol.* 1991;19:245–246.
14. Blondeau KL, Hartnett SD, Pickett MW et al. Nuclear Medicine Technology Certification Board (NMTCB) 1991 Critical Task Analysis Validation Report. *J Nucl Med Technol.* 1992;20:173–176.
15. Wells P and Leahey D. Nuclear Medicine Technology Certification Board 1997 Task Analysis Report. *J Nucl Med Technol.* 1998;26:52–56.
16. Henderson JP. Job analysis. In: Browning AH, Bugbee Jr. AC, and Mullins MA, ed. *Certification: A NOCA Handbook.* Washington, DC: National Organization for Competency Assurance:1996.