Readability of Licensure Examinations

ADA WOO, PhD
Ada Woo, Ph.D., is the psychometrician for the NCLEX Examinations at the National Council of State Boards of Nursing, Chicago, IL

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

WEIWEI LIU, MS
Weiwei Liu, M.S., is statistician for NCLEX Examinations at the National Council of State Boards of Nursing, Chicago, IL

Licensure examinations are intended to measure whether an individual possesses some part of the overall skill set and knowledge necessary for professional practices (Clauser, Margolis & Case, 2006). As a measure of public protection, many professions in the healthcare industry require individuals to obtain licenses prior to granting rights to practice independently. While certain levels of language skills are essential for effective practice in many licensed professions, reading ability is not generally the central focus of licensure examinations. Readability assessment is not part of the routine analyses in many examination programs. A survey of the literature revealed that the majority of readability assessments were conducted on informed consent documents (e.g., Christopher, Foti, Roy-Bujnowski & Appelbaum, 2007; Walters & Hamrell, 2008), legal documents (e.g., Collins, Novotny & Light, 2006) and academic curricula (e.g., Fuller, Horlen, Cisneros & Merz, 2007). The issue of readability of examination items is rarely studied (Hewitt & Homan, 2004).

More than one definition of readability exists in the literature. In a classic text, Gray (1975) defined readability "as the relative ease or difficulty with which a reader understands writing" (p. xi). On the issue of examination readability, the *Standards for Educational and Psychological Testing* states:

> *In testing applications where the level of linguistic or reading ability is not part of the construct of interest, the linguistic or reading demands of the test should be kept to the minimum necessary for the valid assessment of the intended construct* (AERA, APA & NCME, 1999, p. 82).

As the developer of the NCLEX® Examinations, the National Council of State Boards of Nursing (NCSBN) routinely monitors readability of the examination items. The NCLEX is constructed to test knowledge, skills and abilities essential for the safe and effective nursing practice at the entry-level. Currently, all U.S. states and jurisdictions, including the Northern Mariana Islands, Virgin Islands, American Samoa and Guam, require individuals who seek licensure as registered nurses or practical/vocational nurses to pass the NCLEX® Examinations (NCSBN, 2008).

Readability of each NCLEX operational item pool is assessed before the pool is deployed. The NCSBN employs a Computerized Adaptive Testing (CAT) paradigm in delivering the NCLEX. Different items are administered to different candidates on the NCLEX depending on candidates' ability estimates. Essentially, each candidate who takes the NCLEX will receive a different set of items based on his/her ability. To assess

readability levels of the NCLEX, three examinations are simulated from each operational item pool: a minimum-length easy examination, a maximum-length examination of borderline difficulty, and a minimum-length difficult examination. These three simulated examinations represent examinations for low ability failing candidates, borderline ability failing candidates and high ability passing candidates, respectively. Because the items for these examinations came from different difficulty strata of the item pool, it is unlikely that there would be overlapping items across these examinations. These examinations are chosen as representative samples of the operational items to be deployed.

Due to the limitation of most readability formulae on short passages (Fry, 1990), examination items are modified prior to readability assessment. Only complete sentences are included in the analysis. When a sentence is spread across the item's stem and distractors, the beginning fragment from the stem is excluded from analysis. Instead, the beginning fragment is repeated with each distractor to form complete sentences. When the distractors are phrases that do not complete a sentence that began in the stem, distractors are excluded from analysis. However, subsequent complete sentences in the distractors are included in the readability assessment. In addition to these modifications, all technical vocabulary, numerals, and measurements are retained in the analysis. Tables, charts and graphs, on the other hand, are excluded.

From 1993 through 2003, NCSBN used the Fry Readability Index (FRI: Fry, 1968) to evaluate readability of NCLEX examination items. The FRI assesses readability using sentence length and the number of syllables per word, based on the rationale that longer sentences with polysyllabic words are generally higher in readability (require more reading ability) than shorter sentences with words consisting of few syllables. The FRI method is designed for evaluating readability of passages. It is based upon the average number of syllables and sentence length from three randomly selected 100-word segments within the passage. These averages can be plotted on a graph to produce a grade-level readability estimate.

Due to the limitation in passage length of examination items, NCSBN used a modified version of the Fry readability procedure. Instead of selecting three 100-word samples, the number of syllables and sentences for each simulated examination was counted. Since the number of words in the simulated examinations is considerably greater than 100 words, the counts obtained are divided by the number of words and then multiplied by 100 to make them comparables to those in Fry's original procedure (O'Neill, 2004).

Similar to the Fry method, the Lexile® framework (Zakaluk & Samuels, 1995) also considers readability as a function of sentence length and word difficulty. Instead of using number of syllables, however, the Lexile uses word familiarity (frequency with which the word is used in the English language) as indicator of word difficulty. The Lexile formula is based on two assumptions: the passage is easier to read if it contains more familiar words; and the passage is easier to read if it contains shorter sentences. The Lexile system estimates readability based solely on these two factors. Word familiarity is estimated based on the frequency of words in the Lexile corpus of approximately 600 million words taken from more than 37,000 texts (MetaMetrics, 2007). The Lexile Analyzer measures readability of passages ranging from 200 to over 1700 Lexiles. Reading rulers are available for finding corresponding Lexile and grade levels (Fry, 2002).

Starting in 2003, NCSBN used the Lexile framework, in addition to the FRI, for estimating examination items readability. Currently, NCSBN's policy regarding readability states that readability level of the NCLEX-Practical/Vocational Nurse item pool should not exceed 1200 Lexiles (with a corresponding grade level on the FRI scale); and readability level of the NCLEX-Registered Nurse item pool should not exceed 1300 Lexiles (with a corresponding grade level on the FRI scale). It is of interest to note that more than 150 trade and textbook publishers analyze readability of their publications using the Lexile framework (MetaMetrics, 2008). Results of these readability analyses revealed that many nursing textbooks are more difficult to read than NCLEX items (O'Neill, 2004).

As the U.S. workforce becomes increasingly globalized, foreign candidates seeking licensure in the U.S. make up an appreciable portion of the overall candidate volume for many examination programs. For example, internationally educated candidates constituted 27% of NCLEX-RN candidate volume in 2007. While it is essential for all candidates to possess basic language mastery for safe and effective practice, subtle differences between reading ability of foreign and domestic candidates may exist and influence candidates' performance on examinations. With

this in mind, it may be prudent to include readability analysis as part of the routine analyses.

Currently, readability analysis is not widely conducted among licensure testing programs. Some testing professionals argue that readability analysis could add value to examination programs, but is not essential for examination development (e.g., Werner, 1992). The reluctance of testing programs to use a readability formula may be due to the difficulties in applying the formulae to multiple-choice items. The authors hope that the example of NCLEX readability analysis will provide some insight in applying readability formulae to examination items. With the advent of technology and faster computer processing speed, many commercially available readability analytic programs can automate the process of evaluating reading levels of examinations. It may be feasible to include readability analysis as part of routine examination development without posing a great burden on resources.

References

Christopher, P. P., Foti, M. E., Roy-Bujnowski, K., & Appelbaum, P. S. (2007). Consent form readability and educational levels of potential participants in mental health research. *Psychiatric Services (2)*, 58, 227-232.

Clauser, B. E., Margolis, M J., & Case, S. M. (2006). Testing for licensure and certification in the professions. In R. L. Brennan (Ed.), *Educational Measurement* (pp. 701 – 732). Westport, CT: Praeger Publishers.

Collins, N., Novotny, N. L., & Light, A. (2006). A cross-section of read ability of health information portability and accountability act authorizations required with health care research. *Journal of Allied Health (4)*, 35, 223-225.

Fry, E. (1968). A readability formula that saves time. *Journal of Reading (8)*, 12, 639-646.

Fry, E. (1990). A readability formula for short passages. *Journal of Reading (8)*, 33, 594-597.

Fry, E. (2002). Readability versus leveling. *The Reading Teacher (3)*, 56, 286-288.

Fuller, F., Horlen, C., Cisneros, R., & Merz, T. (2007). Pharmacy students' reading ability and the readability of required reading materials. *American Journal of Pharmaceutical Education (6)*, 71, 1-6.

Gray, W. B., Jr. (1975). *How to measure readability.* Philadelphia: Dorrance and Company.

MetaMetrics (2007). *The Lexile Framework: A MetaMetrics White Paper.* Retrieved December 18, 2008, from http://www.sd34research.com/lexilewhitepaper.pdf

MetaMetrics (2008). *Workman Publishing Adopts Lexile Measures: MetaMetrics Collaborates with Respected Publisher to Measure Titles Using the Lexile Framework for Reading.* Retrieved December 18, 2008, from http://www.lexile.com/pdf/11-5-08.pdf

National Council of State Boards of Nursing (2008). *2008 NCLEX® Examination Candidate Bulletin.* Chicago: NCSBN.

O'Neill, T. R. (2004). The readability of NCLEX® examinations. Retrieved December 17, 2008, from https://www.ncsbn.org/NCLEX_Readability_Tech_Brief.pdf

Walters, K. A., & Hamrell, M. R. (2008). Consent forms, lower reading levels, and using Flesch-Kincaid readability software. *Drug Information Journal (4)*, 42, 385-394.

Werner, E. (1992). The answer key. *CLEAR Exam Review (2)*, 3, 7-9.

Wright, B. D., & Stenner, A. J. (1999). One fish, two fish: Rasch measures reading best. *Popular Measurement (1)*, 34-38.

Zakaluk, B. L., & Samuels, S. J. (Eds.) (1995). *The Lexile® Framework for Reading.* Durham, NC: MetaMetrics.