

National Council for Strength & Fitness
Certified Personal Trainer Examination



Annual Report

January 1 to December 31, 2014

January 9th, 2015



Prometric Confidential Information

Unpublished Statistical Report

BACKGROUND

The National Council on Strength and Fitness (NCSF) is a professional, member-driven, education and credentialing organization for personal trainers and exercise science professionals. The NCSF is committed to serving the public through research, service, and advancement of the certified personal trainer profession.

The NCSF sponsors the Certified Personal Trainer examination. The purpose of this report is to document the test and item analysis performed by Prometric Test Development Services in an effort to evaluate the psychometric quality of the examination for the year 2014.

COMPREHENSIVE TEST DEVELOPMENT

In cooperation with Prometric's test development experts, the NCSFBC develops and administers a legally defensible, psychometrically sound examination that follows universally proven procedures summarized in the table 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 &	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	Maintain Exams	Conduct ongoing analysis of item and test statistics. Revise test with update periodically.

TEST DESIGN: CONDUCTING A JOB ANALYSIS STUDY TO DETERMINE TEST SPECIFICATIONS

NCSFBC test design starts by conducting a job analysis. A job analysis refers to procedures designed to obtain descriptive information about the tasks performed on a job and the important knowledge and/or skills needed to adequately perform those tasks. For purposes of developing NCSFBC examinations, the job analysis identified important tasks, knowledge, and/or skills necessary for competent performance. Job analysis is also a highly appropriate and useful methodology to gather information that can inform continuing education and professional development efforts.

ITEM DEVELOPMENT: PROVEN PROCESSES FOR DEVELOPING PROPER QUESTIONS

The NCSFBC and Prometric works together to write meaningful examination items and construct valid test forms according to the approved test specifications. Prometric conducts item writing training using an in-person workshop and web conference whereby a Prometric test developer assists NCSF subject-matter experts in writing exam items to achieve the following outcomes:

1. The option indicated as the item key has been correctly identified.
2. The language of each item is clear and unambiguous.
3. The item is appropriately classified in terms of the test plan or blueprint; and validations (references) have been provided.
4. Items are appropriately difficult and not encumbered with irrelevant sources of difficulty (such as inappropriately complex sentence construction and difficult vocabulary).
5. Items are free from unintended content inaccuracies.
6. Identify language, symbols, words, phrases, or examples that can be regarded as sexist, racist, or otherwise potentially offensive, inappropriate, or negative toward any group. Additionally, each item is reviewed for possible bias in language or social context.

PSYCHOMETRICALLY SOUND TEST CONSTRUCTION PROCESSES YIELD VALID EXAMS

NCSFBC approved examination items are assembled into multiple test forms in accordance with the pre-determined test specifications, ensuring that the appropriate number of items from each knowledge, skill or ability area is incorporated in to each form. As one or more test forms are assembled, pretest and operational test item data is used to anticipate the statistical characteristics of each form. This data is used to verify adequate test functioning 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.

FREQUENT ITEM ANALYSIS WILL KEEP EXAMS SOUND

A key contributor to examination validity is regular analysis of exam-level and item-level statistics. Prometric's staff of expert psychometricians analyze NCSFBC examination items and produce detailed item analysis reports for test review and development 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.

Item analysis generates statistics that assess item and exam performance against the above objectives. These statistics allow our analysts to observe item-level and exam-level characteristics, 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.
- The reliability of each test.

STANDARD SETTING IS A PROVEN PROCESS TO DETERMINE A PASSING SCORE FOR EACH EXAM

The NCSFBC establishes and validates an appropriate minimum passing score for each examination using the Modified Angoff and Beuk Relative-Absolute Compromise methodologies for standard setting. 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. The NCSFBC ensures that the established cut score is reasonable and is based on qualified judgment and empirical evidence.

THE CERTIFIED PERSONAL TRAINER EXAM

The Certified Personal Trainer exam is a computer based test. The candidates need to complete the examination within 180 minutes. Two forms (Form G and Form H) were in the field in 2014. Each form consisted of 150 multiple-choice items, including 125 operational, scored items and 25 pretest, unscored items. The cut score to pass the examination is 77 (i.e., 62% of the total operational items).

TEST FORM ANALYSIS

Table 1 provides the summary statistics of the NCSF examination for each form. The table includes the total number of candidates, pass rates for forms, the number of scored (operational) items in the examination, the score range (i.e., minimum and maximum total raw score), the median score, the mean score, the standard deviation of scores, and the skewness and kurtosis of the score distribution. The overall proportion of candidates passing the NCSF CPT examination (on both forms combined) in 2014 was 0.605 or 60.5% given the cut score of 62% of the total items.

The skewness indicates the degree of asymmetry in the distribution of scores. A positive value indicates that the tail of the distribution stretches toward higher scores; a negative value indicates that the tail extends toward the lower scores. The kurtosis indicates the degree of peakedness in a distribution of scores. The Pearson Kurtosis of a normal distribution is 3.0. As Figure 1 and Figure 2 also show, the score distribution of the NCSF CPT examination forms are slightly negatively skewed (-0.24 and -0.20), but close to normal, and are also close to normal with respect to kurtosis (3.17 and 2.32).

Table 1: NCSF CPT Exam Summary Test Statistics, January 1, 2014 — December 31, 2014

	Form G	Form H
Number of candidates	708	724
Proportion passing	0.62	0.59
# of operational Items	125	125
Maximum score	121	117
Median score	81	81
Minimum score	7	37
Mean score	81.41	79.61
Standard Deviation of scores	16.62	17.56
Skewness	-0.24	-0.20
Kurtosis	3.17	2.32
<u>Summary Item Statistics</u>		
Mean Item Difficulty (P+)	0.65	0.64
St. Dev. of Item Difficulty	0.13	0.13
Mean Item Discrimination (Point Biserial)	0.29	0.30
Mean Item Discrimination (Biserial)	0.38	0.40
St. Dev. of Item Discrimination	0.12	0.13

Table 1 also includes the mean and the standard deviation for the item difficulty index (P+) and item discrimination (point-biserial correlation and biserial) for each form. The mean P+ is the average of the proportions of candidates answering the items correctly averaged across all items included in the score. The standard deviation P+ is the standard measure of dispersion of P+ values around the mean P+.

The point-biserial correlation is the Pearson Product-Moment correlation. It correlates how candidates score on individual dichotomously-scored (correct or incorrect) items with how they score on the exam overall, so it is called an item-total correlation and is an indication of how well individual items discriminate between higher ability and lower ability candidates. A high positive point-biserial correlation suggests that candidates who performed well on the item also performed well overall, while candidates who did not perform well on the item did not perform well overall. The mean point-biserial

correlation is the average of the item-total correlations averaged across all items included in the score. Biserial correlation is another kind of item-total correlation that is used with a dichotomized variable (correct vs. incorrect item scores) and a continuous variable (total scores). It assumes the continuous variable is normally distributed, tends to be systematically larger than the point-biserial correlation, and differs from the point-biserial correlation more at the extremes of the distribution. The standard deviation of a biserial correlation is the standard measure of dispersion of biserial correlations around the mean biserial correlation.

Figure 1: NCSF CPT Exam Form G Score Frequency Distribution 2014

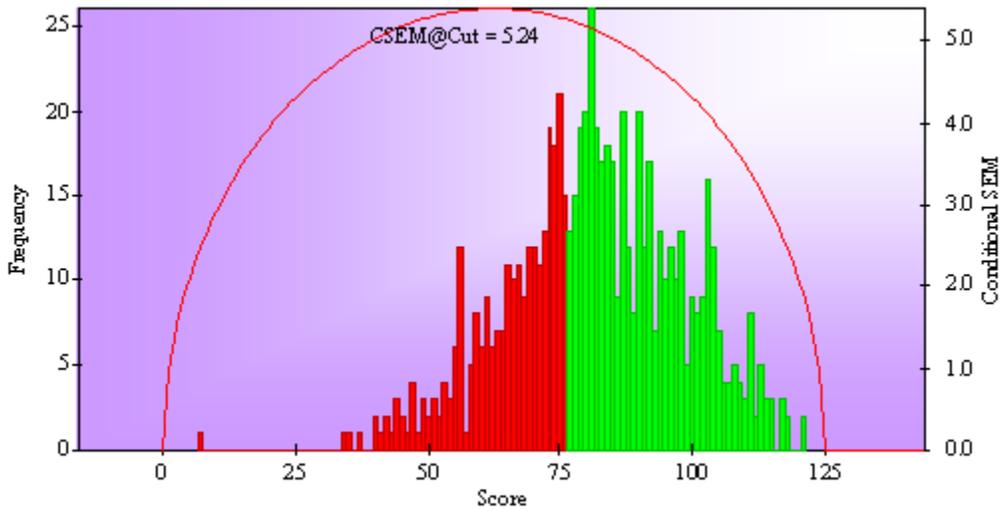
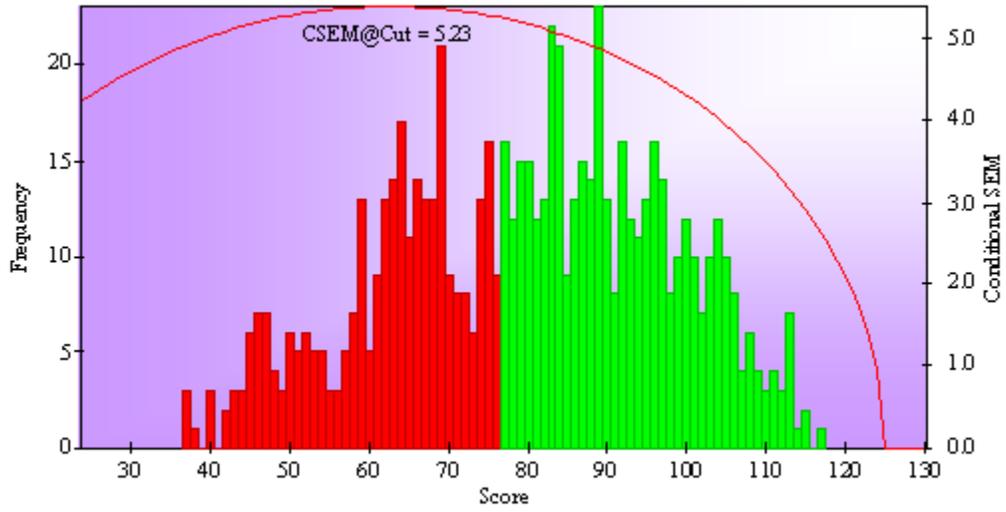


Figure 2: NCSF CPT Exam Form H Score Frequency Distribution 2014



RELIABILITY AND STANDARD ERROR OF MEASUREMENT

Table 2 below presents the internal consistency reliability estimate (KR-20) and the associated standard error of measurement and standard deviation for each form in total and for each of the sub domains.

Internal consistency reliability estimates were computed using Cronbach's coefficient alpha (Cronbach, 1951) for the total forms and for each of the sub-domains. The SEM is an estimate of the standard deviation of the distribution of observed scores around the true score. The SEM can be interpreted as an index of expected variation occurring if the same examinee was tested repeatedly on different forms of the same test without benefiting from practice or being hampered by fatigue.

Table 2: CPT Form G & H Total and Domain Score Reliability Statistics

Domain	# Items	Form G			Form H		
		Reliability	SD	SEM	Reliability	SD	SEM
Functional Anatomy	19	0.655	3.23	1.90	0.689	3.43	1.92
Exercise Physiology	12	0.545	2.29	1.54	0.597	2.39	1.52
Physical Activity and Health Promotion	7	0.462	1.57	1.15	0.512	1.56	1.09
Screening, Evaluation and Professional Practice	14	0.480	2.28	1.64	0.465	2.29	1.68
Nutrition	14	0.554	2.42	1.62	0.67	2.76	1.58
Weight Management	9	0.577	1.96	1.27	0.564	1.98	1.31
Exercise Prescription and Programming Considerations	24	0.693	3.93	2.18	0.672	3.84	2.20
Training Instruction	20	0.595	3.11	1.98	0.628	3.23	1.97
Considerations for Special Populations	6	0.259	1.3	1.12	0.388	1.42	1.11
Total Form Reliability Statistics	125	0.912	16.62	4.93	0.921	17.56	4.94

DECISION CONSISTENCY

Table 3 below presents the decision consistency statistics for CPT forms G & H.

Decision consistency estimates the extent to which classifications based on test scores would match the decisions based on scores from a second, parallel form of the same test. The single decision consistency estimates using formula presented by Subkoviak (1978) and the conditional SEM (CSEM) at the cut score are reviewed.

The Subkoviak statistic provides an estimate of the decision consistency of a pass/fail decision. Using the passing score selected, it estimates the probability that an individual would receive the same decision on two separate administrations of the test. The Subkoviak C statistic uses KR-20 (Alpha) as a reliability estimate and assumes that scores are characterized by the compound binomial distribution. The Subkoviak C is estimated as suggested by Lord and Novick (1968, p. 525).

Table 3: CPT Form G & H Decision Consistency Statistics

Form	Subkoviak C	CSEM at Cut
CPT Form G	0.85	5.24
CPT Form H	0.88	5.23

SUMMARY OF STATISTICAL ANALYSIS

The overall passing rates for the NCSF CPT exam in 2014 are about 60.5%. The average total raw score of the NCSF CPT exam is 81.41 for Form G and 79.61 for Form H, respectively. The standard deviation of the total raw score ranged from 16.62 in Form G to 17.56 in Form H. The reliability coefficients of the NCSF CPT Exam forms in 2014 are both above 0.90 and the SEMs for the two forms showed to be stable and acceptable from the comparison with the previous year's outcomes.

References

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company, Inc.

Subkoviak, M. (1978). Estimating reliability from a single administration of a criterion-referenced test. *Journal of Educational Measurement*, *13*(4), 7-10.