National Council on Strength & Fitness

Certified Personal Trainer Examination

Certified Strength Coach Examination

Annual Report Report

January 1 to December 31, 2019

May 20th 2020

BACKGROUND

The National Council on Strength and Fitness (NCSF) is a professional, member-driven, education and credentialing organization for personal trainers, exercise science as well as strength and conditioning professionals. The NCSF is committed to serving the public through research, service, and advancement of the exercise 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 Solutions in an effort to evaluate the psychometric quality of the examination for the year 2019.

COMPREHENSIVE TEST DEVELOPMENT

In cooperation with Prometric Test Development Solutions, the NCSF develop and administers a legally defensible, psychometrically sound examination. The processes and procedures used to develop and maintain these exams are 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 and revise test periodically

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

NCSFBC test design process starts by conducting a job analysis. Job analysis is designed to determine the tasks performed on a job as well as the critical knowledge and/or skills needed to adequately perform those tasks. For purposes of developing NCSF examinations, the job analysis identified important tasks and knowledge necessary for competent performance as a personal trainer. Job analysis is also a highly appropriate and useful method for gathering information to inform continuing education and professional development efforts.

ITEM DEVELOPMENT: VALID PROCESSES FOR DEVELOPING TEST ITEMS

The NCSF and Prometric work together to write relevant examination items and construct valid test forms according to the approved test specifications. Prometric test developers assists NCSF subject-matter experts in writing and reviewing exam items to achieve the following outcomes:

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

PSYCHOMETRICALLY SOUND TEST CONSTRUCTION PROCESSES FOR VALID EXAMS

NCSF 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.

NCSF Form Assembly and Item Selection Criteria
After the completion of a standard setting, the cut scores for subsequent NCSF (CPT and CSC) exam forms are determined using a section pre-equating method.

Section Pre-Equating Form Assembly Criteria
With section pre-equating, the new form(s) are equated back to the old (base) form with which they have items in common. Prometric psychometric staff members assemble the new forms with a common equating item block of 40% of total operational items from the corresponding base form. Other operational items are selected from the bank of useable items with statistics from prior administrations.

To maintain the reliability and performance of the new forms all operational items are excluded from selection if their item statistics fall outside the following ranges:

a. A point-biserial value less than or equal to 0.15
b. A p-value of less than or equal to 0.25
c. A p-value greater than 0.96

The following item selection criteria are adopted to maintain the comparability across forms:

a. For the common equating block, items are selected so that
    1) they are proportional to the content blueprint and
    2) the distribution of item difficulty follows that of the base form.

b. All operational items (including equater items) are selected to meet the content blueprint. Conscious efforts are made so that the distribution of the item statistics matches that of the base form. The average difficulty of operational items and standard deviation are close to those of operational items in base form.

The common block's average difficulty and standard deviation are the same or close to those of operational items on the base form. The resulting equating block can be treated as a "mini-set" of the operational items from the base form.

THE CERTIFIED PERSONAL TRAINER EXAM

The Certified Personal Trainer exam (CPT) is a computer based test. Candidates need to complete the examination within 180 minutes. A total of two forms (i.e., Form K and Form L) were in the field in 2019. These forms consisted of 150 multiple-choice items each, 125 of which were scored and 25 were unscored. The cut scores to pass the exam were 72 on Form K and 73 on Form L.

TEST FORM ANALYSIS

Table 1 provides the summary statistics of the NCSF CPT 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 all forms combined) in 2019 was 0.7795 or 77.95% given the cut scores.

Table 1: NCSF CPT Exam Summary Test Statistics, January 1, 2019 — December 31, 2019

	Form K	Form L
Number of candidates	1102	998
Proportion passing	0.774	0.786
# of operational Items	125	125
Maximum score	123	125
Median score	93	94
Minimum score	27	32
Mean score	89.18	90.26
Standard Deviation of scores	21.28	20.03
Skewness	-0.49	-0.49
Kurtosis	2.28	2.32
Summary Item Statistics		
Mean Item Difficulty (P+)	0.71	0.72
St. Dev. of Item Difficulty	0.13	0.13
Mean Item Discrimination (Biserial)	0.54	0.52
St. Dev. of Item Discrimination	0.16	0.17

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. The score distributions for each form are shown in Figures 1 and 2.

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 difficulty index indicates the proportion of candidates that answered the item correctly. 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 K Score Frequency Distribution (January 1, 2019 – December 31, 2019)

Max. Freq.	28
Cases	1102
Items	125
Maximum	123.00
Median	93.00
Minimum	27.00
Mean	89.18
SD	21.28
Skew	-0.49
Kurtosis	2.28
Alpha	0.956
SEM	4.47
Subkov. C	0.934
Passing	0.774

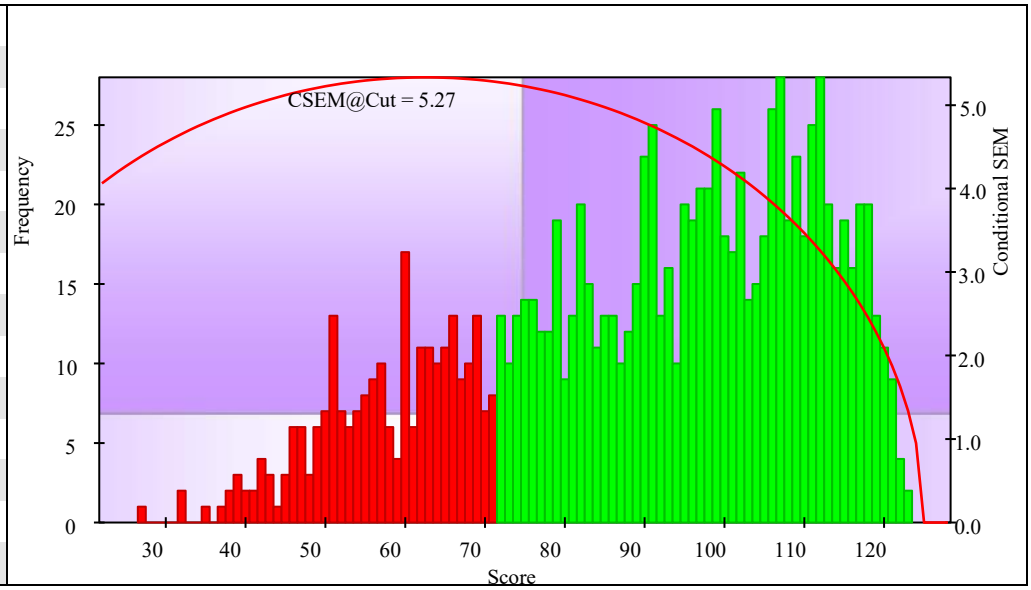
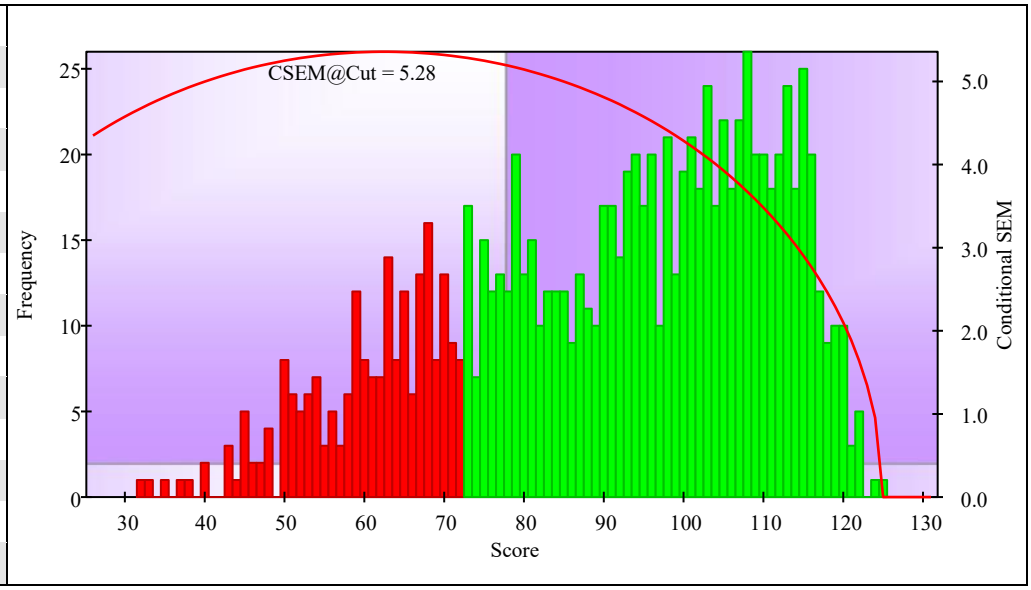


Figure 2. NCSF CPT Exam Form L Score Frequency Distribution (January 1, 2019 - December 31, 2019)

Max. Freq.	26
Cases	998
Items	125
Maximum	125.00
Median	94.00
Minimum	32.00
Mean	90.26
SD	20.03
Skew	-0.49
Kurtosis	2.32
Alpha	0.950
SEM	4.48
Subkov. C	0.925
Passing	0.786



RELIABILITY AND DECISION CONSISTENCY STATISTICS

Internal consistency reliability estimates were computed using Cronbach's coefficient alpha (Cronbach, 1951) among each of the scales in the test, given by:

$$\alpha_X = \frac{n}{n-1}\left(1 - \frac{\sum S_i^2}{S_x^2}\right)$$

where *n* is the number of items, $\sum S_i^2$ is the sum of the item variances, and S_X^2 is the variance of score *X*. Cronbach's alpha is a generalization of the KR20 when data are dichotomous.

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.

The SEM of a raw score is computed from the reliability estimate (a_x) and the standard deviation (SD_x) of the scores by the formula:

$$SEM_X = SD_X\sqrt{1-\alpha_X}$$

Table 2 below presents the internal consistency reliability estimate (KR-20) and the associated standard error of measurement for the total Forms K & Land for each domain in each of the two forms and for the total.

Table 3 below reports the Decision consistency. Decision consistency measures the extent to which classifications based on test scores 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 (1976) and the conditional SEM (CSEM) at the cut score are reviewed.

Decision consistency estimates were computed using the Subkoviak statistic. 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 2. Reliability statistics for Forms K and L in total and by domain

Domain	# Items	Form K			Form L		
		Reliability	SD	SEM	Reliability	SD	SEM
Functional Anatomy	15	0.754	3.17	1.57	0.687	2.83	1.58
Exercise Physiology	10	0.684	2.15	1.21	0.672	2.24	1.28
Health and Physical Fitness	14	0.72	2.72	1.44	0.778	2.93	1.38
Screening and Evaluation	16	0.735	3.03	1.56	0.747	2.97	1.5
Nutrition	9	0.735	2.29	1.18	0.729	2.24	1.16
Weight Management	11	0.703	2.57	1.4	0.68	2.47	1.4
Exercise Programming	24	0.808	4.47	1.96	0.747	3.83	1.92
Training Instruction	18	0.694	3.15	1.74	0.607	2.97	1.86
Considerations for Special Populations	4	0.465	1.09	0.79	0.483	1.1	0.79
Professionalism and Risk Management	4	0.181	0.71	0.64	0.163	0.7	0.64
Total Form	**125**	**0.956**	**21.28**	**4.47**	**0.95**	**20.03**	**4.48**

Table 3. Decision Consistency for Forms K and L

Form	N	Decision Consistency (Subkoviak Index)	Conditional SEM at cut score
Form K	1102	0.93	5.27
Form L	998	0.92	5.28

SUMMARY OF STATISTICAL ANALYSIS

The overall passing rate for the NCSF CPT exam in 2019 was about 77.95%. The average total raw scores of the NCSF CPT exam were 89.18 for Form K, and 90.26 for Form L. The standard deviations of the total raw score were 21.28 and 20.03, respectively. The reliability coefficients of the NCSF CPT exam forms in 2019 were 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.

The total number of NCSF certified personal trainers at the time of this report is 8,487

References

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

Subkoviak, M. (1976). Estimating reliability from a single administration of a criterion referenced test. *Journal of Educational Measurement, 13(4)*, 7-10.

Appendix A

Subkoviak (1976) Documentation

Subkoviak 1976.pdf

CERTIFIED STRENGTH COACH

BACKGROUND

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

The NCSF sponsors the Certified Strength Coach examination. The purpose of this report is to document the test and item analysis performed by Prometric Test Development Solutions in an effort to evaluate the psychometric quality of the examination for the year 2019.

COMPREHENSIVE TEST DEVELOPMENT

In cooperation with Prometric Test Development Solutions, the NCSF develop and administers a legally defensible, psychometrically sound examination. The processes and procedures used to develop and maintain these exams are 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 and revise test periodically

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

NCSFBC test design process starts by conducting a job analysis. Job analysis is designed to determine the tasks performed on a job as well as the critical knowledge and/or skills needed to adequately perform those tasks. For purposes of developing NCSF examinations, the job analysis identified important tasks and knowledge necessary for competent performance as a strength coach. Job analysis is also a highly appropriate and useful method for gathering information to inform continuing education and professional development efforts.

ITEM DEVELOPMENT: VALID PROCESSES FOR DEVELOPING TEST ITEMS

The NCSF and Prometric work together to write relevant examination items and construct valid test forms according to the approved test specifications. Prometric test developers assists NCSF subject-matter experts in writing and reviewing exam items to achieve the following outcomes:

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

PSYCHOMETRICALLY SOUND TEST CONSTRUCTION PROCESSES FOR VALID EXAMS

NCSF 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.

ITEM ANALYSIS FOR EXAM MAINTENANCE

A key contributor to examination validity is regular analysis of exam-level and item-level statistics. Prometric's staff of expert psychometricians analyze NCSF 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 PROCESS TO DETERMINE A PASSING SCORE FOR EACH EXAM

The NCSF 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 NCSF ensures that the established cut score is reasonable and is based on qualified judgment and empirical evidence.

THE CERTIFIED STRENGTH COACH EXAM

The Certified Strength Coach exam is a computer-based test. Candidates need to complete the examination within 180 minutes. One form was in the field in 2019. This form consisted of 150 multiple-choice items, 125 of which were scored and 25 were unscored. The cut score to pass the examination was 75 of operational items.

TEST ANALYSIS

Table 1 provides the summary statistics of the NCSF CSC examination. The table includes the total number of candidates, pass rate, 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 proportion of candidates passing the NCSF CSC examination in the reporting period 2019 was 0.576 or 57.6% given the cut score (percentage cut = 60%; raw cut = 75).

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. A perfectly normal distribution has a kurtosis value of 3.0 and skewness value of 0. Figure 1 shows that the score distribution of the NCSF CSC examination is close to normal (i.e., skewness is 0.07, and kurtosis is 2.77).

Table 1. NCSF CSC Exam Summary Test Statistics (January 1, 2019 - December 31, 2019)

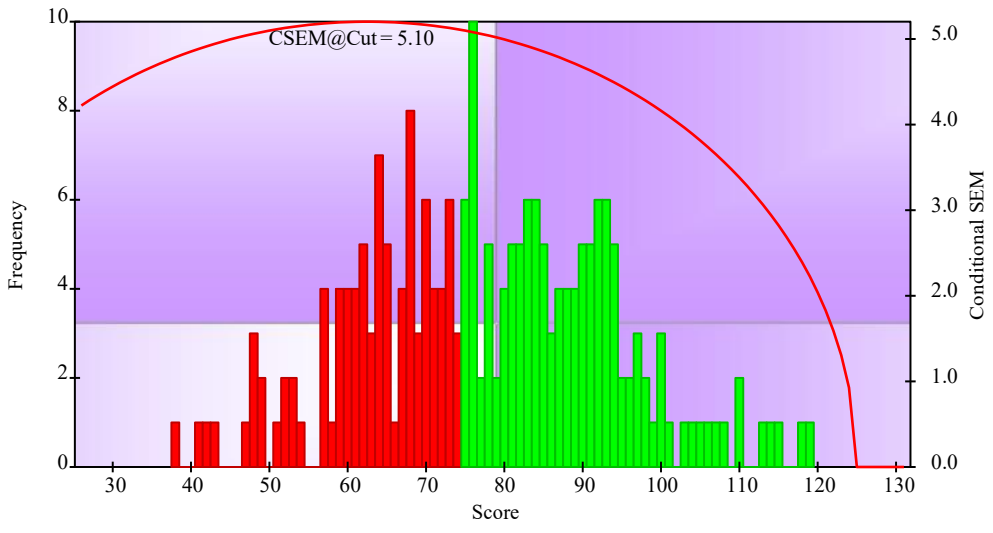
	CSC1
Number of candidates	217
Proportion passing	0.58
# of operational Items	125
Maximum score	119
Median score	77
Minimum score	38
Mean score	77.68
Standard Deviation of scores	15.84
Skewness	0.07
Kurtosis	2.77
Summary Item Statistics	
Mean Item Difficulty (P+)	0.62
St. Dev. of Item Difficulty	0.18
Mean Item Discrimination (Biserial)	0.38
St. Dev. of Item Discrimination	0.15

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 difficulty index indicates the proportion of candidates that answered the item correctly. 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 CSC Exam Score Frequency Distribution (January 1, 2019 – December 31, 2019)

Max. Freq.	10
Cases	217
Items	125
Maximum	119.00
Median	77.00
Minimum	38.00
Mean	77.68
SD	15.84
Skew	0.07
Kurtosis	2.77
Alpha	0.906
SEM	4.87
Subkov. C	0.850
Passing	0.576



RELIABILITY AND DECISION CONSISTENCY STATISTICS

Internal consistency reliability estimates were computed using Cronbach's coefficient alpha (Cronbach, 1951) among each of the scales in the test, given by:

$$\alpha_X = \frac{n}{n-1}\left(1 - \frac{\sum S_i^2}{S_x^2}\right)$$

where n is the number of items, $\sum S_i^2$ is the sum of the item variances, and S_X^2 is the variance of score X. Cronbach's alpha is a generalization of the KR20 when data are dichotomous.

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.

The SEM of a raw score is computed from the reliability estimate (a_x) and the standard deviation (SD_x) of the scores by the formula:

$$SEM_X = SD_X\sqrt{1-\alpha_X}$$

Table 2 below presents the internal consistency reliability estimate (KR-20) and the associated standard error of measurement for the total Form CSC1 and for each domain and for the total.

Table 3 below reports the Decision consistency. Decision consistency measures the extent to which classifications based on test scores 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 (1976) and the conditional SEM (CSEM) at the cut score are reviewed.

Decision consistency estimates were computed using the Subkoviak statistic. 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 2. Reliability statistics for CSC exam in total and by domain

Domain	# Items	CSC1		
		Reliability	SD	SEM
Functional Anatomy and Biomechanics	19	0.656	3.07	1.8
Sport Metabolism	13	0.448	2.21	1.64
Performance Assessment and Evaluation	16	0.648	2.86	1.69
Nutrition and Ergogenic Aids	10	0.532	1.79	1.23
Advanced Programming for Sport	25	0.601	3.59	2.27
Training Techniques for Athletic Performance	25	0.653	3.75	2.21
Injury Prevention and Return to Play	11	0.464	1.99	1.45
Professionalism and Risk Management	6	0.283	1.29	1.09
Total Form	**125**	**0.906**	**15.84**	**4.87**

Table 3. Decision Consistency for CSC exam

Form	N	Decision Consistency (Subkoviak Index)	Conditional SEM at cut score
CSC1	217	0.85	5.1

The passing rate for the NCSF CSC exam in 2019 was about 57.6%. The average total raw score of the NCSF CSC exam was 77.68, and the standard deviation of the total raw score was 15.84. The reliability coefficients of the NCSF CSC exam in 2019 was .906 and the SEM for the exam was appeared to be acceptable.

The total number of NCSF certified strength coachs at the time of this report is 465

References

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

- Subkoviak, M. (1976). Estimating reliability from a single administration of a criterion referenced test. *Journal of Educational Measurement, 13(4)*, 7-10.

Appendix B

Subkoviak (1976) Documentation

Subkoviak 1976.pdf