STATE OF CALIFORNIA

DMV

DEPARTMENT OF MOTOR VEHICLES

STATEWIDE EVALUATION OF THE COMMERCIAL DRIVERS LICENSE WRITTEN KNOWLEDGE TESTS

March 2008

Research and Development Branch
Licensing Operations Division

Authors: Chyan V. Wu and
Sukhvir S. Brar

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE March 2008	3. REPORT TYPE AND DATES COVERED Final Report

4. TITLE AND SUBTITLE
Statewide Evaluation of Commercial Drivers License Written Knowledge Tests

5. FUNDING NUMBERS

6. AUTHOR(S)
Chyan V. Wu and Sukhvir S. Brar

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
California Department of Motor Vehicles
Research and Development Section
P.O. Box 932382
Sacramento, CA 94232-3820

8. PERFORMING ORGANIZATION REPORT NUMBER
CAL-DMV-RSS-08-225

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report presents the results of an evaluation of the written knowledge tests administered to applicants for a commercial driver license or endorsement. The report presents the fail rate, mean number of errors, and internal-consistency reliability coefficient for each test form, as well as the pass rate, item-choice selection rates, and item-total correlation for each test question on each test form. Items that need to be reviewed for possible rewording or replacement are identified. Additional recommendations for improving the testing process are also provided. The results are based on 8,576 test sheets completed in all California Department of Motor Vehicle field offices from November 27, 2007 to December 31, 2007.

14. SUBJECT TERMS
written test, fail rates, reliability, item analysis

15. NUMBER OF PAGES
56

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	None

PREFACE

This report is issued as an internal monograph of the Department of Motor Vehicles' Research and Development Branch rather than an official report of the State of California. The findings and opinions may not represent the views and policies of the State of California.

ACKNOWLDEGEMENTS

EXECUTIVE SUMMARY

Objective

The purpose of this evaluation was to identify any deficiencies in the commercial driver license (CDL) written tests and to make recommendations on how the tests can be improved.

Methods

- Completed CDL tests were collected statewide from November 27, 2006 through February 9, 2007. Tests completed on or after January 1, 2007 were subsequently excluded because most of them were a newer revision than needed for the evaluation.

- The test fail rate, mean number of incorrect items, and internal-consistency reliability coefficient were computed for each test form separately for original and renewal applicants on their first test attempt. Test fail rate and mean number of incorrect items were also computed for each test attempt on all forms combined.

- Item statistics were computed for each test form for original and renewal applicants on their first test attempt.

Results

- The General Knowledge test was failed by 39.5% of original applicants and 58.0% of renewal applicants on the first test attempt. The first-attempt fail rates for the other tests vary widely, ranging from 24.8% (Tank Vehicle) to 78.3% (Air Brakes) for original applicants and from 15.5% (Air Brakes) to 62.5% (Combination Vehicles) for renewal applicants.

- The test fail rates for both applicant groups generally do not decrease appreciably over multiple test attempts.

- The General Knowledge, Combination Vehicles, Air Brakes, and Passenger Transport tests each have test forms that vary considerably in difficulty level.

- The Combination Vehicles, Air Brakes, Passenger Transport, and Tank Vehicle tests completed by renewals have one or more test forms with an internal-consistency reliability coefficient below the .70 standard of acceptability.

- All of the tests have some items that are potentially deficient, as indicated by their having low item-total correlations, pass rates that are too high or too low, or distracter selection rates that are too high or too low.

- Examiners changed some tests from a fail to a pass after exercising their discretion to rephrase one or more questions that were initially answered incorrectly. This resulted in the licensure of a substantial percentage of applicants who failed the test based on their first answer choices.

Recommendations

- Test questions with characteristics that indicate they may be problematic should be reviewed, and revised or replaced as necessary.

- Test forms with internal-consistency reliabilities below .70 should be revised to bring them up to this level or higher.

- All test forms should have at least 20 questions, and preferably more, to ensure that they have an adequate level of reliability.

- The policy that allows examiners to restate missed test questions should be reversed. If the policy is kept, the restriction that it not be used for original applicants should be enforced.

- The passing standard for the Air Brakes test should be raised to 80% (16 out of 20 correct) to match the competency standard set for the other tests.

- The Air Brakes test and the associated content in the CDL driver handbook should be reviewed to determine why it is failed by 80.0% of originals but only 15.7% of renewals.

- The order of answer choices should be randomized at least annually to decrease the opportunity for cheating or rote memorization of the correct answers.

- The department should continue investigating the possible use of computer technology to automate the creation and administration of the knowledge tests.

- A waiting period should be required between test attempts to increase the likelihood that applicants would spend more time studying the CDL driver handbook. At a minimum, applicants should not be allowed to take multiple tests of the same kind on the same day.

TABLE OF CONTENTS

LIST OF TABLES

TABLE OF CONTENTS (continued)

LIST OF TABLES (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES (continued)

INTRODUCTION

This report presents the results of an evaluation of the English versions of the Commercial Driver License (CDL) written tests (DL 653, 654, and 655, Rev. 10/2006). The findings are intended to help guide the department in its ongoing efforts to ensure the reliability and validity of its driver knowledge tests.

The CDL program requires applicants who apply for an original or renewal commercial driver license to be administered a CDL General Knowledge test. In addition, CDL applicants are required to pass one or more of six supplementary tests to add additional commercial driving privileges. The CDL General Knowledge test and the six supplementary tests each have three forms which are meant to be equivalent, but which draw somewhat different questions from a common pool of questions; having different forms is designed to discourage cheating.

Table 1 lists the seven tests that were evaluated, the number of questions on each test, and the maximum number of items an applicant can answer incorrectly and still pass.

Table 1

Number of Questions and Number of Allowed Errors on Each Test

Test type	Number of questions	Number of allowed errors
General Knowledge	50	10
Combination Vehicles	15	3
Air Brakes	20	3
Passenger Transport	20	4
Doubles/Triples	15	3
Tank Vehicle	15	3
Hazardous Materials	20	4

This report presents the fail rate, mean number of errors, and internal-consistency reliability coefficient for each test form for original and renewal applicants. It also includes the pass rate, percentage of applicants choosing each answer choice, and the

1

item-total correlation for each test item on each form for original and renewal applicants. Finally, recommendations for improving the tests are provided based on the statistical characteristics of the tests and test items. It should be noted that the tests were not reviewed to determine the accuracy of subject matter content or the extent to which the questions were representative of material covered in the commercial driver handbook.

METHODS

Data Collection

Commercial driver license test forms completed during November 27, 2006 through February 9, 2007 were collected from all but seven field offices, which were already participating in another research study. The seven excluded offices are: Carmichael, Fairfield, Folsom, Lodi, Sacramento Broadway, South Sacramento, and Vacaville. Test forms completed after January 1, 2007 were subsequently eliminated from the evaluation because many offices began using new versions of test forms after that date.

Field offices were instructed to staple together all test forms a customer completed on a given day and a printout of the Test Results Screen showing the results of those tests and other useful information. In cases where a customer made more than one license application on the same day, materials were collected for each application separately. The collected materials were then mailed to the department's Research and Development Branch (R&D) at the end of each week.

The data collection process yielded far more test forms than necessary for evaluation purposes. Because using all test forms would have created a burdensome key entry workload, samples of test sheets for some of the test types were selected and included in the evaluation. This was done by first randomly selecting about half of the available test packets and using all of the attached usable test sheets. This yielded a sufficient number of forms for all of the tests except for the Passenger Transport, Doubles/Triples, and Tank Vehicle tests completed by renewals. To obtain enough of the latter tests, all such test sheets were taken from the remaining packets and then added to those already selected for inclusion in the evaluation.

Data Processing

R&D processed and coded the test forms using information recorded on the Test Results Screen printout. This printout provided information on all tests completed to date on a customer's current application, and made it possible to determine if the applicant should be considered as an "original" or "renewal" for a particular test. In this study, an applicant is classified as an original if he or she was taking a test to add a new driving privilege, or as a renewal if the test was taken to renew a specific driving

privilege. Thus, it was possible, for example, for a customer to have taken on the same day a General Knowledge test to renew their basic CDL license and a Hazardous Materials test to add the privilege to transport hazardous materials.

Information on the Test Results Screen was also used to determine the number of times a customer took a specific test on the current application. An application terminates 1 year from the application date or when the customer fails three attempts on any of the tests, whichever comes first. The customer could then start a new application after paying the required fee. The Test Results Screen records information for the current application only. Therefore, information on any previous test attempts on earlier applications is not available on the printout. Unfortunately, this resulted in an unknown number of test sheets being assigned a lower attempt number than what was true. For example, a test sheet considered here to be a first attempt test may actually have been an applicant's fourth attempt on that test (three attempts on the first application and one attempt on the second application). Similarly, a test sheet considered as a second attempt may actually have been an applicant's fifth attempt (three attempts on the first application and two attempts on the second application). It is unknown the extent to which the test attempt number was under-counted and the net biasing influence this would have had on the estimation of fail rates.

Each test form was graded by computer using the customer's original answer choices. A question was considered "wrong" if an incorrect answer choice was selected, the item was left blank, or more than one answer choice was marked. All statistics presented on this report are based on computer scoring unless indicated otherwise.

Data Analysis

Test Fail Rate and Mean Errors
Fail rates in this evaluation are based upon the number of incorrect answers determined by computer scoring. A test is considered failed if the number of errors exceeds the passing score cutoff for the test indicated in Table 1. Mean errors refers to the average number of missed questions for a group of applicants.

Examiner Scoring Bias
Previous written test evaluations have found that computer grading of the tests often results in test fail rates that are consistently higher than those based on examiner scores.

This occurs because departmental policy has allowed examiners to rephrase or restate missed questions for a "borderline" failure by a renewal applicant and to award points based on the verbal responses, which often results in changing the outcome of the test to a pass. To determine the extent of examiner scoring bias, the fail rates based on computer scores were compared to those based on examiner scores, using only test sheets on which an examiner score was recorded. This was done separately for originals and renewals for each type of test.

Test Form Fail Rate

Ideally, the three forms of each test should have the same level of difficulty. Analysis of Variance (ANOVA) was used to determine whether there are any statistically significant differences between the test form fail rates. The ANOVA's were performed using SPSS statistical software. Any differences were considered to be statistically significant if the probability (*p*) of their having occurred by chance alone is less than 5 times out of 100 (i.e., .05 or 5%). When significant fail rate differences were found in the omnibus ANOVA, additional (post hoc) statistical significance tests were performed to determine which specific pairs of forms were significantly different.

Test Form Reliability

An internal-consistency reliability coefficient was calculated for each test form using the Kuder-Richardson formula (KR-20). This statistic indicates the degree of uniformity in subject matter content among test items, and commensurately the overall precision of the test as a measurement instrument. If a test has high reliability, an individual should achieve a similar score over repeated administrations of the test (assuming the person's true level of knowledge does not change between tests). The reliability coefficient can range from 0 to 1. A value of 0 indicates no similarities exist between test items. A value of 1, on the other hand, denotes that the test items are perfectly homogeneous in content. A relatively well-designed test should have a reliability coefficient of .70 or higher.

Analysis of Test Questions

Individual test questions were analyzed for first-attempt tests only. Separate item analyses were conducted for originals and renewals. The results were then used to identify suspect items that may need to be revised or replaced.

One of the best indicators of a possibly faulty question is the item-total (or part-whole) correlation coefficient. This measure shows how performance on an individual test item is related to performance on the entire test. This coefficient can range from -1 to 1. A positive value indicates that applicants who score well on the test are also more likely to answer the question correctly. Conversely, a negative value indicates that an applicant who scores well on the test is more likely to fail the item. A value near 0 indicates that answering the question correctly or incorrectly has very little or no relationship to how an applicant scores on the test overall.

In addition to examining the relationship between particular questions and the entire test, it is also useful to inspect the choices applicants make among answers for each question. The answer choice selection rate is the percentage of applicants who selected a particular answer choice. The answer choices are not labeled on the CDL test forms. For this report the answer choices are labeled as "a," "b," or "c" in correspondence to the order in which they appear on the forms. These rates are based on valid responses only. Not represented are instances in which the applicant did not answer the item or selected more than one answer choice.

The results of the item analyses were scrutinized to identify suspect items that might need to be revised or replaced. The following guidelines were used for this purpose:

- Items with an item-total correlation coefficient value below .10 or negative are highly suspect and almost always need to be modified or replaced.

- A very low item pass rate (lower than 60%) often indicates that the question is poorly worded, has ambiguous or misleading answer choices, is not related to the general knowledge domain being tested, or is problematic for some other reason. On the other hand, a very high pass rate (above 95%) may result when the distracters are so illogical that the correct answer is obvious or when the subject matter is common knowledge and therefore doesn't need to be tested.

- A distracter selected by 2% or fewer of the applicants may need to be revised or replaced to make it seem tenable to applicants lacking the knowledge covered by the item. A distracter selected more often than the correct answer, or within 10% of the selection rate for the correct answer, is also of concern.

RESULTS

Data Collection and Screening

More than 15,000 test packets (Test Results Screens with attached test forms) were collected from the participating field offices. As indicated earlier, test packets dated after December 31, 2006 were excluded from the evaluation because many of those tests were a newer revision than needed for the evaluation. With this exclusion, more than 13,000 sample packets remained.

Further screening of the test packets found other problems that resulted in the need to discard additional tests. The whole test packet was rejected if a Test Results Screen printout was not stapled to the test forms, or if all the test forms in a packet had a revision date earlier than that required for the evaluation. If only some of the test forms in a packet were usable, those were kept and the unusable ones were discarded.

It is possible that discarding test packets without a Test Results Screen may have biased the analyses to an unknown degree. This problem was identified and corrected early in the data collection period. It is estimated that no more than 300 packets were excluded for this reason, so any bias resulting from this loss of data is probably negligible.

Each test form that appeared usable was coded with an attempt number and an indicator of whether the customer was to be considered an original or renewal applicant for the particular test. The data on completed forms was then key-entered into an MS-Access electronic database.

Overall Test Statistics

Test Difficulty and Reliability
Table 2 presents the fail rate and internal-consistency reliability coefficient for each test across all forms. These results are for the first test attempt only.

The fail rates vary considerably between the test types. In addition, originals tended to perform better than renewals on all tests except the Air Brakes test, which the vast majority of originals failed and the vast majority of renewals passed.

7

All tests for originals have an acceptable level of reliability (above .70). However, four of the seven tests completed by renewals do not meet this standard.

Table 2

Number of Tests (*n*), Fail Rate, and Internal-Consistency Reliability Coefficient for Each Test on the First Test Attempt for Original and Renewal Applicants

Test type Applicant group	*n*	Fail rate (%)	Reliability
General Knowledge			
Originals	1,653	39.5	.91
Renewals	514	58.0	.87
Combination Vehicles			
Originals	1,019	43.9	.77
Renewals	357	62..5	.66
Air Brakes			
Originals	1,183	78.3	.71
Renewals	821	15.5	.62
Passenger Transport			
Originals	518	36.5	.79
Renewals	187	46.5	.69
Doubles/Triples			
Originals	471	43.1	.77
Renewals	323	57.9	.71
Tank Vehicle			
Originals	541	24.8	.73
Renewals	281	33.5	.58
Hazardous Materials			
Originals	432	31.5	.82
Renewals	276	39.5	.75

Note. Fail rates are based on computer grading of the tests using the applicant's initial answer choices. The sample sizes for renewals on the Passenger Transport, Doubles/Triples, and Tank Vehicle tests represent all usable test sheets collected of these types. All other sample sizes are based on an approximately 50% random selection of test sheets.

Examiner Scoring Bias

Table 3 shows the first-attempt test fail rates based on computer and examiner scoring of the tests. Only test sheets having an examiner score recorded (10.9% of all tests) were included in the comparisons, and therefore the results don't give a complete picture of

the extent of examiner scoring bias. A possible reason for the examiner score not being recorded is that the examiner may have just written "pass" or "fail," instead of the number of errors on the test sheet. Another possible reason is that the examiner may have given credit for rephrased questions answered correctly, but then chose not to record the modified score on the form and just updated the license record with the test result. However, without knowing the precise reasons why some examiner scores were missing, it is not possible to ascertain the influence the exclusion of these cases would have had on the estimation of examiner bias.

Table 3

Number of Tests (*n*) and Fail Rate for Each Test on the First Attempt for Each Applicant Group When the Tests are Graded by Computer Versus the Examiner

Test type Applicant group	*n*	Computer-graded fail rate (%)	Examiner-graded fail rate (%)
General Knowledge			
Originals	1,428	41.1	37.4
Renewals	465	59.1	52.9
Combination Vehicles			
Originals	898	46.1	43.5
Renewals	326	63.5	59.8
Air Brakes			
Originals	1,090	80.0	78.4
Renewals	738	15.7	8.0
Passenger Transport			
Originals	466	36.9	32.0
Renewals	167	46.7	44.3
Doubles/Triples			
Originals	420	44.3	41.2
Renewals	293	58.4	55.6
Tank Vehicle			
Originals	480	26.7	23.5
Renewals	247	35.6	31.6
Hazardous Materials			
Originals	386	33.7	31.3
Renewals	240	40.0	37.1

Note. Only test sheets with an examiner score recorded are included in the comparisons.

Test Form and Item Statistics

This section presents fail rates, mean number of errors, internal-consistency reliability coefficient, and various item statistics for each form of each test type by applicant group. For simplicity, the results for each test are presented in a separate subsection below.

General Knowledge Test

The General Knowledge test is the most widely used of all the CDL tests. All CDL applicants must take this test to obtain an original or renewal CDL license or to add or renew a CDL certificate or endorsement. The test has 50 questions and allows 10 incorrect answers to pass. The results for this test are presented below.

Test Form Difficulty and Reliability. Table 4 presents the fail rate, mean number of errors, and the internal-consistency reliability coefficient for each form of the General Knowledge test on the first attempt for each applicant group.

The fail rates for originals range from 35.9% to 42.6%, but the differences between the rates are not statistically significant (*p* = .07). The renewal fail rates are more variable, ranging from 51.5% to 65.7%. These differences are statistically significant (*p* = .03). Post hoc significance tests found renewals Form 2 to have a significantly lower fail rate than renewals Form 3.

The means for number of errors range from 9.0 to 10.3 for originals and from 12.1 to 13.3 for renewals. (Statistical significance tests were not performed on this measure in the present evaluation.)

All of the test reliability coefficients exceed the .70 standard of acceptability.

Table 4

Number of Tests (*n*), Fail Rate, Mean Number of Errors, and Internal-
Consistency Reliability Coefficient for Each Form of the General Knowledge
Test for Original and Renewal Applicants on the First Test Attempt

Applicant group Test form	*n*	Fail rate (%)	Mean errors	Reliability
Originals[a]				
1	571	40.5	9.8	.91
2	582	35.9	9.0	.91
3	500	42.6	10.3	.91
Total	1,653	39.5	10.0	.91
Renewals[b]				
1	174	56.3	12.1	.89
2	165	51.5	12.1	.91
3	175	65.7	13.3	.82
Total	514	58.0	12.5	.87

Note. The figures presented for "Total" (all forms combined) are weighted averages. All ANOVA's are two-tailed.
[a]The forms do not differ significantly on fail rate ($F = 2.69$, $p = .07$). [b]The forms differ significantly on fail rate ($F = 3.69$, $p = .03$).

Test Difficulty by Attempt. Table 5 presents the fail rate and mean number of errors for each attempt on the General Knowledge test for each applicant group. Test sheets with unknown attempt numbers were not included in the computations. This is unlikely to have biased the estimates because it is improbable that capturing the attempt number is related to test performance.

Of perhaps greatest interest is the fact that the fail rates do not drop substantially on subsequent attempts. One possible reason for this is that the department's existing policy of not requiring a waiting period between test attempts does not encourage applicants to study the driver handbook following a test failure. In fact, it is fairly common practice for applicants to retake the same test within minutes of the previous test failure. It is also possible that applicants have difficulty reading and understanding the handbook, or with taking the test.

Table 5

Number of Tests (*n*), Fail Rate, and Mean Number of Errors for Each Test
Attempt by Applicant Group for the General Knowledge Test

Test type Attempt	*n*	Fail rate (%)	Mean errors
Originals			
First	1,653	39.5	9.7
Second	510	47.1	11.3
Third	163	45.4	10.9
Total	2,326	41.6	10.1
Renewals			
First	514	58.0	12.5
Second	214	49.5	11.3
Third	83	48.2	10.9
Total	811	54.7	12.0

Note. The figures presented for "Total" (all attempts combined) are weighted averages.

Item Statistics. Table 6 presents the item-total correlation coefficient for each question on each form of the General Knowledge test for original and renewal applicants. Items with a correlation coefficient below the .10 criterion are shaded for identification. These items have a performance outcome that has little or no relationship to overall performance. Such items are highly suspect and almost always need to be modified or replaced. There were no such items for originals, but there are four for renewals (three of which are on Form 3).

Table 7 presents the answer choice selection rates for each question on each form of the General Knowledge test for original and renewal applicants. The use of shading, underlining, and boldface type are used in the table to identify special characteristics of some items, as described in the table note.

Table 8 identifies the items with at least one of the following statistical characteristics for identifying bad items: a low item-total correlation coefficient, a low or high pass rate, and/or a distracter with a selection rate that is too low or too high. All of the General Knowledge test forms have a few suspect items.

Table 6

Item-Total Correlation Coefficient for Each Item on Each Form of the General
Knowledge Test for Original and Renewal Applicants on the First Test Attempt

	Originals			Renewals		
	Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
Item	($n = 571$)	($n = 582$)	($n = 500$)	($n = 174$)	($n = 165$)	($n = 175$)
1	.34	.48	.50	.42	.54	.24
2	.27	.51	.35	.36	.40	.11
3	.36	.27	.43	.39	.29	.28
4	.35	.34	.41	.45	.43	.24
5	.37	.32	.30	.22	.35	.21
6	.21	.36	.48	.38	.33	.29
7	.44	.28	.37	.36	.34	.36
8	.33	.46	.22	.09	.34	.12
9	.41	.52	.49	.45	.46	.35
10	.37	.39	.27	.48	.41	.14
11	.41	.39	.28	.31	.42	.09
12	.42	.49	.39	.30	.39	.31
13	.51	.39	.51	.36	.49	.24
14	.41	.33	.42	.17	.29	.42
15	.39	.42	.38	.22	.31	.10
16	.47	.35	.33	.38	.36	.29
17	.45	.33	.56	.39	.14	.33
18	.44	.49	.33	.47	.38	.21
19	.45	.54	.37	.47	.40	.25
20	.44	.51	.41	.38	.41	.25
21	.38	.51	.27	.18	.31	.29
22	.47	.43	.35	.36	.50	.29
23	.13	.28	.33	.13	.40	.21
24	.29	.38	.52	.29	.17	.39
25	.40	.39	.36	.42	.44	.25
26	.39	.37	.35	.32	.45	.14
27	.37	.45	.57	.20	.42	.42
28	.39	.52	.38	.53	.49	.17
29	.40	.49	.46	.39	.49	.31
30	.39	.24	.27	.31	.33	.27

Table 6 (continued)

	Originals			Renewals		
	Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
Item	(n = 571)	(n = 582)	(n = 500)	(n = 174)	(n = 165)	(n = 175)
31	.32	.38	.46	.18	.37	.27
32	.32	.34	.36	.40	.36	.30
33	.39	.47	.45	.38	.65	.33
34	.45	.35	.47	.44	.47	.48
35	.36	.34	.46	.25	.38	.44
36	.30	.29	.50	.26	.33	.03
37	.43	.42	.41	.37	.58	.23
38	.22	.47	.34	.30	.41	.14
39	.42	.45	.40	.34	.31	.39
40	.31	.34	.45	.36	.39	.37
41	.34	.30	.39	.38	.46	.19
42	.39	.35	.47	.32	.57	.17
43	.42	.26	.31	.33	.30	.17
44	.50	.34	.47	.37	.23	.27
45	.53	.36	.23	.43	.39	.05
46	.40	.32	.51	.29	.47	.41
47	.49	.34	.41	.45	.36	.35
48	.53	.38	.40	.52	.49	.21
49	.43	.47	.43	.25	.38	.27
50	.38	.32	.44	.32	.55	.25

Note. The entry for *n* at the top of each column is the number of usable first-attempt tests used to compute the item-total correlation coefficients in the column. Shading identifies items with a value below .10, which indicates the items probably need to be revised or replaced.

Table 7

Percentage of Applicants Selecting Each Answer Choice for Each Item
on Each Form of the General Knowledge Test for Original and
Renewal Applicants on the First Test Attempt

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
	a	8.8	86.4	14.5	12.9	86.6	21.8
1	b	85.5	10.5	67.5	80.6	11.0	46.6
	c	5.7	3.1	18.1	6.5	2.4	31.6

14

Table 7 (continued)

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
2	a	6.2	_66.5_	**0.8**	5.2	_56.7_	**0.6**
	b	_91.7_	5.2	_94.6_	_93.1_	6.1	_95.4_
	c	2.1	28.4	4.6	**1.7**	37.2	4.0
3	a	5.8	10.4	12.9	2.3	9.1	19.4
	b	_82.6_	_84.4_	31.3	_84.5_	_87.9_	**40.6**
	c	11.6	5.2	_55.8_	13.2	3.0	_40.0_
4	a	12.0	8.1	10.9	13.3	11.5	9.2
	b	_84.0_	7.9	9.9	_83.2_	12.7	8.7
	c	4.0	_84.0_	_79.2_	3.5	_75.8_	_82.1_
5	a	5.5	12.5	9.1	12.1	17.2	8.0
	b	_83.0_	7.3	**1.4**	_70.5_	17.2	**1.1**
	c	11.5	_80.2_	_89.5_	17.3	_65.6_	_90.8_
6	a	_94.7_	_84.8_	9.2	_94.3_	_78.0_	9.8
	b	2.6	12.1	_79.7_	5.2	17.7	_71.8_
	c	2.6	3.1	11.0	**0.6**	4.3	18.4
7	a	10.2	11.7	_66.7_	12.8	24.2	_59.8_
	b	_69.0_	_82.8_	11.5	_64.0_	_70.9_	12.6
	c	20.8	5.5	21.8	23.3	4.8	27.6
8	a	12.0	8.9	3.2	17.2	13.3	4.6
	b	3.5	6.3	**1.8**	3.4	9.1	**1.1**
	c	_84.5_	_84.8_	_95.0_	_79.3_	_77.6_	_94.3_
9	a	9.5	22.1	_72.6_	9.2	25.0	_67.1_
	b	8.1	_74.0_	4.8	6.4	_70.7_	**1.7**
	c	_82.4_	3.8	22.5	_84.4_	4.3	31.2
10	a	2.6	6.2	14.9	**1.2**	7.3	22.9
	b	_92.8_	4.7	_73.0_	_90.2_	9.8	_64.0_
	c	4.6	_89.1_	12.1	8.7	_82.9_	13.1
11	a	**1.6**	2.3	3.0	**1.1**	4.3	2.9
	b	_95.1_	_84.6_	3.4	_93.1_	_81.7_	2.9
	c	3.3	13.2	_93.6_	5.7	14.0	_94.3_
12	a	9.9	16.5	2.0	18.6	23.8	2.9
	b	_86.6_	7.6	22.8	_77.3_	13.4	31.8
	c	3.5	_75.9_	_75.2_	4.1	_62.8_	_65.3_
13	a	11.8	4.5	12.1	19.1	4.9	21.8
	b	8.1	_79.5_	_83.9_	12.1	_75.3_	_74.1_
	c	_80.1_	16.0	4.0	_68.8_	19.8	4.0

Table 7 (continued)

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
14	a	84.9	90.9	6.0	83.3	88.3	9.8
	b	6.4	2.9	84.8	7.5	5.5	78.2
	c	8.7	6.2	9.2	9.2	6.1	12.1
15	a	13.7	92.1	81.8	19.8	90.3	78.0
	b	11.3	2.6	8.3	8.1	**1.2**	6.4
	c	75.0	5.3	9.9	72.1	8.5	15.6
16	a	23.8	69.1	79.7		57.6	71.6
	b	8.3	6.7	11.1			17.2
	c	67.9	24.2	9.3	59.6	33.3	11.2
17	a	86.7	68.8	10.0	80.2	57.7	12.1
	b	4.8	11.2	7.5	8.7		9.2
	c	8.5	20.0	82.5	11.0	30.7	78.6
18	a	5.5	13.4	10.3	11.5	20.5	13.2
	b	6.4	4.7	82.6	10.3	3.7	77.0
	c	88.1	81.9	7.1	78.2	75.8	9.8
19	a	6.3	67.9	3.2	7.5	42.1	8.0
	b	88.8	**1.2**	5.2	86.2	4.9	3.4
	c	4.9	30.9	91.6	6.3	**53.0**	88.5
20	a	73.2	28.2	10.6	49.1	**44.4**	12.8
	b	3.7	61.1	6.8	2.9	38.9	9.9
	c	23.1	10.7	82.6	**48.0**	16.7	77.3
21	a	4.8	5.3	3.8	7.0	5.5	5.1
	b	20.2	6.7	7.4	27.9	8.5	9.7
	c	75.0	88.0	88.8	65.1	86.0	85.1
22	a	12.7	84.3	5.0	15.2	76.7	5.7
	b	4.6	10.2	90.8	7.6	14.1	91.4
	c	82.7	5.5	4.2	77.2	9.2	2.9
23	a	62.8	5.0	**0.8**	58.5	5.5	5.1
	b	26.8	7.1	3.6	28.1	5.5	**0.6**
	c	10.5	87.9	95.6	13.5	89.1	94.3
24	a	3.0	15.6	13.5	6.9	15.3	15.5
	b	**1.2**	3.3	15.7	2.3	3.7	22.4
	c	95.8	81.1	70.8	90.8	81.0	62.1
25	a	**1.6**	2.1	91.6	2.3	3.7	93.1
	b	9.0	3.8	6.0	11.0	4.9	4.0
	c	89.4	94.1	2.4	86.7	91.5	2.9
26	a	7.8	3.8	**1.4**	6.4	5.5	**1.7**
	b	6.0	91.2	3.8	5.2	84.2	**1.7**
	c	86.2	5.0	94.8	88.4	10.3	96.6

Table 7 (continued)

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
27	a		6.5	68.7	98.3	12.9	55.3
	b		72.3	18.5		65.6	
	c	2.5	21.2	12.9	0.0	21.5	21.2
28	a	95.4	5.5	7.0	93.1	6.1	14.0
	b		79.5	76.5	5.2	70.9	71.5
	c	1.9	15.0	16.5	1.7	23.0	14.5
29	a	7.2	84.7	3.4	5.2	76.4	4.6
	b	0.5	13.7	16.8	0.6	17.6	30.9
	c	92.3	1.5	79.8	94.3	6.1	64.6
30	a	4.4	2.8	4.6	2.3	1.2	5.7
	b	16.4	4.7	82.1	20.2	8.0	81.0
	c	79.2	92.6	13.3	77.5	90.8	13.2
31	a	21.5	11.2	17.4	27.6	19.5	23.4
	b	42.3	83.6	6.4	50.6	70.1	8.0
	c	36.2	5.2	76.2	21.8	10.4	68.6
32	a	92.4	86.2	83.1	87.9	79.3	76.6
	b	4.8	6.2	5.6	10.9	7.3	5.7
	c	2.8	7.6	11.2	1.1	13.4	17.7
33	a	6.5	95.5	9.1	8.1	92.1	8.1
	b	20.0	1.4	69.8	17.3	1.2	61.3
	c	73.5	3.1	21.1	74.6	6.7	30.6
34	a	16.2	94.3	68.4	18.5	92.7	60.3
	b	4.6	1.0	16.9	2.9	1.8	22.4
	c	79.2	4.7	14.7	78.6	5.5	17.2
35	a	7.8	6.6	13.0	7.5	4.2	23.6
	b	3.6	13.6	9.8	4.0	7.3	11.5
	c	88.6	79.8	77.2	88.5	88.5	64.9
36	a	82.5	11.0	84.1	85.1	9.3	79.9
	b	13.2	9.6	9.2	9.8	11.7	12.6
	c	4.2	79.4	6.6	5.2	79.0	7.5
37	a	23.7	9.5	26.5	29.9	11.7	45.3
	b	13.9	89.1	12.5	10.3	84.0	15.1
	c	62.5	1.4	61.0	59.8	4.3	39.5
38	a	5.6	4.1	4.9	7.5	9.7	9.3
	b	13.9	15.1	6.3	10.4	19.4	8.1
	c	80.4	80.8	88.8	82.1	70.9	82.6

Table 7 (continued)

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
39	a	14.5	7.3	<u>80.8</u>	15.0	13.5	<u>77.6</u>
	b	3.9	13.1	15.8	2.9	17.8	22.4
	c	<u>81.6</u>	<u>79.6</u>	3.4	<u>82.1</u>	<u>68.7</u>	**0.0**
40	a	3.2	16.2	<u>57.1</u>	4.6	12.2	<u>41.0</u>
	b	3.0	**1.0**	37.3	4.6	**1.8**	**54.3**
	c	<u>93.8</u>	<u>82.8</u>	5.6	<u>90.8</u>	<u>86.0</u>	4.6
41	a	7.4	<u>92.2</u>	6.4	11.8	<u>91.5</u>	11.6
	b	6.7	2.6	6.4	12.4	3.6	7.5
	c	<u>85.9</u>	5.2	<u>87.2</u>	<u>75.9</u>	4.8	<u>80.9</u>
42	a	15.8	2.6	<u>86.3</u>	21.5	4.9	<u>89.5</u>
	b	5.3	**1.6**	6.9	6.4	2.4	7.0
	c	<u>78.9</u>	<u>95.9</u>	6.9	<u>72.1</u>	<u>92.7</u>	3.5
43	a	11.5	<u>85.9</u>	8.7	15.8	<u>81.5</u>	4.0
	b	14.1	4.0	<u>87.7</u>	17.5	5.6	<u>94.8</u>
	c	<u>74.4</u>	10.1	3.6	<u>66.7</u>	13.0	**1.2**
44	a	16.1	22.8	16.2	19.5	33.3	18.4
	b	12.4	<u>73.5</u>	<u>65.7</u>	18.4	<u>57.0</u>	<u>56.9</u>
	c	<u>71.5</u>	3.6	18.0	<u>62.1</u>	9.7	24.7
45	a	27.6	7.9	<u>80.1</u>	30.4	5.5	<u>70.9</u>
	b	4.1	<u>87.4</u>	3.0	3.6	<u>87.9</u>	**1.2**
	c	<u>68.3</u>	4.7	16.9	<u>66.1</u>	6.7	27.9
46	a	<u>55.9</u>	88.2	11.8	<u>41.4</u>	80.7	10.4
	b	23.6	**1.2**	21.7	29.6	**1.9**	32.4
	c	20.5	10.6	<u>66.5</u>	29.0	17.4	<u>57.2</u>
47	a	<u>81.6</u>	2.3	4.3	<u>78.1</u>	5.6	**1.2**
	b	15.2	15.3	22.3	19.5	20.5	34.7
	c	3.2	<u>82.5</u>	<u>73.5</u>	2.4	<u>73.9</u>	<u>64.2</u>
48	a	<u>67.9</u>	2.3	5.3	<u>56.5</u>	3.1	9.9
	b	25.4	5.9	11.9	36.5	5.6	9.9
	c	6.8	<u>91.8</u>	<u>82.8</u>	7.1	<u>91.4</u>	<u>80.1</u>
49	a	<u>75.6</u>	32.7	4.1	<u>62.6</u>	**46.5**	5.8
	b	19.1		<u>81.7</u>	29.8	<u>41.5</u>	<u>82.5</u>
	c	5.3	8.0	14.3	7.6	11.9	11.7
50	a	<u>91.5</u>	8.9	9.9	<u>89.3</u>	6.9	8.8
	b	4.1	15.5	<u>83.2</u>	4.7	18.2	<u>85.4</u>
	c	4.4	<u>75.7</u>	6.9	5.9	<u>74.8</u>	5.8

Note. An underlined percentage indicates that the answer choice was the correct response. Shading indicates that the item has a pass rate that is too low or too high and therefore needs to be reviewed and possibly revised or replaced. A boldfaced percentage indicates that the distracter selection rate is too low or too high. Item choice selection rates may not sum to 100% due to rounding.

Table 8

Summary of Problem Items on Each Form of the General
Knowledge Test for Original and Renewal Applicants

Problem indicator	Applicant group	Form 1	Form 2	Form 3
Item-total correlation too low or negative	Originals			
	Renewals	8		11, 36, 45
Pass rate too high	Originals	11, 24, 27, 28	33, 42	23
	Renewals	27		2, 26
Pass rate too low	Originals	31, 46	49	3, 40
	Renewals	16, 20, 23, 31, 37, 46, 48	2, 16, 17, 19, 20, 44, 49	1, 3, 7, 27, 37, 40, 44, 46
Distracter selected too often	Originals	31		
	Renewals	20, 31	19, 20, 49	3, 37, 40
Distracter selected too infrequently	Originals	11, 24, 25, 27, 28, 29	19, 29, 33, 34, 37, 40, 42, 46	2, 5, 8, 23, 26
	Renewals	2, 6, 10, 11, 27, 28, 29, 32	15, 30, 33, 34, 40, 46	2, 5, 8, 9, 23, 26, 39, 43, 45, 47

Combination Vehicles Test

The Combination Vehicles test is required to obtain or renew a Class A license. This test has 15 items and allows up to three incorrect answers for passing. Evaluation results for this test are presented in the following subsections.

Test Form Difficulty and Reliability. Table 9 presents the fail rate, mean number of errors, and internal-consistency reliability coefficient for each form of the Combination Vehicles test on the first attempt for each applicant group.

The fail rates for originals range from 39.2% to 47.1%. However, the differences between these rates are not statistically significant ($p = .09$). The fail rates for renewals are higher, ranging from 51.9% to 70.2%. The differences between these rates are statistically significant ($p = .01$). The post hoc significance tests found that Form 1

differs significantly from Forms 2 and 3 on this measure. No other significant differences were found between the form fail rates.

The means for the number of errors range from 3.5 to 4.1 for originals and from 3.8 to 5.2 for renewals.

All test forms meet or exceed the .70 standard of acceptability, except Form 2 for renewals (.66).

Table 9

Number of Tests (*n*), Fail Rate, Mean Number of Errors, and Internal-Consistency Reliability Coefficient for Each Form of the Combination Vehicles Test for Original and Renewal Applicants on the First Test Attempt

Applicant group Test form	*n*	Fail rate (%)	Mean errors	Reliability
Originals[a]				
1	337	39.2	3.5	.82
2	357	47.1	4.1	.72
3	325	45.2	3.9	.81
Total	1,019	43.9	3.8	.77
Renewals[b]				
1	129	51.9	3.8	.71
2	104	70.2	5.2	.66
3	124	66.9	4.8	.70
Total	357	62.5	4.6	.66

Note. The figures presented for "Total" (all forms combined) are weighted averages. All ANOVA's are two-tailed. [a]The forms do not differ significantly on fail rate ($F = 2.38$, $p = .09$). [b]The forms differ significantly on fail rate ($F = 5.00$, $p = .01$).

Test Difficulty by Attempt. The fail rate and mean number of errors for each attempt and for each applicant group on the Combination Vehicles test are presented in Table 10. There appears to be a general lowering in fail rate over test attempts for this test for renewals.

Table 10

Number of Tests (*n*), Fail Rate, and Mean Number of Errors for Each Test
Attempt by Applicant Group for the Combination Vehicles Test

Test type Attempt	*n*	Fail rate (%)	Mean errors
Originals			
First	1,019	43.9	3.8
Second	365	48.8	4.1
Third	130	36.2	3.1
Total	1,514	44.4	3.8
Renewals			
First	357	62.5	4.6
Second	170	48.8	4.0
Third	52	36.5	3.2
Total	579	56.1	4.3

Note. The figures presented for "Total" (all attempts combined) are weighted averages.

Item Statistics. Table 11 presents the item-total correlation coefficient for each item on each test form completed by each applicant group. Items with a coefficient below .10 are shaded. Such items usually need to be revised or replaced. One item for originals and two items for renewals fall below the .10 criterion.

Table 12 presents the answer choice selection rates for each applicant group.

Table 13 lists the suspect items on each test form. There are several suspect items for each applicant group.

Table 11

Item-Total Correlation Coefficient for Each Item on Each Form of the Combination
Vehicles Test for Original and Renewal Applicants on the First Test Attempt

	Originals			Renewals		
	Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
Item	(n = 337)	(n = 357)	(n = 325)	(n = 129)	(n = 104)	(n = 124)
1	.48	.20	.44	.35	.18	.43
2	.45	.11	.41	.36	.34	.20
3	.40	.54	.51	.41	.45	.27
4	.39	.40	.39	.42	.34	.29
5	.56	.29	.52	.31	.23	.40
6	.41	.43	.47	.46	.30	.45
7	.43	.37	.33	.23	.41	.11
8	.61	.37	.47	.34	.21	.35
9	.39	.29	.36	.38	.20	.01
10	.48	.37	.46	.30	.28	.37
11	.44	.54	.31	.26	.41	.49
12	.41	-.04	.40	.20	-.13	.27
13	.47	.39	.41	.33	.48	.29
14	.41	.47	.42	.14	.22	.34
15	.35	.36	.45	.31	.24	.18

Note. The entry for *n* at the top of each column is the number of usable first-attempt tests used to compute the item-total correlations in the column. Shading identifies items with a value below .10, which indicates the items probably need to be revised or replaced.

Table 12

Percentage of Applicants Selecting Each Answer Choice for Each Item
on Each Form of the Combination Vehicles Test for Original and
Renewal Applicants on the First Test Attempt

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
1	a	7.8	31.6	27.2	6.3	13.2	**39.7**
	b	8.1	16.1	57.0	8.6	23.4	41.3
	c	84.1	52.3	15.8	85.2	51.0	19.0
2	a	63.3	13.0	70.3	54.3	21.4	53.7
	b	26.8	49.6	13.0	35.7	39.8	21.1
	c	9.9	37.5	16.7	10.1	**38.8**	25.2

Table 12 (continued)

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
3	a	13.9	<u>76.4</u>	9.2	10.2		10.5
	b	12.0	2.2	<u>83.7</u>	22.7		<u>83.9</u>
	c	<u>74.1</u>	21.3	7.1	<u>67.2</u>	40.8	5.6
4	a	3.3	2.8	**1.9**	2.3	**1.0**	2.4
	b	<u>87.5</u>	6.7	<u>87.8</u>	<u>81.3</u>	4.9	<u>81.3</u>
	c	9.2	<u>90.4</u>	10.3	16.4	<u>94.2</u>	16.3
5	a	3.9	11.8	<u>63.8</u>	**1.6**	21.4	<u>43.0</u>
	b	<u>78.6</u>	2.5	18.9	<u>76.7</u>	2.9	
	c	17.6	<u>85.7</u>	17.3	21.7	<u>75.7</u>	29.8
6	a	6.0	<u>66.9</u>	19.1	**1.6**	<u>57.7</u>	21.8
	b	7.4	7.6	<u>76.2</u>	4.7	11.5	<u>74.2</u>
	c	<u>86.6</u>	25.5	4.6	<u>93.8</u>	30.8	4.0
7	a	13.0	12.4	<u>85.4</u>	14.3	20.2	<u>94.4</u>
	b	28.9	6.5	5.9	**41.3**	13.5	2.4
	c	<u>58.1</u>	<u>81.1</u>	8.7	<u>44.4</u>	66.3	3.2
8	a	<u>81.7</u>	9.3	4.7	<u>80.6</u>	18.6	**1.6**
	b	11.1	<u>88.8</u>	8.1	10.9	<u>81.4</u>	7.3
	c	7.2	**2.0**	<u>87.3</u>	8.5	**0.0**	<u>91.1</u>
9	a	6.9	5.7	4.0	12.5	12.7	**0.0**
	b	6.0	25.6	6.8	9.4	25.5	**0.8**
	c	<u>87.0</u>	<u>68.7</u>	<u>89.2</u>	<u>78.1</u>	<u>61.8</u>	<u>99.2</u>
10	a	15.1	5.3	9.1	19.7	3.8	18.0
	b	13.3	<u>85.4</u>	<u>72.0</u>	17.3	<u>80.8</u>	<u>55.7</u>
	c	<u>71.7</u>	9.3	18.9	<u>63.0</u>	15.4	26.2
11	a	<u>81.1</u>	22.0	**49.5**	<u>90.7</u>	32.7	**58.5**
	b	15.6	5.9	4.6	5.4	19.2	2.4
	c	3.3	<u>72.0</u>	<u>45.8</u>	3.9	<u>48.1</u>	<u>39.0</u>
12	a	5.7	**60.8**	14.2	3.9	**59.6**	16.4
	b	10.5	<u>32.7</u>	3.7	16.4	<u>35.6</u>	2.5
	c	<u>83.8</u>	6.5	<u>82.0</u>	<u>79.7</u>	4.8	<u>81.1</u>
13	a	10.5	<u>86.0</u>	14.4	7.0	<u>77.9</u>	19.7
	b	21.3	6.7	13.1	31.8	13.5	20.5
	c	<u>68.2</u>	7.3	<u>72.5</u>	<u>61.2</u>	8.7	<u>59.8</u>
14	a	<u>89.6</u>	9.6	25.5	<u>93.8</u>	2.9	32.8
	b	6.0	9.6	<u>61.5</u>	4.7	20.2	<u>50.8</u>
	c	4.5	<u>80.7</u>	13.0	**1.6**	<u>76.9</u>	16.4

Table 12 (continued)

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
	a	<u>65.7</u>	14.2	<u>83.6</u>	<u>72.4</u>	10.6	<u>80.5</u>
15	b	11.4	<u>80.5</u>	4.6	7.1	<u>79.8</u>	3.3
	c	22.9	5.4	11.8	20.5	9.6	16.3

Note. An underlined percentage indicates that the answer choice was the correct response. Shading indicates that the item has a pass rate that is too low or too high and therefore needs to be reviewed and possibly revised or replaced. A boldfaced percentage indicates that the distracter selection rate is too low or too high. Item choice selection rates may not add to 100% due to rounding.

Table 13

Summary of Problem Items on Each Form of the Combination
Vehicles Test for Original and Renewal Applicants

Problem indicator	Applicant group	Form 1	Form 2	Form 3
Item-total correlation too low or negative	Originals		12	
	Renewals		12	9
Pass rate too high	Originals			
	Renewals			9
Pass rate too low	Originals	7	1, 2, 12	1, 11
	Renewals	2, 7	1, 2, 3, 6, 11, 12	1, 2, 5 ,10, 11, 13, 14
Distracter selected too often	Originals		12	11
	Renewals	7	2, 12	1, 11
Distracter selected too infrequently	Originals		8	4
	Renewals	5, 6, 14	3, 4, 8	8, 9

Air Brakes Test

When an applicant applies for a CDL license, a restriction from operating a vehicle with air brakes is automatically placed on the driver record. The applicant must pass the Air Brakes test to remove this restriction. The test has 20 questions and allows four incorrect answers to pass.

The tests were randomly sampled and therefore it was expected that there would be an approximately equal number of test sheets of each form. However, the number of Form 3 test sheets fell far short of this expectation. The procedures for collecting, processing, and analyzing the data were reviewed and no reason for this shortage could be found. It is not known whether this underrepresentation may have biased the parameter estimates.

Test Form Difficulty and Reliability. Table 14 presents the fail rate, mean number of errors, and the internal-consistency reliability coefficient for each test form on the first test attempt for each applicant group. The fail rates are very high for originals and very low for renewals. This is the only test on which originals performed worse than renewals.

Table 14

Number of Tests (*n*), Fail Rate, Mean Number of Errors, and Internal-Consistency Reliability Coefficient for Each Form of the Air Brakes Test for Original and Renewal Applicants on the First Test Attempt

Applicant group Test form	*n*	Fail rate (%)	Mean errors	Reliability
Originals[a]				
1	426	80.3	6.2	.74
2	418	80.4	6.2	.74
3	339	73.2	5.7	.74
Total	1,183	78.3	6.0	.71
Renewals[b]				
1	262	15.6	2.3	.64
2	264	17.0	2.5	.67
3	295	13.9	1.8	.59
Total	821	15.5	2.2	.62

Note. The figures presented for "Total" (all forms combined) are weighted averages. All ANOVAs are two-tailed.
[a]Forms differed significantly on fail rate ($F = 3.68$, $p = .03$). [b]Forms did not differ significantly on fail rate ($F = 0.53$, $p = .60$).

Field office examiners and others who consulted on this project were not surprised by these results. They felt that the Air Brakes test is inherently more difficult for novices than are the other tests because it is difficult to learn about the operation of vehicles with air brakes based only on a review of the driver manual. However, as drivers gain experience from actually operating vehicles with air brakes, they quickly learn the knowledge required to pass the test, thus explaining the superior performance of renewal applicants.

The form fail rates for originals significantly differ from one another (p = .03), with values ranging from 73.2 to 80.4. Post hoc tests found significant differences only between Form 3 and the other two forms. No statistically significant differences were found between form fail rates for renewal applicants (p = .60).

The mean number of errors ranges from 5.7 to 6.2 for originals and from 1.8 to 2.5 for renewals.

All of the test forms for originals have an acceptable level of reliability. However, none of the forms for renewals reach the .70 standard of acceptability, with values ranging from .59 to .67.

Test Difficulty by Attempt. Table 15 presents the fail rate and mean number of errors for each attempt on the Air Brakes test for originals and renewals. There appears to have been a substantial improvement in performance on the third attempt for both applicant groups.

Table 15

Number of Tests (*n*), Fail Rate, and Mean Number of Errors for
Each Test Attempt by Applicant Group for the Air Brakes Test

Test type Attempt	*n*	Fail rate (%)	Mean errors
Originals			
First	1,183	78.3	6.0
Second	498	72.1	5.6
Third	175	63.4	5.0
Total	1,856	75.2	5.8
Renewals			
First	821	15.5	2.2
Second	305	18.0	2.4
Third	122	10.7	2.0
Total	1,248	15.6	2.2

Note. The figures presented for "Total" (all attempts combined) are weighted averages.

Item Statistics. Table 16 presents the item-total correlation coefficient for each question on each Air Brakes test form for original and renewal applicants. All of the correlations for originals are acceptable (above .10 as indicated by shading). However, seven of the questions for renewals have a correlation coefficient that is below .10 or negative, indicating that they likely need modification or replacement.

Table 17 presents the item pass rate and answer choice selection rates for each question on each form of the Air Brakes test for each applicant group. All items with a pass rate that is too low are for originals. Conversely, all items with a pass rate that is too high (except one) are for renewals. This finding is not surprising considering that the vast majority of originals failed the test, while the vast majority of renewals passed it.

Table 16

Item-Total Correlation Coefficient for Each Item on Each Form of the Air
Brakes Test for Original and Renewal Applicants on the First Test Attempt

Item	Originals			Renewals		
	Form 1 (*n* = 426)	Form 2 (*n* = 418)	Form 3 (*n* = 339)	Form 1 (*n* = 262)	Form 2 (*n* = 264)	Form 3 (*n* = 295)
1	.30	.28	.33	.16	.15	.30
2	.21	.18	.24	.17	.39	.18
3	.26	.34	.43	.18	.28	.32
4	.47	.24	.34	.56	.27	.15
5	.44	.19	.39	.58	.22	.11
6	.22	.34	.25	.23	.35	.17
7	.43	.27	.18	.48	.17	.23
8	.33	.34	.26	.16	.19	.33
9	.30	.38	.30	.11	.32	.13
10	.23	.31	.31	.07	.19	.20
11	.34	.15	.32	.53	.12	.19
12	.33	.27	.36	.35	.01	.37
13	.21	.39	.39	-.09	.50	.17
14	.33	.18	.31	.50	.10	.22
15	.31	.36	.28	-.01	.46	.36
16	.39	.33	.30	.23	.38	.07
17	.22	.36	.27	.28	.37	.08
18	.25	.41	.30	.27	.38	-.03
19	.33	.44	.33	.45	.29	.29
20	.20	.36	.14	.20	.36	.14

Note. The entry for *n* at the top of each column is the number of usable first-attempt tests used to compute the item-total correlations in the column. Shading identifies items with a value below .10, which indicates the items probably need to be revised or replaced.

Table 17

Percentage of Applicants Selecting Each Answer Choice for
Each Item on Each Form of the Air Brakes Test For Original
and Renewal Applicants on the First Test Attempt

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
1	a	9.7	3.4	6.8	2.7	**0.8**	**0.7**
	b	8.3	2.6	39.8	2.3	**0.4**	13.9
	c	<u>82.1</u>	<u>94.0</u>	<u>53.4</u>	<u>95.0</u>	<u>98.9</u>	<u>85.4</u>
2	a	21.0	3.6	<u>59.2</u>	9.6	**1.9**	<u>81.3</u>
	b	<u>47.9</u>	92.3	31.5	<u>78.9</u>	96.6	17.7
	c	31.2	4.1	9.2	11.5	**1.5**	**1.0**
3	a	10.5	<u>75.7</u>	11.3	3.5	<u>90.5</u>	**1.4**
	b	6.0	15.4	33.8	**1.2**	6.5	11.0
	c	<u>83.5</u>	8.8	<u>54.9</u>	<u>95.4</u>	3.0	<u>87.7</u>
4	a	6.9	16.8	14.5	**0.0**	3.8	**1.4**
	b	<u>85.3</u>	28.8	6.8	<u>98.5</u>	12.6	**1.4**
	c	7.8	<u>54.3</u>	<u>78.6</u>	**1.5**	<u>83.6</u>	<u>97.3</u>
5	a	5.6	6.2	9.0	**0.8**	**1.1**	**1.7**
	b	9.2	**1.2**	14.9	**1.5**	**0.0**	3.7
	c	<u>85.2</u>	<u>92.6</u>	<u>76.1</u>	<u>97.7</u>	<u>98.9</u>	<u>94.6</u>
6	a	4.3	<u>78.3</u>	<u>86.3</u>	2.7	<u>92.0</u>	<u>95.9</u>
	b	18.8	5.5	6.4	9.6	**1.9**	2.7
	c	<u>77.0</u>	16.1	7.3	<u>87.7</u>	6.1	**1.4**
7	a	11.8	**42.5**	14.9	**0.8**	24.0	6.5
	b	<u>82.8</u>	12.6	25.7	<u>96.2</u>	4.9	7.5
	c	5.4	<u>44.9</u>	<u>59.4</u>	3.1	<u>71.1</u>	<u>86.0</u>
8	a	6.4	33.3	<u>79.9</u>	**1.5**	9.9	<u>95.9</u>
	b	23.2	4.8	4.7	10.7	**1.1**	**0.7**
	c	<u>70.4</u>	<u>61.8</u>	15.3	<u>87.8</u>	<u>89.0</u>	3.4
9	a	3.8	<u>73.6</u>	16.2	**0.8**	<u>90.0</u>	4.8
	b	10.2	19.5	<u>70.4</u>	**1.9**	8.4	<u>93.2</u>
	c	<u>86.0</u>	7.0	13.4	<u>97.3</u>	**1.5**	2.0
10	a	10.9	25.4	22.8	3.5	6.1	8.8
	b	<u>31.6</u>	<u>62.2</u>	<u>63.8</u>	<u>61.2</u>	<u>90.1</u>	<u>85.1</u>
	c	**57.5**	12.4	13.5	35.4	3.8	6.1

Table 17 (continued)

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
11	a	66.0	6.7	**0.6**	92.0	6.1	**0.0**
	b	19.1	46.5	11.6	5.4	55.3	2.4
	c	14.9	**46.8**	87.8	2.7	38.5	97.6
12	a	62.3	57.8	18.8	91.2	87.5	5.8
	b	18.3	3.4	7.4	4.2	**0.0**	**0.3**
	c	19.5	38.8	73.8	4.6	12.5	93.8
13	a	15.1	86.8	30.7	**1.5**	95.4	9.6
	b	**41.2**	10.6	4.8	28.5	3.0	**0.7**
	c	43.8	2.6	64.5	70.0	**1.5**	89.7
14	a	5.5	45.9	5.1	**0.8**	68.1	**0.7**
	b	88.1	**45.0**	34.0	96.9	28.9	15.4
	c	6.4	9.1	60.8	2.3	3.0	83.9
15	a	67.1	17.5	28.6	90.8	4.9	11.3
	b	19.8	14.9	44.0	4.6	3.0	77.3
	c	13.1	67.6	27.4	4.6	92.0	11.3
16	a	68.3	6.5	90.8	84.5	3.4	98.6
	b	22.7	76.9	6.8	11.6	91.3	**1.4**
	c	9.1	16.6	2.4	3.9	5.3	**0.0**
17	a	20.2	50.5	83.9	7.3	81.4	94.5
	b	25.2	**42.5**	4.2	8.8	16.3	**0.0**
	c	54.6	7.0	11.9	83.8	2.3	5.5
18	a	6.4	9.6	5.1	2.7	**0.8**	**0.7**
	b	73.7	2.4	4.2	92.2	**1.5**	**0.0**
	c	19.9	88.0	90.8	5.0	97.7	99.3
19	a	50.2	19.5	87.0	88.3	7.6	98.0
	b	20.3	60.0	10.0	8.6	86.6	**1.0**
	c	29.4	20.5	2.9	3.1	5.7	**1.0**
20	a	5.5	6.5	4.7	3.9	**0.4**	**0.3**
	b	4.3	83.5	16.9	**0.8**	95.8	5.8
	c	90.3	10.1	78.3	95.4	3.8	93.9

Note. An underlined percentage indicates that the answer choice was the correct response. Shading indicates that the item has a pass rate that is too low or too high and therefore needs to be reviewed and possibly revised, or replaced. A boldfaced percentage indicates that the distracter selection rate is too low or too high. Item choice selection rates may not sum to 100% due to rounding.

Table 18 identifies all the suspect items found in the item analysis.

Table 18

Summary of Problem Items on Each Form of the Air
Brakes Test for Original and Renewal Applicants

Problem indicator	Applicant group	Form 1	Form 2	Form 3
Item-total correlation too low or negative	Originals			
	Renewals	10, 13, 15	12	16, 17, 18
Pass rate too high	Originals			
	Renewals	1, 3, 4, 5, 7, 9, 14, 20	1, 2, 5, 13, 18, 20	4 ,6, 8, 11, 16, 18, 19
Pass rate too low	Originals	2, 10, 13, 17, 19	4, 7, 11, 12, 14, 17	1, 2, 3, 7, 15
	Renewals		11	
Distracter selected too often	Originals	10, 13	7, 11, 14, 17	
	Renewals			
Distracter selected too infrequently	Originals		5	11
	Renewals	3, 4, 5, 7, 8, 9, 13, 14, 20	1, 2, 5, 6, 8, 9, 12, 13, 18, 20	1, 2, 3, 4, 5, 6, 8, 11, 12, 13, 14, 16, 17, 18, 19, 20

Passenger Transport Test

The Passenger Transport test is required of drivers who wish add an endorsement to operate vehicles used to transport passengers for commercial purposes. The test has 20 items and allows a maximum of four incorrect answers to pass.

Test Form Difficulty and Reliability. Table 19 presents the fail rates, mean errors, and the internal consistency reliability coefficient for each form of the Passenger Transport test for originals and renewals on the first test attempt.

Fail rates for originals range from 29.1% to 40.4%, but the differences are not statistically significant (p = .06). The variability of fail rates is greater for renewals, with values ranging from 32.8% to 67.2%. The differences between the rates are statistically significant (p = .001). Post hoc tests found the fail rate for Form 3 to differ significantly from the fail rates for the other two renewal forms.

The mean number of errors ranges from 3.6 to 4.4 for originals and from 4.1 to 5.4 for renewals.

For original applicants, the reliability coefficients for all of the forms are above the .70 standard of acceptability. For renewal applicants, only Form 1 has a reliability coefficient higher than .70.

Table 19

Number of Tests (*n*), Fail Rate, Mean Number of Errors, and Internal-Consistency Reliability Coefficient for Each Form of the Passenger Transport Test for Original and Renewal Applicants on the First Test Attempt

Applicant group Test form	*n*	Fail rate (%)	Mean errors	Reliability
Originals[a]				
1	175	39.4	4.4	.80
2	165	29.1	3.6	.82
3	178	40.4	4.2	.79
Total	518	36.5	4.0	.79
Renewals[b]				
1	58	32.8	4.1	.77
2	62	37.1	4.2	.66
3	67	67.2	5.4	.67
Total	187	46.5	4.6	.69

Note. The figures presented for "Total" (all forms combined) are weighted averages. All ANOVAs are two-tailed. [a]Forms do not differ significantly on fail rate (F = 2.89, p = .06). [b]Forms differ significantly on fail rate (F = 9.86, p = .00).

Test Difficulty by Attempt. Table 20 presents the fail rate and mean errors for each test attempt for original and renewal applicants. The fail rate holds steady for both applicant groups through subsequent test attempts except for the increase on the second attempt for originals. The very small sample size for renewals on the third attempt makes the associated fail rate too untrustworthy to interpret.

Table 20

Number of Tests (*n*), Fail Rate, and Mean Number of Errors for Each
Test Attempt by Applicant Group for the Passenger Transport Test

Test type Attempt	*n*	Fail rate (%)	Mean errors
Originals			
First	518	36.5	4.0
Second	140	47.9	5.0
Third	53	35.8	4.0
Total	711	38.7	4.2
Renewals			
First	187	46.5	4.6
Second	67	46.3	4.6
Third	17	47.1	4.5
Total	271	46.5	4.6

Note. The figures presented for "Total" (all attempts combined) are weighted averages.

Item Statistics. Table 21 presents the item-total correlation coefficient for each question on each form of the Passenger Transport test for the two applicant groups. All of the values for originals met the .10 level of acceptability. For renewals, however, five questions have low or negative correlations (two on Form 2 and three on Form 3).

Table 22 presents the answer choice selection rates for the individual questions on each form for each applicant group.

Table 23 identifies the items with one or more statistical characteristics that indicate they need to be reviewed and possibly revised or replaced.

Table 21

Item-Total Correlation Coefficient for Each Item on Each Form of the Passenger
Transport Test for Original and Renewal Applicants on the First Test Attempt

	Originals			Renewals		
	Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
Item	(n = 175)	(n = 165)	(n = 178)	(n = 58)	(n = 62)	(n = 67)
1	.61	.45	.44	.29	.14	.33
2	.26	.22	.40	.51	.45	.16
3	.15	.58	.13	.35	.06	.43
4	.38	.25	.39	.18	.22	.14
5	.30	.48	.38	.26	.35	.20
6	.41	.31	.41	.38	.49	.15
7	.39	.36	.45	.46	.42	.50
8	.60	.41	.40	.41	.29	.35
9	.31	.24	.29	.45	.20	.08
10	.44	.54	.40	.33	-.01	.28
11	.49	.48	.23	.14	.23	.27
12	.36	.34	.47	.32	.17	.52
13	.60	.45	.29	.20	.29	.04
14	.29	.29	.44	.35	.35	.28
15	.39	.41	.35	.44	.49	.18
16	.28	.49	.40	.28	.31	.36
17	.15	.40	.16	.33	.17	-.01
18	.55	.43	.45	.46	.60	.33
19	.33	.36	.21	.30	.11	.15
20	.36	.40	.21	.46	.10	.15

Note. The entry for n at the top of each column is the number of usable first-attempt tests used to compute the item-total correlations in the column. Shading identifies items with a value below .10, which indicates the items probably need to be revised or replaced.

Table 22

Percentage of Applicants Selecting Each Answer Choice for
Each Item on Each Form of the Passenger Transport Test for
Original and Renewal Applicants on the First Test Attempt

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
1	a	**54.4**	87.2	31.6	**45.6**	77.4	**50.0**
	b	15.8	4.3	5.7	12.3	3.2	9.1
	c	29.8	8.5	62.6	42.1	19.4	40.9
2	a	6.9	22.7	12.4	5.2	27.4	16.7
	b	1.7	69.3	10.7	1.7		15.2
	c	91.4	8.0	77.0	93.1	14.5	68.2
3	a	4.6	14.8	16.0	3.4	6.7	23.9
	b	75.7	5.6	80.6	77.6	6.7	74.6
	c	19.7	79.6	3.4	19.0	86.7	1.5
4	a	89.6	84.8	**0.6**	89.7	83.9	**0.0**
	b	8.1	7.3		8.6	12.9	
	c	2.3	7.9	99.4	1.7	3.2	100.0
5	a	8.1	6.7	**1.1**	3.4	**1.6**	3.0
	b	7.0	13.5	7.3	6.9	24.2	4.5
	c	84.9	79.8	91.5	89.7	74.2	92.5
6	a	8.0	**1.2**	4.5	15.8	**0.0**	3.0
	b	7.4	94.5	7.3	8.8	96.8	4.5
	c	84.6	4.3	88.2	75.4	3.2	92.4
7	a	68.8	16.0	**1.1**	67.2	19.4	**1.5**
	b	17.9	76.7	84.3	10.3	71.0	80.6
	c	13.3	7.4	14.6	22.4	9.7	17.9
8	a	19.0	73.0	**1.7**	17.2	82.3	**1.5**
	b	4.6	4.9	41.6	5.2	**1.6**	**55.2**
	c	76.4	22.1	56.7	77.6	16.1	43.3
9	a	4.6	13.3	65.5	3.5	26.7	55.2
	b	82.9	81.8	17.5	91.2	73.3	25.4
	c	12.6	4.8	16.9	5.3	**0.0**	19.4
10	a	93.7	81.7	10.2	98.3	75.8	7.5
	b	2.3	10.4	16.5	**0.0**	19.4	29.9
	c	4.0	7.9	73.3	**1.7**	4.8	62.7

Table 22 (continued)

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
11	a	84.5	6.2	10.7	91.4	**1.6**	13.4
	b	4.6	8.6	85.3	**1.7**	14.5	86.6
	c	10.9	85.2	4.0	6.9	83.9	**0.0**
12	a	12.7	23.6	10.2	22.8	21.3	10.4
	b	73.4	4.2	75.6	63.2	**1.6**	65.7
	c	13.9	72.1	14.2	14.0	77.0	23.9
13	a	9.9	**1.8**	24.3	19.0	**0.0**	**40.3**
	b	64.9	3.0	63.8	48.3	3.2	44.8
	c	25.1	95.1	11.9	32.8	96.8	14.9
14	a	85.1	10.4	92.7	91.2	14.5	89.4
	b	6.9	76.2	3.4	3.5	62.9	3.0
	c	8.0	13.4	4.0	5.3	22.6	7.6
15	a	82.5	3.0	90.4	86.2	**1.6**	87.9
	b	8.2	3.0	2.2	5.2	3.2	4.5
	c	9.4	93.9	7.3	8.6	95.2	7.6
16	a	2.3	5.5	4.0	**0.0**	**1.6**	6.0
	b	2.9	3.7	6.8	5.3	6.5	**1.5**
	c	94.9	90.9	89.3	94.7	91.9	92.5
17	a	5.2	9.1	16.5	**1.7**	17.7	20.0
	b	14.0	22.6	11.9	8.6	32.3	20.0
	c	80.8	68.3	71.6	89.7	50.0	60.0
18	a	9.2	3.1	74.6	14.0	**0.0**	75.8
	b	83.8	3.7	5.6	80.7	3.3	9.1
	c	6.9	93.3	19.8	5.3	96.7	15.2
19	a	4.1	89.7	5.7	5.2	91.9	6.1
	b	86.0	**1.2**	77.3	89.7	**1.6**	66.7
	c	9.9	9.1	17.0	5.2	6.5	27.3
20	a	13.7	12.2	5.1	17.2	32.3	6.2
	b	21.1	4.3	**1.7**	15.5	6.5	**1.5**
	c	65.1	83.5	93.2	67.2	61.3	92.3

Note. An underlined percentage indicates that the answer choice was the correct response. Shading indicates that the item has a pass rate that is too low or too high and therefore needs to be reviewed and possibly revised or replaced. A boldfaced percentage indicates that the distracter selection rate is too low or too high.

Table 23

Summary of Problem Items on Each Form of the Passenger
Transport Test for Original and Renewal Applicants

Problem indicator	Applicant group	Form 1	Form 2	Form 3
Item-total correlation too low or negative	Originals			
	Renewals		3, 10	9, 13, 17
Pass rate too high	Originals		13	4
	Renewals	10	6, 13, 15, 18	4
Pass rate too low	Originals	1		8
	Renewals	1, 13	2, 17	1, 8, 9, 13
Distracter selected too often	Originals	1		
	Renewals	1		1, 8, 13
Distracter selected too infrequently	Originals	2	6, 13, 19	4, 5, 7, 8, 20
	Renewals	2, 4, 10, 11, 16, 17	5, 6, 8, 9, 11, 12, 13, 15, 16, 18, 19	3, 4, 7, 8, 11, 16, 20

Doubles/Triples Test

The Doubles/Triples test has 15 items and allows the applicant to miss no more than three questions to pass. The results for this test are presented below.

Test Form Difficulty and Reliability. Table 24 presents the fail rate, mean number of errors, and internal-consistency reliability coefficient for each test form for original and renewal applicants on the first test attempt.

The test form fail rates range from 40.8% to 46.2% for originals and from 55.4% to 63.8% for renewals. The differences in the form fail rates were not statistically significant for either originals ($p = .64$) or renewals ($p = .39$).

The mean number of errors ranges from 3.6 to 3.8 for originals and from 4.3 to 4.7 for renewals.

Table 24

Number of Tests (*n*), Fail Rate, Mean Number of Errors, and Internal-Consistency Reliability Coefficient for Each Form of the Doubles/Triples Test for Original and Renewal Applicants on the First Test Attempt

Applicant group Test form	*n*	Fail rate (%)	Mean errors	Reliability
Originals[a]				
1	159	42.8	3.8	.81
2	169	40.8	3.6	.74
3	143	46.2	3.8	.78
Total	471	43.1	3.7	.77
Renewals[b]				
1	117	55.6	4.4	.78
2	94	63.8	4.7	.67
3	112	55.4	4.3	.71
Total	323	57.9	4.5	.71

Note. The figures presented for "Total" (all forms combined) are weighted averages. All ANOVAs are two-tailed. [a]Forms do not differ significantly on fail rate ($F = 0.45$, $p = .64$). [b]Forms do not differ significantly on fail rate ($F = 0.96$, $p = .39$).

All of the reliability coefficients are above the .70 standard of acceptability, with the exception of the .67 value for Form 2 for renewals.

Test Difficulty by Attempt. Table 25 presents the fail rate and mean number of errors for each test attempt on the Doubles/Triples test for each applicant group. There is little if any improvement on performance between the first and second test attempts. The very small sample sizes used to compute the third-attempt fail rates makes these estimates somewhat unreliable and therefore risky to interpret.

Table 25

Number of Tests (*n*), Fail Rate, and Mean Number of Errors for Each
Test Attempt by Applicant Group for the Doubles/Triples Test

Test type Attempt	*n*	Fail rate (%)	Mean errors
Originals			
First	471	43.1	3.7
Second	137	46.7	3.9
Third	39	56.4	4.9
Total	647	44.7	3.9
Renewals			
First	323	57.9	4.5
Second	114	50.9	4.0
Third	26	38.5	3.3
Total	463	55.1	4.3

Note. The figures presented for "Total" (all attempts combined) are weighted averages.

Item Statistics. Table 26 presents the item-total correlation coefficients for each question on each Doubles/Triples test form for the two applicant groups. Only one item for originals and three items for renewals fall below the .10 level of acceptability.

Table 27 presents the answer choice selection rates for each question on each form by applicant group.

Table 28 lists the items that should be reviewed. Once again, most of the suspect items are for renewals.

Table 26

Item-Total Correlation Coefficient for Each Item on Each Form of the
Doubles/Triples Test for Original and Renewal Applicants on the First Test Attempt

| Item | Originals | | | Renewals | | |
	Form 1 (n = 159)	Form 2 (n = 169)	Form 3 (n = 143)	Form 1 (n = 117)	Form 2 (n = 94)	Form 3 (n = 112)
1	.38	.51	.49	.28	.23	.58
2	.58	.30	.36	.45	.20	.33
3	.29	.19	.46	.32	-.14	.41
4	.37	.48	.30	.43	.31	.38
5	.37	.32	.43	.41	.28	.44
6	.47	.38	.46	.40	.39	.21
7	.36	.47	.15	.38	.44	.06
8	.49	.03	.38	.42	.07	.16
9	.50	.28	.45	.28	.24	.38
10	.31	.37	.40	.48	.30	.22
11	.22	.19	.46	.35	.22	.25
12	.58	.38	.39	.45	.43	.26
13	.44	.39	.44	.29	.35	.28
14	.49	.59	.46	.42	.52	.49
15	.20	.10	.13	.39	.49	.37

Note. The entry for *n* at the top of each column is the number of usable first-attempt tests used to compute the item-total correlations in the column. Shading identifies items with a value below .10, which indicates the items probably need to be revised or replaced.

Table 27

Percentage of Applicants Selecting Each Answer Choice for
Each Item on Each Form of the Doubles/Triples Test for
Original and Renewal Applicants on the First Test Attempt

| Item | Answer choice | Originals | | | Renewals | | |
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
	a	10.7	10.7	6.3	9.5	23.4	7.2
1	b	22.0	34.9	66.2	34.5	**55.3**	58.6
	c	67.3	54.4	27.5	56.0	21.3	34.2

Table 27 (continued)

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
2	a		16.0	12.0	_40.5_	10.8	11.6
	b		8.3	**0.7**		15.1	2.7
	c	30.6	_75.7_	_87.3_	**44.0**	_74.2_	_85.7_
3	a	3.1	2.4	16.9	5.1	**1.1**	18.0
	b	32.1	25.0	6.3	**46.2**	41.5	**0.9**
	c	_64.8_	_72.6_	_76.8_	_48.7_	_57.4_	_81.1_
4	a	_81.0_	4.2	13.4	_81.9_	16.1	11.7
	b	11.4	15.8	15.5	10.3	12.9	10.8
	c	7.6	_80.0_	_71.1_	7.8	_71.0_	_77.5_
5	a	6.3	23.8	5.6	3.5	23.7	4.5
	b	_90.5_	13.1	34.5	_92.2_	15.1	**43.8**
	c	3.2	_63.1_	_59.9_	4.3	_61.3_	_51.8_
6	a	_65.6_	5.9	11.2	_60.7_	5.4	4.6
	b	13.4	_74.6_	5.6	12.0	_71.0_	3.7
	c	21.0	19.5	_83.2_	27.4	23.7	_91.7_
7	a	_76.3_	_74.4_	_56.7_	_67.2_	_72.3_	_50.9_
	b	**0.0**	7.7	22.0	4.3	6.4	19.4
	c	23.7	17.9	21.3	28.4	21.3	29.6
8	a	_81.4_	21.3	16.8	_80.3_	24.7	28.6
	b	9.0	_75.7_	_79.7_	9.4	_74.2_	_67.0_
	c	9.6	3.0	3.5	10.3	**1.1**	4.5
9	a	8.9	2.4	_87.4_	13.2	6.4	_81.3_
	b	_82.8_	12.4	5.6	_78.1_	17.0	9.8
	c	8.3	_85.2_	7.0	8.8	_76.6_	8.9
10	a	16.6	_77.3_	_79.6_	10.4	_67.4_	_76.8_
	b	3.8	12.9	12.0	3.5	14.1	11.6
	c	_79.6_	9.8	8.5	_86.1_	18.5	11.6
11	a	_79.7_	19.3	17.5	_82.6_	20.2	29.4
	b	17.1	9.6	_69.2_	13.9	6.7	_56.9_
	c	3.2	_71.1_	13.3	3.5	_73.0_	13.8

Table 27 (continued)

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
12	a	17.3	9.6	11.9	6.2	12.9	9.8
	b	63.5	3.6	72.7	69.0	3.2	81.3
	c	19.2	86.7	15.4	24.8	83.9	8.9
13	a	12.7	3.0	71.1	19.7	7.5	61.3
	b	2.5	82.7	7.0	6.0	73.1	17.1
	c	84.8	14.3	21.8	74.4	19.4	21.6
14	a	7.6	82.2	14.0	4.3	75.3	20.5
	b	15.8	7.7	75.5	23.5	7.5	76.8
	c	76.6	10.1	10.5	72.2	17.2	2.7
15	a	5.7	91.0	10.6	2.6	88.3	15.3
	b	79.9	4.2	87.3	82.6	9.6	81.1
	c	14.5	4.8	2.1	14.8	2.1	3.6

Note. An underlined percentage indicates that the answer choice was the correct response. Shading indicates that the item has a pass rate that is too low or too high and therefore needs to be reviewed and possibly revised or replaced. A boldfaced percentage indicates that the distracter selection rate is too low or too high.

Table 28

Summary of Problem Items on Each Form of the
Doubles/Triples Test for Original and Renewal Applicants

Problem indicator	Applicant group	Form 1	Form 2	Form 3
Item-total correlation too low or negative	Originals		8	
	Renewals		3, 8	7
Pass rate too high	Originals			
	Renewals			
Pass rate too low	Originals	2	1	5, 7
	Renewals	1, 2, 3	1, 3	1, 5, 7, 11
Distracter selected too often	Originals			
	Renewals	2, 3	1	5
Distracter selected too infrequently	Originals	7		2
	Renewals		3, 8	3

Tank Vehicle Test

The Tank Vehicle test has 15 items and allows a maximum of 3 errors to pass. The results for this test are presented below.

Test Form Difficulty and Reliability. Table 29 presents the fail rate, mean number of errors, and internal-consistency reliability coefficient for each Tank Vehicle test form for originals and renewals on the first test attempt.

The test form fail rates range from 23.5% to 25.7% for originals and from 30.3% to 35.0% for renewals. No statistically significant differences in form fail rates were found for either originals (*p* = .89) or renewals (*p* = .76).

Table 29

Number of Tests (*n*), Fail Rate, Mean Number of Errors, and Internal-Consistency Reliability Coefficient for Each Form of the Tank Vehicle Test for Original and Renewal Applicants on the First Test Attempt

Applicant group Test form	*n*	Fail rate (%)	Mean errors	Reliability
Originals[a]				
1	183	25.7	2.4	.75
2	188	25.0	2.4	.79
3	170	23.5	2.2	.70
Total	541	24.8	2.3	.73
Renewals[b]				
1	100	35.0	2.9	.62
2	89	30.3	2.6	.59
3	92	34.8	2.9	.67
Total	281	33.5	2.8	.58

Note. The figures presented for "Total" (all forms combined) are weighted averages. All ANOVAs are two-tailed. [a]Forms do not differ significantly on fail rate (*F* = 0.11, *p* = .89). [b]Forms do not differ significantly on fail rate (*F* = 0.28, *p* = .76).

The mean number of errors ranges from 2.2 to 2.4 for originals and from 2.6 to 2.9 for renewals.

Only Forms 1, 2, and 3 for originals meet or exceed the .70 standard of acceptability. All of the other forms are deficient on this measure.

Test Difficulty by Attempt. Table 30 presents the fail rate and mean number of errors for each attempt on the Tank Vehicle test for each applicant group. The fail rate doesn't change much between the first and second attempts for either applicant group. The small sample sizes for the third attempt rates make those estimates too untrustworthy to interpret.

Table 30

Number of Tests (*n*), Fail Rate, and Mean Number of Errors for Each
Test Attempt by Applicant Group for the Tank Vehicle Test

Test type Attempt	*n*	Fail rate (%)	Mean errors
Originals			
First	541	24.8	2.3
Second	86	29.1	2.9
Third	8	12.5	3.0
Total	635	25.2	2.4
Renewals			
First	281	33.5	2.8
Second	62	30.6	2.8
Third	8	12.5	2.4
Total	351	32.5	2.8

Note. The figures presented for "Total" (all attempts combined) are weighted averages.

Item Statistics. Table 31 presents the item-total correlation coefficients for each question on each test form of the Tank Vehicle test for original and renewal applicants. Only three items, all for renewals, had a coefficient value below the .10 standard of acceptability.

Table 31

Item-Total Correlation Coefficient for Each Item on Each Form of the Tank
Vehicle Test for Original and Renewal Applicants on the First Test Attempt

	Originals			Renewals		
	Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
Item	(n = 183)	(n = 188)	(n = 170)	(n = 100)	(n = 89)	(n = 92)
1	.30	.38	.28	.39	.20	.53
2	.44	.55	.31	.41	.39	.37
3	.25	.12	.27	.19	.05	.22
4	.29	.48	.24	.40	.30	.13
5	.48	.54	.31	.34	-.08	.17
6	.24	.34	.39	.19	.10	.42
7	.52	.39	.29	.34	.25	.09
8	.32	.37	.33	.36	.35	.36
9	.35	.40	.49	.10	.24	.52
10	.44	.43	.18	.23	.18	.10
11	.48	.42	.19	.13	.30	.13
12	.30	.44	.19	.23	.29	.13
13	.41	.43	.35	.37	.33	.35
14	.31	.35	.33	.34	.26	.27
15	.46	.43	.34	.40	.16	.35

Note. The entry for *n* at the top of each column is the number of usable first-attempt tests used to compute the item-total correlations in the column. Shading identifies items with a value below .10, which indicates the items probably need to be revised or replaced.

The responses to the individual items on each Tank Vehicle test form by applicant group are presented in Table 32. Most of the suspect (shaded) items have a pass rate that is too high.

Table 33 presents the items with one or more statistical characteristics that indicate they may be problematic.

Table 32

Percentage of Applicants Selecting Each Answer Choice for
Each Item on Each Form of the Tank Vehicle Test for
Original and Renewal Applicants on the First Test Attempt

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
1	a	90.7	13.3	7.1	95.0	28.1	6.5
	b	3.3	2.7	87.1	2.0	3.4	88.0
	c	6.0	84.0	5.9	3.0	68.5	5.4
2	a	3.8	10.2	2.9	4.0	6.7	8.7
	b	1.1	80.6	89.4	0.0	76.4	81.5
	c	95.1	9.1	7.6	96.0	16.9	9.8
3	a	27.8	27.3	5.3	54.0	31.5	8.7
	b	5.6	69.0	1.2	4.0	61.8	2.2
	c	66.7	3.7	93.5	42.0	6.7	89.1
4	a	1.6	6.4	82.0	1.0	12.4	66.3
	b	0.5	89.3	1.8	0.0	85.4	2.2
	c	97.8	4.3	16.2	99.0	2.2	31.5
5	a	86.3	2.7	1.2	92.0	1.1	0.0
	b	0.5	2.1	90.0	1.0	0.0	83.7
	c	13.2	95.2	8.8	7.0	98.9	16.3
6	a	86.2	5.9	20.8	87.0	2.2	23.1
	b	7.7	86.1	2.4	7.0	89.9	2.2
	c	6.1	8.0	76.8	6.0	7.9	74.7
7	a	17.1	17.6	86.5	31.3	13.5	89.1
	b	74.0	2.7	9.4	61.6	1.1	8.7
	c	8.8	79.7	4.1	7.1	85.4	2.2
8	a	1.1	87.6	83.4	1.0	95.5	77.2
	b	97.3	8.6	1.8	98.0	4.5	2.2
	c	1.6	3.8	14.8	1.0	0.0	20.7
9	a	3.3	9.7	81.2	3.1	8.0	77.2
	b	75.6	0.5	12.9	71.4	1.1	12.0
	c	21.1	89.8	5.9	25.5	90.9	10.9
10	a	8.2	4.3	3.5	9.0	3.4	3.3
	b	5.5	83.3	0.6	10.0	85.4	1.1
	c	86.3	12.4	95.9	81.0	11.2	95.7
11	a	81.1	84.0	7.6	67.0	80.9	2.2
	b	11.7	8.6	91.8	19.0	5.6	97.8
	c	7.2	7.5	0.6	14.0	13.5	0.0

Table 32 (continued)

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
12	a		4.3	_93.5_		4.5	_88.0_
	b		_81.8_	4.7		_77.5_	6.5
	c	_50.0_	13.9	**1.8**	_39.0_	18.0	5.4
13	a	_92.3_	_78.1_	_77.5_	_91.0_	_70.5_	_72.8_
	b	2.2	10.7	11.8	5.0	10.2	8.7
	c	5.5	11.2	10.7	4.0	19.3	18.5
14	a	**1.1**	4.3	_77.5_	3.0	4.5	_71.9_
	b	_94.5_	_94.7_	7.1	_93.0_	_92.0_	9.0
	c	4.4	**1.1**	15.4	4.0	3.4	19.1
15	a	**0.0**	_87.7_	2.4	2.0	_85.4_	11.0
	b	4.9	11.2	_77.5_	2.0	14.6	_59.3_
	c	_95.1_	**1.1**	20.1	_96.0_	**0.0**	29.7

Note. An underlined percentage indicates that the answer choice was the correct response. Shading indicates that the item has a pass rate that is too low or too high and therefore needs to be reviewed and possibly revised or replaced. A boldfaced percentage indicates that the distracter selection rate is too low or too high.

Table 33

Summary of Problem Items on Each Form of the Tank
Vehicle Test for Original and Renewal Applicants

Problem indicator	Applicant group	Form 1	Form 2	Form 3
Item-total correlation too low or negative	Originals			
	Renewals		3, 5	7
Pass rate too high	Originals	2, 4, 8, 15	5	10
	Renewals	2, 4, 8, 15	5, 8	10, 11
Pass rate too low	Originals	12		
	Renewals	3, 12		15
Distracter selected too often	Originals			
	Renewals	3, 12		
Distracter selected too infrequently	Originals	2, 4, 5, 8, 14, 15	9, 14, 15	3, 4, 5, 8, 10, 11, 12
	Renewals	2, 4, 5, 8	5, 7, 8, 9, 15	5, 10, 11

Hazardous Materials Test

The Hazardous Materials test has 20 items and allows a maximum of four errors to pass. The results for each form of this test are presented below.

Test Form Difficulty and Reliability. Table 34 presents the fail rate, mean number of errors, and internal-consistency reliability coefficient for each form of the Hazardous Materials test for originals and renewals on the first attempt.

The form fail rates range from 29.2% to 32.7% for originals and from 34.7% to 44.1% for renewals. The differences between the rates were not statistically significant for either originals ($p = .77$) or renewals ($p = .42$).

The mean number of errors ranges from 3.2 to 3.4 for originals and from 3.3 to 4.0 for renewals.

All of the test form reliability coefficients are above the .70 standard of acceptability.

Table 34

Number of Tests (*n*), Fail Rate, Mean Number of Errors, and Internal-Consistency Reliability Coefficient for Each Form of the Hazardous Material Test for Original and Renewal Applicants on the First Test Attempt

Applicant group Test form	*n*	Fail rate (%)	Mean errors	Reliability
Originals[a]				
1	156	32.7	3.4	.86
2	144	29.2	3.2	.82
3	132	32.6	3.3	.78
Total	432	31.5	3.3	.82
Renewals[b]				
1	93	44.1	4.0	.79
2	98	34.7	3.3	.72
3	85	40.0	4.0	.77
Total	276	39.5	3.8	.75

Note. The figures presented for "Total" (all forms combined) are weighted averages. All ANOVAs are two-tailed. [a]Forms do not differ significantly on fail rate ($F = 0.27$, $p = .77$). [b]Forms do not differ significantly on fail rate ($F = 0.88$, $p = .42$).

Test Difficulty by Attempt. Table 35 presents the fail rates and mean number of errors for each attempt by applicant group on the Hazardous Materials test. The fail rate increased between the first and second test attempts for both applicant groups. The third attempt fail rates are of questionable accuracy due to the small sample sizes on which they are based, and therefore are not interpreted here.

Table 35

Number of Tests (*n*), Fail Rate, and Mean Number of Errors for Each Test Attempt by Applicant Group for the Hazardous Materials Test

Test type Attempt	*n*	Fail rate (%)	Mean errors
Originals			
First	432	31.5	3.3
Second	105	42.9	4.1
Third	30	46.7	4.0
Total	567	34.6	3.5
Renewals			
First	276	39.5	3.8
Second	76	50.0	4.4
Third	26	30.8	3.8
Total	378	41.0	3.9

Note. The figures presented for "Total" (all attempts combined) are weighted averages.

Item Statistics. Table 36 presents the item-total correlation coefficient for each item on each Hazardous Materials test form for original and renewal applicants. Each applicant group has only one question with a value below the .10 standard of acceptability.

Table 37 presents the item responses for each question on each form by applicant group.

Table 38 lists the suspect items that need to be reviewed and possibly revised or replaced.

Table 36

Item-Total Correlation Coefficient for Each Item on Each Form of the Hazardous Materials Test for Original and Renewal Applicants on the First Test Attempt

	Originals			Renewals		
Item	Form 1 (n = 156)	Form 2 (n = 144)	Form 3 (n = 132)	Form 1 (n = 156)	Form 2 (n = 144)	Form 3 (n = 132)
1	.64	.30	.48	.30	.13	.43
2	.53	.27	.41	.31	.38	.39
3	.51	.54	.33	.41	.33	.43
4	.52	.43	.29	.39	.16	.31
5	.66	.59	.23	.61	.35	.15
6	.42	.44	.30	.42	.20	.19
7	.46	.34	.16	.34	.50	.41
8	.43	.41	.11	.40	.13	.11
9	.19	.48	.54	.46	-.07	.38
10	.29	.54	.56	.29	.45	.57
11	.47	.27	.20	.44	.31	.25
12	.52	.49	.45	.37	.18	.24
13	.11	.58	.42	.31	.33	.45
14	.56	.33	.09	.36	.57	.30
15	.57	.27	.37	.46	.11	.37
16	.25	.40	.44	.35	.46	.38
17	.60	.37	.33	.32	.39	.31
18	.23	.27	.42	.17	.37	.32
19	.40	.47	.35	.44	.16	.17
20	.50	.43	.34	.30	.22	.45

Note. The entry for n at the top of each column is the number of usable first-attempt tests used to compute the item-total correlations in the column. Shading identifies items with a value below .10, which indicates the items probably need to be revised or replaced.

Table 37

Percentage of Applicants Selecting Each Answer Choice for Each
Item on Each Form of the Hazardous Materials Test for Original
and Renewal Applicants on the First Test Attempt

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
1	a	83.9	7.7	84.7	82.8	8.2	82.1
	b	3.2	90.9	6.9	2.2	88.8	9.5
	c	12.9	**1.4**	8.4	15.1	3.1	8.3
2	a	8.3	7.0	18.3	7.5	14.3	18.8
	b	80.1	11.3	3.1	77.4	9.2	8.2
	c	11.5	81.7	78.6	15.1	76.5	72.9
3	a	2.6	92.3	3.0	2.2	93.9	**1.2**
	b	81.4	**1.4**	10.6	78.3	**1.0**	14.1
	c	16.0	6.3	86.4	19.6	5.1	84.7
4	a	87.8	4.2	**0.0**	90.3	**0.0**	3.5
	b	3.8	87.4	3.1	2.2	92.9	**1.2**
	c	8.3	8.4	96.9	7.5	7.1	95.3
5	a	67.1	2.8	**1.5**	62.4	**1.0**	**1.2**
	b	14.2	95.8	4.6	20.4	99.0	7.1
	c	18.7	**1.4**	93.9	17.2	**0.0**	91.8
6	a	**1.9**	7.7	**0.0**	**0.0**	6.1	**1.2**
	b	93.6	3.5	89.3	96.8	8.2	88.2
	c	4.5	88.8	10.7	3.2	85.7	10.6
7	a	18.1	21.7	10.7	32.3	20.4	32.1
	b	76.8	5.6	84.0	62.4	14.3	65.5
	c	5.2	72.7	5.3	5.4	65.3	2.4
8	a	3.2	4.2	**0.0**	3.2	3.1	5.9
	b	94.2	92.3	96.9	95.7	92.9	87.1
	c	2.6	3.5	3.1	**1.1**	4.1	7.1
9	a	**0.0**	88.1	6.1	**1.1**	87.8	3.5
	b	98.1	2.1	17.4	96.8	**1.0**	20.0
	c	**1.9**	9.8	76.5	2.2	11.2	76.5
10	a	7.7	70.6	20.8	6.5	68.4	24.7
	b	9.0	2.8	70.0	20.4	**0.0**	70.6
	c	83.3	26.6	9.2	73.1	31.6	4.7

51

Table 37 (continued)

Item	Answer choice	Originals			Renewals		
		Form 1	Form 2	Form 3	Form 1	Form 2	Form 3
11	a	4.5	78.3	15.3	7.5	72.4	8.2
	b	16.0	10.5	5.3	18.3	19.4	10.6
	c	79.5	11.2	79.4	74.2	8.2	81.2
12	a	9.0	93.7	8.3	10.8	94.9	8.3
	b	10.3	5.6	70.5	8.6	2.0	65.5
	c	80.8	**0.7**	21.2	80.6	3.1	26.2
13	a		11.9	8.5		8.2	
	b		83.2	28.5		88.8	
	c	98.7	4.9	63.1	98.9	3.1	48.8
14	a	75.5	11.2	6.1	73.1	7.1	9.4
	b	**1.9**	76.9	90.2	7.5	73.5	88.2
	c	22.6	11.9	3.8	19.4	19.4	2.4
15	a	26.3	10.5	82.2	39.1	10.4	84.7
	b	71.2	14.0	11.6	58.7	5.2	5.9
	c	2.6	75.5	6.2	2.2	84.4	9.4
16	a	**1.3**	70.8	65.6	3.2	58.8	50.6
	b	5.2	3.5	**0.8**	4.3	3.1	9.4
	c	93.5	25.7	33.6	92.5	38.1	40.0
17	a	8.4	2.1	6.8	12.9	5.2	6.0
	b	74.0	90.1	**1.5**	62.4	91.8	4.8
	c	17.5	7.7	91.7	24.7	3.1	89.2
18	a	87.1	83.7	93.1	89.2	77.6	92.9
	b	3.9	5.7	2.3	5.4	6.1	3.5
	c	9.0	10.6	4.6	5.4	16.3	3.5
19	a	91.6	6.9	90.2	92.5	5.1	92.9
	b	7.7	**0.0**	**0.8**	4.3	**1.0**	**1.2**
	c	**0.6**	93.1	9.1	3.2	93.9	5.9
20	a	29.2	2.8	95.4	32.3	4.1	91.8
	b	68.2	86.1	**1.5**	61.3	85.7	2.4
	c	2.6	11.1	3.1	6.5	10.2	5.9

Note. An underlined percentage indicates that the answer choice was the correct response. Shading indicates that the item has a pass rate that is too low or too high and therefore needs to be reviewed and possibly revised or replaced. A boldfaced percentage indicates that the distracter selection rate is too low or too high.

Table 38

Summary of Problem Items on Each Form of the Hazardous
Materials Test for Original and Renewal Applicants

Problem indicator	Applicant group	Form 1	Form 2	Form 3
Item-total correlation too low or negative	Originals			14
	Renewals		9	
Pass rate too high	Originals	9, 13	5	4, 8, 20
	Renewals	6, 8, 9, 13	5	4
Pass rate too low	Originals			
	Renewals	15	16	13, 16
Distracter selected too often	Originals			
	Renewals			
Distracter selected too infrequently	Originals	6, 9, 13, 14, 16, 19	1, 3, 5, 12, 19	4, 5, 6, 8, 16, 17, 19, 20
	Renewals	6, 8, 9, 13	3, 4, 5, 9, 10, 19	3, 4, 5, 6, 19

RECOMMENDATIONS

This statewide evaluation of the department's CDL written knowledge tests has revealed deficiencies in some of the tests and how the tests are administered. There are several steps the department can take to improve the quality of the CDL written knowledge tests. The following specific actions are recommended:

1. Questions with any of the following statistical characteristics should be reviewed and revised or replaced as necessary: (a) an item-total correlation less than .10, (b) a distracter selected more often or within 10% of the correct answer, or by fewer than 2% of the respondents, or (c) a pass rate that is too high (95% or higher) or too low (60% or lower).

2. Test forms with an internal-consistency reliability coefficient less than .70 for either applicant group should be revised to bring them up to this standard. This can be accomplished to some extent by correcting problem items, especially those with low or negative item-total correlations. Another way would be to increase the number of questions on the tests, since longer tests tend to be more reliable than shorter ones. It is advisable that all tests have at least 20 questions, and preferably more if the subject matter content is extensive enough in scope to support additional questions. Certainly, 15 items are not sufficient to ensure an adequate level of test reliability.

3. Test forms with fail rates that differ significantly from one another by more than 10 percentage points should be revised to make them more equal in difficulty to ensure fairness and consistency of outcomes. Specifically, the General Knowledge, Combination Vehicles, and Passenger Transport tests have form fail rates that exceed this tolerance limit for one or other of the applicant groups. The disparity between form fail rates is especially large for the Passenger Transport test for renewals, with Form 3 having a fail rate that is more than 30 percentage points higher than the fail rates for both Forms 1 and 2. Ideally, an applicant should have the same likelihood of failing any of the three forms of a test. Correcting items that are too easy or too difficult, as recommended above, might bring these forms into greater equivalence, but this is not guaranteed. Therefore, those responsible for revising the tests should be mindful of the likely effect that changing the difficulty of items would have on the uniformity of test form difficulties. If requested, R&D can provide guidance on how to make the test form fail rates more balanced.

4. Unless there is a compelling reason to do otherwise, the minimum passing score for the Air Brakes test should be increased to four allowed errors (from the three now allowed) to be consistent with the 80% passing standard set for the other tests. Of course, the passing scores of the tests would need to be changed if the numbers of questions on the tests are increased, as recommended above.

5. The Air Brakes test and associated content in the CDL handbook should be reviewed to determine why originals do so poorly on the test when the renewals do so well, which is the reverse of the findings for the other tests showing originals to generally outperform renewals. Departmental consultants on this project believe that originals do more poorly than renewals because it is difficult to learn about the operation of air brakes by reading the handbook without having operational experience. (Apparently, the subject matter for the other tests can be more easily learned from reading the handbook.) However, it is possible that the handbook material covering this subject matter can be better developed and presented to facilitate greater learning and hopefully better performance on the test.

6. The order of answer choices should be randomized at least annually to decrease the opportunity for cheating or rote memorization of the correct answers. Computer applications are available that can do this efficiently and cost-effectively. R&D can provide guidance in this matter if requested.

7. The department should reverse the policy that allows examiners to restate or rephrase missed test questions when the applicant initially misses too many questions to pass. This would increase the integrity of the testing process by eliminating examiner subjectivity in determining whether the customer has an adequate understanding of traffic laws and safe driving practices. It would also protect the department against charges of discrimination based on the gender, age, race, or other characteristics of the applicant rather than his or her level of knowledge. This recommendation is also one of the safety-improvement strategies proposed in the Strategic Highway Safety Plan approved by the Governor. At a minimum, the practice of rephrasing missed questions for originals should be stopped because it is inconsistent with current policy (allowing this only for renewals).

8. Applicants who fail a test should have to wait a minimum period of time before retesting. At a minimum, applicants should not be permitted to take more than one knowledge test of the same type per day. This one-test-per-day limitation could probably be implemented through departmental policy, but longer waiting periods may require enabling legislation. Instituting a mandatory waiting period of even one day would make it more likely that applicants would study the commercial driver handbook more thoroughly before their first test attempt to avoid possible repeat field office visits. It should also increase the amount of study before retaking the test following a test failure, which should improve their performance on subsequent test attempts. This recommendation was also included in the Strategic Highway Safety Plan.