Technical Manual for the ATI TEAS®

Table of Contents

List of Tables

List of Figures

Executive Summary

The purpose of the ATI TEAS® assessment is to assess an examinee's overall academic preparedness for a health science program. The assessment contains 150 scored items and 20 unscored, pretest items. It is intended for use with adult nursing and allied health program applicant populations. All items on the ATI TEAS are scored as correct or incorrect, with no partial credit awarded on any item and no penalty for guessing. The ATI TEAS is a fixed-length test with individual section time limits (see Table 1) and an overall time limit of 209 minutes. The intended population for the test is students who are applying to or have been recently admitted into a nursing or allied health program.

This technical manual includes topics such as the nature of the test and intended use, content validation and the test development process, equating, scoring, generation of norm tables and score reports, reliability and validity evidence, candidate demographic data, item analyses, and appropriate and inappropriate test score use.

The blueprint for ATI TEAS was developed through a series of meetings that were conducted by Alpine Testing Solutions. The assessment contains 150 items divided among the four content areas of reading, mathematics, science, and English and language usage. Items were written by external content experts on a contract basis and were reviewed by additional content experts, as well as a bias review committee. Items also underwent pilot testing during a pretesting process and were required to meet a set of statistical criteria prior to being used as scored items on the test.

ATI TEAS is available in both paper-pencil and web-based formats. ATI provides clients with extensive guidelines regarding test security and proctoring under standardized conditions, including a detailed proctoring process guide and script (ATI, 2016a). The scores for the test are reported on the same statistical scale as TEAS V and are adjusted to account for any potential differences in terms of difficulty through the equating process. Means and percentile ranks are also provided, along with an academic preparedness category based on the total score. The score report also includes a list of Topics to Review, and a Focused Review is available to examinees.

Introduction

This report is organized by chapter to specifically address the standards cited by the APA, AERA, and NCME Standards for Educational and Psychological Testing (2014) as most critical in validation documentation for a test related to development tasks completed to release the ATI TEAS. According to the Standards:

> Evaluating the acceptability of a test or test application does not rest on the literal satisfaction of every standard in this document, and the acceptability of a test or test application cannot be determined by using a checklist. Specific circumstances affect the importance of individual standards, and individual standards should not be considered in isolation. Therefore, evaluating acceptability depends on (a) professional judgment that is based on a knowledge of behavioral science, psychometrics, and the relevant standards in the professional field to which the test applies; (b) the degree to which the intent of the standard has been satisfied by the test developer and user; (c) the alternative measurement devices that are readily available; (d) research and experiential evidence regarding the feasibility of meeting the standard; and (e) applicable laws and regulations. (AERA, APA, NCME, 2014, p. 7)

Accordingly, authors of this technical report make no claim to meet the Standards as a whole or to address all standards in that publication. Specific standards judged by the report authors to be most relevant to the topic at hand are quoted; however, this does not imply that they are fully met or that unquoted standards are disregarded.

Nature of the Test/Intended Use

Standard 1.1 — The test developer should set forth clearly how test scores are intended to be interpreted and consequently used. The population(s) for which a test is intended should be delimited clearly, and the construct or constructs that the test is intended to assess should be described clearly.

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 use interpretation.

Standard 1.3 — If validity for some common or likely interpretation for a given use has not been evaluated, or if such an interpretation is inconsistent with available evidence, that fact should be made clear and potential users should be strongly cautioned about making unsupported interpretations.

(AERA, APA, NCME, 2014, p. 23)

Standard 12.10 — In educational settings, a decision or characterization that will have major impact on a student should take into consideration not just scores from a single test but other relevant information.

(AERA, APA, NCME, 2014, p. 198)

ATI is responsible for the development, delivery, and scoring of the ATI TEAS, while the client institutions using the test are responsible for setting policy governing the decisions based on the test scores. In order for institutions to properly interpret these scores and make sound decisions, it is first necessary to understand the nature and intended use of the ATI TEAS. The purpose of the ATI TEAS assessment is to assess an examinee's overall academic preparedness for a health science program. It is intended for use with adult nursing and allied health program applicant populations. All construction of the ATI TEAS has been completed to support the assumption that the entire assessment (all four content areas) would be administered to a candidate and that the total score based upon the four content areas would be the score used in making decisions about a candidate's entry into a program. While sub-content area scores are provided, these scores are not equated scores and are not comparable across candidates or administrations. Sub-content area scores can be used to determine content areas where candidates did not perform well and may need additional study.

The ATI TEAS assessment is intended to be used as a tool for admissions. If the ATI TEAS is used to make decisions regarding whom to admit, institutions should use the test results in conjunction with other admission criteria to determine whether or not to admit an applicant. The ATI TEAS should not be used as the sole determining factor as to whether or not an examinee is admitted into a program. Furthermore, it is recommended that schools use the overall score from the ATI TEAS, instead of individual content area scores, when evaluating an applicant.

Test Development Process

Standard 1.11 — When the rationale for test score interpretation for a given use rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified.

(AERA, APA, NCME, 2016, p. 26)

ATI TEAS Blueprint Development Process

The ATI TEAS blueprint was created through a series of activities conducted by Alpine Testing Solutions (hereafter referred to as Alpine) for Assessment Technologies Institute (ATI). The following sections summarize this process.

Content analysis workshop. Alpine staff first conducted a content analysis meeting in July 2014 and provided a full content analysis workshop detailed report summarizing all activities of the content analysis process for the ATI TEAS assessment to ATI (Alpine Testing Solutions, 2014).

Three content workshop steering committees (reading and English and language usage, mathematics, and science) comprised of 16 nurse educators and 18 subject matter experts in the areas of reading, English and language usage, mathematics, and science were asked to produce a list of sub-content areas, objectives, and knowledge and skills and abilities (KSAs) a student needed to be successful in the first year of a nursing program. The subject matter experts were also asked to ensure these objectives and KSAs aligned with the Common Core State Standards for Reading, English and Language Usage, Mathematics, and the Next Generation Science Standards. Each steering committee was trained on the procedures of the proceedings and familiarized with the current definitions of minimally qualified candidates at each of the TEAS academic preparedness categories so they could complete the conversation about the content that needed to be assessed on the TEAS. After a full listing of exhaustive (and mutually exclusive) objectives existed, each was assigned either a foundational or critical thinking level. Finally, the steering committees provided preliminary judgments as to the percentage of critical thinking items that should be on the section of the exam on which they worked. Before adjourning the meeting, the steering committees finalized draft rating scales that would be used with the subsequent content analysis survey and completed a process evaluation form for the proceedings.

Blueprint survey. The second activity of the content validation process was a large-scale blueprint survey. ATI provided Alpine with a list of e-mail addresses for potential survey participants. The intended target population of the survey was individuals in the nursing field who have taught first-year nursing students; respondents not meeting this requirement were excluded from the analysis. The full set of objectives was split into two surveys (Reading and English and Language Usage, Science and Mathematics) so that respondents were not responding to an extremely long survey. Respondents were asked to rate each objective on two Likert scales: the frequency[1] and the importance[2] of each objective. Survey respondents were also given the opportunity to provide objectives they felt should be included, but were omitted from the survey. Recommended weights by content area, sub-content area, and objective were computed from the survey results using

[1] The frequency scale included the following options: at least once a day, at least once per week but not daily, at least once per month but not weekly, not monthly but have performed at least once, and never performed.

[2] The importance scale included the following options: highly important, moderately important, slightly important, not at all important, and never performed.

the mean ratings. Weights were determined using a multiplicative model (Kane, et al., 1989) in which the importance and frequency ratings contributed equally to the overall weights. Any recommendations for modifications to the content areas, sub-content areas, or objectives were discussed at the subsequent content analysis meeting.

Final content analysis meeting. The second content analysis meeting (the final step in the process) was held in November 2014. Only the nurse educator subject matter experts participated in this second meeting as it was determined that these experts were more knowledgeable about the purpose of the exam than non-users and would be the best group to decide upon the final blueprint recommendations. Of the 18 nurse educators invited to the July meeting, 13 participated in the November meeting. This steering committee was asked to, "(1) finalize the list of recommended sub-content areas, objectives, and KSAs; (2) produce a recommendation for test length and the weight of each content area and sub-content area; and (3) produce a recommendation for the percentage of critical thinking items that should be administered within each content area." (Alpine Testing Solutions, 2014, p. 26). Prior to adjourning this meeting, the steering committee provided a final recommendation on the list of content areas, sub-content areas, and KSAs and completed an evaluation of the content analysis meeting process.

The content analysis was completed under the assumption that the entire assessment would be administered to a candidate and that the total score be based upon all four content sections. The next section provides details of the ATI TEAS test content blueprint specifications.

ATI TEAS Blueprint Specifications

Content areas. The ATI TEAS is comprised of four content areas: reading, mathematics, science, and English and language usage. The assessment contains 47 reading, 32 mathematics, 47 science, and 24 English and language usage items. Table 1 displays the breakdown of each content area into sub-content areas and the number of items in each subcategory. As seen in the table, there are 170 total items assessed on the ATI TEAS, which includes 150 scored items and 20 unscored, pretest items.

Reading. The reading content area focuses on the assessment of functional literacy skills. The domain is divided into three main sub-content areas: key ideas and details, craft and structure, and integration of knowledge and ideas. The key ideas and details sub-content area includes items related to the comprehension of reading selections, complex text, printed communications, and graphical representations of information. The sub-content area of craft and structure includes items related to evaluating an author's purpose and point of view, as well as interpreting the meanings of words phrases. Also included are items related to recognizing the structure of texts, distinguishing between fact and opinion, and identifying biases and stereotypes. Finally, the sub-content area of integration of knowledge and ideas includes items assessing a student's ability to compare and contrast themes from different sources and to evaluate an argument and its specific claims. Students will also encounter items requiring them to use evidence from the text to make predictions and inferences and draw conclusions about a piece of writing, as well as integrate data from multiple sources.

The targeted word count for the reading section is between 2,677 and 3,289 with the optimal range being the low or middle part of the range. This includes both scored and pretest items. The pretest word count should be between 430 and 560 words.

Mathematics. The mathematics content area assesses basic mathematical skills. Calculators are allowed on this content area of the ATI TEAS. While calculators are allowed, the questions are written at a level for which a calculator is not necessary to answer the questions. The mathematics content area is divided into two main sub-content areas: numbers and algebra, and measurement and data. Items pertaining to numbers and algebra cover fractions, decimals, percentages, rational and irrational numbers, and the operations of adding, subtracting, multiplying, and dividing. Also included in this sub-content area are items related to translating phrases and sentences into expressions, equations, and inequalities. The sub-content area of measurement and data includes items related to interpreting and evaluating information in tables, charts, and graphs using statistics, as well as explaining the relationship between variables and calculating geometric quantities.

Science. The science content area is divided into three major sub-content areas: human anatomy and physiology, life and physical sciences, and scientific reasoning. Items in the human anatomy and physiology sub-content area relate to describing the anatomy and physiology of a human and specifically the respiratory, cardiovascular, gastrointestinal, neuromuscular, reproductive, integumentary, endocrine, genitourinary, immune, and skeletal systems. Life and physical sciences has items related to describing the basic macromolecule and biological system, as well as comparing and contrasting chromosomes, genes, and DNA. Additionally, items cover Mendel's laws of heredity, basic atomic structure, properties of substances, and chemical reactions. Students are also asked to respond to items comparing and contrasting changes in states of matter. The scientific reasoning sub-content area is comprised of items related to identifying basic scientific measurements using laboratory tools; explaining relationships among events, objects, and processes; and analyzing the design of a scientific investigation, as well as using logic and evidence to critique a scientific explanation.

English and language usage. The English and language usage area is divided into three sub-content areas: conventions of standard English, knowledge of language, and vocabulary acquisition. Conventions of standard English relates to using the conventions of standard English spelling, punctuation, and sentence structure. The sub-content area of knowledge of language includes items related to applying basic knowledge of elements of the writing process, using grammar to enhance clarity in writing, distinguishing between formal and informal language, and developing a well-organized paragraph.

TABLE 1. *ATI TEAS Content Specifications*

Content Area *Sub-content area*	*Scored items* *Percent*	*Number*	*Unscored Items*	*Total Number of Items*	*Time Allowed Per Content Area*
Reading	31%	47	6	53	64 min
Key ideas and details	15%	22			
Craft and structure	9%	14			
Integration of knowledge and ideas	7%	11			
Math	22%	32	4	36	54 min
Numbers and algebra	16%	23			
Measurement and data	6%	9			
Science	31%	47	6	53	63 min
Human anatomy and physiology	21%	32			
Life and physical sciences	5%	8			
Scientific reasoning	5%	7			
English and Language Usage	16%	24	4	28	28 min
Conventions of standard English	6%	9			
Knowledge of language	6%	9			
Vocabulary acquisition	4%	6			
Total	**100%**	**150**	**20**	**170**	**209 min**

Foundational and critical thinking items. The ATI TEAS consists of foundational and critical thinking items. Items requiring foundational thinking are those at the knowledge or comprehension level. These items require the recollection or comprehension of information foundational to the specified content area. Items requiring critical thinking are those which require application or analysis of presented information (i.e., they require problem solving in a given context.) The approximate percentage of each type is listed in Table 2.

Item format. The ATI TEAS item format is multiple-choice with four response options, one of which is identified as the correct response.

TABLE 2. *Approximate Percentage of Items to be Classified as Foundational and Critical Thinking by Content Area*

Content Area	*Foundational Thinking*	*Critical Thinking*
Reading	45%	55%
English and Language Usage	45%	55%
Science	65%	35%
Mathematics	55%	45%
Total	**54%**	**46%**

Item Development Process

Standard 3.2 — Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics.

(AERA, APA, NCME, 2016, p. 64)

Standard 4.7 — The procedures used to develop, review, and try out items, and to select items from the item pool should be documented.

(AERA, APA, NCME, 2016, p. 87)

Item writing. All items on the ATI TEAS were written to assess one of the objectives on the detailed test blueprint. Most of the items on the ATI TEAS were written by outside content experts on a contract basis. The items associated with the objectives were developed and reviewed by content experts in the areas of reading, mathematics, science, and English. ATI test developers entered all of these items into the item bank in preparation for the item review process. Participants in the item review meetings included two content experts and a test developer. The content experts held master's or doctoral degrees and had teaching experience in the specific content area of the items under review (i.e., reading, mathematics, science, or English). The test developers held degrees in English, journalism, or a related field.

Each item is reviewed by independent content experts who verify the item for accuracy and alignment to the intended objective, as well as assigning a foundational or critical thinking level. Final edits are made with input from the test developers.

Sensitivity review. An item may be biased if it contains material unrelated to the objective being assessed or that is unfamiliar to an examinee subgroup, thus inadvertently making the item more difficult for this subgroup. For example, an item may be biased if it uses terminology that is not commonly used across ethnic groups. Consequently, it is common for large-scale tests to be subjected to careful judgmental review and empirical checks to minimize bias.

ATI's bias review committee is composed of an independent and diverse panel of individuals representing historically affected groups. ATI bias review committee members typically come from a variety of backgrounds and are not usually reading, mathematics, science, or English educators because their role is to identify potential bias, not to address the actual content of the items. Committee members were selected to represent a cross-section of historically impacted groups. Ethnicity, gender, age, and disability status were all considerations in selecting the members of this committee. The committee was given training on item bias and its role in the review process. Then the reviewers were asked to individually read each item, flagging those that they considered potentially biased. Reviewers inspect test items looking for material that might be interpreted differently by an examinee subgroup. For example, a test intended to measure verbal reasoning should include words in general use, not words and expressions associated with particular disciplines, occupations, ethnic groups, or geographical regions. Reviewers also examine test items for material that might be offensive, demeaning, or emotionally disturbing. After completing the individual review, the group reconvened to discuss any flagged items. If the group agreed that a flagged item was potentially biased, it was given to a content expert to edit. After the bias review edits were completed, items were ready for pretesting.

Item pretesting procedures. Items were pretested using an embedded pretest design on the TEAS V proctored assessment. Pretest item sets were rotated with sets of proctored scored items. On the ATI TEAS, there are 6 unscored (pretest) items in both the reading and science sections and 4 unscored pretest items in both the mathematics and English and language usage sections. New pretest items are rotated on a web version of an existing set of scored items on a regular basis. A rotation occurs when a minimum of 200 examinees from a minimum of 15 different institutions have completed the exam. In addition, no one institution can account for more than 25% of the examinees completing the exam. Once sampling requirements were met, item analyses were completed on pretest items to assess candidate performance, and a new set of pretest items was published for data collection.

Psychometricians analyzed pretest item data and sent items flagged for questionable performance to the test development team with statistical interpretation. Classical statistics of proportion correct (p-value) and point-biserial correlation (pbs) are used to flag items for inclusion in the item pool. An item is flagged if the item difficulty falls outside the range of p-values from 0.30 to 0.95. Items are also flagged if the point-biserial correlation is below 0.10. The difficulty (p-value) of an item corresponds to the proportion of examinees that correctly answered an item; therefore, the higher the difficulty value, the easier the item. The discrimination statistic (pbs) provides an estimate of the correlation between an individual item (correct or incorrect) and an examinee's total test score. The higher the discrimination, the more the item differentiates between those examinees who received a high total score on the exam versus those who obtained a low total score. Higher values of item discrimination are associated with greater numbers of students responding correctly to the item who also score well on the whole test. Once the classical item analyses were completed and the psychometrics team sent the results to the test development team, the test developers met with two content experts to review flagged items and either edited items for re-pretesting or removed the items from the proctored item bank.

Technical Characteristics

This chapter discusses the technical characteristics of the ATI TEAS. This includes topics such as the test scale, item calibration and equating, reliability, validity, and speededness. Several of these sections involve evaluation of administration data. The first section will describe the data that will be used throughout this chapter.

Description of Administration Data

For the sections within this chapter that involve analysis of administration data, student and item data collected during the first year of the test's administration will be used (Aug. 31, 2016 to Aug. 31, 2017). In order to exclude anomalous data, additional filters were also used. Students that did not complete all four content areas or who scored less than 30% at the total score level were excluded from the sample. Finally, only data from a student's first attempt on TEAS was included.

The total sample sizes and breakdown by demographic characteristic—such as ethnicity, gender, program type, and primary language—for the analysis sample are shown in Table 3. As displayed in Table 3, a majority of the students taking the ATI TEAS identify as Caucasian/white and female, are applying to or recently enrolled in ADN nursing schools, and speak English as their primary language. Geographical and age-related information are also displayed in Figures 1 and 2, respectively. Figure 1 shows that test takers come from across the U.S., with the largest numbers coming from California and Texas. Additionally, Figure 2 shows that the test-takers comprise a wide range of ages, although the majority are young adults.

TABLE 3. *Demographic Profile for Examinees Taking the ATI TEAS*

	ATI TEAS
Ethnicity*	
African American/black	14.2%
Asian	6.4%
Caucasian/white	53.1%
Hispanic	12.5%
Native American	0.7%
Other	2.5%
Gender*	
Female	79.0%
Male	14.9%

	ATI TEAS
Program Type	
ADN	54.7%
BSN	20.9%
Diploma	1.1%
PN	14.9%
Allied Health	6.7%
Other	1.7%
Language*	
English	88.4%
French	0.3%
Spanish	1.3%
Other	2.3%

TOTAL SAMPLE SIZE: 153,704

*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 ATI TEAS*

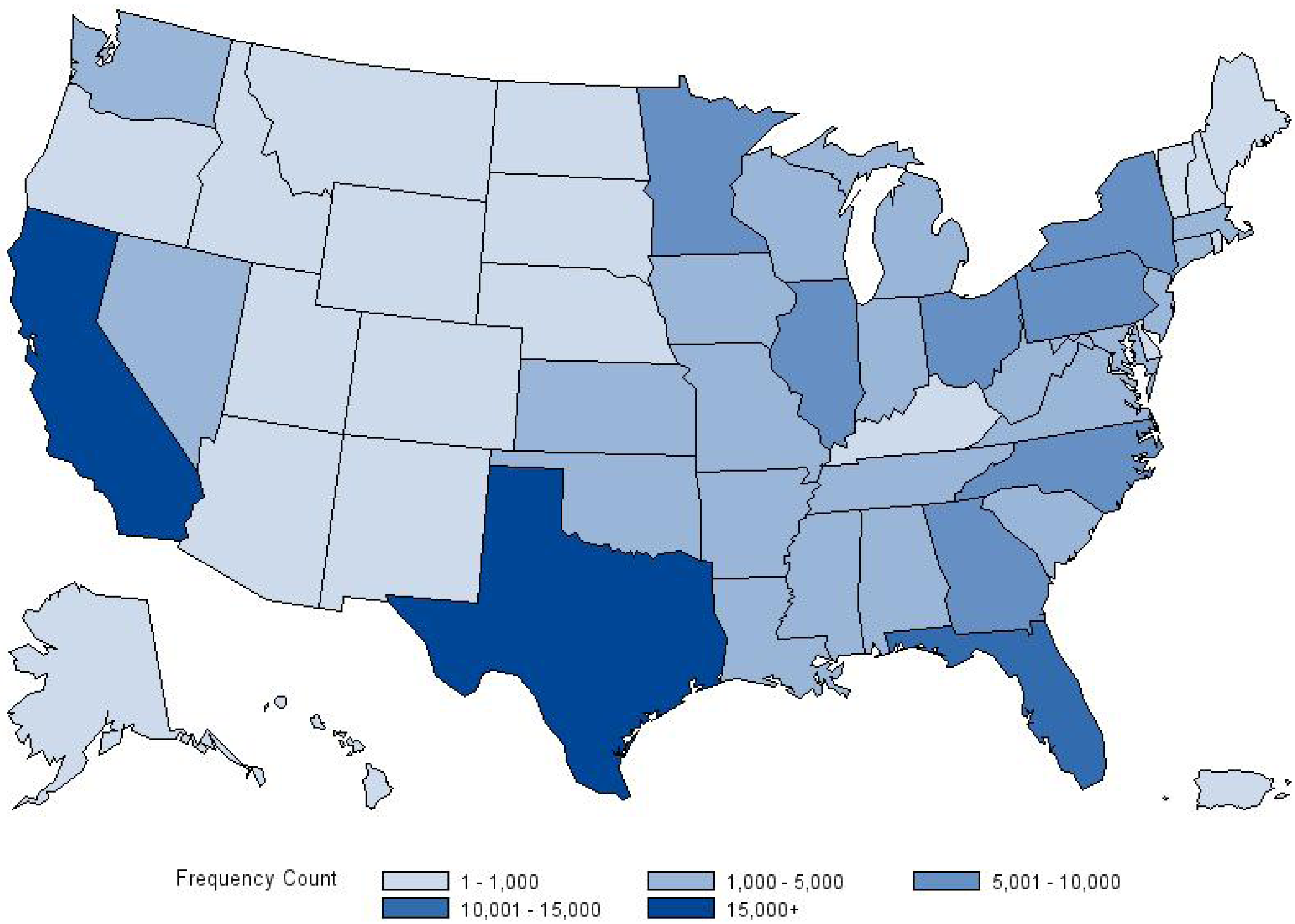
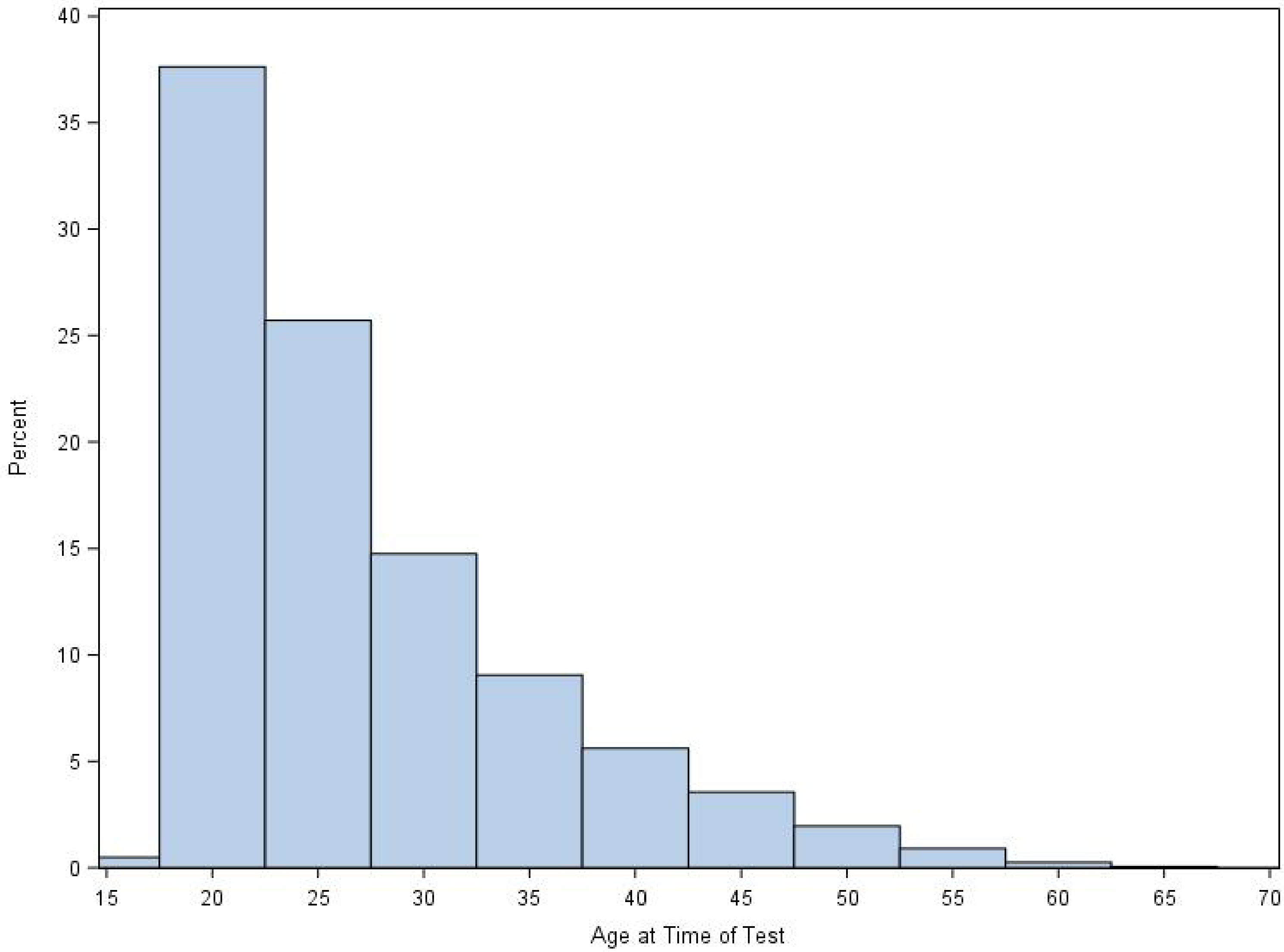


FIGURE 2. *Age Profile for Examinees Taking the ATI TEAS*



Test Scale

When a new blueprint is released, a decision has to be made regarding the statistical scale on which the scores will be presented. Maintaining the same statistical scale between TEAS V and ATI TEAS makes the transition across blueprints fairly seamless both in terms of client usage and also item development. However, the ability to maintain scale requires that the content and test-taking populations be comparable across blueprints. As a result, ATI conducted several investigations to determine whether the statistical scale could be maintained between TEAS V and ATI TEAS.

Once ATI received the final set of objectives for the ATI TEAS assessment, the development team created a crosswalk of the TEAS V content outline to the ATI TEAS content outline at the objective level. This was done to assess the similarity in the content test plans. Content experts in the areas of mathematics, science, English, and reading verified the alignment of the crosswalk, noting that at this finer level of detail (the objective level) there was very little difference between the two outlines.

A second question that needed to be addressed before making a final decision about continuing the scale was to assess the homogeneity of the target population. The ATI TEAS target population is health science program candidates. For previous versions of the TEAS, it was thought that the nursing and allied health candidates were separate sub-populations needing their own versions. ATI investigated evidence supporting the claim that the properties of equating as described in Kolen and Brennan (2014, p. 8–12) hold across these subgroups, and therefore, are one population.

The hypothesis investigated asserted that all health science candidates were one population, high school graduates, when taking the TEAS assessment and that the divergence into separate populations occurs after acceptance to and experience in a particular health science educational program. Data for two TEAS V sets of items were used in this analysis. The content specifications for the forms across the two groups were identical for the allied health and nursing populations. For example, allied health and nursing students who saw form X saw the exact same set of items and the same can be said for form Y. Because the content of the forms administered to these two groups was identical for the purposes of this investigation, we can say the same content specifications property for equating was met. To investigate subgroup performance across the forms, students' first attempt data[3] were analyzed. Classical summary statistics were calculated to determine if the reliability estimates for the forms were similar across the groups. If the same forms given to both groups were not similarly reliable, then this would suggest that the two groups were from distinct populations. The means, standard deviations, reliabilities, and standard errors of measurement (SEM) were calculated for each form for both the Nursing and Allied Health groups (Table 4).

TABLE 4. *Classical Summary Statistics for Two TEAS V Forms Using Allied Health and Nursing Data*

Subgroup	Form	N (1st Attempt)	Average Number Items Correct	SD	Reliability	SEM
AH	X	1459	94.830	18.579	0.920	5.255
Nursing	X	985	93.950	20.420	0.932	5.325
AH	Y	1585	91.488	18.359	0.917	5.289
Nursing	Y	9459	93.740	19.821	0.930	5.244

The raw score mean differences between form Y for Nursing and Allied Health is 2.25, meaning that on average, nursing students answer 2.25 more questions correctly than allied health students. Since the TEAS V exam has 150 questions, this corresponds to approximately a 1.5% difference in percent correct. The mean difference for form X is −0.880, meaning that on average, nursing students answer 0.880 fewer questions correctly than allied health students. This corresponds to approximately a −0.59% difference in percent correct. The reliabilities and SEMs for these two forms for these two groups also do not differ substantially between forms or groups.

As a result, the decision was made to maintain the statistical scale from the TEAS V test plan to the ATI TEAS test plan. Thus, many items could be reused and maintained their item IRT parameter estimates. Later sections summarizing item pretesting, equating, and norming summarize how new item content is pretested and item parameter estimates are calibrated within the context of this decision.

[3] *The date ranges used are as follows. AH X, AH Y, and Nursing Y were 7/1/14–1/1/15. Because Nursing X was retired around that time, the sample size was 134. As a result, 1/1/13–1/1/15 was used, which yielded the sample size of 985.*

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.

(AERA, APA, NCME, 2016, p. 105)

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 ATI TEAS items (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) of the residuals from the Rasch model can be used to assess the unidimensionality assumption. The purpose of the analysis is to verify whether any dominant components exist among the items apart from the latent trait being assessed through the measurement model. If other dimensions are found, the unidimensionality assumption would be violated.

Many different guidelines have been proposed for determining whether the PCA results indicate a violation of the unidimensionality assumption (Hattie, 1985). It is important to note that in practical applications unidimensionality is a strong assumption that is often violated, at least to some degree (de Ayala, 2009). The important consideration is whether the degree of violation is enough to be detrimental to the model. Research has shown that the IRT model tends to be fairly robust to moderate violations of unidimensionality (Dorans & Kingston, 1985; Bolt, 1999).

The eigenvalues and explained variances from the PCA of a typical set of items on ATI TEAS are displayed in Table 5. The Rasch model accounts for 22.1% of the total variance in the data. By comparison, all of the subsequent dimensions have small eigenvalues and account for minimal amounts of the remaining variance. As a result, it is reasonable to conclude that there is one dominant dimension for this test.

TABLE 5. *Principal Components Analysis*

Component	Eigenvalue	Explained Variance
Total	**192.5**	**100.0%**
Measures	42.5	22.1%
1st contrast	3.6	1.9%
2nd contrast	2.9	1.5%
3rd contrast	2.0	1.0%
4th contrast	1.8	0.9%
5th contrast	1.6	0.8%

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 response 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). However, 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 6 summarizes this analysis for ATI TEAS. The vast majority of the correlations were very small, suggesting that local item independence generally holds for these assessments.

TABLE 6. *Summary of Item Residual Correlations to Evaluate Local Independence of Items*

	N	Mean	SD	Minimum	Maximum	> 0.2
ATI TEAS	98,193	0.00	0.02	-0.12	0.36	10

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 of these 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 7 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. All the items had infit and outfit values less than 2.0. Overall, these results indicate that the Rasch model fits the ATI TEAS item data well.

TABLE 7. *Summary of Item Fit Statistics*

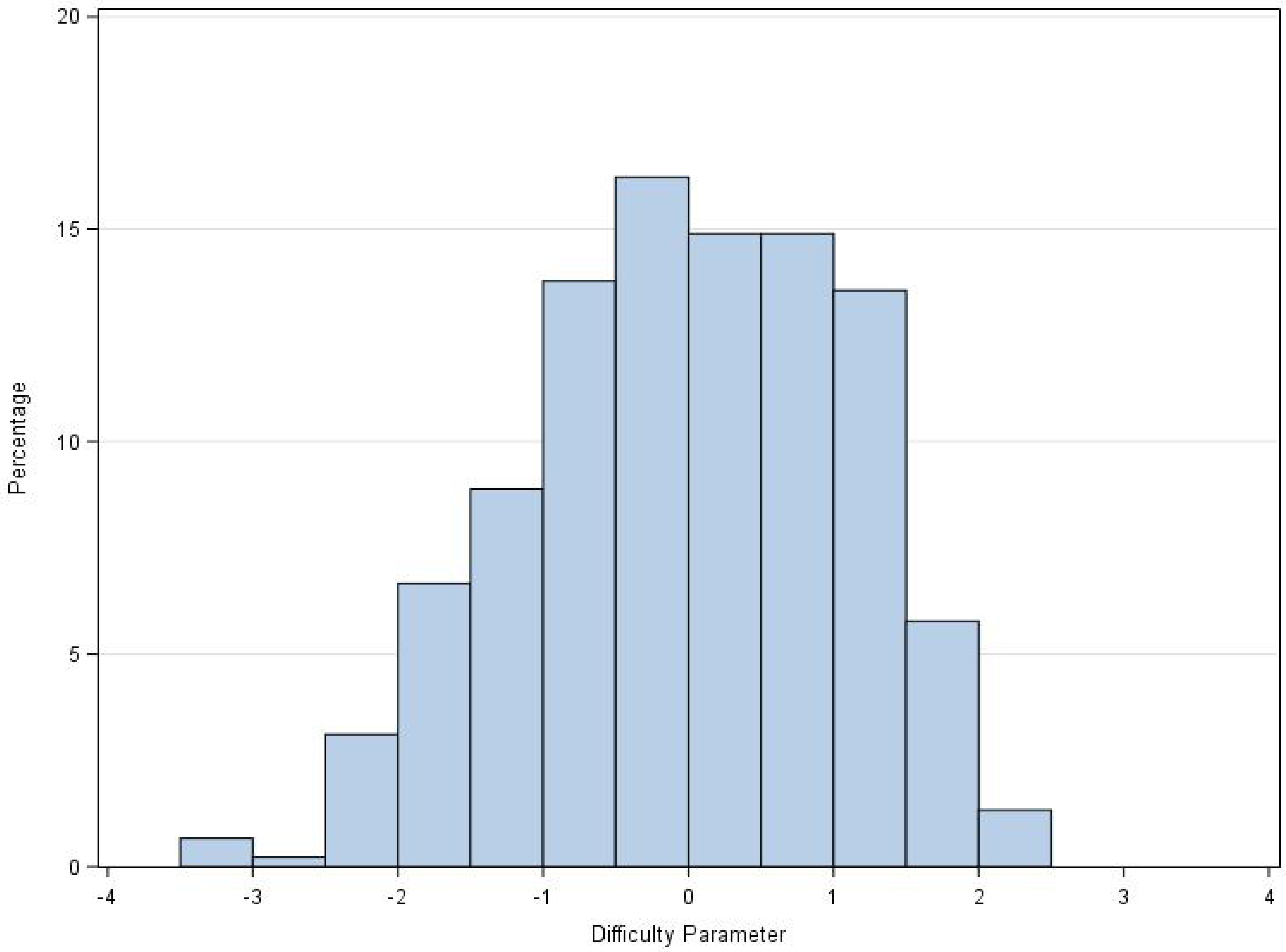
	Infit					Outfit				
	Mean	*SD*	*Min*	*Max*	*> 2.0*	*Mean*	*SD*	*Min*	*Max*	*> 2.0*
ATI TEAS	1.00	0.05	0.87	1.23	0	0.98	0.12	0.66	1.46	0

Item calibration. To ensure comparable scores for candidates across sets of items, total scores for the ATI TEAS assessments must be equated to a base set of items. ATI uses a Rasch, one-parameter Item Response Theory (IRT) procedure (Kolen & Brennan, 2014; Linacre, 2009) to calibrate item response theory parameter estimates using data collected in an embedded pretest design. 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 items are freely calibrated, 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 3 shows the distribution of the Rasch difficulty parameters for ATI TEAS. 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 3. *Histogram of Difficulty Parameters for ATI TEAS*



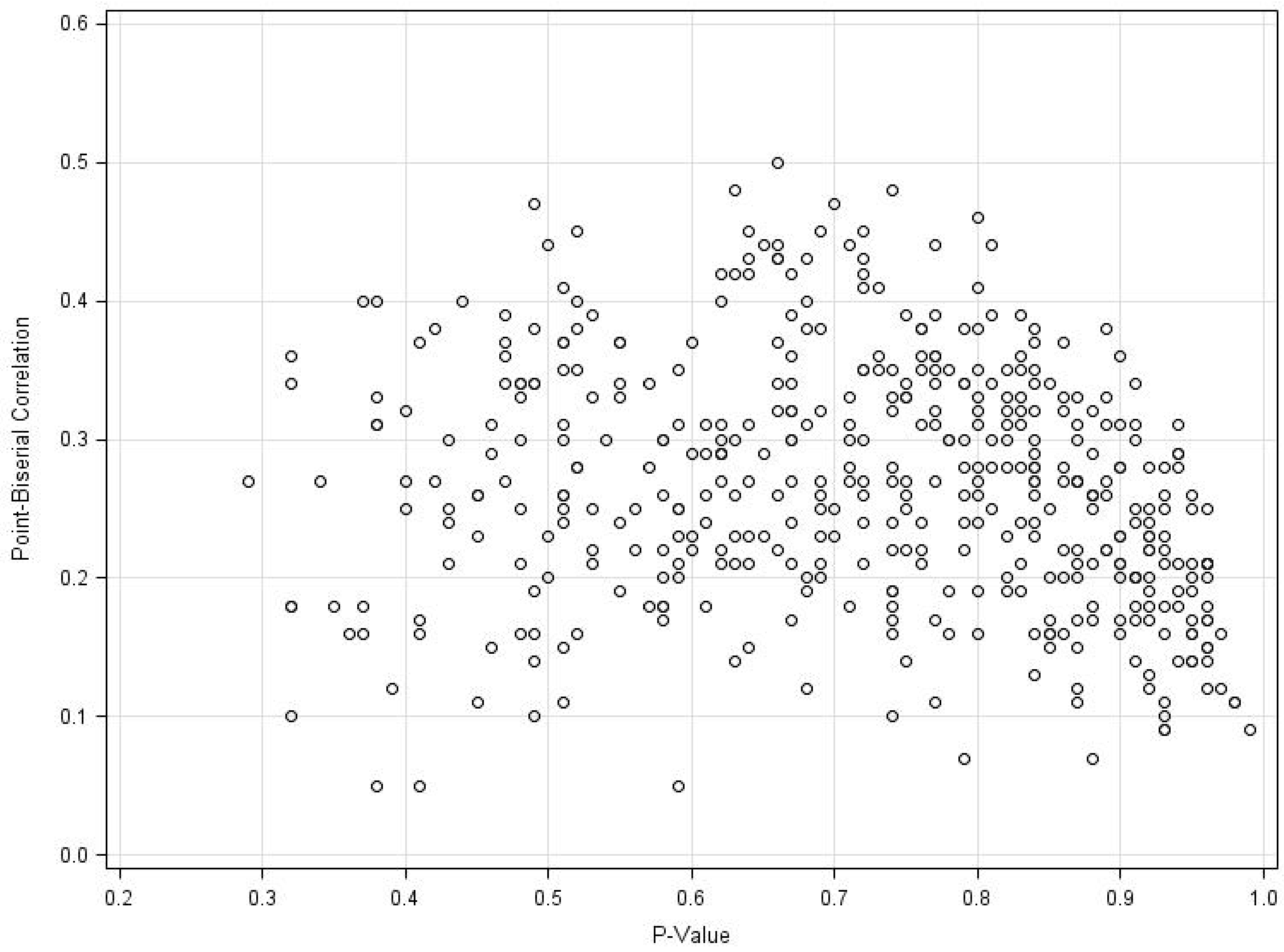
Scoring table development. To ensure that scores for different sets of items are comparable, the content balance of each group of items is carefully matched to a test blueprint to ensure all candidates taking the test are demonstrating their knowledge of the relevant content for the measure. Test developers 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 (Hambleton, Swaminathan & Rogers, 1991) from the newly assembled form to the test characteristic curve for the base form. Using this information, the adjusted percent correct table for the new proctored set of items is created.

Test Item Analyses

Item difficulty and discrimination. The distributions of item difficulty (p-value) and discrimination (pbs correlation) statistics are shown in Figure 4. 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 point-biserial 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 4. *P-value and Point-Biserial Correlations for Items on ATI TEAS*



Reliability Analysis

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

(AERA, APA, NCME, 2016, p. 42)

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

(AERA, APA, NCME, 2016, p. 45)

The reliability coefficients and standard errors of measurement for each content area and the total score on ATI TEAS can be found in Table 8. The standard errors of measurement are reported in the percent correct metric. Results show a fairly high reliability index around 0.96 at the total score level, which indicates that the total score is reliable for the population. The content area reliability indices are slightly lower, which is to be expected as scores based on many items tend to be more reliable than those based on fewer items

TABLE 8. *Number of Items, Reliability Coefficients, and Standard Errors of Measurement*

	Number of Items	*Reliability*	*Standard Error of Measurement*
Math	32	0.69	8.44
English and Language Usage	24	0.70	7.52
Reading	47	0.82	5.47
Science	47	0.88	5.27
Total	**150**	**0.96**	**2.50**

Conditional Standard Error of Measurement

Standard 2.14 — Conditional standard errors of measurement should be reported at several score levels if constancy cannot be assumed. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score.

(AERA, APA, NCME, 2016, p. 46)

The conditional standard error of measurement (CSEM) calculated at the cut scores allows the test user to gauge the expected stability of scores at the levels of greatest interest. This is of particular interest for examinations in which cut scores are used to make decisions.

The CSEM for a typical set of items on ATI TEAS are calculated using an IRT method. The CSEM is conditioned on a student's ability level and can be defined as the multiplicative inverse of the square root of the test information function, $I(\theta)$, at a given ability (Hambleton, Swaminathan, and Rogers, 1991): $CSEM = 1/\sqrt{I(\theta)}$. The ability level on the logit scale (θ) was identified for each of the four cut scores that determine the academic preparedness levels. The 95% confidence interval around each of the cut scores was then calculated from the CSEM as follows: $\theta \pm 1.96(CSEM)$. The resulting values are in the logit scale, but were converted to the adjusted percent correct metric via the test characteristic curve. The 95% confidence intervals for each of the four cut scores are displayed in Table 9.

TABLE 9. *Confidence intervals for academic preparedness levels using the conditional standard error of measurement*

Academic Preparedness Level	*Cut Score*	*95% Confidence Interval*
Basic	41.3%	[34.6%, 48.4%]
Proficient	58.7%	[51.6%, 65.5%]
Advanced	78.0%	[71.4%, 83.6%]
Exemplary	90.7%	[85.3%, 94.3%]

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 entire technical report describes the technical aspects of the ATI TEAS in support of its score interpretation and use. Each of the previous chapters 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 ATI TEAS 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 ATI TEAS is a criterion-referenced assessment. The criteria referenced are the ATI TEAS blueprint. Evidence supporting the alignment among the ATI TEAS objectives and the ATI TEAS blueprint should be provided.

For the ATI TEAS, strong content validity evidence is derived directly from the test construction process. Each item was based on and directly aligned to the ATI TEAS blueprint to ensure good content validity. The item development and test construction process ensures that every item aligns directly to one content category. ATI selected qualified item writers and provided training to help ensure they wrote high-quality items. This alignment is foremost in the minds of the item writers and editors. As a routine part of item selection prior to an item appearing on a test form, the review committees checked the alignment of the items with the test plan and made any adjustments necessary. Meanwhile, the test development team established detailed test- and item-development specifications and ensured the items were sufficient in number and adequately distributed across content and levels of cognitive complexity and difficulty. Items were also submitted to bias review for issues related to diversity, gender, and other pertinent factors. Items passing all the prior hurdles 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 4. All values are positive and of acceptable magnitude.

Item response theory dimensionality. Results from principal components analyses are presented in Table 5. The ATI TEAS is 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 ATI TEAS, 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. ATI TEAS essentially met the underlying assumptions of Rasch models, 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.

Predictive Validity. By measuring an individual's overall academic preparedness for a health sciences program, the ATI TEAS purports to positively correlate with early health science program success. For potential nursing-school students, this relationship can be evaluated by looking at the correlation between scores on ATI TEAS and either the ATI RN Content Mastery Series (CMS) Fundamentals or ATI PN CMS Fundamentals tests. The CMS Fundamentals tests are usually administered during the first year of nursing school, and are therefore a good indication of early performance.

For this analysis, all students who completed ATI RN CMS Fundamentals 2016 or ATI PN CMS Fundamentals 2014 after completing the ATI TEAS assessment were included in the sample. The correlation coefficients between ATI TEAS (both total score and content areas) and Fundamentals are displayed in Table 10. The correlations are positive and of moderate strength, indicating that ATI TEAS is a good indicator of early program success in nursing school.

TABLE 10. *Correlation coefficients between ATI TEAS and RN and PN Fundamentals*

ATI TEAS Content Area	*RN Fundamentals 2016 (N = 3,058)*	*PN Fundamentals 2014 (N = 1,402)*
Math	0.25	0.26
English and Language Usage	0.31	0.30
Reading	0.37	0.35
Science	0.34	0.34
Total	**0.41**	**0.41**

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.

(AERA, APA, NCME, 2016, p. 90)

The ATI TEAS is 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 ATI TEAS. These analyses were conducted at the content area level since the test times for each content area are independent.

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 11, all of the ATI TEAS content area tests 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 11 shows the ATI TEAS content area tests 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 comprehensive indication when looking across all metrics is that there is a very low likelihood that student ATI TEAS scores are affected in a meaningful way by the amount of time allowed.

TABLE 11. *Test Speededness Analyses*

	% of students answering the last item	*% of students answering at least 75% of items*	*Speededness Quotient*	*Std (unreached) / Std (incorrect)*
Math	93.14%	100.00%	0.03	0.22
English and Language Usage	99.48%	100.00%	0.00	0.07
Reading	95.28%	100.00%	0.02	0.28
Science	99.84%	100.00%	0.00	0.04

Test Administration

Standard 7.8 — Test documentation should include detailed instructions of how a test is to be administered and scored.

(AERA, APA, NCME, 2016, p. 127)

Standard 6.1 — Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer and any instructions from the test user.

(AERA, APA, NCME, 2016, p. 114)

Administration Setting

The ATI TEAS is an entrance examination administered directly by the client institutions that have purchased the exam or by a PSI testing center. The test is available in both paper-pencil and web-based formats.

Administration Procedures

In order to ensure a standardized testing experience across administration settings, proctors are required to take a proctor training course. Proctors also administer the assessments from directions provided in the ATI Proctor Process Guide, including a proctor script to read to examinees before and during the administration. Proctors must be physically present in the room or able to have live and engaged interactions throughout the test administration process with the examinees. They are advised on all guidelines in regard to restricted items, scratch paper, restroom breaks, and stopping the administration in the event of suspected or observed misconduct—including test fraud or test theft.

Time Limits

Time is not intended to be a limiting factor on the assessment; consequently, the administration time per content area was set to allow examinees ample time to complete each section: approximately 1.2 minutes per question is allotted for reading; 1.5 minutes for mathematics; 1.2 minutes for science; and 1 minute per question for the English and language usage section. Table 1 identifies the time allotted for examinees to complete each portion of the assessment. Extensive analyses are completed to assess the speededness of each of the ATI TEAS content areas. There is no indication that student performance is adversely affected by the time limitations.

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.

(AERA, APA, NCME, 2016, p. 23)

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.

(AERA, APA, NCME, 2016, p. 102)

Score Interpretation

ATI provides two ways to interpret the ATI TEAS 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 ATI TEAS 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 academic preparedness 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 ATI TEAS are scored as correct or incorrect, with no partial credit awarded on any item and no penalty for guessing. Appendix A contains a sample of the score report and explanation page that examinees receive after completing the assessment. So that total test scores and content area 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 and content area scores are equated to adjust for slight differences that might exist across different groups of items making up a test.

The reported total score is called Total Score on the score report. It can be interpreted as the percentage of items answered correctly on the whole 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. The Total Score is provided along with both norm-referenced data (means and percentile ranks) and criterion-referenced data (academic preparedness level), each of which will be discussed in further detail later in this chapter.

Adjusted percent correct scores are also reported for each of the four content areas, along with percentile rank information. In addition to the score, normative data, and category, the score report also includes a list of Topics to Review and a Focused Review. This list includes topic descriptors and section references to the ATI TEAS Study Manual for each incorrectly answered item.

Summary Descriptive Statistics and Normative Data Reported

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.

(AERA, APA, NCME, 2016, p. 104)

Normative comparisons are not the primary purpose of the ATI TEAS; 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, as well as program-type means and percentile ranks for students in ADN, BSN, PN, or Diploma nursing programs. The national and program type mean adjusted percent correct scores are reported on student and group score reports along with percentile rank information (see score report example, Appendix A.) These tables are available on the ATI website under the faculty help page, assessments section.

As noted previously, the decision was made with the release of ATI TEAS to maintain the statistical scale between the TEAS V and ATI TEAS. This means that the scores for TEAS V are comparable to scores on the ATI TEAS. Additionally, empirical evidence supported the claim that the allied health program candidates and nursing program candidates performed as one population prior to being admitted to a health science program. As a result, the summary descriptive statistics and norm tables were updated to reflect the ATI TEAS target population of test takers. These statistics for the national health science and nursing program types are shown in Table 12.

TABLE 12. *Program Type ATI TEAS Adjusted Percent Correct* Means and Sample Sizes (Nursing)*

Group (Program Type(s) or Category)	Total Test Score Content Area	Adjusted Percent Correct Mean	Sample Size
National Health Science *(Allied Health + Nursing)*	Total Test	65.6%	244,453
	Reading	72.4%	
	Mathematics	68.6%	
	Science	57.2%	
	English and Language Usage	66.3%	
Nursing *(BSN)*	Total Test	70.4%	55,010
	Reading	75.6%	
	Mathematics	73.9%	
	Science	63.4%	
	English and Language Usage	70.4%	
Nursing *(ADN)*	Total Test	66.1%	130,967
	Reading	72.8%	
	Mathematics	68.8%	
	Science	58.0%	
	English and Language Usage	66.8%	
Nursing *(Diploma)*	Total Test	63.8%	2,471
	Reading	72.6%	
	Mathematics	66.2%	
	Science	54.0%	
	English and Language Usage	64.6%	
Nursing *(PN)*	Total Test	58.4%	42,268
	Reading	66.9%	
	Mathematics	61.7%	
	Science	48.0%	
	English and Language Usage	60.3%	

* Adjusted Percent Correct Scores are equated scores and are comparable across sets of items and administrations.

Academic Preparedness 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.

(AERA, APA, NCME, 2016, p. 107)

To provide a numerical indication of a student's preparedness level, ATI developed a set of recommended criterion-referenced proficiency levels that schools could choose to adopt as benchmarks for student performance on the ATI TEAS. The academic preparedness category is a level assigned to examinees based on their adjusted percent correct total score. These cut scores were established during the TEAS V National Cut Score Study (Wolkowitz, 2010), which describes five categories of overall academic preparedness for a nursing program: Developmental, Basic, Proficient, Advanced, and Exemplary. The names of the categories were determined by the national cut score study committee as the least stigmatizing, yet accurate representation of the level of preparedness for examinees. Details of the cut score study are available in the TEAS National Cut Score Study report (Wolkowitz, 2010).

Setting Institutional Cut Scores

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

(AERA, APA, NCME, 2016, p. 107)

To provide a numerical indication of a student's preparedness level, ATI developed a set of recommended criterion-referenced proficiency levels that schools could choose to adopt as benchmarks for student performance on the ATI TEAS. The academic preparedness category is a level assigned to examinees based on their adjusted percent correct total score. These cut scores were established during the TEAS V National Cut Score Study (Wolkowitz, 2010), which describes five categories of overall academic preparedness for a nursing program: Developmental, Basic, Proficient, Advanced, and Exemplary. The names of the categories were determined by the national cut score study committee as the least stigmatizing, yet accurate representation of the level of preparedness for examinees.

Institutions choosing to use a cut score associated with the ATI TEAS should do so in a defensible manner. ATI recommends setting cut scores based upon the total test scores as these scores are equated and are comparable across students and across administrations. For institutions using the ATI recommended cut score and academic preparedness categories, the National Cut Score Study (Wolkowitz, 2010) provides documentation of the cut score study process. The academic preparedness categories are also described in the integration document, ATI TEAS Academic Preparedness Levels Summary (ATI, 2016b).

Although the academic preparedness levels were originally developed to be used specifically for nursing programs on the previous blueprint of the TEAS assessment (TEAS V), these academic preparedness levels can also be used on the updated blueprint of TEAS and applied more broadly to health science programs. The reasons for this are twofold.

First, the statistical scale has been maintained between TEAS V and ATI TEAS. This means that the scores for TEAS V are comparable to scores on the ATI TEAS. This decision was based upon empirical evidence, including a data analysis of TEAS assessments, as well as judgments from content experts familiar with health science fields and TEAS content. As a result of the scale continuity between the blueprints, the standards that were developed on TEAS V can be directly mapped to ATI TEAS.

Second, the decision to extend the academic preparedness levels from nursing students to the broader population of health science students was based on several factors. Most notably, data analytic comparisons of performance on TEAS between the nursing and allied health populations yielded no evidence that these two populations perform differently at the time of TEAS administration. Similarly, the percentage of students that fall into each academic preparedness level remains consistent whether the populations are considered together or separately. Overall, nursing and allied health students appear to be one population at the time of TEAS administration and diverge into separate populations after being exposed to course work in a particular program.

Alternatively, institutions may consider conducting their own studies to set a cut score for their institution using one of many established methodologies, such as the Angoff method or a contrasting group's method as implemented in the ATI TEAS Cut Score Analysis Tool. Regardless of the methodology used, institutions should carefully weigh the impact any cut score choice is likely to have on their applicant and student population before the cut score is implemented. Cut scores should be periodically re-examined and adjusted after ATI TEAS scores from admitted students and their subsequent performance in the program is available. Applicants who fail to meet the cut score should be allowed to compensate with higher performance on other criteria.

Institutions may also consider using ATI TEAS without a cut score in an admission formula design. Such a design allows programs to place weights on ATI TEAS and other criteria according to their value in the admission decision. Great care should be taken to ensure that the criteria chosen and weights assigned for an admission formula result in valid selection decisions. Admission formulas should be regularly re-examined and adjusted as data are accumulated.

Appropriate and Inappropriate Test Use

Standard 1.3 — If validity for some common or likely interpretation for a given use has not been evaluated, or if such an interpretation is inconsistent with available evidence, that fact should made clear and potential users should be strongly cautioned about making unsupported interpretations.

(AERA, APA, NCME, 2016, p. 23)

Standard 12.8 — When test results contribute substantially to decisions about student promotion or graduation, evidence should be provided that students have had an opportunity to learn the content and skills measured by the test.

(AERA, APA, NCME, 2016, p. 197)

Standard 12.10 — In educational settings, a decision or characterization that will have major impact on a student should take into consideration not just scores from a single test but other relevant information.

(AERA, APA, NCME, 2016, p. 198)

The purpose of the ATI TEAS assessment is to assess an examinee's overall academic preparedness for a health science program. It is marketed as low- to moderate-stakes test. ATI TEAS is not designed for high-stakes purposes, such as being the sole criterion for determining admission into a health science program. Under no circumstances is it recommended that the ATI TEAS be used as a sole criterion for any high-stakes decision. A test can be considered a sole criterion if failure is possible based on test performance, regardless of how the student performs on other measures. For additional guidance on the stakes associated with exam scores, refer to the ATI Position on High Stakes Testing at www.atitesting.com under "Help/Policies & Research."

Examinees completing the ATI TEAS, or any other standardized examination, should be offered an opportunity to retake the assessment. With more than one active form of the TEAS available, examinees have the opportunity to retest with new questions. Although a program may wish to allow more than two test administrations, this should be done with careful consideration of the amount of time specified between administrations of the same form. An ideal test retake policy would specify a waiting period that encourages remediation between testing and ensures that students are not "practicing" the test by repeatedly taking proctored test forms, while still allowing students a reasonable opportunity to retest to improve scores when necessary.

References

Alpine Testing Solutions. (2014). *Assessment technologies institute (ATI): Test of essential academic skills (TEAS) content analysis workshop.* Leawood, KS: Ascend Learning.

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.

Assessment Technologies Institute. (2012). *TEAS V for allied health content validation study.* Leawood, KS: Ascend Learning.

Assessment Technologies Institute. (2016b). *ATI academic preparedness levels summary.* Retrieved from: http://www.atitesting.com/ati_next_gen/Faculty/FacultyLanding.aspx

Assessment Technologies Institute. (2016a). *ATI proctor process guide.* Retrieved from: http://www.atitesting.com/ati_next_gen/Includes/PDF/en-US/Nursing/HowTo/Faculty/ProctorManuals/HowToProctorProcessGuide-TEAS-Online.pdf

Bolt D. M. (1999). Evaluating the effects of multidimensionality on IRT true-score equating. *Applied Measurement in Education, 12,* pp. 383–407.

Chen, W., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*(3), pp. 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), pp. 178–194.

de Ayala R. J. (2009). *The theory and practice of item response theory.* New York, NY: Guilford Press.

Dorans, N. J., & Kingston, N. M. (1985). The effects of violations of unidimensionality on the estimation of item and ability parameters and on item response theory equating of the GRE verbal scale. *Journal of Educational Measurement, 22*(4), pp. 249-262.

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 Publications.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*(2), pp. 139–164.

Kane, M.T., Kingsbury, C., Colton, D., & Estes, C. (1989). Combining data on criticality and frequency in developing test plans for licensure and certification examinations. *Journal of Educational Measurement, 26*(1), pp. 17–27.

Kolen, M.J., & Brennan, R.L. (2014). *Test Equating, scaling and linking; Methods and practice.* New York, NY: Springer.

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,* pp. 29–37.

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

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

Wolkowitz, A. (2010). *TEAS V national standard setting study.* Stilwell, KS: Author.

Wright, B., & Linacre, J. M. (n.d.). *Reasonable mean-square fit values.* Retrieved from 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.

Appendix A — Sample Score Report

Individual Performance Profile
ATI TEAS Retake 12

Individual Name:	**SAMPLE STUDENT**	Test Date	**8/26/2016**
Institution:	**ATI**	Attempt	**2 of 2**
Program Type:	**BSN**	Days Since Last Attempt:	**2102**

Scores

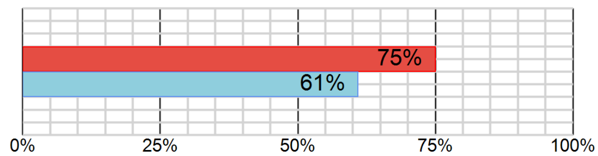
Total Score:

74.0%

Academic Preparedness Level: Proficient

National Mean: 65.6% **All BSN Programs Mean: 70.4%**

Percent of students who scored at or below your score



Percent of students who scored at or below your score

Reading Score:

51.1%

Math Score:

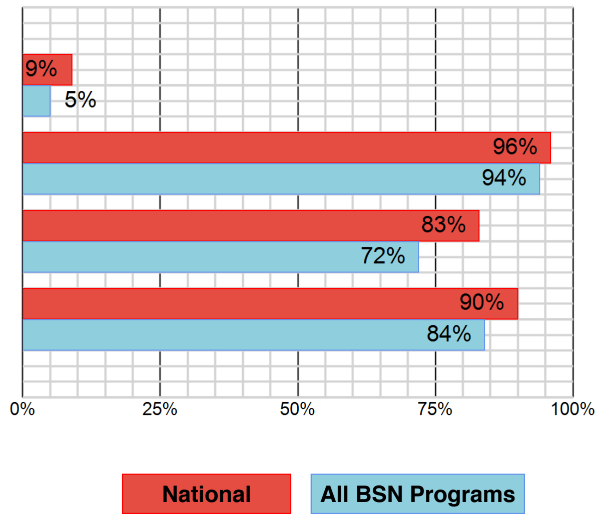
93.8%

Science Score:

72.3%

English and Language Usage Score:

83.3%



Content areas do not add up to the total score

Topics To Review

Reading (47 items)

Key Ideas and Details (22 items, 68.2% answered correctly)

Summarize a complex text.

Identify information from a graphic representation of information.

Follow a given set of directions.

Recognize events in a sequence.

Recognize events in a sequence.

Please see page 3 for an explanation of the Scores

Page 1 of 3

Report Created on: 8/26/2016 09:16 AM EDT

ATI_TEAS_Individual

Recognize events in a sequence.

Identify the topic, main idea, and supporting details.

Craft and Structure (14 items, 64.3% answered correctly)

Evaluate the author's point of view in a given text.

Utilize text features.

Interpret the meaning of words and phrases using context.

Evaluate the author's purpose in a given text.

Evaluate the author's point of view in a given text.

Integration of Knowledge and Ideas (11 items, 36.4% answered correctly)

Evaluate and integrate data from multiple sources in various formats including media.

Evaluate and integrate data from multiple sources in various formats including media.

Compare and contrast themes from print and non-print sources.

Evaluate an argument and its specific claims.

Compare and contrast themes from print and non-print sources.

Identify primary sources in various media.

Evaluate an argument and its specific claims.

Math (32 items)

Numbers and Algebra (23 items, 95.7% answered correctly)

Solve real world problems involving proportions.

Measurement and Data (9 items, 88.9% answered correctly)

Explain the relationship between two variables.

Science (47 items)

Human Anatomy and Physiology (32 items, 75.0% answered correctly)

Describe the anatomy and physiology of the cardiovascular system.

Describe the anatomy and physiology of the integumentary system.

Describe the anatomy and physiology of the cardiovascular system.

Describe the anatomy and physiology of the integumentary system.

Describe the anatomy and physiology of the neuromuscular system.

Describe the anatomy and physiology of the reproductive system.

Describe the anatomy and physiology of the skeletal system.

Describe the general anatomy and physiology of a human.

Scientific Reasoning (7 items, 57.1% answered correctly)

Analyze the design of a scientific investigation.

Identify basic scientific measurements using laboratory tools.

Analyze the design of a scientific investigation.

English and Language Usage (24 items)

Conventions of Standard English (9 items, 88.9% answered correctly)

Use conventions of standard English spelling.

Knowledge of Language (9 items, 88.9% answered correctly)

Apply basic knowledge of the elements of the writing process.

Vocabulary Acquisition (6 items, 83.3% answered correctly)

Determine the meaning of words by analyzing word parts.

Please see page 3 for an explanation of the Scores

Report Created on: 8/26/2016 09:16 AM EDT

ATI_TEAS_Individual

ati NURSING EDUCATION

Score Explanation and Interpretation
Individual Performance Profile

TOTAL SCORE

To adjust for possible differences in difficulty among the forms of this assessment, the raw score (the total number of items correct) is converted to the total score through a process known as equating. The total score is on a scale of 0% to 100%.

CONTENT AREA SCORES

The content area scores appear below the total score on the score report. Like the adjusted individual total score, the scores for the four content areas are adjusted to account for possible differences in difficulty among the forms of this assessment. The content area scores do not add up to the total score.

ATI ACADEMIC PREPAREDNESS LEVELS

Academic Preparedness Level	Academic Preparedness Level Definition	Score Range
Developmental	Developmental scores generally indicate a very low level of overall academic preparedness necessary to support learning of health sciences-related content. Students at this level will require additional preparation for most objectives assessed on ATI TEAS. (See Topics to Review on this score report.)	0.0% to 40.7%
Basic	Basic scores generally indicate a low level of overall academic preparedness necessary to support learning of health sciences-related content. Students at this level are likely to require additional preparation for many objectives assessed on ATI TEAS. (See Topics to Review on this score report.)	41.3% to 58.0%
Proficient	Proficient scores generally indicate a moderate level of overall academic preparedness necessary to support learning of health sciences-related content. Students at this level can require additional preparation for some objectives assessed on ATI TEAS. (See Topics to Review on this score report.)	58.7% to 77.3%
Advanced	Advanced scores generally indicate a high level of overall academic preparedness necessary to support learning of health sciences-related content. Students at this level are not likely to require additional preparation for the objectives assessed on ATI TEAS.	78.0% to 90.0%
Exemplary	Exemplary scores generally indicate a very high level of overall academic preparedness necessary to support learning of health sciences-related content. Students at this level are not likely to require additional preparation for the objectives assessed on ATI TEAS.	90.7% to 100.0%

NATIONAL MEAN

This is the average score of all examinees.

PROGRAM MEAN

This is the average score of all examinees within your specified program type.

NATIONAL PERCENTILE RANK

This is the percentage of examinees who scored at or below your score.

PROGRAM PERCENTILE RANK

This is the percentage of examinees within your program type who scored at or below your score.

TOPICS TO REVIEW

Based on the questions missed on this assessment, a listing of content areas and topics to review is provided. The *ATI TEAS Study Manual* and ATI TEAS Online Practice assessments A and B are available to aid in the review process and can be found at the ATI website.

DISCLAIMER

Total scores are calculated using items from all four content areas. Therefore, total scores are not provided for test takers who do not complete all four content areas of the assessment.

ATI TEAS Content Areas	Number of questions
Reading	*47*
Key ideas & details	22
Craft & Structure	14
Integration of knowledge & ideas	11
Mathematics	*32*
Numbers & algebra	23
Measurement & data	9
Science	*47*
Human anatomy & physiology	32
Life & physical sciences	8
Scientific reasoning	7
English & language usage	*24*
Conventions of standard English	9
Knowledge of language	9
Vocabulary acquisition	6
	150

Exp_Ind_ATI_TEAS

www.atitesting.com

Appendix B — Allied Health Occupational Field Titles & Rigor Categories for Reporting Normative Data

With the decision to assess all health science program candidates as one population with the ATI TEAS assessment, ATI then moved to updating the allied health occupational field titles for the 2016 release. The development team wanted to ensure the titles reflected contemporary health science language when listed in the end-of-assessment survey for the categories of rigor normative data analysis. This section first summarizes how the categories of rigor were developed for TEAS V and then summarizes the survey process to update the allied health occupational title language.

When the allied health TEAS V assessment outline was finalized, consideration turned to the practical use of the exam. Since the allied health programs were not easily separated into program types like nursing programs (e.g., PN, ADN, BSN), a sample of allied health educators was asked to assist to create categories for the allied health fields to be used for reporting normative data. In December 2009, nine allied health professionals were sent a list of 46 allied health professions and asked to rate the rigor of the academic program based on a 4-point scale.[4] During February 2010, this committee met via web conference to discuss the results of the survey. Eight of the nine (88.9%) original committee members responded to the survey and 7 committee members participated in the web conference. During the web conference, the mean rating for each of the 46 allied health professions were discussed and each committee member voiced his or her opinion as to the appropriate category for that profession. Any disagreement amongst the committee members was discussed until a consensus was reached. Two additional professions were added: an EKG technician and a clinical medical assistant (ATI, 2012, p.10).

With the release of ATI TEAS in 2016, updated occupational titles were included in the category of rigor listing. ATI developed a two-question survey and distributed this to 25 ATI clients who were serving as deans or directors for allied health programs across the country. The survey questions along with the final list of programs are listed in Appendix C.

[4] *The scale: very rigorous (Only top students will be academically prepared for success in this program), rigorous (Above-average students will be academically prepared for success in this program), somewhat rigorous (Average students will be academically prepared for success in this program), not very rigorous (Almost all students will be academically prepared for success in this program).*

Appendix C — Allied Health Program Title Survey and Final Program Titles

Survey participants were asked: "Below is a list of common names for allied health professions. We're updating the list to ensure it represents how each profession is described today by health care practitioners. Please read through the list. If you believe a name is outdated or no longer appropriate, please provide your preferred/suggested name." and "Which Allied Health professions would you like us to consider adding to the TEAS? (Add up to five additional professions.)"

Final List of Occupational Field Titles	
Athletic Trainer	Magnetic Resonance Technologist/Imagist
Biomedical Engineer	Massage Therapist
Biomedical Equipment Technician	Medical Administrative Assistant
Blood Bank Technologist	Clinical Medical Assistant
Cardiovascular Technician	Medical Coder/Billing; Coding Specialist
Cardiovascular Technologist/Interventional Cardiovascular Technologist	Medical Laboratory Assistant/Technician
Central Sterile Processing Technician	Medical Office Assistant
Clinical Laboratory Technician	Medical Records (Health Information) Administrator
Clinical (Medical) Laboratory Technologist/scientist	Medical Records (Health Information) Specialist
Computed Tomography Technologist/Computed Tomographer	Medical Records (Health Information) Technician
Dental Assistant	Medical Transcriptionist
Dental Hygienist	Nuclear Medicine Technologist
Diagnostic Cardiovascular Sonographer	Occupational Therapist Assistant
Diagnostic Medical Sonographer	Ophthalmic Medical Technician/Ophthalmic Assistant
Dialysis Technician	Ophthalmology Technologist
EKG Technician	Paramedic
Emergency Medical Technician	Patient Care Technician
Health Information Technologist	Perfusionist
Interventional Vascular Technologist	Pharmacy Technician
Athletic Trainer	Phlebotomist/Phlebotomy Technician
Biomedical Engineer	Physical Therapist Assistant
Biomedical Equipment Technician	Radiation Therapist/Radiation Therapy Technologist
Blood Bank Technologist	Radiographic Technologist/Radiographer
Cardiovascular Technician	Certified Respiratory Therapist (CRT)
Cardiovascular Technologist/Interventional Cardiovascular Technologist	Registered Respiratory Therapist (RRT)
Central Sterile Processing Technician	Surgical Assistant
Clinical Laboratory Technician	Surgical Technologist