

2008 ANNUAL SUMMARY

NCSF-CPT EXAMINATION

Introduction

The National Council on Strength and Fitness Board for Certification (NCSFBC) is the authority over the certification program and NCSF-CPT credential for personal trainers. The Board sets eligibility and competency requirements for entry into the personal trainer profession as well as practice standards and guidelines for certified personal trainers. The NCSFBC has partnered with Prometric Testing Services to oversee the test development and administration of current certification exams as well as new and enhanced examinations for candidates seeking certification in the personal trainer profession. Prometric's expertise and extensive experience in test development enables the NCSFBC to provide valid and reliable assessments of competency for entry level personal trainers. Prometric's industry-leading psychometricians and test development professionals provide the NCSFBC with the necessary expertise to provide legally defensible exams and credible and accurate outcomes in testing. Holding to these administrative and procedural rigors supports the NCSFBC's mission to protect certification stakeholders including the general public served by the profession.

The quality of a certification exam depends on the quality of the development process (and the expertise of the developers) of that test. Although finished examinations may look substantially alike, the way those examinations perform and the extent to which they do what they are designed to do depend on the attention to detail and the commitment to quality reflected in the test development process. The NCSFBC Prometric partnership ensures the certification program meets the high standards embodied in the *ETS Standards for Quality and Fairness*². The *ETS Standards* are designed to ensure that all Prometric services demonstrably meet explicit criteria in areas of basic importance, including quality control for accuracy and timeliness, validity, assessment creation, test administration, reliability, scale definition, equating and score interpretation. The *ETS Standards* reflect the *Standards for Educational and Psychological Testing*, which are considered industry standards.

Test Development

Prometric's test development experts work with NCSFBC selected subject matter experts in the creation of legally defensible, psychometrically sound examinations. This process starts with an extensive Job Task Analysis (JTA). The JTA is used to define the specific tasks and duties of the personal trainer, the knowledge and skills necessary to successfully complete the tasks competently, as well as the relevance and frequency of the task in the daily activities of the profession. The JTA employs input from stakeholders of all levels to thoroughly research the profession from different perspectives and from different parts of the country so a national scope is properly established. Once a JTA is complete, test developers can use the data to form test blueprints, which encompass the requisite competency domains and the depth of knowledge necessary for each area within the scope of the profession.

Test developers must then create items for the examination that validly measure the knowledge or skill set within each competency domain. Subject matter experts must create numerous test items for each testable category in compliance with strict item writing principles. The new items are then edited on several levels and researched for validity. A large number of test items are placed in the test bank so

those questions that are identified as flawed are replaced with those that perform properly to form a valid assessment. Flawed items are designated as such during the research process so that they are deleted and never reach the assessment instrument. The NCSFBC, like other credible organizations, utilizes unscored research items on current examinations so items are psychometrically analyzed and validated before entering the test bank allowing tests to be updated in an ongoing manner. The NCSFBC test development process is tabled below.

Test Design	Job Analysis	Define the tasks, knowledge, and skill important for performing the specified role.
	Test Specifications	Subject matter experts (SMEs) review the importance and determine how many items should be written to each objective.
	Test Definition	Defines the purpose, scope, target population, general topics, duration, number of forms, number of items and types of items.
Item Development	Item Writing	Provide training on item writing to meet the test specifications and amounts listed in the blueprint.
	Technical Item Reviews	Review items for language and technical accuracy
Psychometrics & Test Construction	Item Analysis	Compute statistics that measure item performance.
	Item Selection	Assign items for inclusion on final forms, discarding or rewriting.
	Form Assembly	Distribute items across forms so that each form meets the specifications of the blueprint plan and remain equally difficult.
	Beta Test	Evaluate items and examinations before scored use.
	Standard Setting	Set the cut score.
In-service Analysis	Maintain Exams	Conduct ongoing analysis of item and test statistics. Revise test with updates periodically.

Once the items have been created and researched, test development experts assemble the approved examination items into multiple test forms in accordance with the pre-determined test specifications, ensuring that the appropriate number of items from each knowledge, skill and ability area is incorporated in to each form. As one or more test forms are assembled, the psychometricians use pretest and operational test item data to anticipate the statistical characteristics of each form. This data is used to verify adequate test functionality and test form comparability. Test analysis assures that test forms exhibit expected pass rates, adequate reliability and pass/fail decision consistency, tolerable measurement error, expected item response consistencies and interrelationships between test parts, adequate response times and comparability between forms. These attributes are critical to valid interpretation and use of test scores.

In subsequent examination development cycles, test developers assemble draft forms of each test to present for committee review. This process assures that every test form developed is parallel to all others and meets test specifications. Each form will consist of the appropriate number of operational (scored) and pretest (unscored) items. Prometric test experts then conduct a form review prior to publishing and distributing a new form of each examination. Recent updates to the NCSF-CPT examinations occurred in 2004, 2006, and 2008.

Standard Setting - Determining a Passing Score for Each Exam

Prometric psychometricians provide expertise in assisting subject matter experts in establishing and validating an appropriate minimum passing score for each examination. It is extremely important to set the cut score appropriately for each examination. If the cut score is set too high, qualified people will fail the examination. If the cut score is set too low, unqualified candidates will pass. The cut score is a policy judgment, but it must be defensible because of the societal and personal consequences that flow from it. For this reason, the NCSFBC uses Prometric's high level standard setting services, all of which meet the professional approval of measurement standards organizations. Prometric test development experts ensure that the established cut score is reasonable and is based on qualified judgment and empirical evidence.

The standard setting study panel members come from a diverse mix of qualified professionals. All panel members are familiar with the job's knowledge requirements but are different than those subject matter experts who participated in item writing or review for the evaluated examination. The NCSFBC has approved a modified Angoff approach to standard setting. This process requires several rating levels where recommendations for the cut score are computed. The subject matter experts collective average cut score recommendation is also computed, as is the standard error of judgment and Beuk adjustments¹ associated with the panel recommended cut. At the conclusion of the meeting, the data is checked for accuracy and analyzed for rating errors of measurement, reliability of the judges, and standard errors of judgments (SEJ). In addition, a Beuk compromise adjustment analysis is performed to evaluate whether the standard produced from the Angoff procedure is reasonable and consistent with the panel's expectations of examinee ability. Six adjustments are based on the SEJ and one adjustment is based on the compromise procedure. Ultimately the final score is established by the (Board's) exam committee based on the presented psychometric data to ensure the decision comes from those who have the fiduciary responsibility to the organization and the public for the veracity of the credential. The current cut score used on the two active forms of the examination is 62% of scored items.

Frequent Item Analysis – Keeping exams valid

A key contributor to examination validity is regular analysis of exam-level and item-level statistics. Prometric's staff of expert psychometricians routinely analyze the NCSF-CPT examination items and produce detailed item analysis reports for the NCSFBC. The analysis is used during test review and development meetings and is presented by the committee as a set line item on the Board's agenda during annual meetings. There are four main objectives of classical item analysis:

1. Confirm that each item has an expected and appropriate level of difficulty.
 2. Determine the degree to which performance on each item is an indicator of performance on the overall test (discrimination).
-

3. Determine if candidates are selecting or constructing an answer to the item in a way that is consistent with a well-formed test item.
4. Establish item parameter estimates for assigning items appropriately to test forms.

Psychometricians perform item analysis to generate statistics that assess item and exam performance against these objectives. These statistics allow the analysts to observe item-level and exam-level characteristics of NCSFBC exams, such as:

- the proportion of candidates answering each question correctly;
- the correlation between the question score (correct or incorrect) and the total test score;
- the correlation between distracters and the total test score;
- the average score for the total test and each of its subsections;
- the pass ratio for each test; and
- the reliability of each test

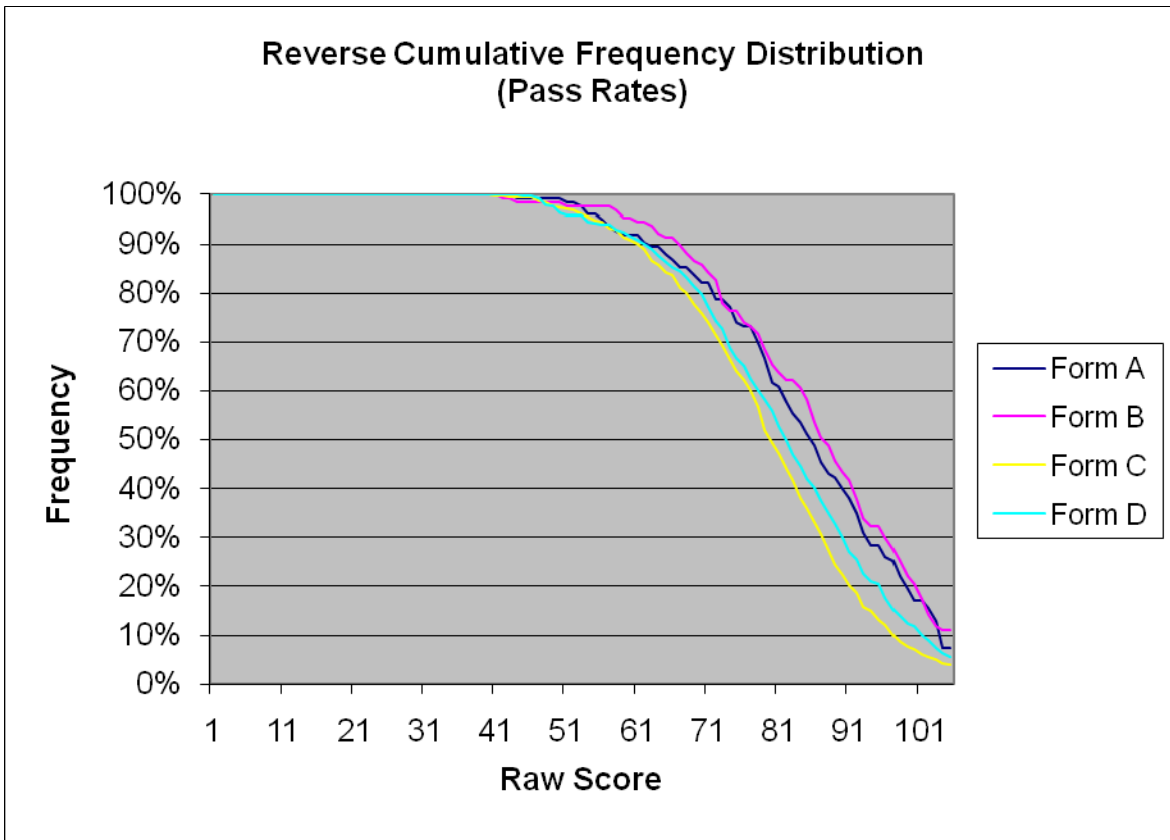
Psychometricians perform high level computer analysis on each item and generate a number of automated “flags” that are triggered when statistics reach predetermined threshold values that indicate an item requires attention. In this manner items can be removed from a test and replaced by new research items that 1) have demonstrated proper performance, 2) are of equal difficulty rating, and 3) that are consistent with the content domain on the test blueprint; thereby maintaining valid and reliable examinations.

Item and Examination Performance

Prometric provides the examination committee regular in-service analysis reports which outline score distributions (frequencies) as well as summary statistics that help describe the performance of items and exams over a given period. The NCSFBC reports these values to the National Commission for Certifying Agencies to ensure the standard industry statistics comply with statistical performance expectations for accreditation renewal each year. Statistics reported may include:

- the number of candidates tested;
 - observed and possible raw score ranges;
 - raw and scaled score means and standard deviations;
 - internal consistency and pass/fail classification consistency reliability coefficients;
 - scaled standard error of measurement at the passing score;
 - information regarding the speediness of the test; and
 - sub score statistics, such as:
 - number of items;
 - observed score range; and
 - average item difficulty
-

Prometric’s statistical system securely maintains all the data, so analyses can be run for all forms of each exam as well as on the individual items. Item performance is maintained separately so that a variety of item level statistics can be generated, which allows informed decisions regarding the continued use of items and to target time for equating blocks to be made as updates are made and new forms enter circulation. The goal is to consistently offer only valid and reliable forms that perform with clear predictability and that accurately assess candidates regardless of the form used for competency assessment. The NCSF runs two forms of the exam per cycle and as demonstrated below the exams’ performance mirror the partnered assessment instrument. In this case Forms A and B are partnered as are Forms C and D. The close proximity of the lines suggests similarity in test outcomes.



Forms	n candidates	Mean	SD	Alpha	SEM	Pass Rate
Form A	123	83.11	15.41	0.908	4.68	69%
Form B	127	84.86	15.35	0.908	4.67	71%

The chart above represents a sample of performance between the forms retired in 2008. These forms were replaced with new forms of the exam that were equated with statistically valid test items by content section. The examination committee is presented with the overall data to keep exams consistent between test takers but each exam must also be reviewed at the item level. Psychometricians provide a list of items that present data that merits additional review based on candidate performances along with the actual item display table. An example of an item table is provided below. A number of other statistics related to the item’s performance is provided in graphical and tabular format, and different indices related to the difficulty and discrimination of the item are also provided to ensure a discriminate decision-making process.

Statistics for Item 1 relative to the Total Raw Score

Alt.	Lowest	Low	Medium	High	Highest	%	N	Mean Scr	Biserial
1	0.37		0.29		0.20	28.8%	673	380.5	-0.22
2	0.16		0.07		0.04	9.1%	213	359.7	-0.34
3	0.34		0.51		0.70	51.6%	1204	413.1	0.38
4	0.13		0.12		0.06	10.4%	243	382.2	-0.13
OM	0.00		0.00		0.00	0.1%	2	404.0	0.04
Mn Scr	330.38		398.27		459.87				
Grp N	790		775		770				

Test Form Summary

In 2008, four forms were used to assess competency for the Certified Personal Trainer credential. The statistics associated with the exam forms suggest that the quality of the NCSF-CPT examinations comply with the psychometric rigors that discern effective testing instruments. The following data was reported to the National Commission for Certifying Agencies in compliance with annual reporting requirements and represent internal reliability estimates for all candidates taking the exams.

N = CANDIDATES TESTED FOR EACH FORM OF A CERTIFICATION EXAM IN THE REPORTING YEAR	Percent of Candidates Passing in 2008	Passing Point	Average Score	Standard Deviation	Standard Error of Measurement	Reliability Estimate
1,156	61	77	81.52	14.94	4.76	.90

Conclusion

Quality certification programs promote high standards in assessment and continued learning to ensure positive outcomes for all stakeholders of the profession. The NCSFBC serves its mission by challenging fitness professionals with a valid certification program aimed at raising the competency and subsequent ability of the professionals. This suggests holding professionals to higher standards and expecting more from the profession. In return for this effort the NCSFBC serves a strong advocacy roll, provides resources for professional success and represents the best interests of the profession.



¹ *Beuk, C. H. (1984). A method of reaching a compromise between absolute and relative standards in examinations. Journal of Educational Measurement, 21, 147-152.*

² *Educational Testing Service (2000) ETS Standards for Quality and Fairness. Princeton, NJ: Author.*

