Technical Summary for Concept-Based Assessments

Psychometrics

Table of Contents

Introduction

Assessment Technologies Institute (ATI) has developed a set of four standardized, nationally benchmarked concept-based assessments to address the current and evolving needs of nursing schools who host a concept-based curriculum (CBC). The CBC assessments feature 46 concepts and 245 exemplars commonly shared among concept-based curricula across the U.S. Because of their standardization, these assessments can be used to compare learning outcomes across cohorts and programs. The assessments are designed to be administered at four points, or levels, in a nursing program and provide robust reports for various outcomes, which are essential for state board approval and accreditation. Each set of level assessments includes online practice and proctored assessments supported by focused remediation for the concepts and exemplars that a student answered incorrectly.

The CBC assessments include four assessment levels which should be assessed, respectively, at the following time points in the CBC program.

- At the 25% completion point in the program
- At the midpoint of the program
- At the 75% completion point in the program
- At the end of the CBC program

The number of scored items, pretest items and total items as well as the total testing time for each level of the CBC assessments are provided in Table 1.

TABLE 1 **Number of Items and Testing Time by CBC Assessment Level**

LEVEL (IN MINUTES)	TOTAL ITEMS	SCORED ITEMS	PRETEST ITEMS	TOTAL TIME
RN Concept-Based Assessment Level 1	130	100	30	130
RN Concept-Based Assessment Level 2	140	110	30	140
RN Concept-Based Assessment Level 3	140	110	30	140
RN Concept-Based Assessment Level 4	140	110	30	140

Technical Characteristics

This report discusses the technical characteristics of the Concept-Based Curriculum assessments (CBC). This includes topics such as item calibration and equating, reliability, validity, and speededness. Several of these sections involve evaluation of administrative data. The first section describes the data used throughout this report.

DESCRIPTION OF ADMINISTRATION DATA

Student and item data collected between July 1, 2018 and March 7, 2021 were used in the analysis for this report. The total sample sizes and demographic characteristics of the analysis sample are shown in Table 2. As displayed in Table 2, a majority of the students taking the CBC assessments identify as Caucasian/White, female, are enrolled in a BSN or ADN nursing program, and speak English as their primary language. Geographical and age-related information are also displayed in Figures 1 and 2, respectively. The information presented in these figures is based on the combined data across all CBC assessment levels because the variation between the assessment levels was minimal.

TABLE 2 **Demographic Profile for Examinees Taking the CBC Assessments by Level and across All Levels**

	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4	ALL LEVELS
ETHNICITY*					
African American/Black	10.0%	12.0%	9.0%	8.0%	10.0%
Asian	4.0%	3.0%	3.0%	3.0%	3.0%
Caucasian/White	63.0%	59.0%	64.0%	64.0%	62.0%
Hispanic	12.0%	13.0%	12.0%	12.0%	12.0%
Native American	1.0%	1.0%	1.0%	1.0%	1.0%
Other	1.0%	1.0%	1.0%	1.0%	1.0%
GENDER*					
Female	82.0%	81.0%	81.0%	81.0%	81.0%
Male	13.0%	14.0%	14.0%	14.0%	13.0%
PROGRAM TYPE					
ADN	65.0%	70.0%	67.0%	51.0%	65.0%
BSN	30.0%	28.0%	30.0%	46.0%	32.0%
Diploma	5.0%	1.0%	3.0%	3.0%	3.0%
PN	0.0%	0.0%	0.0%	--	0.0%

TABLE 2 **Demographic Profile for Examinees Taking the CBC Assessments by Level and across All Levels** *(continued)*

	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4	ALL LEVELS
LANGUAGE*					
English	91.0%	91.0%	92.0%	92.0%	91.0%
French	0.0%	0.0%	0.0%	0.0%	0.0%
Spanish	1.0%	1.0%	1.0%	1.0%	1.0%
Other	1.0%	1.0%	1.0%	1.0%	1.0%
Total Sample Size	16,392	11,650	11,666	6,635	46,343

*The percentages in these categories do not sum to 100% because some students chose not to disclose this information.

FIGURE 1 **Geographic Profile for Examinees Taking the CBC Assessments (All levels)**

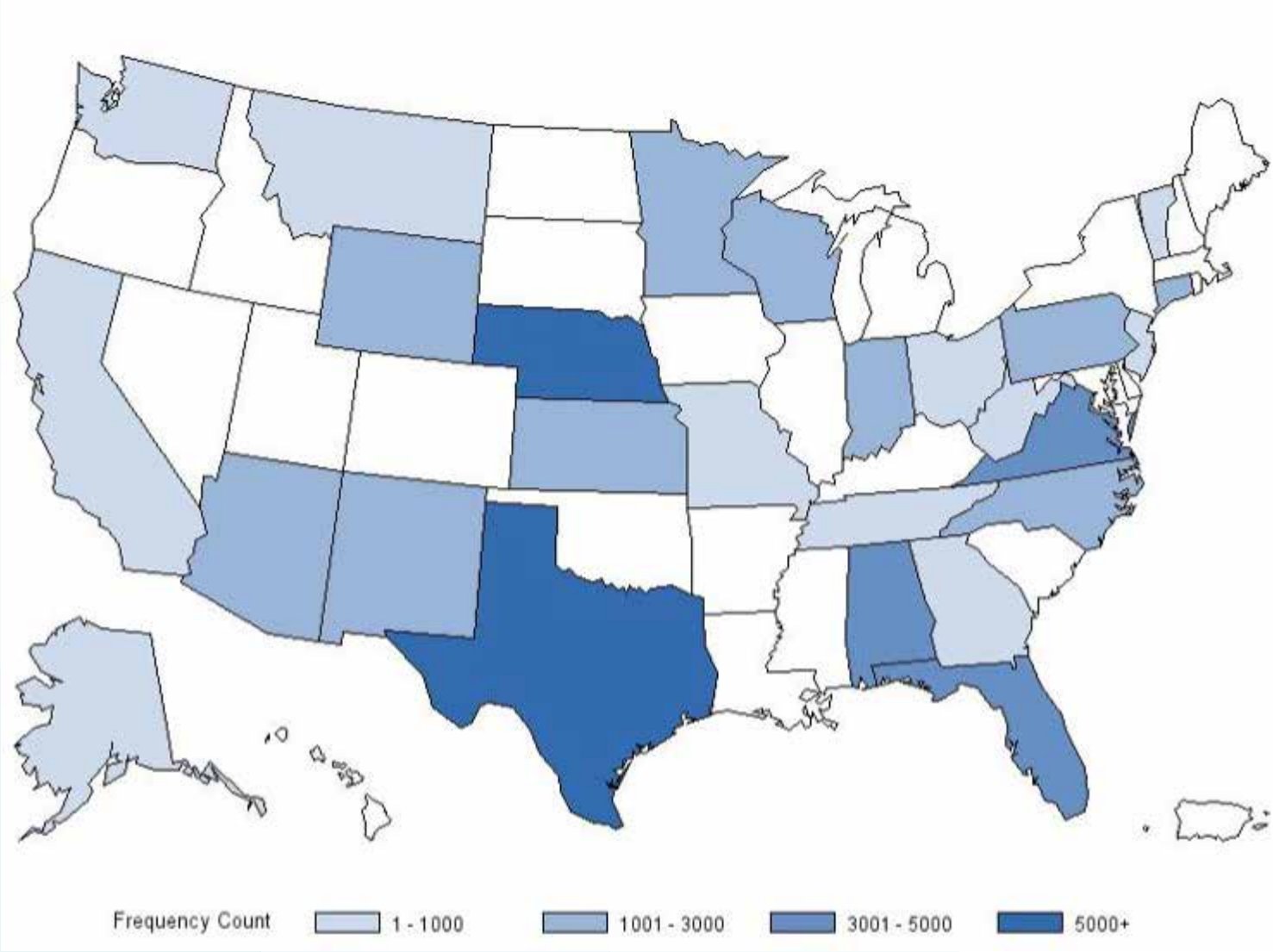
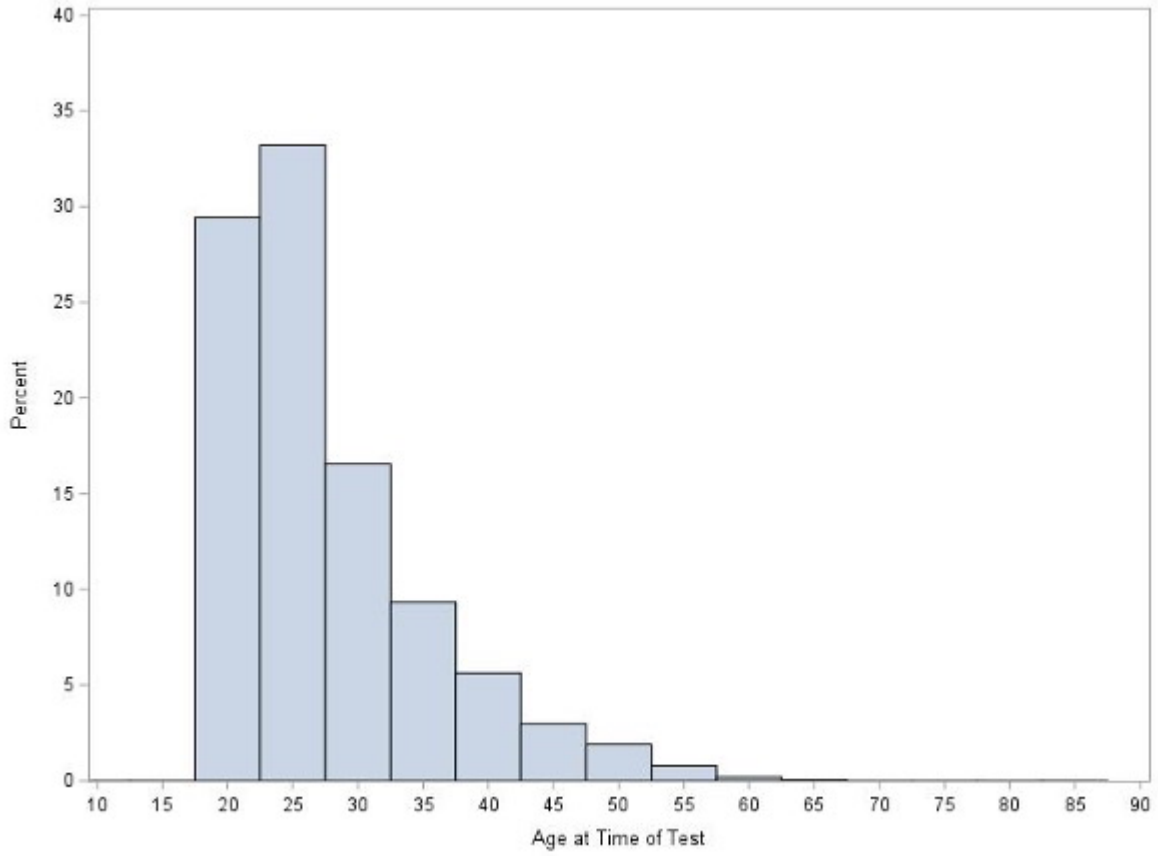


FIGURE 2 Age Profile for Examinees Taking the CBC Assessments (All levels)



ITEM CALIBRATION AND TEST EQUATING

Standard 5.13 – When claims of form-to-form equivalence are based on equating procedures, detailed technical information should be provided on the method by which equating functions were established and on the accuracy of equating functions.

Standard 5.15 – In equating studies that employ an anchor test design, the characteristics of the anchor test and its similarity to the forms being equated should be presented, including both content specifications and empirically determined relationships among test scores. If anchor items are used in the equating study, the representativeness and psychometric characteristics of anchor items should be presented.

Generally, item calibration is the process of assigning a difficulty-parameter estimate to each item on an assessment so that all items are placed onto a common scale. The one parameter logistic item response theory (1-PL IRT) model—also known as the Rasch rating scale model—was used to calibrate the items comprising the CBC assessments (Rasch, 1960; Wright & Panchapakesan, 1969). Item response theory (IRT) has several advantages over classical test theory, so it has become a common procedure for analyzing item response data in applied testing programs. However, IRT models make a number of strong assumptions related to dimensionality, local independence, and model-data fit. Resulting inferences derived from any application of IRT rests strongly on the degree to which the underlying assumptions are met.

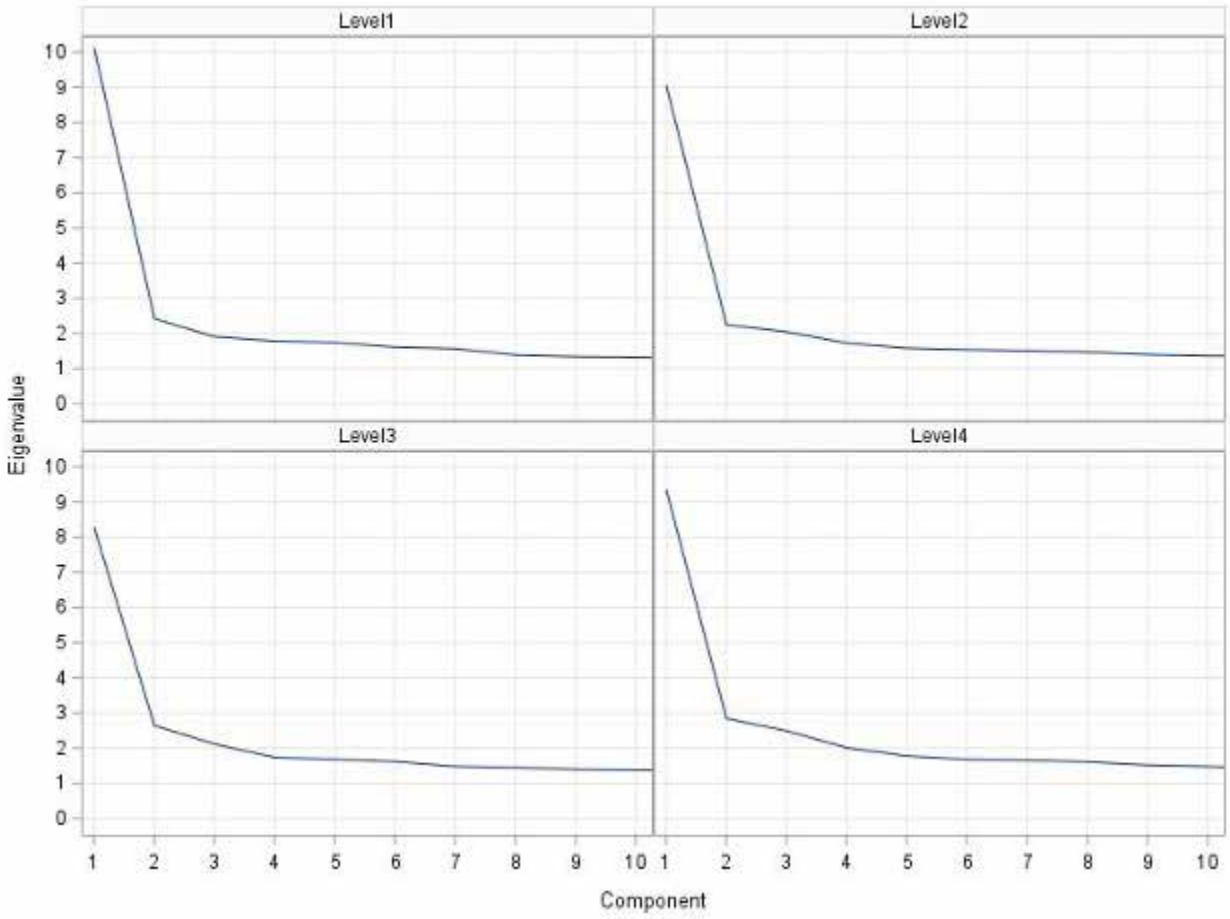
This section briefly introduces the Rasch model, reports the results from evaluations of the adequacy of the Rasch assumptions, summarizes Rasch item statistics, and briefly describes the process of test equating.

Model assumptions. This section evaluates the dimensionality of the data, local item independence, and item fit. It should be noted that only operational items were analyzed because they are the basis of student scores.

Unidimensionality. Rasch models assume that one dominant dimension determines the difference among students' performances. Principal components analysis (PCA) can be used to assess the unidimensionality assumption. The purpose of the analysis is to verify whether any other dominant components exist among the items. If other dimensions are found, the unidimensionality assumption would be violated.

A scree plot for a typical set of items for each of the CBC assessment levels is displayed in Figure 3. Based on the eigenvalue patterns for each of the components, it is apparent that there is one dominant dimension for each of the CBC assessment levels.

FIGURE 3 **Scree Plots of Eigenvalues from Principal Components Analysis by CBC Assessment Level**



Local independence (LI). Another assumption of the Rasch model is that of local independence of items. This means that the probability of a correct response to any item is independent of the responses to other items after controlling for ability level. As indicated in the item development section, the first step taken to ensure local independence is to evaluate all items for enemy status based on content. In order to mitigate the chance of any issues with local dependence in forms administered to candidates, enemy pairs are kept off of the same examination forms through the form development process.

As an additional check of the local independence assumption, residual correlations between all item pairs are evaluated. The residual correlation among item pairs essentially corresponds to Yen's Q3 index, a popular LI statistic. Many critical value standards for Q3 have been proposed in the literature and are used in practice (Christensen, Makransky & Horton, 2017). One of the most common is to examine any item pairs with residual correlations greater than 0.2 in absolute value (Chen & Thissen, 1997).

Table 3 summarizes this analysis for each of the CBC assessment levels. The vast majority of the item pairs had weak negative correlations, suggesting that local item independence generally holds for these assessments.

TABLE 3 Summary of Item Residual Correlations to Evaluate Local Independence of Items by CBC Assessment Level

CBC LEVEL	N	MEAN	SD	MIN	MAX	GREATER THAN 0.2
Level 1	4,950	-0.01	0.02	-0.18	0.37	1
Level 2	5,995	-0.01	0.02	-0.09	0.34	4
Level 3	5,995	-0.01	0.02	-0.08	0.41	1
Level 4	5,995	-0.01	0.02	-0.16	0.21	1

Item fit. Another way to assess appropriateness of the model is through item fit statistics. Two common fit statistics are the infit and outfit mean square statistics, which are oriented toward practical significance (Linacre, 2009). Both statistics have an expected value of 1.0 and range from 0.0 to infinity. Values greater than 1.0 indicate lack of fit between the data and the model. Values less than 1.0 indicate overfit between the data and the model. Although there are many opinions about what values should cause concern, it is reasonable to focus attention of items with mean square values greater than 2.0, as this is the level where items begin to degrade the measurement model (Wright & Linacre, n.d.).

Table 4 presents the summary statistics of the infit and outfit mean square statistics, including the mean, standard deviation, minimum, maximum, and number of items with values greater than 2.0. The mean values for both fit statistics were close to 1.00 for all CBC assessment levels. The majority of the items had infit and outfit values less than 2.0. Overall, these results indicate that the Rasch model fits the item data well for the CBC assessments.

TABLE 4 Summary of Item Fit Statistics by CBC Assessment Level

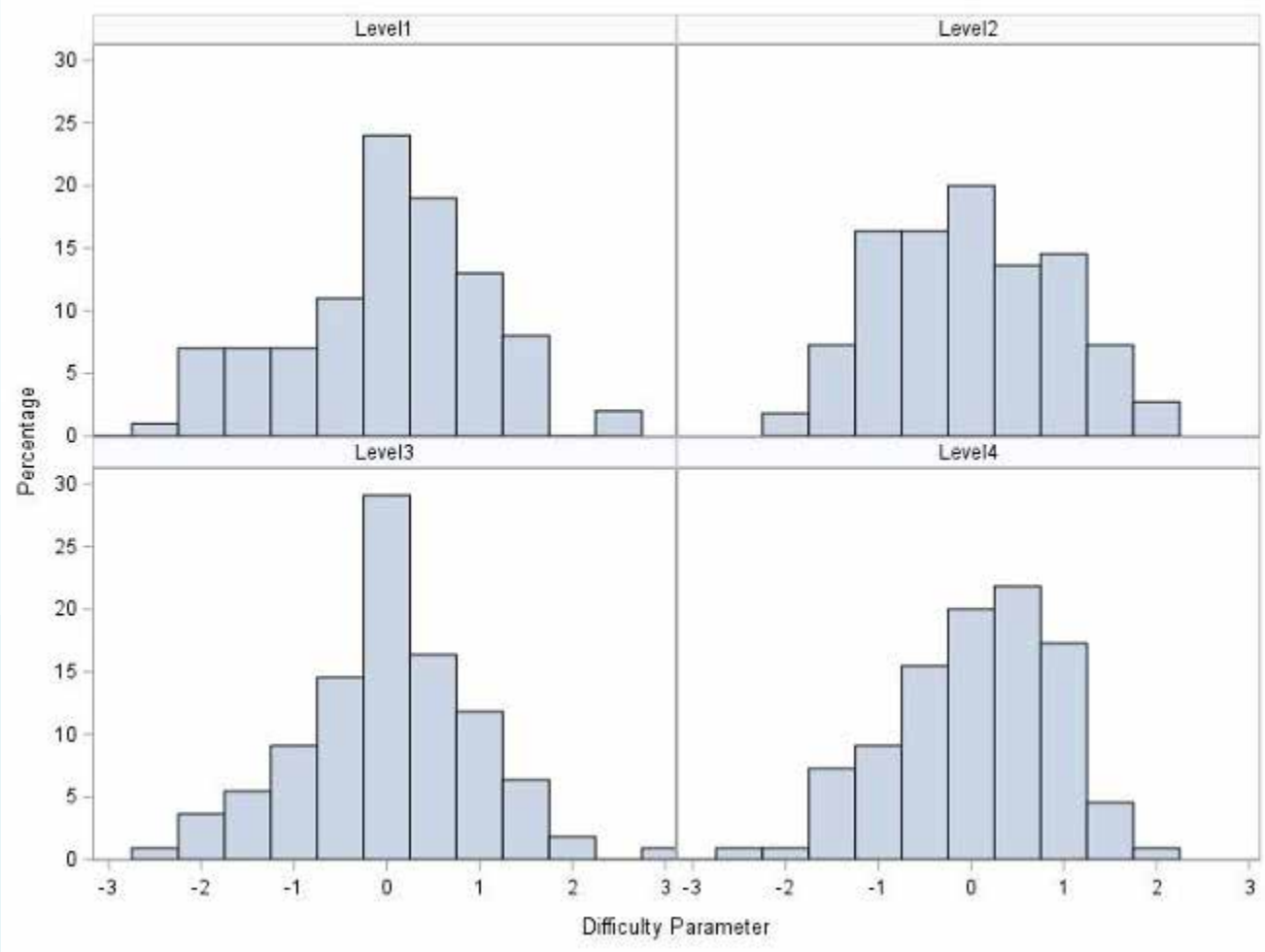
		INFIT					OUTFIT				
CBC LEVEL	N_ ITEMS	MEAN	SD	MIN	MAX	GREATER THAN 2.0	MEAN	SD	MIN	MAX	GREATER THAN 2.0
Level 1	100	1.00	0.23	0.00	1.90	0	0.99	0.24	0.00	1.85	0
Level 2	110	1.02	0.14	0.45	1.47	0	1.01	0.16	0.42	1.51	0
Level 3	110	1.06	0.19	0.52	2.09	1	1.06	0.21	0.44	2.11	2
Level 4	110	1.00	0.14	0.63	1.40	0	1.00	0.17	0.59	1.49	0

Item calibration. To ensure comparable scores for candidates across sets of items, total scores for the CBC assessments must be equated to a base set of items. ATI uses an embedded pretest design. This means that the proctored set of items (scored items) serve as the anchor set to calibrate and anchor pretest items to the proctored assessment IRT scale.

For each pretest set rotation, the proctored set of items serves as the anchor block. Anchor block equating designs rest in part on the assumption that the items comprising the anchor block are representative of the total test in terms of content and statistical properties (Kolen and Brennan, 2014). As the entire set of proctored items is used as the anchor block, the anchor set for the pre-equated design is a precise representation of the content and statistical specifications of the assessment. Consequently, the anchor block was deemed sufficiently representative of the total test for calibration of the pretest item sets to proceed with equating.

To assess the stability of item parameter estimates, IRT displacement statistics are calculated for each proctored item each time a pretest set is rotated for analysis. Items flagged with displacement values greater than or equal to 1.0 in absolute value are removed from the anchor set for the purpose of anchoring the pretest item parameter estimates to the equating statistical scale. The pretest item measures are estimated in the frame of reference defined by the anchor set, and the item pool is then updated with the statistics for those items. The Rasch model expresses item difficulty in logits rather than the percent-correct metric. Large negative logits represent easier items, while large positive logits represent more difficult items. Figure 4 shows the distribution of the Rasch difficulty parameters for each of the CBC assessment levels. The items cover a wide range of difficulty levels. This is important because it ensures that the test can be used to assess a wide range of ability levels.

FIGURE 4 **Histogram of Item Difficulty Parameters by CBC Assessment Level**



Test equating. To ensure that scores for different sets of items are comparable, the content balance of each group of items is carefully matched to a test blueprint to ensure all students taking the test are demonstrating their knowledge of the relevant content for the measure. Test developers assembled sets of successfully pretested items to strictly align to the test content specifications. Items were also selected using statistical specifications, which ensured that the differences in difficulty across sets of items were minimal.

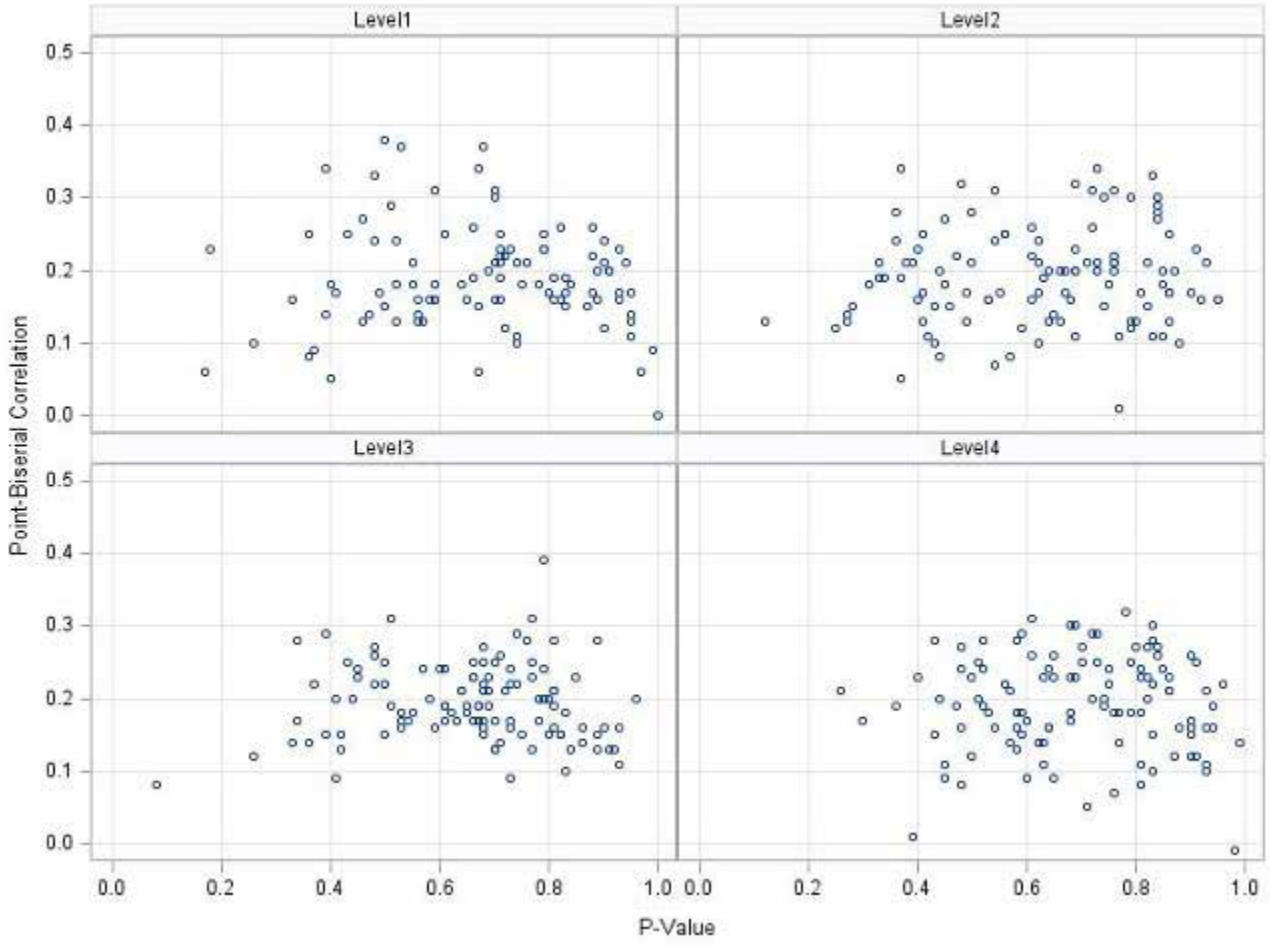
To adjust for any remaining differences in difficulty across sets of items, a statistical equating procedure was used. Once the full set of proctored items was identified, psychometricians pulled the pre-equated item parameter estimates and created the final scoring tables through the process of mapping the test characteristic curves from the newly assembled form to the test characteristic curve for the base form (Hambleton, Swaminathan & Rogers, 1991). Using this information, the adjusted percent correct table for the new proctored set of items was created.

This resulted in a set of fully comparable scores between the two sets of items. ATI adopted this pre-equating design due to the need for scoring tables prior to test administration.

TEST ITEM ANALYSES

Item difficulty and discrimination. The distributions of item difficulty (p-value) and discrimination (pbs correlation) statistics are shown in Figure 5. The difficulty of an item corresponds to the proportion of students who correctly answered an item—the higher the difficulty value index, the easier the item. The discrimination index represents the pbs correlation, or the Pearson product-moment correlation between the dichotomous score on an individual item (correct or incorrect) and a student's scaled score. The higher the discrimination index, the more the item differentiates or discriminates between upper- and lower-ability examinees.

FIGURE 5 **P-Value and Point-Biserial Correlations for Items by CBC Assessment Level**



RELIABILITY ANALYSIS

Standard 2.0 – Appropriate evidence of reliability/precision should be provided for the interpretation for each intended score use.

Standard 2.13 – The standard error of measurement, both overall and conditional (if reported), should be provided in units of each reported score.

The reliability coefficients and standard errors of measurement for the total score on each of the CBC assessment levels are provided in Table 5. The standard errors of measurement are reported in the percent correct metric. Results show a fairly high reliability index around 0.94 or 0.95 at the total score level for each of the tests, which suggests that the total score is reliable for the population.

TABLE 5 Number of Items, Reliability Coefficients, and Standard Errors of Measurement for Total Scores by CBC Assessment Level

CBC LEVEL	NUMBER OF ITEMS	RELIABILITY	STANDARD ERROR OF MEASUREMENT
Level 1	100	0.94	2.01
Level 2	110	0.95	1.86
Level 3	110	0.95	1.92
Level 4	110	0.95	1.88

VALIDITY

As defined in the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014), validity refers to, "the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests" (p. 9). The validity process involves the collection of a variety of evidence to support the proposed test score interpretations and uses. This technical report describes the technical aspects of the CBC assessments in support of their score interpretations and uses. Each of the previous sections contributes important evidence components that pertain to score validation: test development, test scoring, item analysis, Rasch calibration, equating, and reliability. This section summarizes and synthesizes the evidence based on the framework presented in the Standards.

Evidence based on test content. Content validity addresses whether the test adequately samples the relevant material it purports to cover. Test content validity of the CBC assessments rests greatly on establishing a link between each piece of the assessment (i.e., the items) and what the students should know and be able to do as required by the test plan. The CBC assessments are criterion-referenced assessments. The set of concepts and exemplars for the CBC curriculum and test plan were developed through interviews, multiple rounds of review meetings, independent reviews and validation survey (ProExam, 2016).

For the CBC assessments, strong content validity evidence is derived directly from the test construction process. Each item was based on and was directly aligned to the CBC test plan to ensure good content validity. The item development and test construction process ensured that every item aligned directly to one, and only one, content category.

The Test Development team began by establishing detailed test and item development specification, ensuring that the items were sufficient in number and were adequately distributed across content domains and levels of cognitive complexity and difficulty. ATI selected qualified item writers and provided training to help ensure that high-quality items were written. As a routine part of the item development review committees checked the alignment of the items with the test plan, the accuracy of the item content, and made any adjustments as necessary. Items were also submitted to bias review for issues related to diversity, gender, and other pertinent factors. Only items that passed all the prior checks were tried out in a pretesting event.

Several statistical analyses were conducted on the pretesting data. Items flagged by the statistical criteria were sent to content specialists for further evaluation. The result is consensus among the content specialists that the assessment does in fact assess what was intended.

Evidence based on internal structure. As described in the Standards (2014), internal-structure evidence refers to the degree to which the relationships between test items and test components conform to the construct on which the proposed test interpretations are based.

Item-test correlations. Item-test correlations are reviewed in Figure 5. All values are positive and of acceptable magnitude.

Item response theory dimensionality. Results from principal components analyses are presented in Figure 3. Each of the CBC assessment level tests are essentially unidimensional, providing evidence supporting interpretations based on the total scores for the test.

Evidence related to the use of the Rasch model. Because the Rasch model is the basis of all calibration and equating analyses associated with the CBC assessments, the validity of the inferences from these results depends on the degree to which the assumptions of the model are met as well as the fit between the model and test data. The underlying assumptions of Rasch models were essentially met for all of the CBC assessment data, indicating the appropriateness of using the Rasch models to analyze the data.

In addition, the Rasch model was used to link different operational sets of the test.

The accuracy of the linking also affects the accuracy of student scores and the validity of score uses. The Ascend psychometric staff conducted verifications to check the accuracy of the procedures, including item calibration, and conversions from the raw score to the Rasch ability estimate.

TEST SPEEDEDNESS

Standard 4.14 – For a test that has a time limit, test development research should examine the degree to which scores include a speed component and evaluate the appropriateness of that component, given the domain the test is designed to measure.

The CBC assessments are designed to measure the knowledge of an examinee without regard to response speed. According to Lu and Sireci (2007), "When speededness is unintended, it introduces construct-irrelevant variance into the test scores and thus changes the construct the test intends to measure" (p. 31). The presence of test speededness can undermine the test reliability and validity because a portion of the examinees' scores is not solely a result of their ability. The analyses described below were conducted to verify that speededness was not a significant source of construct irrelevant variance on the CBC assessments.

Swineford (1974) presented a rule stating that if 80% of students answer the last item and all students answer at least 75% of the items, then the test can be considered unspeeded. As seen from Table 6, almost all of the CBC assessments appear to have met this standard. Stafford (1971) proposed a speededness quotient (SQ) based on a simple ratio of the number of unreached items to total number of incorrect items, to include wrong, unreached, and omitted items. A purely speeded test would have an SQ of 1.0. Table 6 shows the CBC assessments have a low SQ, indicating that the proportion of total errors due to speededness is quite low. The Gulliksen (1950) formula compares the standard deviation of the number of unreached items to the standard deviation of total number of incorrect items, with ratios less than 0.25 considered indicative of an unspeeded test (Swineford, 1974). The results from all three methods indicate a very low likelihood that student CBC assessment scores are affected in a meaningful way by the amount of time allowed.

TABLE 6 **Test Speededness Analyses by CBC Assessment Level**

CBC LEVEL	% OF STUDENTS ANSWERING THE LAST ITEM	% OF STUDENTS ANSWERING AT LEAST 75% OF ITEMS	SPEEDEDNESS QUOTIENT	STD (UNREACHED) / STD (INCORRECT)
Level 1	98.63%	99.98%	0.01	0.40
Level 2	98.81%	99.99%	0.00	0.23
Level 3	99.44%	99.99%	0.00	0.16
Level 4	99.64%	99.98%	0.00	0.21

Interpretation of Scores

Standard 1.2 – A rationale should be presented for each intended interpretation of test scores for a given use, together with a summary of the evidence and theory bearing on the intended interpretation.

Standard 5.1 – Test users should be provided with clear explanations of the characteristics, meaning, and intended interpretation of derived scale scores, as well as their limitations.

SCORE INTERPRETATION

ATI provides two ways to interpret the CBC assessment test scores. Norm-referenced data provide examinees and institutions with the ability to compare their score with others taking the exam nationwide. ATI provides normative data for the CBC assessments on the score reports—namely, means and percentile ranks. Criterion-referenced data provide examinees with the ability to compare their performance with the actual test objectives and not with others who have taken the exam. The ATI proficiency levels (as discussed later in this section) are an example of a criterion-referenced interpretation of a test score.

SCORE REPORTING

All items on the CBC assessments are scored as correct or incorrect, with no partial credit awarded on any item and no penalty for guessing. To ensure the total test scores are comparable from one testing experience to another, ATI carefully controls two characteristics of each group of items students are given. First, the content balance of each set of items is carefully matched to a test blueprint to ensure all students are being tested on the relevant content for the measure. Second, the total score is equated to adjust for slight differences that might exist across different groups of items making up a test. The equating adjustment makes sure that students are not unfairly advantaged or disadvantaged based on the particular group of items they are administered relative to other students. A reported total score is a comprehensive description of student performance on the whole test. This score, referred to as the Adjusted Individual Total Score on the score report, is provided along with both norm-referenced data (means and percentile ranks) and criterion-referenced data (proficiency levels), each of which will be discussed in further detail later in this chapter.

Based on the data collected from July 1, 2018 to March 7, 2021, the student population Adjusted Individual Total Scores are approximately normally distributed for each of the CBC assessment levels (see Figure 6). Table 7 provides the mean, standard deviation, minimum, maximum and quartiles of the Adjusted Individual Total Score by CBC assessment level.

FIGURE 6 **Score Distribution by CBC Assessment Level**

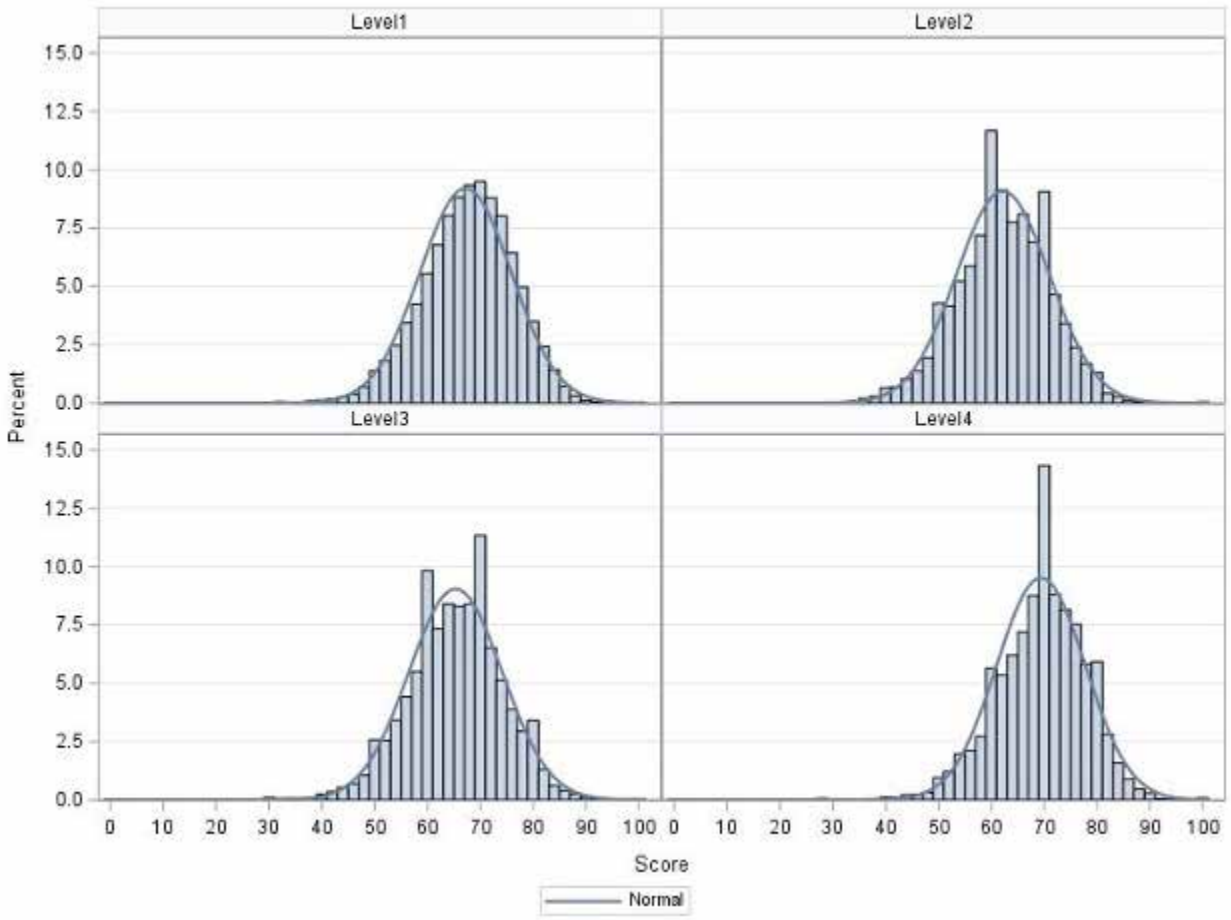


TABLE 7 **Descriptive Statistics of the Adjusted Individual Total Scores by CBC Assessment Level**

CBC LEVEL	N	MEAN	SD	MIN	Q1	MEDIAN	Q3	MAX
Level 1	16,392	67.2	8.7	0.0	62.0	68.0	73.0	100.0
Level 2	11,650	62.2	8.8	0.0	56.4	62.7	68.2	100.0
Level 3	11,666	65.3	8.8	0.0	60.0	65.5	70.9	100.0
Level 4	6,635	69.4	8.4	1.8	64.5	70.0	75.5	100.0

REPORTED NORMATIVE DATA

Standard 5.8 – Norms, if used, should refer to clearly described populations. These populations should include individuals or groups with whom test users will ordinarily wish to compare their own examinees.

Standard 5.9 – Reports of norming studies should include precise specification of the population that was sampled, sampling procedures and participation rates, any weighting of the sample, the dates of testing, and descriptive statistics. Technical documentation should indicate the precision of the norms themselves.

Standard 5.11 – If a test publisher provides norms for use in test score interpretation, then as long as the test remains in print, it is the publisher's responsibility to renorm the test with sufficient frequency to permit continued accurate and appropriate score interpretations.

Normative comparisons are not the primary purpose of the CBC assessments. However, it is important that examinees and programs be able to assess their relative standing. For comparative purposes, ATI provides national mean and percentile rank data for all students. The national mean adjusted percent correct scores are reported on student and group score reports along with percentile rank information. The norm tables by CBC assessment level are available on the ATI website under the faculty landing page, integration resources section .

"National mean" is the average percent correct score of all individual test takers in a national sample of RN program students. The way to interpret "national percentile rank" is to think of it as a comparison to the national sample of RN students. For example, if your national percentile rank is 32, then you have scored the same as or better than 32% of the RN students nationwide who also took the test.

PROFICIENCY LEVELS

Standard 5.21 – When proposed score interpretations involve one or more cut scores, the rationale and procedures use for establishing cut scores should be documented clearly.

To provide a numerical indication of a student's proficiency level, ATI developed a set of recommended criterion-referenced proficiency levels that nursing schools could choose to adopt as benchmarks for student performance on each of the four CBC assessment levels. These cut scores were established through the RN Concept-Based Curriculum Assessments National Standard Setting Studies.

Each of the CBC assessments has four proficiency levels. The descriptions of the proficiency levels can be found in Table 8. The score ranges for each proficiency level on each of the CBC assessment levels can be found in Table 9, and the percentage of students in each proficiency level is reported in Table 10.

Nursing education institutions should recognize that the cut scores are recommended (rather than required) for those using the CBC assessments. Institutions using the CBC assessments for high-stakes purposes (e.g., a portion of course grade or progression) should be extremely cautious, regardless of whether the ATI recommended cut scores are adopted. If these assessments are used for a portion of a course grade, institutions should ensure that the content of the assessments is fully aligned with program curriculum. ATI strongly discourages the use of the CBC assessments as a sole criterion for progression or graduation.

TABLE 8 Criterion–Referenced Proficiency Level Definitions for CBC Assessments

BELOW LEVEL 1

A student not meeting the criterion for Proficiency Level 1

- Does not yet demonstrate proficiency in the knowledge and skills related to the assessed concepts and exemplars required to deliver safe, effective nursing care at the entry-level for the profession.
- Needs substantial academic support in the assessed concepts and exemplars to ensure successful performance in subsequent curricular material.

LEVEL 1

A student meeting the criterion for Proficiency Level 1

- Demonstrates a partial proficiency in the knowledge and skills related to the assessed concepts and exemplars required to deliver safe, effective nursing care at the entry-level for the profession.
- Needs additional academic support in the assessed concepts and exemplars to ensure successful performance in subsequent curricular material.

LEVEL 2

A student meeting the criterion for Proficiency Level 2

- Demonstrates a proficiency in the knowledge and skills related to assess concepts and exemplars required to deliver safe, effective nursing care at the entry-level profession.
- Is prepared for successful performance in subsequent curricular material.

LEVEL 3

A student meeting the criterion for Proficiency Level 3

- Demonstrates advanced proficiency in the knowledge and skills related to assess concepts and exemplars required to deliver safe, effective nursing care at the entry-level profession.
- Is well-prepared for successful performance in subsequent curricular material.

TABLE 9 Score Ranges for ATI Recommended Cut Scores by CBC Assessment Level

CBC LEVEL	BELOW LEVEL 1	LEVEL 1	LEVEL 2	LEVEL 3
Level 1	0.0% to 54.0%	55.0% to 67.0%	68.0% to 75.0%	76.0% to 100.0%
Level 2	0.0% to 50.9%	51.8% to 59.1%	60.0% to 68.2%	69.1% to 100.0%
Level 3	0.0% to 51.8%	52.7% to 61.8%	62.7% to 73.6%	74.5% to 100.0%
Level 4	0.0% to 52.7%	53.6% to 63.6%	64.5% to 75.5%	76.4% to 100.0%

TABLE 10 Percentage of Students in Each Proficiency Level by CBC Assessment Level

CBC LEVEL	BELOW LEVEL 1	LEVEL 1	LEVEL 2	LEVEL 3
Level 1	7.57%	41.43%	34.38%	16.62%
Level 2	10.64%	26.10%	39.87%	23.38%
Level 3	6.97%	28.10%	49.75%	15.18%
Level 4	3.33%	20.77%	54.45%	21.45%

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: AERA, APA, NCME.

Chen, W., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*(3), 265–289.

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q3: Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement, 41*(3), 178-194.

Gulliksen, H. (1950). *Theory of mental tests.* New York: John Wiley.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Kolen, M.J. & Brennan, R.L. (2014). Test equating, scaling, and linking: Methods and practices (3rd Ed.). New York, NY: Springer-Verlag.

Linacre, J. M. (2009). *A user's guide to WINSTEPS MININSTEP Rasch-model computer programs.* Chicago, IL: Winsteps.

Lu, Y. & Sireci, S.G. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice 26*, 29-37.

National Council of State Boards of Nursing. (2018a). *Report of findings from the 2017 RN practice analysis: Linking the NCLEX-RN® examination to practice.* https://www.ncsbn.org/17_RN_US_Canada_Practice_Analysis.pdf: Author.

National Council of State Boards of Nursing. (2018b). *2019 NCLEX-RN test plan.* Chicago, IL.

National Council of State Boards of Nursing. (2020). *What the exam looks like.* https://www.ncsbn.org/9010.htm

ProExam. (2016). *Report to ATI education on the development and validation of a test content outline for Concept-Based Curriculum (CBC) assessments.* New York, NY: ProExam.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: Danish Institute for Educational Research.

Stafford, R.E. (1971). The speed quotient: A new descriptive statistic for tests. *Journal of Educational Measurement, 8*, 275-278.

Swineford, F. (1974). *The test analysis manual (SR-74-06).* Princeton, NJ: Educational Testing Service.

Wright, B., & Linacre, J. M. (n.d.). *Reasonable mean-square fit values.* https://www.rasch.org/rmt/rmt83b.htm

Wright, B. & Panchapakesan, N. (1969). *A procedure of sample-free item analysis. Educational and Psychological Measurement, 29*, pp. 23-48.