

GUIDE
FOR INTERPRETING RESULTS
and
TECHNICAL DATA
for the
NATIONAL AUTOMOTIVE STUDENT SKILLS
STANDARDS ASSESSMENT



Prepared by
NATIONAL INSTITUTE FOR
AUTOMOTIVE SERVICE EXCELLENCE (ASE)

2009

PREFACE

This guide contains information on interpreting your students' results on the National Automotive Student Skills Standard Assessment (NA3SA). Also, it includes documentation of the technical adequacy of the assessment program for its intended purposes.

The Automotive Youth Educational Systems (AYES), the National Automotive Technicians Education Foundation (NATEF), and SkillsUSA jointly offer the NA3SA assessment. These tests, developed by the National Institute for Automotive Service Excellence (ASE), are appropriate for evaluating students who are near the end of their studies in the areas of Automobile Service and Repair, and Collision Repair and Refinish.

NATEF administers the industry's certification program for career-entry Automobile, Collision Repair and Refinish, and Medium/Heavy Duty Truck Training Programs. The standards for becoming an ASE (NATEF) certified program include specifications covering the content of instruction, tools and equipment, hours, and instructor qualifications. Concurrently, the National Institute for Automotive Service Excellence (ASE) conducts periodic analyses of the tasks and knowledges required to successfully perform many of the vehicle service jobs in the automotive industry. NATEF policy stipulates that the task lists developed by ASE are to serve as the basis for the NATEF task lists. In this way, the content of the NATEF exams are kept current, consistent with ASE, and linked to the specific tasks and knowledges requisite to the successful performance of the various automotive service occupations.

The examinations are intended for students completing two-year secondary or post-secondary automotive technician training programs.

Questions pertaining to this program should be directed to NA3SA, c/o NATEF at 101 Blue Seal Drive, SE, Suite 101, Leesburg, VA 20175. Phone 800-362-0544. Or, go to www.na3sa.com for more information.

TABLE OF CONTENTS

PREFACE	2
TABLE OF CONTENTS	3
NATIONAL AUTOMOTIVE STUDENT SKILLS STANDARDS ASSESSMENT	4
Description of the Battery	4
Automobile	4
Collision Repair And Refinish	4
Test Development Procedures	4
Content Specifications.....	4
Question Writing.....	4
Test Assembly	5
INTERPRETING RESULTS.....	5
Performance Comparisons	5
Comparing Your Students to Another Group.....	5
Comparing Means.....	6
SCORE REPORTS	6
Who Gets Reports.....	6
Score Reports Retention and Replacement	6
Automobile Percentile Rank Table – Spring 2009	7
How To Use This Table.....	7
Collision Repair and Refinishing Percentile Rank Table – Spring 2009	8
How To Use This Table.....	8
TECHNICAL DATA – GLOSSARY OF TERMS.....	9
N of Items	9
Mean.....	9
Variance and Std. Dev. (Standard Deviation).....	9
Alpha (Coefficient Alpha, or Test Reliability)	9
SEM (Standard Error of Measurement).....	10
Mean Pcnt Corr (Mean Percent Correct, or Item Difficulty).....	10
Mean Biserial (Item Discrimination)	10
Validity	10
NA3SA Test Form Statistics - Spring 2009.....	11

NATIONAL AUTOMOTIVE STUDENT SKILLS STANDARDS ASSESSMENT

Description of the Battery

The NA3SA assessment consists of two series of secure multiple-choice examinations: Automobile Service and Repair, and Collision Repair and Refinish.

Automobile

Suspension and Steering
Brakes
Electrical/Electronic Systems
Engine Performance
Engine Repair
Automatic Transmission/Transaxle
Manual Drive Train and Axles
Heating and Air Conditioning

Collision Repair And Refinish

Painting and Refinishing
Structural Analysis and Damage Repair
Non-structural Analysis and Damage Repair
Mechanical and Electrical Components

Each series is comprised of individual tests that relate to one or more of the technical areas under the NATEF Standards. Students may be assigned a single examination, all examinations, or any combination of them. The examinations emphasize the application of knowledge and theory to tasks actually performed by automotive technicians.

The examinations are currently administered twice annually, in the Fall and in the Spring. Separate student score reports are prepared for each of the examinations. There are 40 scored questions in each examination, but the tests as given will be longer because of the inclusion of nonscored "pretest" questions. Administration time is recommended to be 60 minutes per exam. Each student will be given a pass/fail status on each test attempted. For each test passed, students earn a National Automotive Student Skills Standards Assessment certificate.

Test Development Procedures

Content Specifications

ASE periodically conducts analyses of the work of the automotive technician in the various subject areas. Job Analysis Workshops involving subject matter experts from around the country are convened specifically for this purpose. The task lists contained in the standards for NATEF's certification program are tied to ASE's task lists derived from these job analyses. The task lists are then organized into content outlines. These subject areas are then weighted according to judgments of frequency and criticality, and these weights are translated into numbers of questions in each content area. This provides the content specifications for the examinations. As described earlier, the task lists are designed to correspond to the tasks required to successfully perform the various automotive service procedures.

Question Writing

Items (test questions) are written by groups of subject matter experts (SME's) who are selected and trained by the ASE staff. The item writing teams include faculty members of educational institutions as well as experienced, working automotive technicians.

After the SME's draft the items and assign content codes, the items are reviewed by other SME's for accuracy. They are then edited, formatted, and entered into a permanent item bank. SME's then review and approve all the text changes.

Newly written items are tried out as nonscored "pretest" items embedded into the test forms. Data collected in this manner are then used to identify any items that may not function properly so that they can be rewritten or discarded if necessary. All data are banked with the item text in the item banks.

Test Assembly

Subject matter experts begin test assembly by selecting pretested items from the bank for each of the examinations. Items are selected to meet both content and statistical (performance) specifications. Items selected for the examinations are appropriately distributed among the NATEF tasks. Each form of the examination will sample the NATEF tasks, however not all tasks will be tested by each form of the examination. Relevant item statistics include discrimination (item-test correlation) indices that exceed 0.20 and a difficulty level (P-value) within the range of 0.20 to 0.90. Items with unsatisfactory statistics are discarded or rewritten.

Each annual form may contain a combination of pre-tested and new items. Before final scoring, statistical and content analysis is conducted on all items as a final check to detect flaws.

Passing Standards

Passing standards are individually set for each of the examinations. The determination of passing scores for high-stakes examinations like the NA3SA tests must be done systematically and with care. Several methods are possible, but the one chosen as most appropriate is called a contrasting-groups approach. This method is based on actual performance of real students, not judgments of how students are likely to perform. Criterion groups of "should-pass," "borderline," and "should-not pass" students are selected in advance of testing. These selections are made by instructors with detailed knowledge of the level of preparedness of the students. After testing, a passing score is selected that minimizes the false-positive and false-negative classifications in the obtained score distributions of these groups. Passing standards set this way are generally regarded by instructors and administrators as more appropriate and more realistic than test-based judgmental approaches. These same passing standards are then carried forward to future forms of the NA3SA Exams.

INTERPRETING RESULTS

The NA3SA score reports allow comparisons of a school's or student's performance with that of others participating in the program during the same year. Mean scores and pass/fail proportions are calculated for each of the examinations. These are reported at the instructor and school level. State reports comparing all the schools in a state are provided to the designated state level supervisor.

Performance Comparisons

Percentile Rank Tables

Following this narrative are tables of percentile ranks of the national population of examinees who took the current year's test forms in the spring administration. This is useful for comparing your students' performance to the national sample. Instructions for using the table are presented below each table.

Comparing Your Students to Another Group

The statistics reported for each administration are based upon the group taking the examinations in that testing period, and do not include prior year's administrations. Total group statistics are given for comparison purposes.

A critical issue is the extent to which the composition of your examinee group resembles that of any other group to which they are being compared. If population characteristics (e.g. age, amount of prior experience, etc.) account for differences between your students and another group, then the comparison may be of less use to you. You must make a judgment about any other characteristics that may contribute to differences in achievement, then decide how to interpret the comparison.

Comparing Means

Mean scores of groups can be compared if they were tested in the same year. However, the means of small groups can be expected to contain increased sampling error, and so should not be interpreted to accurately represent the performance of any larger population. For example, if only a few students from a school take a particular test, their performance should not be assumed to represent all the students in that school. Also, year-to-year differences between the means of groups, especially small groups, should be interpreted with caution. These statistics will include sampling error, as described above, plus error resulting from any differences in test form difficulty across administrations.

SCORE REPORTS

Who Gets Reports

Reports are prepared for students, instructors, and state supervisors. Student reports include number correct in each of the content areas, the total score, and pass/fail. The instructor report shows a summary of the information contained on that instructor's student score reports. Copies of the student reports are also provided to instructors. State reports summarize the results in terms of mean scores and pass/fail rates from each school in that state and are available to the designated state level supervisor.

Score Reports Retention and Replacement

All recipients, including students, are allowed to keep their score reports. The NA3SA partner organizations do not provide a records-maintenance service, so duplicate or replacement copies of these reports are not normally available.

Automobile Percentile Rank Table – Spring 2009

Number Correct	Suspension & Steering	Brakes	Elec/Elec Systems	Engine Performance	Engine Repair	Auto Trans & Transaxle	Manual Drive Train & Axles	Heating & A/C	Number Correct
0-5	1	1	1	1	1	1	1	1	0-5
6	1	1	1	1	1	1	1	1	6
7	1	1	1	2	1	1	2	2	7
8	2	2	2	3	2	1	3	3	8
9	3	3	3	4	3	2	4	5	9
10	5	5	5	6	4	3	6	8	10
11	7	7	7	8	5	5	8	11	11
12	9	11	10	12	7	8	10	16	12
13	13	15	13	15	9	10	13	20	13
14	17	19	16	19	12	14	17	24	14
15	21	23	21	24	15	17	22	28	15
16	26	28	25	28	18	21	26	32	16
17	31	34	29	33	21	24	30	37	17
18	37	40	35	39	26	28	34	41	18
19	43	45	40	44	31	32	39	46	19
20	48	50	45	50	35	37	44	51	20
21	54	56	51	54	40	42	48	55	21
22	60	61	56	59	45	46	51	59	22
23	66	66	62	63	49	52	56	63	23
24	72	71	66	67	53	57	60	67	24
25	77	75	72	71	58	62	64	72	25
26	81	78	76	75	63	68	68	76	26
27	85	82	81	79	67	72	73	81	27
28	89	85	85	83	72	77	78	84	28
29	91	88	88	86	76	81	82	87	29
30	94	91	91	89	81	85	84	90	30
31	96	93	93	92	84	88	88	92	31
32	98	95	95	93	88	92	91	94	32
33	99	96	97	95	91	95	93	96	33
34	99	97	98	97	94	96	95	97	34
35	99	98	99	98	97	98	97	99	35
36	99	99	99	99	99	99	98	99	36
37-40	99	99	99	99	99	99	99	99	37-40

How To Use This Table

A percentile is the percentage of students who scored at or below a given score point. To use the table, find the student's Number Correct score for a given test in the left (or far right) column, and then look over to that test's column to find the percentile equivalent. For example, if a student scored 25 correct on Engine Repair, first find 25 in the left column. Then look to the right under the Engine Repair heading, and you will find 58. Thus, a score of 25 on the Engine Repair test is at the 58th percentile of the national population of students who took this exam in the Spring of 2009.

Collision Repair and Refinishing Percentile Rank Table – Spring 2009

Number Correct	Non-structural Analysis & Damage Repair	Painting & Refinishing	Structural Analysis & Damage Repair	Mechanical & Electrical Components	Number Correct
0-5	1	1	1	1	0-5
6	2	1	2	1	6
7	2	1	2	2	7
8	2	1	2	4	8
9	2	1	4	4	9
10	2	1	6	5	10
11	3	2	8	7	11
12	4	3	10	9	12
13	6	4	13	11	13
14	8	5	15	13	14
15	11	7	16	17	15
16	14	9	19	24	16
17	18	13	22	28	17
18	22	15	28	37	18
19	27	18	31	41	19
20	32	24	35	46	20
21	40	28	38	54	21
22	45	33	44	57	22
23	51	38	51	64	23
24	56	44	54	70	24
25	62	50	62	74	25
26	69	54	66	75	26
27	75	63	69	77	27
28	80	68	74	81	28
29	85	75	79	88	29
30	90	81	82	91	30
31	92	86	85	94	31
32	95	90	88	96	32
33	98	93	92	97	33
34	99	95	95	98	34
35	99	98	97	99	35
36	99	99	99	99	36
37-40	99	99	99	99	37-40

How To Use This Table

A percentile is the percentage of students who scored at or below a given score point. To use the table, find the student's Number Correct score for a given test in the left (or far right) column, and then look over to that test's column to find the percentile equivalent. For example, if a student scored 25 correct on Non-structural Analysis and Damage Repair, first find 25 in the left column. Then look to the right under the Non-structural Analysis and Damage Repair heading, and you will find 62. Thus, a score of 25 on the Non-structural Analysis and Damage Repair test is at the 62nd percentile of the national population of students who took this exam in the Spring of 2009.

TECHNICAL DATA – GLOSSARY OF TERMS

ASE computes both item- and test-level statistics as well as candidate performance statistics separately for each form of each of the examinations. Following this narrative are the data tables for the current forms of the exams. The information below is intended to help interpret the technical data in these tables.

N of Items

This refers to the number of *scored* items (questions) in the test form. Students may have been administered more questions than this, however. ASE "pretests" newly written or revised questions by embedding them into test forms as non-scored items. Most often, test forms will contain about 10 non-scored pretest items.

Mean

The mean of a set of scores is commonly referred to as the average. This is the sum of all scores divided by the number of scores.

Variance and Std. Dev. (Standard Deviation)

These statistics convey the spread of a set of scores. The variance is the average of the squared deviations of the scores about the mean. The standard deviation (Std. Dev.) is a more easily interpretable statistic. It can be thought of as the average amount that scores differ from the mean score (although this definition is not precisely correct). When the standard deviation is larger the scores are more spread out. As a rule of thumb, about two-thirds of the scores of a sample are likely to fall within one standard deviation of the mean.

Skew

Distributions of scores are rarely bilaterally symmetrical. If the scores are clustered more to the left with a longer tail to the right, skew is positive. If the tail is longer to the left, skew is negative.

Kurtosis

Compared to the standard normal shape, score distributions may be more peaked (positive kurtosis) or more flat (negative kurtosis). Skew and kurtosis are included in these reports in the interest of completeness, but this information has limited practical value for the instructor in interpreting student scores.

Median

This is the test score above (and below) which 50% of the students scored. In other words, it is the middle score in the group of scores. Because score distributions are rarely perfectly symmetrical, the median will seldom be exactly equal to the Mean.

Alpha (Coefficient Alpha, or Test Reliability)

The measurement of any cognitive characteristic contains some degree of inconsistency or error. For example, an examinee taking parallel forms of the same examination would likely earn somewhat different scores on the two forms. These differences might be due to sources of error originating with the examinee, the testing environment, or the examination itself. Reliability as considered here refers to freedom from random error originating in the test itself.

The reliability coefficients reported for the NA3SA examinations are measures of internal consistency computed by the Coefficient Alpha formula (also known as KR-20 in the dichotomous case such as this). Reliability coefficients range from zero to one, with a value of one indicating perfect reliability. The size of a reliability coefficient is affected by several factors including the degree to which the test items are measuring the same cognitive construct and the number of items in the test. All other things being equal, longer tests generally have higher reliability.

SEM (Standard Error of Measurement)

Error of measurement results from unreliability and refers to random error associated with a test score. Such error may inflate or depress an examinee's score. As measurement error goes up, reliability goes down and the standard error of measurement goes up. The SEM represents the standard deviation of a theoretical distribution of obtained scores scattered about the theoretical true score of the candidate. As such, it is a function of both reliability and the standard deviation of test scores. Standard error of measurement may be thought of as a "margin of error" that can be used to express the degree of confidence in the accuracy of a test score.

Mean Pcnt Corr (Mean Percent Correct, or Item Difficulty)

The item difficulty, defined as the percentage of examinees answering the item correctly, is computed for each item. Items that are either too difficult (20% or lower) or too easy (90% or higher) are flagged and examined by subject matter experts for flaws. The mean item difficulty expressed as mean percent correct (Mean Pcnt Corr) is provided for each test form.

Mean Biserial (Item Discrimination)

This is the mean Biserial correlation between the selection of the correct option and total test scores. Biserial correlation coefficients are used as indices of the discriminating power of the options within the items. The correct option should correlate positively with total score. Any items that fail to discriminate between examinees having high and low ability are subject to content review and may be either (1) eliminated or (2) rewritten and subsequently pilot tested as new items. The mean biserial of the correct options of the items in each test are provided in the statistical tables, indicated by "Mean Biserial."

Validity

Validity refers to the degree to which interpretations of test scores are appropriate. For exams such as these, evidence of the appropriateness of the test content is the central validity argument, and proper test construction methods are the primary assurance that the exams can support the intended interpretations.

The NA3SA End-of Program Examinations are designed and constructed to assess examinees' mastery of the NATEF task lists. The participation of subject matter experts on the item-writing teams and the item and test review processes are designed to ensure conformity of the tests with the approved NATEF task list. Following this, ASE staff select test items that are (1) appropriate to the purpose of the test, (2) suitably balanced over topics and skills, (3) free from irrelevant sources of difficulty, and (4) as a group, comparable with previous test forms in difficulty and other performance characteristics. These, plus other rigorous psychometric procedures for item development and test construction, provide excellent assurance of content appropriateness of the exams. NA3SA examinations are not intended to predict future success on any other test or endeavor.

NA3SA Test Form Statistics - Spring 2009

Test: Suspension and Steering (SS)

N of Items	40
Mean	20.731
Variance	38.769
Std. Dev.	6.226
Skew	-0.074
Kurtosis	-0.280
Median	21.000
Alpha	0.781
SEM	2.916
Mean Pcnt Corr	52
Mean Biserial	0.416

Test: Brakes (BR)

N of Items	40
Mean	20.850
Variance	45.939
Std. Dev.	6.778
Skew	0.211
Kurtosis	-0.483
Median	20.000
Alpha	0.821
SEM	2.871
Mean Pcnt Corr	52
Mean Biserial	0.456

Test: Electrical/Electronic Systems (EE)

N of Items	40
Mean	21.401
Variance	44.318
Std. Dev.	6.657
Skew	0.019
Kurtosis	-0.507
Median	21.000
Alpha	0.810
SEM	2.899
Mean Pcnt Corr	54
Mean Biserial	0.446

Test: Engine Performance (EP)

N of Items	40
Mean	21.089
Variance	50.721
Std. Dev.	7.122
Skew	0.110
Kurtosis	-0.598
Median	21.000
Alpha	0.839
SEM	2.858
Mean Pcnt Corr	53
Mean Biserial	0.476

Test: Engine Repair (ER)

N of Items	40
Mean	23.527
Variance	52.871
Std. Dev.	7.271
Skew	-0.229
Kurtosis	-0.475
Median	24.000
Alpha	0.852
SEM	2.798
Mean Pcnt Corr	59
Mean Biserial	0.505

Test: Automatic Transmission & Transaxle (AT)

N of Items	40
Mean	22.846
Variance	48.396
Std. Dev.	6.957
Skew	-0.158
Kurtosis	-0.528
Median	23.000
Alpha	0.833
SEM	2.844
Mean Pcnt Corr	57
Mean Biserial	0.472

Test: Manual Drive Train & Axles (MD)

N of Items	40
Mean	22.160
Variance	55.522
Std. Dev.	7.451
Skew	0.019
Kurtosis	-0.829
Median	22.000
Alpha	0.861
SEM	2.781
Mean Pcnt Corr	55
Mean Biserial	0.513

Test: Heating and Air Conditioning (AC)

N of Items	40
Mean	20.599
Variance	53.858
Std. Dev.	7.339
Skew	0.095
Kurtosis	-0.743
Median	20.000
Alpha	0.849
SEM	2.852
Mean Pcnt Corr	51
Mean Biserial	0.492

Test: Painting and Refinishing (PR)

N of Items	40
Mean	25.022
Variance	37.290
Std. Dev.	6.107
Skew	-0.368
Kurtosis	-0.244
Median	26.000
Alpha	0.798
SEM	2.747
Mean Pcnt Corr	63
Mean Biserial	0.459

Test: Nonstructural Analysis & Damage Repair (NS)

N of Items	40
Mean	23.017
Variance	38.043
Std. Dev.	6.168
Skew	-0.536
Kurtosis	0.579
Median	23.000
Alpha	0.789
SEM	2.831
Mean Pcnt Corr	58
Mean Biserial	0.439

Test: Structural Analysis & Damage Repair (SR)

N of Items	40
Mean	23.073
Variance	57.015
Std. Dev.	7.551
Skew	-0.374
Kurtosis	-0.176
Median	23.000
Alpha	0.868
SEM	2.742
Mean Pcnt Corr	58
Mean Biserial	0.535

Test: Mechanical & Electrical Components (ME)

N of Items	40
Mean	21.309
Variance	46.223
Std. Dev.	6.799
Skew	-0.141
Kurtosis	-0.008
Median	21.000
Alpha	0.829
SEM	2.808
Mean Pcnt Corr	53
Mean Biserial	0.476
