STATE OF CALIFORNIA
DEPARTMENT OF MOTOR VEHICLES

EVALUATION OF THE CLASS C DRIVER LICENSE WRITTEN KNOWLEDGE TESTS

By
Scott V. Masten

March 1999

Research and Development Branch
California Department of Motor Vehicles
Technical Monograph 182

PREFACE

This report is being issued as an internal monograph of the Department of Motor Vehicles's 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.

ACKNOWLEDGMENTS

This project was conducted under the general direction of Raymond C. Peck, Research Chief, and the supervision of Robert A. Hagge, Research Manager I. Douglas Luong, Management Services Technician, Paul Choate, Research Program Specialist I, and Tony Woo Associate Information Systems Analyst, helped with the sorting, coding, and preparation of data for system entry and also assisted in entering table data.

SUMMARY

Introduction

- This report presents the results of an evaluation of the English DL 5 (Rev. 10/98), Spanish DL 5 (Rev. 5/98), and English DL 5T (Rev. 8/98) Class C license written knowledge examinations. Specifically, the study assessed the fail rate, mean number of errors, and internal-consistency reliability for each test form, as well as the pass rate, percentage of applicants selecting each answer choice, and item-total correlation for each item on each test form for the English tests. Also presented is an assessment of the randomness of the answer choice assignment for the English DL 5.

- As part of the department's Driver Competency Project in the early 1990's it was proposed that written knowledge examinations with a perforated answer sheet be used, and that applicants only have this answer strip returned when the tests were graded, not the test questions themselves (see January 5, 1994 memo from Carol Bedwell to Anne Bersinger, Division Chiefs). This answer strip would include the page numbers where the knowledge domains for each item are presented in the driver handbook so that applicants could study the content area for items missed. This concept was never implemented because of fear that the public would react negatively and that it would result in increased customer interaction time. In an attempt to simulate the worst-case public relations scenario of using written tests with a perforated answer strip, the impact of not returning graded tests to customers at all was evaluated.

- The tests have been extensively modified since the 1998 evaluation, and the intention of this evaluation is to assess the effects of these changes, particularly with regard to the test fail rates (Masten, 1998). Because of the creation of additional test forms and a large number of new items, it was anticipated that this analysis would identify a significant number of faulty items and that further revisions of the tests would be needed.

- The results are based on 10,259 completed test forms that were collected statewide on November 5, 1998.

Results

- The fail rates of the tests, except those for English renewal applicants, are higher than those reported in the 1998 written test evaluation. The overall fail rate for applicants on their first-attempt is 67.0% for English originals, 45.7% for English renewals, 86.6% for Spanish originals, 84.9% for Spanish renewals, and 63.3% for English provisionals. The differences between the 1998 and 1999 written test evaluation fail rates for these groups are illustrated in Figure 1.

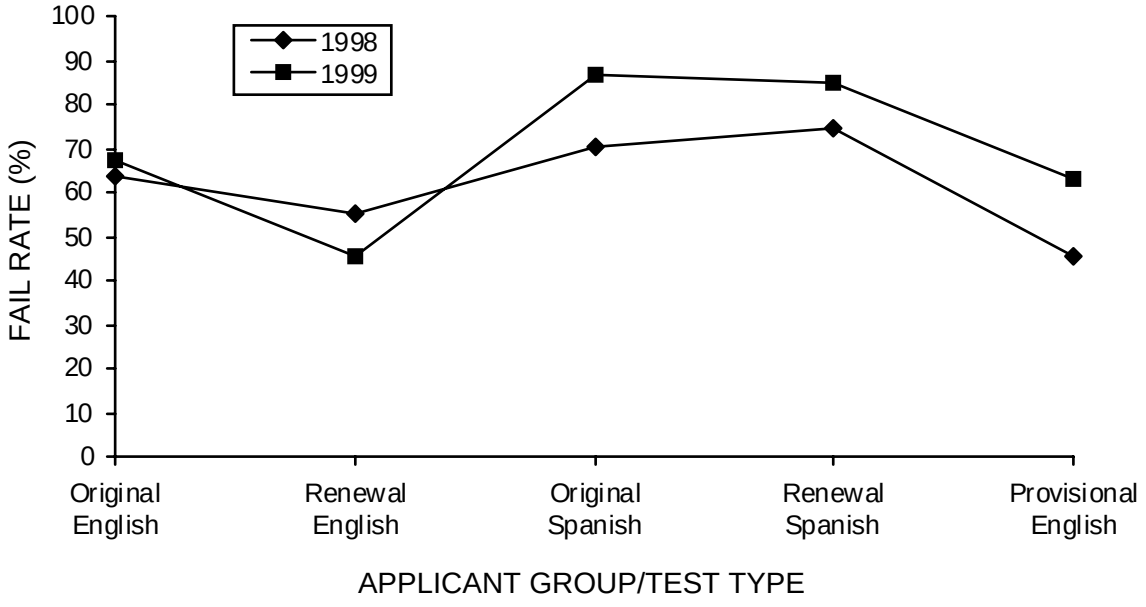

Figure 1. 1998 and 1999 written test evaluation first-attempt fail rates for English and Spanish original and renewal applicants and English provisional applicants.

- For all tests and applicant groups, the fail rates remain surprisingly high on second, third, and fourth or higher attempts. The fail rates for English renewals tend to increase slightly over successive attempts, while those for the other tests and applicant groups tend to decrease slightly. The high failure rate on repeated attempts is extremely discouraging because it suggests that applicants are not reviewing the driver license handbook and items missed before retaking the test. The test and applicant group fail rates over successive attempts are illustrated in Figure 2.

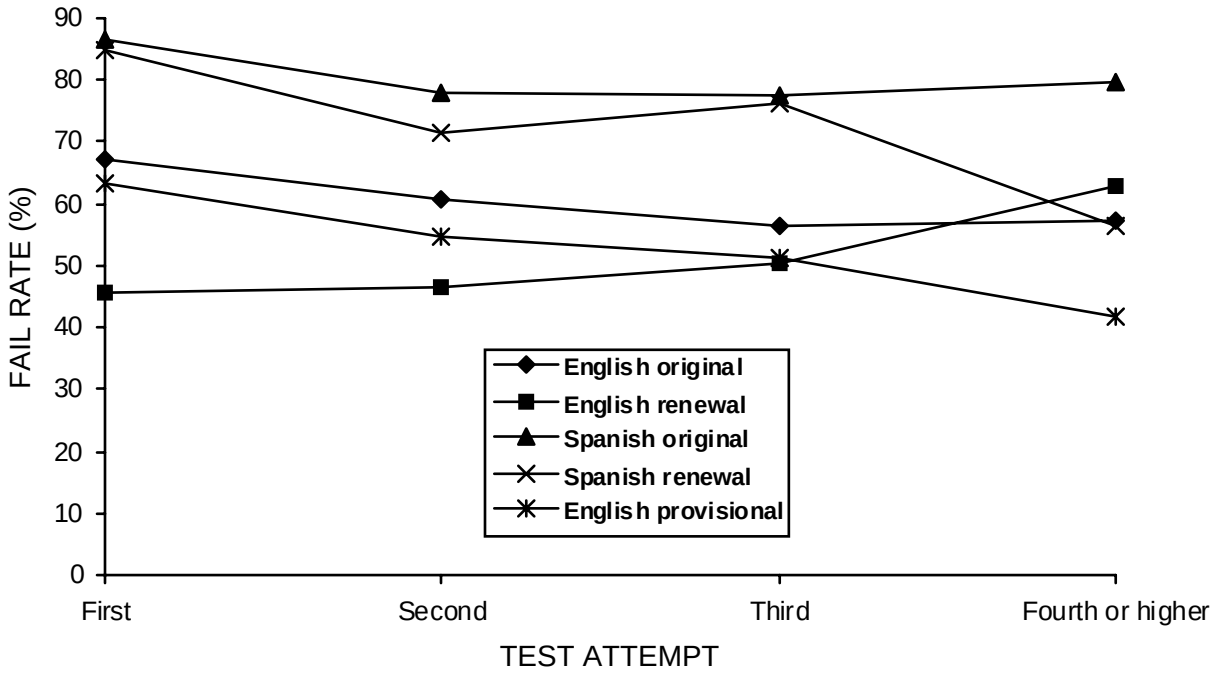

Figure 2. Test fail rates for English and Spanish original and renewal applicants and English provisional applicants over successive attempts.

- Some forms of the same test are more difficult than others. The fail rates for some forms of the same test differ by as much as 41 percentage points.

- Almost all of the internal-consistency test reliabilities for the English and Spanish DL 5 for original applicants and the English DL 5T for provisional applicants are

adequate, and some are even "good" or "excellent." Any of the test forms with reliabilities below .70 should be reviewed and revised to improve the reliability. These reliabilities can be increased and made more homogenous across different forms of the same test by correcting or replacing problem items, particularly those with low or negative item-total correlations.

- None of the English and Spanish DL 5 test form reliabilities for renewal applicants exceed the .70 whole-test reliability standard. Those for Form 6 of the English DL 5 for renewal applicants and Form 2 of the Spanish DL 5 for renewal applicants are truly inadequate and require serious review and revision. Forms 6 and 7 of the English DL 5 for renewals have the lowest internal-consistency reliabilities and also the highest number of items with zero or negative item-total correlations. These two forms are in particularly bad shape and require special attention to correct the items with low item-total correlations which diminish their reliability. Overall, the renewal test reliability findings indicate that 18 may be an insufficient number of items to produce renewal driver license knowledge tests with adequate reliability.

- As would be expected in the first iteration of evaluating so many new test items, a large number of items on each test form are potentially deficient due to a low item-total correlation, a pass rate that is too high or too low, or a distractor selection rate that is too high or too low. Items having two or more problem characteristics, such as those with poor item-total correlations and pass rates that are too high or too low, are the most likely to be deficient, and special care should be taken to review and correct them.

- The much higher fail rate for English originals compared to that for English renewals is partially caused by different passing standards for the groups. If the number of allowable incorrect responses (misses) for original applicants were relaxed to be proportionally the same as the number of allowed misses for renewal applicants (i.e., allowing six rather than five errors), their fail rate would be expected to decrease by 10 percentage points or more. If the proportion of allowable misses for provisional applicants was similarly equated to that for renewal applicants (i.e., allowing eight rather than seven errors), their fail rate would be expected to decrease by 16 percentage points or more.

- Answer choice "b" is the correct answer much more often than are choices "a" or "c" for both the original and the renewal English DL 5 tests.

- Some field offices administered the back side of the DL 5 test sheet to renewal applicants, which is inconsistent with department policy in the Driver License Manual.

- Examiners often did not count all missed items when calculating the total test score for original, renewal, and provisional applicants. This resulted in the computer-graded fail rates reported in this evaluation being slightly higher than the true operational fail rates. The true operational fail rate across all test attempts is 64.9% for English originals, 39.1% for English renewals, and 60.4% for English provisionals. The practice of awarding points to original and provisional applicants based on their verbal responses is inconsistent with department policy.

- Many field offices are still using older revisions of the Class C license tests. This practice diminishes the effectiveness of the current policy of randomizing the English DL 5 every 3 months to curtail applicant cheating.

- The vast majority of customers who were tested on the day of data collection did not have a question, concern, or complaint about not having their graded test returned to them. About 13 of every 100 customers tested raised a question, concern, or complaint, and the field office personnel spent an average of roughly one additional minute dealing with each of these customers.

- The volumes reported in this study underestimate the true daily testing volumes because 27 field offices did not submit any test forms for the evaluation. In addition, eight field offices did not submit any English DL 5 forms of the revision being evaluated, which resulted in the data they provided being excluded from the estimations.

<u>Recommendations</u>

- The tests should be reviewed and revised to make them more sound measurement instruments. In particular, the difficulty level and reliability should be made more homogenous across different forms of the same test, and items with deficient characteristics should be reviewed and modified as necessary. The following five recommendations should help accomplish these goals:

    1. Items with item-total correlations that are below .10 or negative need to be modified or replaced, particularly those with other item deficiencies such as pass rates that are too high or too low. Items with this characteristic are usually

poorly worded and confusing, and definitely need to be revised or replaced. Doing so is likely to increase the internal-consistency reliability of the tests overall, and make them more homogenous across different forms of the same test, particularly in those cases where the reliability was inadequate or borderline.

2. Items with pass rates that are too high or too low, or with distractor selection rates that are too high or too low, should be reviewed for possible problems and modified as necessary. The low pass rate items should be revised to eradicate confusing wording, and those that have little relevance to safety and are conceptually difficult to grasp should be replaced. This would help bring the test fail rates closer to historical levels.

3. Items should be checked to ensure that the knowledge covered by them is contained in the driver handbook and has relevancy to safety, mobility, or other knowledge domains deemed important to driving or safety.

4. The answer choices on each test should be truly randomized and balanced on each test form to decrease the applicants' chances of guessing the correct answers. There are computer applications available that can efficiently and cheaply accomplish this goal. Rewriting item distractors in which none, or almost none, of the applicants chose should also be completed to increase test reliability.

5. A task force composed of knowledge experts from various areas of the department should be convened to accomplish the test revisions.

- The department might want to consider allowing one additional error for original and provisional applicants. This would equalize the expected competency levels of all applicant types as well as lower the test fail rates for original and provisional applicants.

- In light of the high fail rates for the tests, it is increasingly important that the tests have adequate reliability. The low reliability of the renewal tests indicates that applicants taking the test on two occasions or different forms of the test are likely to get disparate scores affecting licensure status (pass vs. fail). The department should therefore consider increasing the length of the renewal tests to increase their reliability, possibly by having renewal applicants complete all 36 items on the DL 5

instead of only the first 18. It is estimated that this doubling of the test length would increase the overall English DL 5 renewal test internal-consistency reliability from .51 to .68, and the overall Spanish DL 5 renewal test reliability from .45 to .62. These reliabilities would be expected to increase even more if items with poor item-total correlations are also reviewed and revised.

- Considering the small percentage of customers raising a question, concern, or complaint about not having their graded test returned at all, and the negligible increase in field office processing time associated with handling these customers, management should reconsider implementing the use of written tests with a perforated answer strip. This would enable applicants to still receive some feedback about the content area where their knowledge was deficient, increase test security, and make it more likely that customers would study the driver handbook.

- Steps should be taken to ensure that field office personnel are administering only current versions of the tests in accordance with the department procedures stated in the Driver License Manual. Doing so would ensure the department's compliance with California statute and increase the effectiveness of randomizing the English DL 5 test every 3 months to reduce the possibility of cheating.

- After the tests are revised, they should be evaluated to determine whether the changes have produced the desired results.

- The department should expedite consideration of technology for computer generating a unique test for each applicant from a large item pool data base. Vendor interest proposals have been solicited for accomplishing this goal and AAMVA has already developed a computerized commercial driver license item pool.

- The following strategies should be considered for improving test scores:

    1. The department should prepare and distribute information brochures and press releases that publicize the content areas and principles that applicants most frequently have problems with.

    2. The department should consider initiating legislation to charge applicants for repeat testing. This could be done by charging for each written test attempt or allowing three attempts as under current law and then imposing a much larger fee for a second application after three written test failures.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

APPENDICES

LIST OF TABLES

TABLE OF CONTENTS (continued)

LIST OF TABLES (continued)

LIST OF FIGURES

TABLE OF CONTENTS (continued)

LIST OF FIGURES (continued)

INTRODUCTION

This report presents the results of an evaluation of the English DL 5 (Rev. 10/98) and English DL 5T (Rev. 8/98) Class C license written knowledge examinations. Test statistics are also presented for the Spanish DL 5 (Rev. 5/98), which is based on a translation of the English DL 5 (Rev. 4/98).

The last written test evaluation, completed in January 1998, indicated that the test fail rates were substantially higher than was previously the case (Masten, S. V. [1998]. *Evaluation of the Class C driver license written knowledge tests* [Report No. 173]. Sacramento: California Department of Motor Vehicles). In order to remove any contribution to the high fail rates that may have been caused by confusing item wording, the tests were subsequently extensively modified based on the report recommendations. Items with low item-total correlations or pass rates that were too high or too low were rewritten or removed. One of the four answer choices for each item was also removed, which resulted in all the current tests having only three answer choices. The remaining answer choices were reworded, removed, or modified if their selection rates were too high or too low, or if it was believed that there was some other reason to do so.

Five additional versions of the English DL 5 and two additional versions of the English DL 5T were created. This was accomplished by pooling all the older items (modified and unmodified) with a batch of newly created items, and then randomly assigning the items in the pool to the test forms (with the exception of items relating to blood alcohol level and reporting the sale of a vehicle, which are required to be on the front of each test form). Existing policy requires the Communications Services Branch to randomize the order (sequence) of the items on each test form of the English DL 5 and the answer choices for each item every 3 months. The purpose of periodically randomizing the tests is to reduce the ability of applicants to compile and use crib sheets for cheating.

The objective of this evaluation is to assess the effects of these changes, particularly with regard to the test fail rates. In addition, item statistics needed to be determined for the large number of new items which were added to the tests. The findings are intended to assist the Department of Motor Vehicles (DMV) in creating future revisions of the examinations that are more reliable and valid.

There are 10 different versions or forms of the English DL 5 examination and five forms of the Spanish DL 5 examination, each consisting of 36 questions. These tests are administered to Class C license applicants who are 18 years of age or older. License renewal applicants complete only the first 18 items, while applicants for an original license complete all 36 items. There are five different forms of the English DL 5T examination, each consisting of 46 questions. This test is administered to original Class C license applicants who are younger than 18 years of age (provisional licensees). Copies of the evaluated test forms are in Appendix E.

This report presents the fail rate, mean number of errors, and internal-consistency reliability for each form of the English and Spanish DL 5 tests and the English DL 5T test. Results for the English and Spanish DL 5 tests are presented separately for original and renewal applicants, and all test fail rates are presented separately for first, second, third, and fourth or higher attempts on the tests. In addition, the pass rate, percentage of applicants choosing each answer choice, and item-total correlation for each item on each test form, and also the percentage of applicants who would pass each test form at different cut-points, are presented for the English DL 5 and DL 5T tests. All of the item statistics are based on first-attempt applicants (those taking the test for the first time on the current application). Item statistics for the Spanish DL 5 test are not presented because an insufficient number of test forms of this type was collected for accurate computation. Similarly, neither item nor test statistics are presented for the Spanish DL 5T (Rev. 5/98) because an insufficient number of tests were collected. Assessments of the randomness of the answer choice assignment for the English DL 5 and the impact of not returning graded tests to customers are also presented.

METHODS

Data Collection

The completed tests used in this evaluation were collected from all field offices on November 5, 1998. Offices were asked to collect all first, second, third, and fourth or higher attempts of all English and Spanish DL 5 and DL 5T examinations administered on that day. The field office personnel were instructed to write "1st," "2nd," "3rd," or "4th or higher" in the upper-left hand corner of each test given to indicate the test attempt on the current license application. If the test was a DL 5, the field office personnel were instructed to write "ORIG" at the top of the test if the applicant was applying for an original license, or "REN" if he or she was applying for a renewal license. Field office personnel were further instructed to administer and collect only the most current revisions of the examinations, score tests in the usual manner, circle the correct answers to missed questions, write the total number of items missed and field office number on the front of the test sheet, and not return graded tests or copies of them to applicants.

The current practice in the majority of field offices is to return graded tests to customers for studying or other purposes. It was suggested in the early 1990's as part of the department's Driver Competency Project that one way of tightening test security and reducing cheating would be to use written examinations with a perforated answer strip (see January 5, 1994 memo from Carol Bedwell to Anne Bersinger, Division Chiefs). This answer strip would include the page number where the knowledge domain for each item is presented in the driver handbook, and would be removed and returned to the applicants instead of the test questions themselves. This suggestion was not implemented because it was believed that the public would react negatively, and increased processing time would result from handling customer complaints and verbally reviewing missed items with the applicants. Because graded tests were not to be returned during data collection for this evaluation, it was decided to study whether not returning the tests at all would indeed result in public disfavor and increased processing time during the study data collection day. Although this is different than the suggested use of a perforated answer strip, the results of not returning anything to the applicants would be expected to demonstrate a worst-case customer relations testing scenario. Hence, the results would likely overestimate the percentage of complaints and amount of increased processing time that would occur if applicants had received a perforated answer strip listing the page number of the content area of each

item. The data collection for this pilot study was accomplished in the following manner.

To determine the impact of not returning graded tests to the customers, the field office personnel were instructed to write "YES" on the bottom-front of the test sheet if the customer raised a question, concern, or complaint about not getting his or her graded test sheet back. In addition, they were told to write the amount of increased processing time associated with handling the question, concern, or complaint (to the nearest 5-second interval). If the customer did not raise a question, concern, or complaint about not receiving the test sheet back, the field office personnel were instructed to write "NO" at the bottom-front of the test sheet. At the end of the day on November 5, 1998 the offices were to package all the completed tests, identify the field office on the envelope, and forward the packages to DMV headquarters.

The tests were screened by the Research and Development Branch (R&D) and electronically keyed by the Data Entry unit within the Registration Services Branch.

Data Analysis

A statistical technique known as analysis of variance (ANOVA) was used to determine if statistically significant differences exist between the test form fail rates or average scores. An alpha level of .05 is used to evaluate the statistical significance of differences between test forms, which means that the differences are deemed significant if the probability of their occurrence by chance alone is less than 5 times in 100. Games and Howell multiple comparison tests were used for post hoc analyses when a significant omnibus ANOVA was encountered. A family-wise alpha level of .05 is maintained in the post hoc comparisons for all dependent measures.

The internal-consistency reliability of each test form was computed using the Kuder-Richardson (K-R 20) formula. In general, this type of reliability indicates the degree of uniformity among test items and the extent to which the test items measure a common domain of knowledge. It also serves as a gauge of the overall precision of the test as a measurement instrument. A test that is highly reliable is likely to result in very similar scores across repeated testings of the same people (assuming a fixed knowledge level between test administrations). The reliability coefficient can range from 0 to 1, where a value of 0 indicates no similarity between the test items and a value of 1 denotes that the items are perfectly homogenous. Coefficients closer to 1 are more desirable, and any test form with a coefficient below the .70 whole-test standard is probably in need of revision.

The item-total correlation indicates the degree to which performance on the item coincides with performance on the test. The item-total correlation coefficient can theoretically range from -1.00 to 1.00, with the zero midpoint representing no relationship at all between performance on the item and performance on the test. In other words, performance on items with correlations approaching zero (between -.10 and .10) has very little or no relationship to whether the applicant scored high or low on the examination. Items with positive item-total correlations are more likely to have been answered correctly by applicants who scored higher on the test, while items with negative item-total correlations are more likely to have been answered correctly by applicants who scored lower on the test.

Because the answer choices for each item on the English DL 5 are randomized every 3 months, the proportion of times that each answer choice (a, b, or c) is the correct answer should be approximately equal (33% each) across the entire item pool and for each test form. Having randomized answer choices is desirable because it assures that the chance of an applicant guessing the correct answer is truly one in three. The chi-square (χ^2) Goodness-of-Fit test was used to determine if answer choices "a," "b," and "c" are the correct answers in similar proportions across the entire item pool for the English DL 5 tests for both original and renewal applicants. The chi-square (χ^2) Test of Independence was used to determine if the pattern of these proportions was the same for each test form of the English DL 5 for both original and renewal applicants. An alpha level of .05 was used to determine statistical significance.

A paired-samples *t* test was used to compare the mean number of errors that original applicants made on the first half of the English DL 5 to the mean number of errors they made on the second half. The purpose of this comparison was to ascertain whether the first half of the test was easier than the second half, which could help explain any differential in fail rates between renewal and original applicants. Independent-samples *t* tests were used to determine the statistical significance of differences between the test scores and the fail rates for original and renewal applicants on the first 18 items of the English DL 5. The purpose of these comparisons was to evaluate whether any obtained differences in fail rates between original and renewals may have been due to differing knowledge levels and/or differences in passing standards. An alpha level of .05 was used to determine statistical significance in all *t* tests.

RESULTS

Data Collection and Screening

A total of 10,259 test forms were received by R&D. The number of each type of form received is presented in Table 1. It is readily apparent from examining the table that outdated revisions of the licensing exams are still being used in many field offices. This, of course, greatly diminishes the effectiveness of randomizing the English DL 5 items and answer choices every 3 months to curtail cheating.

Table 1

Description of Test Forms Received by Research and Development

Test type	Language	Revision	Total
DL 5	English	3/96	82
DL 5	English	7/98	375
DL 5	English	10/98	7,248
DL 5T	English	3/96	105
DL 5T	English	8/98	805
DL 5	Spanish	7/95	175
DL 5	Spanish	10/97	288
DL 5	Spanish	5/98	1,148
DL 5T	Spanish	6/96	18
DL 5T	Spanish	5/98	15
Total			10,259

Note. Because 35 field offices did not report useable data for this evaluation, the volumes reported in the table should not be used to estimate operational testing volumes.

The volumes reported in Table 1 underestimate the true daily testing volume because 27 field offices did not submit any test forms for the evaluation. In addition, eight field offices did not submit any English DL 5 forms of the revision being evaluated, which resulted in the data they provided being excluded from the estimation. (These 35 field offices are identified in Appendix D.)

To determine the degree to which the findings of this evaluation might have been biased by the exclusion of these 35 nonreporting field offices, the percentage of total California driver license activity for November 1998 contributed by these offices was computed. (These volumes were compiled by the Field Office Division and include all driver licensing transactions.) Because these offices were found to account for 24% of all driver license activity, it is possible that their offices biased the test and item statistics presented in this evaluation. It is more likely that the test and item fail rates would be biased by the exclusion of these offices than would the internal-consistency reliability measures, although the true impact of the bias is unknown. However, there is no reason to believe that the excluded field offices differ systematically from those that reported useable data, because the excluded offices are fairly dispersed throughout California and it is unlikely that their lack of reporting would be related systematically to the various test performance measures. Nonetheless, the number of tests received is clearly an underestimate of the daily testing volume, and should therefore not be used to extrapolate quarterly or annual volumes.

A review of the test forms submitted from each field office revealed that there are no offices that submitted a highly disproportionate number of passes or failures that would have indicated a possible over or underreporting of test forms based on test results (which, if present, would have biased the fail rate estimates). The problems introduced by the poor data reporting underscore the importance of having all field offices comply with data collection procedures during department evaluations.

Ordinarily it is desirable that at least 100 first-attempt test forms of a given type be analyzed to produce reasonably accurate estimates of item statistics. This standard was clearly met for the English DL 5 (Rev. 10/98) for original and renewal applicants, and was nearly met for the English DL 5T (Rev. 8/98) for provisional applicants. The volumes for the Spanish DL 5 (Rev. 5/98) and Spanish DL 5T (Rev. 5/98) were far smaller than is necessary to compute accurate estimates of their item statistics. Therefore, item statistics were computed for only the English test forms and applicant groups. However, the volume of Spanish DL 5 tests was sufficient enough to compute total test statistics, so these are presented for original and renewal applicants.

The test forms were screened and 1,043 were removed because they were older versions of the tests, five were removed because they had one or more answers torn off the test,

and 118 were removed because they were missing information on the revision date or form number. Ten tests were also discarded because they had more than five double-marked test questions for original applicants or more than four double-marked questions for renewal applicants, and three more were discarded because they had an extremely high number of answers marked as incorrect (indicating that the tests were likely to have been incorrectly coded or keyed). As indicated earlier, the current Spanish DL 5T test forms received were too few in number (15) to calculate even test statistics, and these were therefore also removed from the evaluation. Finally, 93 renewal tests were removed because the applicant was administered the back of a DL 5 test instead of the front (which violates departmental policy stated in the Driver License Manual). The screening process resulted in 8,972 usable test forms for the evaluation.

The test forms were graded by computer and items were counted as incorrect if the wrong answer choice was marked, the item was left blank, or more than one answer choice was marked. The computer graded test scores were used to compute the test fail rates and internal-consistency reliabilities. All fail rates presented within this report are based on the current passing score of five errors or fewer for DL 5 original applicants, three or fewer for DL 5 renewal applicants, and seven or fewer for DL 5T provisional applicants.

Examiner Scoring Bias

The computer grading of the tests revealed that the actual number of errors made by the applicant often differs from the error score written by the examiner on the front of the test form. The scoring bias for these forms is almost always in the applicant's favor, and often affects the test result. To demonstrate the difference between the computer-graded and examiner-graded fail rates, three different fail rates are presented in Table 2 for English DL 5 original and renewal applicants and English DL 5T provisional applicants. The first fail rate was calculated from computer grading of the tests. The second fail rate was calculated from the scores that the field office examiners wrote on the tests, which were not available for all the tests. The third fail rate represents a combination of the previous grading methods; the examiner score was used to calculate the fail rate if a score was available, otherwise the computer-graded score was used. The differences in the fail rates are illustrated in Figure 3.

Table 2

Number of Subjects and Fail Rate for First-Attempt English Original, Renewal, and Provisional Applicants when Graded by Computer, Examiner, and Examiner/Computer in Combination

	Computer graded		Examiner graded		Examiner/computer graded	
Test type	*n*	Fail rate	*n*	Fail rate	*n*	Fail rate
English DL 5 original	1,530	67.0	1,380	64.9	1,530	64.0
English DL 5 renewal	2,590	45.7	2,258	39.1	2,590	40.6
English DL 5T provisional	474	63.3	449	60.4	474	59.9

Note. Examiner/computer grading involved the use of the examiner score if available and the computer score otherwise.

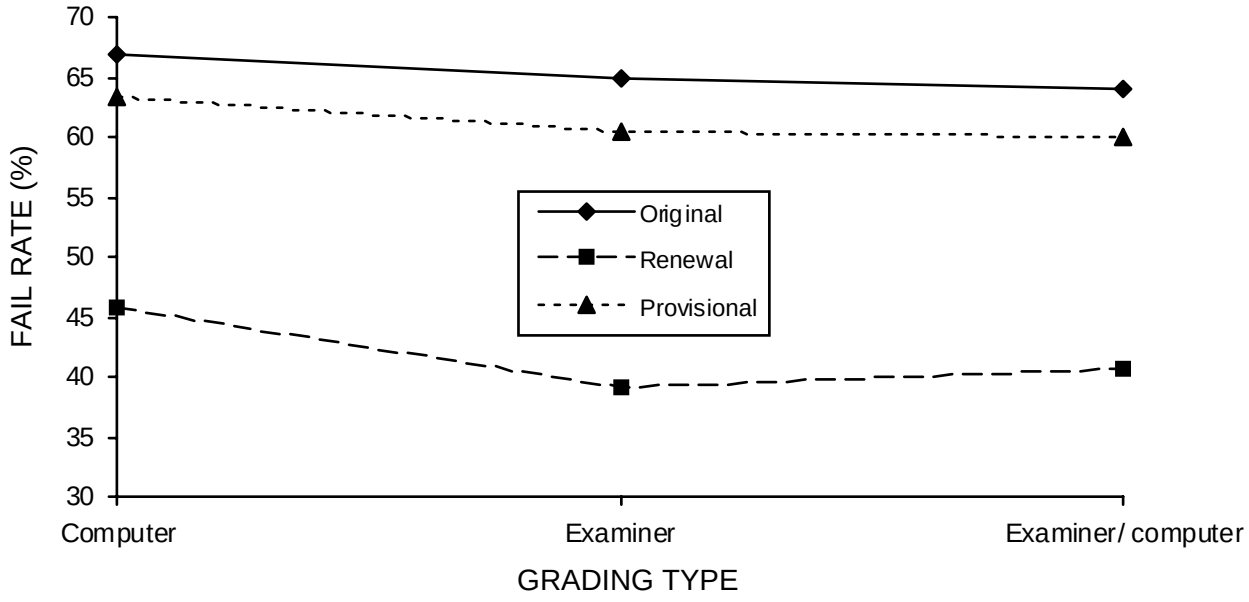

Figure 3. Fail rate for English original, renewal, and provisional first-attempt applicants as a function of type of grading used.

The difference between the examiner and computer scores is primarily due to some examiners' having discussed missed items with examinees and having awarded points based on their verbal responses. Department policy allows this practice only for renewal applicants, and the largest fail rate difference (6.6 percentage points) between the computer and examiner grading was for renewal applicants. The rationale for this policy is that renewals are experienced drivers who have previously passed a written examination and that some of those with marginal failure scores really know the material but missed some items due to nervousness, literacy problems, or ambiguities in item wording. The existence of differences between the computer and examiner fail rates for original and provisional applicants indicates that some field office personnel are also applying this policy to these applicant types. In all cases, however, the examiner scoring bias has caused the test fail rates presented in this report (which are based on computer-graded scoring) to be slightly higher than the true operational fail rates (i.e., examiner fail rates) occurring in the field. The true operational fail rate across all test attempts is 64.9% for English original applicants, 39.1% for English renewal applicants, and 60.4% for English provisional applicants.

Test Statistics

Test Form Difficulty and Reliability

The frequency, fail rate, mean number of errors, and internal-consistency reliability coefficient for each test form and applicant type are presented in Table 3. The differences in the form fail rates and mean errors, and the pattern of internal-consistency reliabilities for the forms are illustrated in Figures 4, 5, and 6, respectively, for each applicant type. The test statistics shown in the table and figures are discussed in the following five subsections.

Table 3

Frequency (*n*), Fail Rate, Mean Number of Errors, and Internal-Consistency Reliability Coefficient for Each Form of the English DL 5 (Rev. 10/98), Spanish DL 5 (Rev. 5/98), and English DL 5T (Rev. 8/98) for First-Attempt Applicants

Test form	*n*	Fail rate	Mean errors	Reliability
Original DL 5 English[a]				
1	149	62.4	7.8	.76
2	145	62.8	7.3	.76
3	157	64.3	7.4	.74
4	148	70.9	8.7	.77
5	156	84.6	9.9	.75
6	167	59.3	6.9	.71
7	166	72.3	7.9	.67
8	139	66.2	7.4	.74
9	163	53.4	6.4	.74
10	140	75.0	8.3	.72
Total	1,530	67.0	7.8	.74
Renewal DL 5 English[b]				
1	259	23.2	2.5	.50
2	263	24.7	2.4	.47
3	283	51.9	3.9	.62
4	258	56.2	4.0	.56
5	247	62.3	4.3	.57
6	242	44.6	3.4	.33
7	248	64.5	4.5	.46
8	271	40.2	3.3	.51
9	258	32.6	3.0	.54
10	261	58.2	4.3	.56
Total	2,590	45.7	3.5	.51
Original DL 5 Spanish[c]				
1	64	84.4	10.5	.79
2	63	80.9	10.9	.84
3	64	92.2	12.5	.81
4	57	91.2	12.1	.78
5	58	84.5	12.4	.82
Total	306	86.6	11.7	.81
Renewal DL 5 Spanish[d]				
1	29	72.4	5.6	.68
2	38	89.5	6.9	.34
3	33	81.8	5.9	.53
4	34	85.3	5.8	.36
5	45	91.1	6.2	.39
Total	179	84.9	6.1	.45
Provisional DL 5T English[e]				
1	98	65.3	9.7	.76
2	102	62.7	9.0	.73
3	88	55.7	9.2	.68
4	91	56.0	8.5	.70
5	95	75.8	10.6	.73
Total	474	63.3	9.4	.72
Grand total	5,079	57.6	6.0	.61

Note. The figures presented for total and grand total fail rate, mean errors, and reliability are weighted averages. All ANOVAs are two-tailed. [a]Forms differ significantly on fail rate ($F = 5.75$, $p < .001$) and mean errors ($F = 8.40$, $p < .001$). [b] Forms differ significantly on fail rate ($F = 26.14$, $p < .001$) and mean errors ($F = 31.08$, $p < .001$). [c]Forms did not differ significantly on fail rate ($F = 1.25$, $p = .29$) or mean errors ($F = 1.66$, $p = .16$). [d]Forms did not differ significantly on fail rate ($F = 1.44$, $p = .22$) or mean errors ($F = 1.49$, $p = .21$). [e]Forms differ significantly on fail rate ($F = 2.74$, $p = .03$) and mean errors ($F = 2.98$, $p = .02$).

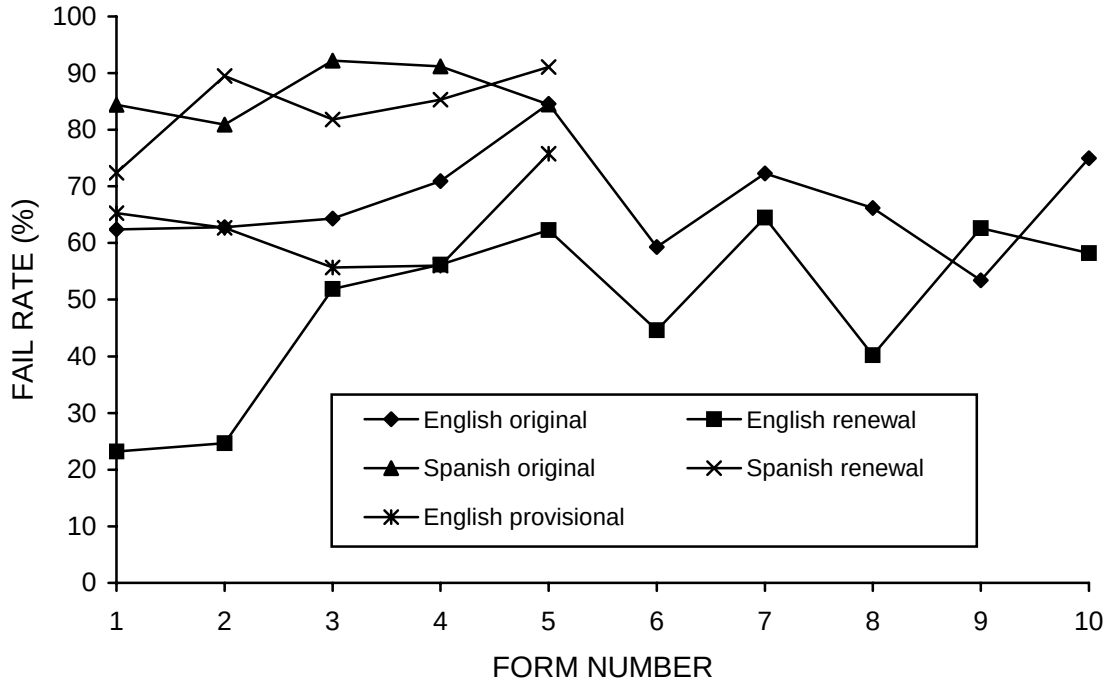

Figure 4. Form fail rate for first-attempt English and Spanish original and renewal applicants (DL 5) and English provisional applicants (DL 5T).

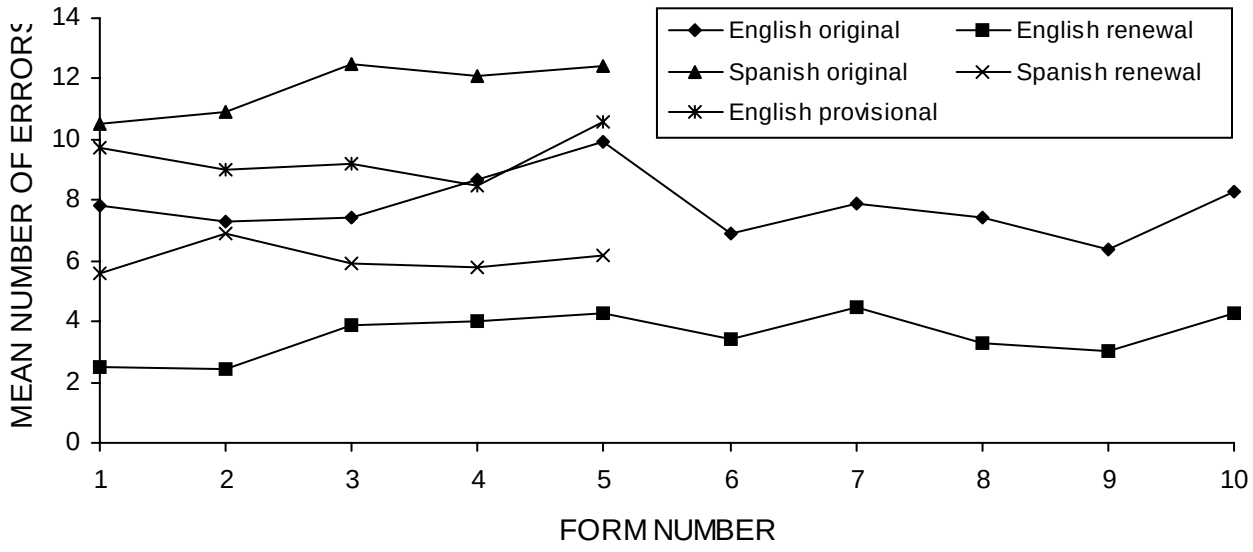

Figure 5. Form mean number of errors for first-attempt English and Spanish original and renewal applicants (DL 5) and English provisional applicants (DL 5T).

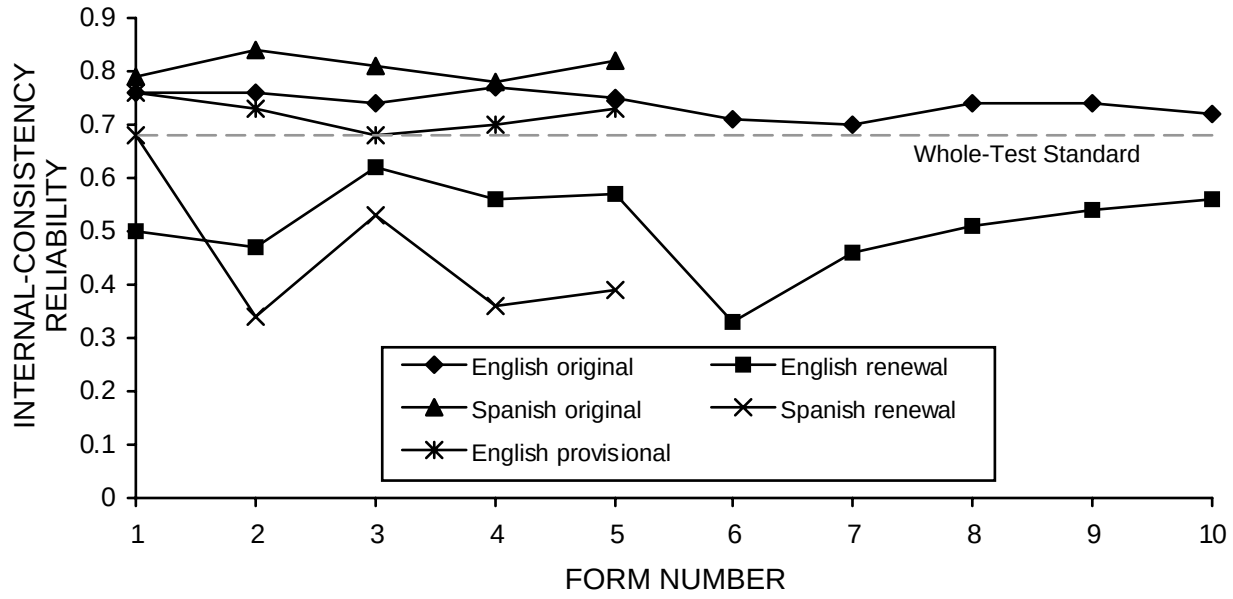

Figure 6. Form internal-consistency reliability for first-attempt English and Spanish original and renewal applicants (DL 5) and English provisional applicants (DL 5T).

English DL 5 original applicants. For original applicants taking the English DL 5, the 10 form fail rates range from 53.4% to 84.6%. The differences between the form fail rates are significant ($p < .001$). The fail rate for Form 5 is higher than the rates for Forms 1, 2, 3, 6, 8, and 9. In addition, the fail rate for Form 9 is lower than the rates for Forms 4, 7, and 10. The differences between the other form fail rates are not significant.

The mean number of errors for the 10 forms range from 6.4 to 9.9, and the differences are significant ($p < .001$). Applicants who completed Form 5 made significantly more errors than did applicants who completed Forms 1, 2, 3, 6, 7, 8, or 9. In addition, applicants who completed Form 9 made significantly fewer errors than did applicants who completed Forms 4, 7, or 10, and those who completed Form 4 made significantly more errors than did those who completed Form 6. Differences between the mean errors for the other forms are not significant.

The K-R 20 reliabilities for the 10 forms range from .67 to .77. These coefficients all fall within the .66 to .79 range of "good" reliability for 36-item test segments, although the reliability for Form 7 is below the .70 whole-test reliability standard.

English DL 5 renewal applicants. The fail rates for the 10 test forms of the English DL 5 for renewal applicants range from 23.2% to 64.5%. The differences between the form

fail rates are significant ($p < .001$). The fail rates for Forms 1 and 2 are significantly lower than the rates for all the other forms, except those for Form 9 and each other. The fail rate for Form 9 is significantly lower than the rates for Forms 3, 4, 5, 7, and 10. In addition, the fail rate for Form 8 is significantly lower than the rates for Forms 4, 5, 7, and 10, and the fail rate for Form 6 is significantly lower than the rates for Forms 5 and 7. The differences between the other form fail rates are not significant.

The mean number of errors for the 10 forms range from 2.4 to 4.5, and the differences are significant ($p < .001$). Applicants who completed Form 2 missed significantly fewer items than did applicants who completed any of the other test forms, and those who completed Form 1 missed significantly fewer items than did applicants who completed any of the other test forms, with the exception of Forms 9 and 2. Applicants who completed Form 9 missed significantly fewer items than did applicants who completed Forms 3, 4, 5, 7, or 10; those who completed Form 8 missed significantly fewer items than did those who completed Forms 4, 5, 7, or 10; those who completed Form 6 missed significantly fewer items than did those who completed Forms 4, 5, 7, or 10; and those who completed Form 3 missed fewer items than did those who completed Form 7. Differences between the other form mean errors are not significant.

The K-R 20 reliabilities for the 10 forms range from .33 to .62. All the coefficients, except that for Form 6, fall within or exceed the .35 to .49 cut-off of acceptability for test segments with 18 items. The reliability of Form 6 falls within the "questionable" range, while the values for Forms 1, 3, 4, 5, 8, 9, and 10 fall within the .50 to .66 range of "good" reliability for test segments of this length. The fact that these reliability coefficients are much lower than those for the 36-item test is due primarily to the difference in test length rather than to any defect in the content of the items (because reliability generally increases as the number of items increases). Because all of the reliabilities are below the .70 whole-test standard, the data indicate that 18 may be an inadequate number of items for producing driver license knowledge tests with adequate reliability.

Spanish DL 5 original applicants. The fail rates for the five forms of the Spanish DL 5 for original applicants range from 80.9% to 92.2%. The differences between the form fail rates are not significant ($p = .29$). The mean number of errors for the forms range from 10.5 to 12.5, and the differences are also not significant ($p = .16$).

The five form K-R 20 reliabilities range from .78 to .84. All these reliabilities fall within or exceed the .66 to .79 range of "good" reliability for 36-item test segments, and the values for Forms 2, 3, and 5 exceed the .80+ lower limit considered "excellent."

Spanish DL 5 renewal applicants. The fail rates for the five forms of the Spanish DL 5 for renewal applicants range from 72.4% to 91.1%. The differences between the form fail rates are not significant (p = .22). The mean number of errors for the forms range from 5.6 to 6.9, and the differences are also not significant (p = .21).

The five form K-R 20 reliabilities range from .34 to .68. All these reliabilities, except that for Form 2, fall within or exceed the .35 to .49 cut-off of acceptability for 18-item test segments. The reliability of Form 2 falls within the "questionable" range, while that for Form 3 falls within the .50 to .66 range considered "good," and the value for Form 1 exceeds the .67+ value considered "excellent." Because all of the reliabilities are below the .70 whole-test standard, the data once again indicate that 18 items may be too few to produce driver license knowledge tests with adequate reliability.

English DL 5T provisional applicants. The fail rates for provisional applicants taking the English DL 5T range from 55.7% to 75.8%, and the differences between them are significant (p = .03). The fail rate for Form 5 is significantly higher than the rates for Forms 3 and 4. The differences between the other form fail rates are not significant.

The mean number of errors for the five test forms of the English DL 5T for provisional applicants range from 8.5 to 10.6, and the differences are significant (p = .02). Applicants who completed Form 5 made significantly more errors than did applicants who completed Form 4. The differences between the other form mean errors are not significant.

The five form K-R 20 reliabilities range from .68 to .76. All these reliabilities fall within or exceed the .62 to .74 range of acceptable reliability for 46-item test segments, and that for Form 1 falls within the .75 to .85 "good" range. However, the reliability for Form 3 falls below the .70 whole-test reliability standard.

Test Difficulty by Attempt

The frequency (n), fail rate, and mean number of errors for first, second, third, fourth or higher, and missing attempt applicants are presented in Table 4.

Table 4

Frequency (*n*), Fail Rate, and Mean Number of Errors for the English DL 5 (Rev. 10/98), Spanish DL 5 (Rev. 5/98), and English DL 5T (Rev. 8/98) for First, Second, Third, Fourth or Higher, and Missing Attempt Applicants

Attempt	*n*	Fail rate	Mean errors
Original DL 5 English			
First	1,530	67.0	7.8
Second	961	60.7	6.9
Third	380	56.3	6.8
Fourth or higher	98	57.1	6.3
Missing	143	74.8	8.2
Total	3,112	63.8	7.4
Renewal DL 5 English			
First	2,590	45.7	3.5
Second	929	46.3	3.7
Third	214	50.5	3.8
Fourth or higher	43	62.8	4.7
Missing	168	45.8	3.7
Total	3,944	46.3	3.6
Original DL 5 Spanish			
First	306	86.6	11.7
Second	240	77.9	9.1
Third	125	77.6	10.0
Fourth or higher	54	79.6	9.2
Missing	28	85.7	10.2
Total	753	81.8	10.4
Renewal DL 5 Spanish			
First	179	84.9	6.1
Second	115	71.3	5.0
Third	46	76.1	5.3
Fourth or higher	16	56.2	4.2
Missing	11	63.6	5.3
Total	367	77.6	5.5
Provisional DL 5T English			
First	474	63.3	9.4
Second	196	54.6	8.3
Third	82	51.2	8.2
Fourth or higher	12	41.7	7.0
Missing	32	59.4	9.1
Total	796	59.4	9.0
Grand total	8,972	57.8	6.0

Note. All figures presented for total fail rate and total mean errors are weighted averages. Missing attempt cases did not have the applicant's attempt number written on the front of the test form.

The examination fail rate averaged over successive and missing attempts is 63.8% for English DL 5 original applicants and 46.3% for English DL 5 renewal applicants, 81.8% for Spanish DL 5 original applicants and 77.6% for Spanish DL 5 renewal applicants, and 59.4% for English DL 5T provisional applicants. For all the tests and applicant groups, the fail rates remain surprisingly high on the second, third, and fourth or higher attempts. The English DL 5 original, English DL 5T provisional, and Spanish DL 5 original and renewal fail rates tend to decrease slightly over successive attempts, while those for English DL 5 renewals increase slightly over successive attempts. The nonreporting of attempt number is not likely to be associated with test performance, and therefore the exclusion of cases without an attempt number specified is not believed to have significantly biased the attempt fail rate estimates. In any event, it is clear that many of these applicants either do little or no preparation or have severe literacy problems.

Disparity Between Original and Renewal Fail Rates

The much lower fail rates for English renewal applicants compared to those for English original applicants is a curiosity, given that both applicant groups complete the DL 5 test. Because renewals only complete the first half of the exam, while originals complete all 36 items, the difference in their fail rates could be the result of the second half of the test being more difficult than the first half. To test this hypothesis, the mean number of errors that the original applicants made on the first 18 items of the DL 5 was compared the mean number of errors they made on the second 18 items using a two-way paired-samples *t* test. The results indicate that the original applicants did not make a significantly different mean number of errors on the first 18 items ($M = 3.9$) than they did on the second 18 items ($M = 3.9$), $t(1,529) = 0.32$, $p = .75$. Therefore, a difference in difficulty between the first and second halves of the English DL 5 does not appear to be the cause of the discrepancy between the original and renewal fail rates.

The mean number of errors made by the original applicants on the first 18 items was then compared to the mean number of errors made by renewal applicants on the same items to determine if the two applicant groups differed in level of knowledge. Results of a two-tailed independent-samples *t* test indicate that the original applicants made more errors on the first 18 items ($M = 3.9$) than did the renewal applicants ($M = 3.5$), $t(4,118) = 4.70$, $p < .001$. The fail rate for the English original applicants based on only the first 18 items of the test and the renewal cut-point (three allowable errors) was also compared to that for the renewal applicants. The results of a two-tailed independent-

samples *t* test indicate that the fail rate for the original applicants on the first 18 items (50.8%) was significantly higher than that for renewal applicants (45.7%), $t(4,118) = 3.19$, $p = .001$.

Note, however, that the fail rate for the original applicants when the more relaxed renewal passing criterion is applied to the first 18 items is more than 16 percentage points lower than their fail rate for the entire 36-item test when the more stringent criterion for originals (5 allowable errors) is applied. Although the 18-item fail rate for original applicants is still 5.1 percentage points higher than that for renewals, the previous exposure of renewal applicants to the written test and practical driving experience on California roads probably accounts for their lower fail rate compared to original applicants, who are generally applying for a California license for the first time. These data indicate that, although a lack of knowledge accounts for some of the difference between the actual original and renewal fail rates, the most influential factor is that the renewal applicants are permitted to fail a higher percentage of the items (3 out of 18 x 100 = 17%) than are original applicants (5 out of 36 x 100 = 14%). If the original applicants were allowed one more error on the entire 36-item test, they would have the same expected competency (6 out of 36 x 100 = 17%) as renewal applicants on the 18-item test, and the fail rate for first attempts would be only 57.8% (a decrease of almost 10 percentage points).

Test Fail Rates by Field Office

The field office fail rate and number of tests received for English DL 5 original and renewal applicants over all test attempts are presented in Appendix D. Field office fail rates that are computed from fewer than 20 test forms are highly unstable, and therefore may not be very accurate. Field office fail rates are not presented for Spanish DL 5 original and renewal applicants, or for English DL 5T provisional applicants, because too few test forms were collected to compute accurate estimates for the majority of the field offices. Also noted in the table are the 27 nonreporting field offices and the eight offices that did not submit any current revision test forms. The fail rates for offices with 20 or more forms range from 34% to 84% for original applicants, and 23% to 74% for renewal applicants. These numbers suggest wide variation between the field office fail rates for these applicant types.

Test of Answer Choice Randomness

The number and percentage of items for which choices "a," "b," and "c" are the correct answers on each form of the English DL 5 are presented in Table 5 for original and renewal applicants. If the assignment of the correct answers was truly randomized across the answer choices, it would be expected that choices "a," "b," and "c" would be the correct answer for about 33% of the items across the entire item pool and also on each test form.

Table 5

Number (*n*) and Percentage of Times that Each Answer Choice was the Correct Answer for Each Form of the English DL 5 (10/98) for Original and Renewal Applicants

Test form	Number of test items	Answer choice					
		a		b		c	
		n	%	*n*	%	*n*	%
Originals[a]							
1	36	14	38.9	15	41.7	7	19.4
2	36	8	22.2	21	58.3	7	19.4
3	36	10	27.8	16	44.4	10	27.8
4	36	12	33.3	15	41.7	9	25.0
5	36	7	19.4	22	61.1	7	19.4
6	36	9	25.0	13	36.1	14	38.9
7	36	8	22.2	19	52.8	9	25.0
8	36	9	25.0	16	44.4	11	30.6
9	36	15	41.7	10	27.8	11	30.6
10	36	10	27.8	12	33.3	14	38.9
Total	360	102	28.3	159	44.2	99	27.5
Renewals[b]							
1	18	7	38.9	5	27.8	6	33.3
2	18	4	22.2	11	61.1	3	16.7
3	18	3	16.7	8	44.4	7	38.9
4	18	7	38.9	6	33.3	5	27.8
5	18	3	16.7	10	55.6	5	27.8
6	18	2	11.1	8	44.4	8	44.4
7	18	5	27.8	8	44.4	5	27.8
8	18	5	27.8	8	44.4	5	27.8
9	18	6	33.3	5	27.8	7	38.9
10	18	5	27.8	7	38.9	6	33.3
Total	180	47	26.1	76	42.2	57	31.7

Note. All χ^2 tests were two-tailed. [a]The percentages of items for which choices "a," "b," and "c" were the correct answers differed significantly ($\chi^2[2, N = 360] = 19.05$, $p < .001$), and the pattern of these percentages did not significantly vary across the 10 forms ($\chi^2[18, N = 360] = 20.95$, $p = .28$). [b]The percentages of items for which choices "a," "b," and "c" were the correct answer differed significantly ($\chi^2[2, N = 180] = 7.23$, $p = .03$), and the pattern of these percentages did not significantly vary across the 10 forms ($\chi^2[18, N = 180] = 13.25$, $p = .78$).

For original applicants taking the English DL 5, the percentages of items for which choices "a," "b," and "c" are the correct answers differ significantly ($p < .001$), and the pattern of the percentages does not vary significantly across the 10 test forms ($p = .28$). Answer choice "b" is the correct answer much more often (44.2%) than are choices "a" and "c" (28.3% and 27.5%, respectively). For three of the test forms, choice "b" is the correct answer for more than 50% of the items. Although there is some variation in the percentages across the test forms, the differences are not significant ($p > .05$)

For renewal applicants taking the English DL 5, the percentages of items for which choices "a," "b," and "c" are the correct answers also differ significantly ($p = .03$), and the pattern of the percentages does not vary significantly across the test forms ($p = .78$). Answer choice "b" is again the correct answer much more often than are choices "a" and "c." For two of the test forms, choice "b" is the correct answer for more than 50% of the items. Although the percentages do vary between the test forms, the differences once again are not significant.

Item Statistics

All item statistics are based on first-attempt applicants to eliminate potential learning effects of previous exposure to the test items. The results of the analyses of test items are presented in Appendices A, B, and C. Appendix A contains the results for the English DL 5 for original applicants, Appendix B contains the results for the English DL 5 for renewal applicants, and Appendix C contains the results for the English DL 5T for provisional applicants. Item statistics are not presented for the Spanish DL 5 applicant groups because too few tests were received to accurately compute the statistics. Each of the appendices contains four tables that are always in the same order. The tables contain the following information, respectively: item pass rates and answer choice selection rates; item-total correlations; percentage of applicants who would pass at different cut-points; and summary of problem items on each test form. The results reflected in the first three tables in each appendix are discussed in the next three subsections of this report.

Item Pass Rate and Answer Choice Selection Rates

The percentage of applicants selecting each item answer choice on each test form appears in the first table of Appendices A, B, and C (Tables A1, B1, and C1). The percentages for the correct answer choices are underlined in the tables. The percentages for each item are based on people who selected a valid item choice, and excludes people who did not choose an answer or chose more than one answer per item. Less than 3% of applicants are excluded per test item for these reasons.

The item pass rate refers to the percentage of applicants who correctly answered the item. A very low item pass rate may indicate that the item is poorly worded, has ambiguous or misleading answer choices, is not related to the general knowledge domain that is being tested, or is problematic for some other reason. Items that 60% or fewer of the applicants answered correctly are generally considered suspect, and should therefore be reviewed for clarity and accuracy. Items with extremely high pass rates (95% or higher) are also questionable and should be reviewed as well, because they may not discriminate between people with different levels of knowledge. These "freebie" items often occur when the distractor choices are so illogical that the correct answer is obvious, or when the knowledge required for a question has become common knowledge. It is also desirable that the alternative choices be tenable enough to actually attract the responses of a small but nontrivial percentage of the applicants (generally those who try to guess the answer). Therefore, items with individual distractors that were selected by a very small percentage of applicants, say 2% or less, may also need to be revised. Items with pass rates rounded to the nearest integer that are either too high (more than 95%) or too low (less than 60%) are shaded in the tables. Distractors that were selected by 2% or fewer applicants, or which were selected too often (i.e., selected more often than the correct answer or within 10% of the selection rate for the correct answer), are in boldface in the tables.

The 60% and 95% pass rate criteria for items, and the greater than 2% selection rate criterion for distractors, are provided only as guideline indicators of potential item deficiency. Standard item difficulty levels for personnel selection tests generally range from 40% to 70%, while makers of academic tests often attempt to obtain items which 50% of applicants pass and 50% fail. In personnel and academic testing, the purpose is to use the test as a screening device for predicting future achievement, whereas the written driver license knowledge tests are used to encourage applicants to master the information contained in the driver handbook. Hence, the 60% to 95% item difficulty

standards used in this evaluation are more relaxed than those used for these other purposes. Although these statistical standards are useful for pinpointing items that may be inadequate, it is not recommended that items be revised or replaced on the basis of these criteria alone. The relative importance of knowledge covered by an item, possible wording problems, and other relevant factors should always be weighed when deciding to revise or replace an item. For example, almost all applicants may have understanding and command of certain laws and principles. If this content is critical to safe driving, the item should not be discarded simply because 98% of the population correctly answers the item.

Item-Total Correlation

The item-total correlations for the test items are presented in Tables A2, B2, and C2. Items that tended to be answered correctly by applicants who scored low on the test overall (i.e., items with negative item-total correlations), or that had very little or no relationship to the other items on the test (i.e., those with item-total correlations between -.10 and .10), are highly undesirable and should be modified or replaced. All items with either of these problems are shaded in the tables. A weak item-total correlation indicates that the knowledge measured by the item is very disparate from the knowledge measured by the other items on the test. Items with this characteristic are usually poorly worded and confusing, and definitely need to be revised or replaced because they tend to lower test reliability. Although items with high positive item-total correlations are the most desirable, any with a correlation equal to or above .10 has at least some relationship with the other items on the test, and therefore may be acceptable.

Items that have pass rates that are too high or too low that also have poor item-total correlations require special consideration. These items are the most likely to be deficient and should definitely be reviewed and rewritten because of the negative effect they tend to have on test reliability. Items with high or low pass rates that also have poor item-total correlations are specifically identified in the problem item summary table for each test type.

Percentage of Applicants Who Would Pass at Different Cut-Points

The percentage of applicants who would pass at different cut-points on the tests is presented in Tables A3, B3, and C3. The tables present the percentage of applicants who missed the number of items indicated in the leftmost column of each table row or

fewer, and would therefore pass if that number was used as the test cut-point. For instance, Table A3 indicates that 45.0% of original applicants who took Form 1 of the English DL 5 (Rev. 10/98) missed six or fewer items and would have passed if the number of allowable errors had been six, while 51.7% of original applicants taking the same test would have passed if seven errors had been allowed. The shaded row in each table shows the pass rate for each form of each test at the current cut-off score of allowable misses (five for DL 5 original applicants, three for DL 5 renewal applicants, and seven for DL 5T provisional applicants).

The Impact of Not Returning Graded Tests to Applicants

Recall that during the day that data were collected for this evaluation, the field office personnel were instructed to indicate on each test form whether or not the customer had a question, concern, or complaint regarding not having his or her graded test returned, and if so, to record the amount of increased processing time associated with handling the question, concern, or complaint. The purpose of doing so was to evaluate a worst-case customer relations testing scenario of using tests with a perforated answer strip by evaluating the impact of not returning anything at all to the customers. Any test sheets that did not have either "YES" or "NO" written at the bottom of them were excluded from this portion of the evaluation, because it was uncertain whether or not the applicants had a question, concern, or complaint. About 22% of the test forms were excluded for this reason. Any bias introduced by the exclusion of these forms would likely inflate the percentage of applicants indicating a problem, because it seems more likely that the field office personnel would have forgotten to record the status of contented customers than those who were resentful. In addition, the amount of increased processing time was not recorded for 18.6% of forms on which it was indicated that the applicant had a question, concern, or complaint. These forms are included in the percentage of customers with a question, concern, or complaint, but excluded from the mean increased processing time calculation. All other test forms are included in the figures, including those that were older revisions of the tests, or that were too few in number for computing the test and item statistics presented earlier.

The number and percentage of forms on which it was indicated that the applicant had a question, concern, or complaint about not having the test returned, or that were

excluded from the analysis, are presented in Table 6. The average amount of increased processing time associated with handling each question, concern, or complaint is also shown in the table.

Table 6

Number and Percentage of Test Forms Indicating a Question, Concern, or Complaint About Not Having Graded Test Returned and the Associated Average Increased Processing Time for English and Spanish Originals, Renewals, and Provisionals

Language Applicant group	Total received	Forms excluded		Corrected total received	Question, concern, or complaint		Mean increased processing time
		n	%		*n*	%	(min:sec)
English							
Originals	3,399	794	23.4	2,605	350	13.4	1:15
Renewals	4,306	899	20.9	3,407	467	13.7	0:59
Provisionals	910	234	25.7	676	92	13.6	1:06
Spanish							
Originals	1,078	231	21.4	847	121	14.3	0:59
Renewals	533	95	17.8	438	43	9.8	0:58
Provisionals	33	2	6.1	31	1	3.2	0:05
Total	10,259	2,255	22.0	8,004	1,074	13.4	1:05

Note. The volumes include all test attempts. Forms were excluded if neither "YES" nor "NO" was written on the test in response to the question of whether they had a question, concern, or complaint about not having their test returned. Increased processing time is based on the 81.4% of "YES" forms that indicated a time.

As can be seen in the table, the overwhelming majority of customers were apparently agreeable to not having their graded test returned. About 13 of every 100 customers had a question, concern, or complaint about not having their graded test returned, and the field office personnel spent an average of roughly one additional minute dealing with each of these customers. Across all the reporting field offices, a total of about 19 hours of additional time (1,074 customers total at an additional 1 minute and 5 seconds each) were spent on the data collection day dealing with customer questions, comments, and concerns. Because the time period of the study only consisted of one day, and therefore more represents what would be expected on the first day of implementing a "no-return" policy rather than what would happen after such a policy was in place for a longer period of time, the results probably overestimate the volume of complaints and the amount of increased processing time. These figures also probably

overestimate the actual negative impact of using tests with a perforated answer strip listing the driver handbook page numbers relating to the content area of each item. Any negative impact caused by customers expecting to have the test questions returned would diminish over time if tests with a perforated answer strip were used, because the customers' expectations would also change.

DISCUSSION

The fail rates of the tests, except those for English renewal applicants, are higher than those reported in the 1998 written test evaluation. Although the fail rate for first-attempt English renewal applicants is about nine percentage points lower than that reported in the 1998 written test evaluation, the fail rate for first-attempt English original applicants increased about three percentage points and that for first-attempt English provisional applicants increased about 18 percentage points. The fail rates for Spanish applicants also increased, by about 16 percentage points for original applicants and 10 percentage points for renewal applicants.

It is unknown whether the increase in most of the test fail rates is due to tighter test security, a lack of knowledge on the part of applicants, or deficiencies in the tests themselves. However, many new items were generated to create new forms of the tests since the last evaluation, and the results of this study show that some of these items are very difficult. This underscores the need for subjecting new tests and item pools to a psychometric analysis of the type used here, and for using the results to improve the tests through an ongoing iterative analytic process. One of the contributing factors to the high fail rates is the large number of low pass rate (or difficult) items on each form of each test. (Recall that an item was considered likely to be too difficult if 60% or fewer applicants correctly answered the item). For instance, every form of the English DL 5 36-item test except one has four to eight highly difficult items. There are enough of these difficult items on five of the original test forms and four of the renewal test forms to cause applicants who miss only these difficult items to fail the test.

Most of these difficult items test knowledge that is related to traffic safety, defensive driving, or laws in the California Vehicle Code. However, some are numerical and have correct answers that can be learned only by rote memorization. Whether the low pass rate for individual items are the result of poor item characteristics and wording, a lack

of coverage of the knowledge area in the driver manual, or a lack of knowledge competency among applicants is also unknown, and will have to be ascertained during a review of the items. It is expected, however, that revising the low pass rate items to eradicate confusing wording, and replacing those that have little relevance to safety or which are conceptually difficult to grasp, would bring the test fail rates closer to historical levels.

Items with high or low pass rates which also have poor item-total correlations are the most likely to be deficient. These items not only contribute to the fail rates, they also lower the internal-consistency reliability of the tests. Therefore, items with high or low pass rates that also have poor item-total correlations should definitely be reviewed and revised.

The higher fail rate for original applicants is, to some extent, an artifact of the more stringent cut-point passing criterion that is applied to them. The current standard for original applicants allows them to miss proportionally fewer items than is permitted for renewal applicants. Increasing the number of allowable errors for original applicants by one would not only equate the expected competency levels for original and renewal applicants, it would also lower their fail rate by 10 percentage points or more. Similarly, provisional applicants are also allowed proportionally fewer errors than are renewal applicants. Raising the number of allowable misses for provisional applicants by one would also equate their expected competency levels, and would lower their failure rate by 16 percentage points or more. However, these changes would have the undesirable effect of lowering the level of competency expected of original and provisional applicants. Raising the cut-points for originals and provisionals should not be the only strategy for reducing the high fail rates. It is also necessary that problem items be reviewed and revised to eliminate confusing wording and item choices in order to maintain the integrity of the tests.

Another problem with the tests, in addition to the high fail rates, is that some forms of the same test are much more difficult than others. The fail rates for some forms of the same test differ by as much as 41 percentage points. This result is largely due to the fact that some forms had more low pass rate items and fewer high pass rate items ("freebies") than did other forms. The correction of problem items should help equalize the fail rates between the forms of each test.

The test fail rates remain surprisingly high on the second, third, and fourth or higher attempts. The fail rates for English renewal applicants tend to increase slightly over successive attempts, while those for the other applicant and test-type groups tend to decrease slightly. Even those rates that did decrease over successive test attempts still remained high, which indicates that many applicants either do little or no preparation or have severe literacy problems.

Original and renewal applicants taking the Spanish DL 5 tended to make more errors and fail more often than did applicants taking the English DL 5. This may be due in part to imprecise translation of some test items, or to differences between the current version of the English test and the revision of the English test from which the Spanish test was translated. Differences between the language groups in the availability of driver license manuals and instruction opportunities could also have contributed to the performance difference.

Almost all of the internal-consistency test reliabilities for the English and Spanish DL 5 for original applicants and the English DL 5T for provisional applicants are adequate, and some are even "good" or "excellent." Any of the test forms with reliabilities below .70 should be reviewed and revised to improve the reliability. These reliabilities can be increased and made more homogenous across different forms of the same tests by correcting or replacing problem items, particularly those with low or negative item-total correlations.

None of the English and Spanish DL 5 test form reliabilities for renewal applicants exceeded the .70 whole-test reliability standard. Those for Form 6 of the English DL 5 for renewal applicants and Form 2 of the Spanish DL 5 for renewal applicants are truly inadequate and require serious review and revision. Note also that Forms 6 and 7 of the English DL 5 for renewals have the lowest internal-consistency reliabilities and also the highest number of items with zero or negative item-total correlations. These two forms are in particularly bad shape and require special attention to correct the items with low item-total correlations which diminish their reliability. Overall, the renewal test reliability findings indicate that 18 may be an insufficient number of items to produce renewal driver license knowledge tests with adequate reliability. In light of the high fail rates for the tests, it is increasingly important that the tests have adequate reliability. The low reliability of the renewal tests indicates that applicants taking the tests on two occasions or different forms of the test are likely to get highly disparate scores affecting

licensure status (pass vs. fail). The department should therefore consider increasing the length of the renewal tests to increase their reliability, possibly by having renewal applicants complete all 36 items on the DL 5 instead of only the first 18.

As would be expected given the large number of new items on each test form and the fact that they have never before been evaluated, many of the items on each test form have an item-total correlation that is too low, a pass rate that is too high or too low, or a distractor that was selected either too often or too infrequently. Tables A4, B4, and C4 identify these items on the test forms. These statistical characteristics indicate that there may be a problem with the items, and they should therefore be reviewed and modified as necessary. In particular, items with weak item-total correlations show the strongest evidence of item deficiency and warrant immediate attention. Items having two or more problem characteristics, such as those with poor item-total correlations and pass rates that are too high or too low, are the most likely to be deficient, and special care should be taken to review and correct them.

Answer choice "b" is the correct answer much more often than are choices "a" or "c" for both the original and renewal English DL 5 tests. On some of the test forms, choice "b" is the correct answer for more than 50% of the items. The assignment of the answer choices should be truly randomized and balanced within each test form to decrease the applicants' chances of guessing the correct answers. Rewriting item distractors that none, or almost none, of the applicants chose is also important for decreasing the applicants' chances of guessing.

While screening the test forms it became apparent that some field office personnel were administering the back side of the DL 5 to renewal applicants. This deviance from department policy results in applicants not being exposed to the mandatory items pertaining to blood alcohol level and time to report the sale of a vehicle, which appear on the front of each test form. Because the small amount of tests with this problem were removed from the pool of usable test forms during the screening process, they were unable to bias the results of the evaluation.

Some field offices also deviated from department procedure by awarding points to original and provisional license applicants who were borderline test failures, and/or by using older revisions of the Class C license test. The use of older revisions of the tests

greatly diminishes the effectiveness of randomizing the English DL 5 every 3 months to curtail applicant cheating.

To simulate the worst-case negative impact of using tests with a perforated answer strip listing the driver license handbook page numbers for the content area of each item instead of returning the test questions themselves, the current study evaluated the effects of not returning graded tests to customers at all. The overwhelming majority of customers in this evaluation were agreeable to not having their graded test returned at all. About 13 of every 100 customers had a question, concern, or complaint about not having their graded test returned, and the field office personnel spent an average of roughly one additional minute dealing with each of these applicants. The number of complaints would be expected to be lower than these figures if the perforated answer strip was used because the applicants would still receive feedback on the content area of missed items. Any negative impact would also decrease over time if the department were to use tests with a perforated answer strip, because the customers' expectations would also change. Because the amount of increased processing time is so negligible, the department should reconsider the use of written tests with a perforated answer strip. Not only would it increase test security by limiting outside circulation of the test questions, it would also increase the likelihood that applicants study for the tests using the driver handbook. However, it might also result in a temporary increase in the fail rates, because those applicants who would have passed the test by cheating would be less likely to be able to obtain the test answers when the questions are not returned.

RECOMMENDATIONS

- The tests should be reviewed and revised to make them more sound measurement instruments. In particular, the difficulty level and reliability should be made more homogenous across different forms of the same test, and items with deficient characteristics should be reviewed and modified as necessary. The following five recommendations should help accomplish these goals:

    1. Items with item-total correlations that are below .10 or negative need to be modified or replaced, particularly those with other item deficiencies such as pass rates that are too high or too low. Items with this characteristic are usually poorly worded and confusing, and definitely need to be revised or replaced. Doing so is likely to increase the internal-consistency reliability of the tests

overall, and make them more homogenous across different forms of the same test, particularly in those cases where the reliability was inadequate or borderline.

2. Items with pass rates that are too high or too low, or with distractor selection rates that are too high or too low, should be reviewed for possible problems and modified as necessary. The low pass rate items should be revised to eradicate confusing wording, and those that have little relevance to safety and are conceptually difficult to grasp should be replaced. This would help bring the test fail rates closer to historical levels.

3. Items should be checked to ensure that the knowledge covered by them is contained in the driver handbook and has relevancy to safety, mobility, or other knowledge domains deemed important to driving or safety.

4. The answer choices on each test should be truly randomized and balanced on each test form to decrease the applicants' chances of guessing the correct answers. There are computer applications available that can efficiently and cheaply accomplish this goal. Rewriting item distractors in which none, or almost none, of the applicants chose should also be completed to increase test reliability.

5. A task force composed of knowledge experts from various areas of the department should be convened to accomplish the test revisions.

- The department might want to consider allowing one additional error for original and provisional applicants. This would equalize the expected competency levels of all applicant types as well as lower the test fail rates for original and provisional applicants.

- In light of the high fail rates for the tests, it is increasingly important that the tests have adequate reliability. The low reliability of the renewal tests indicates that applicants taking the test on two occasions or different forms of the test are likely to get disparate scores affecting licensure status (pass vs. fail). The department should therefore consider increasing the length of the renewal tests to increase their reliability, possibly by having renewal applicants complete all 36 items on the DL 5 instead of only the first 18. It is estimated that this doubling of the test length would increase the overall English DL 5 renewal test internal-consistency reliability from .51 to .68, and the overall Spanish DL 5 renewal test reliability from .45 to .62. These

reliabilities would be expected to increase even more if items with poor item-total correlations are also reviewed and revised.

- Considering the small percentage of customers raising a question, concern, or complaint about not having their graded test returned at all, and the negligible increase in field office processing time associated with handling these customers, management should reconsider implementing the use of written tests with a perforated answer strip. This would enable applicants to still receive some feedback about the content area where their knowledge was deficient, increase test security, and make it more likely that customers would study the driver handbook.

- Steps should be taken to ensure that field office personnel are administering only current versions of the tests in accordance with the department procedures stated in the Driver License Manual. Doing so would ensure the department's compliance with California statute and increase the effectiveness of randomizing the English DL 5 test every 3 months to reduce the possibility of cheating.

- After the tests are revised, they should be evaluated to determine whether the changes have produced the desired results.

- The department should expedite consideration of technology for computer generating a unique test for each applicant from a large item pool data base. Vendor interest proposals have been solicited for accomplishing this goal and AAMVA has already developed a computerized commercial driver license item pool.

- The following strategies should be considered for improving test scores:

    1. The department should prepare and distribute information brochures and press releases that publicize the content areas and principles that applicants most frequently have problems with.

    2. The department should consider initiating legislation to charge applicants for repeat testing. This could be done by charging for each written test attempt or allowing three attempts as under current law and then imposing a much larger fee for a second application after three written test failures.

Appendix A

Item Statistics for the English DL 5 (Rev. 10/98) for First-Attempt Original Applicants

Table A1

Percentage of Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5 (Rev. 10/98) Test for First-Attempt Original Applicants

Item	Answer choice	Form 1 (*n* = 149)	Form 2 (*n* = 145)	Form 3 (*n* = 157)	Form 4 (*n* = 148)	Form 5 (*n* = 156)	Form 6 (*n* = 167)	Form 7 (*n* = 166)	Form 8 (*n* = 139)	Form 9 (*n* = 163)	Form 10 (*n* = 140)
1	a	14.3	14.7	**31.0**	90.4	**0.6**	10.8	2.4	2.2	33.8	2.9
	b	77.6	74.8	**31.6**	**0.7**	94.2	83.7	91.5	12.3	58.8	40.1
	c	8.2	10.5	37.4	8.9	5.2	5.4	6.1	85.5	7.5	56.9
2	a	**0.0**	63.2	5.1	3.4	5.8	13.9	52.7	23.4	14.8	36.4
	b	90.6	4.9	5.8	84.4	20.5	13.9	26.1	71.5	79.6	53.6
	c	9.4	31.9	89.1	12.2	73.7	72.3	21.2	5.1	5.6	10.0
3	a	77.2	3.4	9.6	12.2	**23.2**	96.4	30.1	5.0	**0.0**	7.9
	b	11.7	86.9	**0.6**	83.0	32.3	3.0	66.3	95.0	**0.0**	59.0
	c	11.0	9.7	89.8	4.8	**44.5**	**0.6**	3.6	**0.0**	100.0	33.1
4	a	85.7	**1.4**	24.5	87.8	3.2	13.9	14.6	15.3	**1.2**	**0.7**
	b	**0.7**	86.8	5.8	4.1	94.2	4.8	12.8	76.6	9.3	2.9
	c	13.6	11.8	69.7	8.1	2.6	81.2	72.6	8.0	89.5	96.4
5	a	27.6	7.6	5.7	**0.7**	**0.6**	2.4	7.3	94.2	23.0	15.0
	b	13.8	81.3	93.6	**2.0**	10.3	88.0	79.4	2.2	16.8	67.1
	c	58.6	11.1	**0.6**	97.3	89.0	9.6	13.3	3.6	60.2	17.9
6	a	21.8	10.3	2.5	11.5	16.7	3.0	67.1	**0.0**	**0.0**	58.0
	b	74.1	87.6	86.0	76.4	67.9	94.6	3.7	95.7	2.5	8.7
	c	4.1	2.1	11.5	12.2	15.4	2.4	29.3	4.3	97.5	33.3
7	a	2.7	95.9	11.0	19.2	28.1	29.5	84.2	15.1	12.3	88.6
	b	35.1	2.1	81.9	**0.7**	3.3	9.0	10.3	71.9	76.1	7.9
	c	62.2	2.1	7.1	80.1	**68.6**	61.4	5.5	12.9	11.7	3.6
8	a	**2.0**	2.1	91.7	69.9	22.7	71.5	95.8	**0.0**	70.6	87.8
	b	95.3	97.2	8.3	11.0	67.5	13.3	2.4	2.9	15.3	5.0
	c	2.7	**0.7**	**0.0**	19.2	9.7	15.2	**1.8**	97.1	14.1	7.2
9	a	6.1	**0.7**	11.5	27.0	**1.3**	2.4	63.3	**0.0**	7.4	4.3
	b	**0.0**	92.4	87.2	51.4	94.2	11.4	31.9	4.4	14.8	7.9
	c	93.9	6.9	**1.3**	21.6	4.5	86.2	4.8	95.6	77.8	87.9
10	a	93.3	9.7	9.7	43.2	4.5	10.2	8.5	10.1	93.8	14.3
	b	6.7	6.2	62.6	**0.7**	87.8	84.4	**1.8**	79.7	6.2	84.3
	c	**0.0**	84.1	27.7	56.1	7.7	5.4	89.7	10.1	**0.0**	**1.4**
11	a	89.9	8.3	3.2	15.6	35.1	8.4	29.0	16.9	93.2	75.0
	b	2.7	89.7	21.8	**49.7**	3.9	91.0	57.4	14.0	3.7	20.7
	c	7.4	2.1	75.0	34.7	61.0	**0.6**	13.6	69.1	3.1	4.3
12	a	2.7	76.9	12.2	90.5	65.1	**1.2**	**0.0**	64.5	71.4	24.5
	b	6.7	7.0	82.7	**1.4**	15.1	**0.0**	93.4	10.1	14.9	72.7
	c	90.6	16.1	5.1	8.1	19.7	98.8	6.6	25.4	13.7	2.9
13	a	5.4	11.8	89.2	14.3	29.4	9.0	10.2	35.5	89.0	30.0
	b	10.1	65.3	10.2	78.9	7.8	4.2	9.0	**35.5**	9.2	10.7
	c	84.5	22.9	**0.6**	6.8	62.7	86.7	80.7	**29.0**	**1.8**	59.3
14	a	96.6	5.5	5.8	63.9	2.6	14.5	20.0	**0.7**	3.1	10.7
	b	**1.3**	8.3	91.6	25.9	93.5	72.3	43.0	99.3	7.4	89.3
	c	**2.0**	86.2	2.6	10.2	3.9	13.3	**37.0**	**0.0**	89.5	**0.0**
15	a	91.9	11.2	15.9	93.9	64.7	10.2	2.4	31.7	17.8	82.9
	b	3.4	17.5	8.3	**0.7**	26.8	66.3	3.6	**0.7**	68.7	12.9
	c	4.7	71.3	75.8	5.4	8.5	23.5	94.0	67.6	13.5	4.3
16	a	6.8	13.2	18.6	2.7	7.7	3.6	**1.2**	95.7	4.3	**1.4**
	b	**1.4**	76.4	79.5	10.9	85.2	74.3	98.8	**0.7**	83.3	13.6
	c	91.9	10.4	**1.9**	86.4	7.1	22.2	**0.0**	3.6	12.3	85.0
17	a	**1.3**	81.8	3.8	11.0	2.6	7.2	34.5	71.2	6.1	27.3
	b	94.0	5.6	**1.9**	66.4	2.6	5.4	13.9	19.4	3.1	10.8
	c	4.7	12.6	94.2	22.6	94.8	87.4	51.5	9.4	90.8	61.9
18	a	81.6	4.9	52.3	70.9	11.6	**1.2**	30.7	6.5	73.0	17.3
	b	7.5	94.4	22.2	17.6	78.7	98.2	65.7	84.2	3.7	53.2
	c	10.9	**0.7**	25.5	11.5	9.7	**0.6**	3.6	9.4	23.3	29.5

Table A1 (continued)

Item	Answer choice	Form 1 (*n* = 149)	Form 2 (*n* = 145)	Form 3 (*n* = 157)	Form 4 (*n* = 148)	Form 5 (*n* = 156)	Form 6 (*n* = 167)	Form 7 (*n* = 166)	Form 8 (*n* = 139)	Form 9 (*n* = 163)	Form 10 (*n* = 140)
19	a	70.1	20.1	2.5	15.9	12.3	66.5	17.6	13.8	96.3	**0.7**
	b	19.0	7.6	96.2	69.0	80.0	16.2	77.6	59.4	3.1	97.1
	c	10.9	72.2	**1.3**	15.2	7.7	17.4	4.8	26.8	**0.6**	2.1
20	a	98.7	6.3	6.4	84.9	72.3	75.3	10.2	96.4	87.1	91.4
	b	**0.0**	9.0	89.8	3.4	21.3	10.8	86.7	**1.4**	**1.2**	5.8
	c	**1.3**	84.7	3.8	11.6	6.5	13.9	3.0	2.2	11.7	2.9
21	a	87.8	8.3	33.1	7.5	6.5	24.1	**1.2**	5.0	5.0	96.4
	b	2.7	91.0	19.7	7.5	61.4	70.5	14.0	10.1	15.0	**1.4**
	c	9.5	**0.7**	47.1	84.9	32.0	5.4	84.8	84.9	80.0	2.1
22	a	**49.7**	**0.7**	8.4	8.8	6.6	**1.2**	21.8	12.3	9.9	**47.9**
	b	44.9	98.6	86.5	91.2	**2.0**	**0.6**	65.5	17.4	87.7	41.4
	c	5.4	**0.7**	5.2	**0.0**	91.4	98.2	12.7	70.3	2.5	10.7
23	a	6.8	**29.4**	8.3	8.9	**0.6**	94.0	**1.2**	18.4	95.1	2.9
	b	6.1	33.6	91.7	87.0	99.4	4.2	92.2	30.1	**1.2**	11.5
	c	87.2	**37.1**	**0.0**	4.1	**0.0**	**1.8**	6.6	51.5	3.7	85.6
24	a	30.9	9.1	57.7	36.5	18.8	5.4	**0.0**	86.3	47.2	10.0
	b	53.0	76.9	14.1	7.4	64.3	4.8	**0.0**	2.2	**44.2**	**0.0**
	c	16.1	14.0	28.2	56.1	16.9	89.8	100.0	11.5	8.6	90.0
25	a	26.7	86.7	16.0	98.6	3.2	96.4	6.7	16.5	96.9	5.8
	b	67.1	10.5	10.9	**1.4**	90.9	2.4	92.7	4.3	**1.8**	81.3
	c	6.2	2.8	73.1	**0.0**	5.8	**1.2**	**0.6**	79.1	**1.2**	12.9
26	a	72.5	83.1	96.2	92.6	33.8	45.2	**0.0**	95.6	90.7	7.1
	b	21.5	6.3	3.2	3.4	57.8	**50.6**	100.0	2.9	7.5	91.4
	c	6.0	10.6	**0.6**	4.1	8.4	4.2	**0.0**	**1.5**	**1.9**	**1.4**
27	a	10.3	**1.4**	94.9	87.2	34.4	**0.6**	**1.2**	15.1	3.1	7.9
	b	51.7	97.2	**0.6**	8.8	59.1	6.0	91.0	2.9	91.4	12.1
	c	37.9	**1.4**	4.5	4.1	6.5	93.4	7.8	82.0	5.5	80.0
28	a	8.1	**64.8**	9.0	**1.4**	66.9	11.4	25.0	7.2	27.2	**0.7**
	b	82.6	**1.4**	59.6	2.7	16.2	84.3	7.3	87.1	4.9	99.3
	c	9.4	33.8	31.4	95.9	16.9	4.2	67.7	5.8	67.9	**0.0**
29	a	94.0	2.1	3.2	9.5	**1.9**	78.7	7.2	**0.0**	9.9	25.2
	b	4.7	84.8	62.2	89.2	96.2	3.7	89.8	**0.7**	58.0	12.9
	c	**1.3**	13.1	34.6	**1.4**	**1.9**	17.7	3.0	99.3	32.1	61.9
30	a	30.9	4.2	**1.3**	7.5	12.4	3.0	12.1	27.3	**0.6**	82.7
	b	63.1	86.7	98.7	67.3	68.0	2.4	60.6	64.7	96.3	10.1
	c	6.0	9.1	**0.0**	25.2	19.6	94.6	27.3	7.9	3.1	7.2
31	a	20.1	**1.4**	14.2	36.6	65.4	92.2	**1.2**	10.8	6.2	**41.4**
	b	76.5	97.2	75.5	55.9	10.5	6.6	**0.6**	48.2	92.0	**26.4**
	c	3.4	**1.4**	10.3	7.6	24.2	**1.2**	98.2	**41.0**	**1.9**	32.1
32	a	27.9	81.4	94.9	26.7	**0.6**	18.0	3.0	**0.0**	92.6	73.4
	b	4.1	13.1	**0.6**	44.5	98.7	76.0	**0.6**	99.3	**1.8**	5.0
	c	**68.0**	5.5	4.5	28.8	**0.6**	6.0	96.4	**0.7**	5.5	21.6
33	a	80.5	**1.4**	94.9	17.1	66.0	7.2	67.7	97.1	10.4	**0.7**
	b	**0.0**	93.1	**1.9**	67.8	27.5	56.6	21.3	**0.0**	11.7	**0.0**
	c	19.5	5.5	3.2	15.1	6.5	36.1	11.0	2.9	77.9	99.3
34	a	4.7	11.9	6.4	11.7	14.9	9.0	95.2	25.2	95.1	98.6
	b	77.2	30.1	6.4	80.0	19.5	6.6	**0.6**	69.1	3.1	**0.0**
	c	18.1	58.0	87.2	8.3	65.6	84.4	4.2	5.8	**1.8**	**1.4**
35	a	**1.3**	73.8	66.5	**59.5**	12.2	6.6	2.4	23.7	68.1	**1.4**
	b	96.6	11.0	**1.9**	**0.0**	53.8	83.1	95.8	63.3	23.9	5.0
	c	**2.0**	15.2	31.6	40.5	34.0	10.2	**1.8**	12.9	8.0	93.6
36	a	8.7	7.7	68.8	94.6	16.7	2.4	80.0	**0.0**	**1.2**	13.7
	b	79.2	58.0	13.4	**0.0**	39.7	**1.2**	12.7	91.4	12.3	23.0
	c	12.1	34.3	17.8	5.4	**43.6**	96.4	7.3	8.6	86.4	63.3

Note. Underlining of a percentage indicates that the answer choice was the correct response according to the official answer key. Shading indicates that the item needs to be reviewed and possibly revised due to the item pass rate being too low or too high. A boldface percentage indicates that the distractor selection rate is too low or too high.

Table A2

Item-Total Correlation for Each Item on Each Form of the English DL 5 (Rev. 10/98) Test for First-Attempt Original Applicants

Item	Form 1 (*n* = 149)	Form 2 (*n* = 145)	Form 3 (*n* = 157)	Form 4 (*n* = 148)	Form 5 (*n* = 156)	Form 6 (*n* = 167)	Form 7 (*n* = 166)	Form 8 (*n* = 139)	Form 9 (*n* = 163)	Form 10 (*n* = 140)
1	.37	.08	.31	.23	.26	.11	.31	.24	.28	.44
2	.24	.14	.19	.20	.37	.20	.27	.28	.02	.23
3	.32	.05	.19	.49	.25	.06	.37	-.03	*	.16
4	.37	.36	.38	.05	.23	.26	.26	.40	.07	.11
5	.16	.26	-.01	.27	.21	.19	.11	.36	.37	.36
6	.24	.29	.28	.12	.29	.42	.20	.25	.13	.11
7	.29	.29	.28	.38	.22	.28	.12	.28	.41	.37
8	.21	.20	.06	.31	.31	.11	.18	.13	.19	.31
9	.13	.34	.39	.31	.43	.36	.16	.08	.31	.35
10	.04	.29	.33	.08	.20	.17	.12	.43	.20	.04
11	.32	.30	.31	.41	.23	.29	.39	.29	.07	.24
12	.29	.17	.18	.41	.23	.26	.02	.32	.13	.41
13	.40	.39	.02	.03	.28	.37	.31	.13	.35	.22
14	.24	.20	.26	.21	.37	.35	.06	.19	.41	.01
15	.19	.36	.32	.08	.24	.23	.16	.20	.29	.21
16	.22	.22	.27	.39	.19	.10	.14	.27	.28	.28
17	.23	.37	.34	.42	.26	.10	.21	.21	.20	.17
18	.28	.14	.11	.16	.23	.13	.21	.19	.17	.32
19	.17	.43	.42	.25	.21	.32	.30	.24	.20	-.05
20	.11	.15	.21	.22	.21	.18	.29	.36	.21	.10
21	.42	.29	.19	.30	.36	.19	.28	.11	.19	.18
22	.36	.13	.09	.19	.11	.26	.05	.14	.28	.30
23	.33	.31	.28	.26	.04	.27	.21	.17	.26	.03
24	.30	.32	.34	.23	.05	.25	*	.29	.34	.14
25	.26	.26	.21	.07	.38	.32	.24	.03	.37	.31
26	.35	.26	.20	.37	.09	.19	*	.39	.44	.48
27	-.03	.28	.30	.33	.33	.25	.24	.25	.27	.34
28	.21	.14	.15	.40	.33	.17	.17	.50	.13	.36
29	.36	.48	.23	.37	.48	.40	.27	.34	.02	.29
30	.30	.35	.45	.29	.32	.08	.24	.25	.20	.24
31	.11	.16	.21	.24	.30	.15	.25	.30	.07	.10
32	.37	.04	.36	.29	.13	.14	.20	.34	.39	.06
33	.09	.21	.42	.27	.09	.15	.04	.47	.12	.32
34	.52	.26	.36	.40	.10	.16	.20	.18	.35	.03
35	.17	.37	.05	.28	.24	.39	.14	.23	.39	.11
36	.13	.19	.22	.21	.39	.27	.25	.36	.50	.40

Note. Shading indicates that an item needs to be reviewed and revised or replaced because the item-total correlation is negative or less than .10. *Item-total correlation could not be calculated because the item had no variance.

Table A3

Percentage of First-Attempt Original Applicants Who Would Pass if Different Cut-Points Were Used for Each Form of the English DL 5 (Rev. 10/98)

Number Missed	Form 1 (*n* = 149)	Form 2 (*n* = 145)	Form 3 (*n* = 157)	Form 4 (*n* = 148)	Form 5 (*n* = 156)	Form 6 (*n* = 167)	Form 7 (*n* = 166)	Form 8 (*n* = 139)	Form 9 (*n* = 163)	Form 10 (*n* = 140)	Total (*N* = 1,530)
0	1.3	0.0	3.2	1.4	1.9	1.2	1.2	2.9	1.2	1.4	1.6
1	6.7	7.6	4.5	3.4	3.8	4.2	4.8	4.3	6.7	4.3	5.0
2	10.1	11.7	8.9	5.4	5.1	7.8	6.6	7.2	17.2	7.1	8.8
3	16.8	19.3	14.6	11.5	5.8	16.2	11.4	15.8	25.2	10.7	14.8
4	26.2	29.0	21.0	19.6	8.3	30.5	19.9	23.7	36.8	15.7	23.2
5	37.6	37.2	35.7	29.1	15.4	40.7	27.7	33.8	46.6	25.0	33.0
6	45.0	45.5	46.5	35.8	21.2	53.9	36.1	46.0	58.3	32.1	42.2
7	51.7	55.9	56.7	43.9	31.4	62.3	42.8	56.8	66.3	45.7	51.4
8	60.4	65.5	66.9	54.1	39.1	71.9	58.4	68.3	76.1	55.7	61.8
9	64.4	75.2	72.0	60.1	51.9	79.0	66.9	73.4	79.8	66.4	69.0
10	71.8	79.3	80.3	68.9	59.6	82.6	78.9	83.5	82.8	72.9	76.1
11	77.9	83.4	85.4	76.4	66.0	86.8	82.5	86.3	87.7	81.4	81.4
12	83.2	86.9	89.2	79.1	73.7	90.4	88.6	89.9	92.0	85.7	85.9
13	89.3	91.7	93.0	84.5	78.8	92.8	91.6	93.5	95.1	90.7	90.1
14	93.3	94.5	94.9	89.2	84.0	95.8	95.2	93.5	96.3	94.3	93.1
15	94.6	96.6	96.8	91.9	87.8	97.6	97.0	96.4	96.9	95.7	95.2
16+	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. The shaded line represents the pass rate at the current passing standard, which allows up to five errors.

Table A4

Summary of Problem Items on Each Form of the English DL 5 (Rev. 10/98) Test for First-Attempt Original Applicants

Problem indicator	Form 1 items	Form 2 items	Form 3 items	Form 4 items	Form 5 items	Form 6 items	Form 7 items	Form 8 items	Form 9 items	Form 10 items
Item-total correlation too low or negative[a]	10, 27, 33	01, 03, 32	05, 08, 13, 22, 35	04, 10, 13, 15, 25	23, 24, 26, 33	03, 30	12, 14, 22, 24, 26, 33	03, 09, 25	02, 03, 04, 11, 29, 31	10, 14, 19, 23, 32, 34
Pass rate too high[b]	08, 14, 20	07, 08, 22, 27, 31	19, 26, 27, 30, 32, 33	05, 25, 28, 36	17, 23, 29, 32	03, 06, 12, 18, 22, 25, 30, 36	08, 16, 24, 26, 31, 32, 34, 35	03, 06, 08, 09, 14, 16, 20, 26, 29, 32, 33	03, 06, 19, 23, 25, 30, 34	04, 19, 21, 28, 33, 34
Pass rate too low[c]	05, 22, 24, 27, 32	23, 28, 34, 36	01, 18, 21, 24	09, 10, 11, 24, 31, 32, 35	03, 07, 26, 27, 35, 36	16, 26, 33	02, 11, 14, 17, 21	13, 19, 23, 31	01, 05, 24, 29	01, 02, 03, 06, 13, 18, 22, 31
Pass rate too high or low and item-total correlation too low or negative	27	none	none	10, 25	23, 26	03, 30	14, 24, 26	03, 09	03, 29	19, 34
Distractor selected too often[d]	22, 32	23, 28	01	11, 35	03, 07, 36	26	14	13, 31	24	22, 31
Distractor selected too infrequently[e]	02, 04, 08, 09, 10, 14, 16, 17, 20, 29, 33, 35	04, 08, 09, 18, 21, 22, 27, 28, 31, 33	03, 05, 08, 09, 13, 16, 17, 19, 23, 26, 27, 30, 32, 33, 35	01, 05, 07, 10, 12, 15, 22, 25, 28, 29, 35, 36	01, 05, 09, 22, 23, 29, 32	03, 11, 12, 18, 22, 23, 25, 27, 31, 36	08, 10, 12, 16, 21, 23, 24, 25, 26, 27, 31, 32, 34, 35	03, 06, 08, 09, 14, 15, 16, 20, 26, 29, 32, 33, 36	03, 04, 06, 10, 13, 19, 20, 23, 25, 26, 30, 31, 32, 34, 36	04, 10, 14, 16, 19, 21, 24, 26, 28, 33, 34, 35

[a]The item-total correlation was negative or less than .10. [b]More than 95% of applicants answered the item correctly. [c]Less than 60% of applicants answered the item correctly. [d]A distractor was chosen more, or almost, as often as the correct answer to the item. [e]A distractor was selected by 2% or fewer applicants.

Appendix B

Item Statistics for the English DL 5 (Rev. 10/98) for First-Attempt Renewal Applicants

Table B1

Percentage of Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5 (Rev. 10/98) Test for First-Attempt Renewal Applicants

Item	Answer choice	Form 1 (*n* = 259)	Form 2 (*n* = 263)	Form 3 (*n* = 283)	Form 4 (*n* = 258)	Form 5 (*n* = 247)	Form 6 (*n* = 242)	Form 7 (*n* = 248)	Form 8 (*n* = 271)	Form 9 (*n* = 258)	Form 10 (*n* = 261)
1	a	17.4	8.4	**27.3**	86.3	**0.0**	8.4	**1.6**	**1.1**	22.6	5.9
	b	76.7	74.8	**50.7**	**0.0**	97.2	83.9	93.1	10.4	74.3	38.3
	c	5.8	16.8	22.0	13.7	2.8	7.6	5.3	88.5	3.1	55.9
2	a	**1.9**	78.2	5.3	**2.0**	6.9	15.8	46.6	13.0	14.3	26.4
	b	93.0	4.2	2.8	87.9	17.9	10.4	27.3	83.6	83.7	66.7
	c	5.1	17.6	91.8	10.2	75.2	73.8	26.1	3.3	**1.9**	6.9
3	a	72.2	**1.9**	9.6	6.6	**27.2**	98.3	26.0	3.0	3.1	6.2
	b	8.1	87.5	**0.0**	90.3	31.7	**1.7**	67.5	97.0	**1.6**	53.5
	c	19.7	10.6	90.4	3.1	**41.1**	**0.0**	6.5	**0.0**	95.3	40.3
4	a	83.4	**0.0**	26.7	85.9	**1.6**	16.2	10.2	20.0	**0.0**	4.2
	b	**1.5**	95.8	5.1	4.7	98.0	7.1	9.3	74.1	3.5	2.3
	c	15.1	4.2	68.2	9.4	**0.4**	76.8	80.5	5.9	96.5	93.4
5	a	27.4	5.7	4.6	**0.4**	**0.0**	**0.0**	4.9	96.3	23.4	9.3
	b	20.8	80.5	95.4	**0.4**	3.6	86.7	75.5	**1.5**	15.6	81.1
	c	51.7	13.8	**0.0**	99.2	96.4	13.3	19.6	2.2	60.9	9.7
6	a	32.8	4.9	4.4	13.0	17.2	4.5	65.9	**1.5**	**0.8**	78.0
	b	64.5	90.5	80.1	77.2	74.6	93.8	**1.2**	97.4	5.4	**1.9**
	c	2.7	4.6	15.5	9.8	8.2	**1.7**	32.9	**1.1**	93.8	20.1
7	a	**0.4**	99.2	19.1	16.3	35.9	33.8	80.9	8.2	5.8	95.0
	b	29.1	**0.8**	75.1	**0.8**	5.3	5.8	13.8	82.1	83.7	3.5
	c	70.5	**0.0**	5.8	82.9	**58.8**	60.4	5.3	9.7	10.5	**1.5**
8	a	**0.4**	**0.8**	96.5	65.6	26.4	75.0	97.2	**1.1**	57.8	94.6
	b	92.3	99.2	3.5	9.0	67.9	11.7	**1.6**	**1.5**	32.5	**1.1**
	c	7.3	**0.0**	**0.0**	25.0	5.7	13.3	**1.2**	97.4	9.6	4.2
9	a	3.1	**1.2**	13.2	11.1	**0.8**	**0.8**	64.9	3.7	9.0	**1.1**
	b	**0.0**	92.7	81.8	59.3	98.4	8.8	32.2	4.4	11.8	4.2
	c	96.9	6.2	5.0	29.6	**0.8**	90.4	2.9	91.9	79.2	94.6
10	a	97.7	13.8	7.9	37.2	8.5	12.4	8.9	8.1	96.9	11.5
	b	2.3	3.4	57.6	**2.0**	86.2	85.9	**2.0**	87.1	2.7	87.7
	c	**0.0**	82.8	34.5	60.8	5.3	**1.7**	89.0	4.8	**0.4**	**0.8**
11	a	96.5	7.2	3.5	7.9	32.8	3.3	20.8	23.4	95.7	61.8
	b	**1.2**	91.6	9.6	**44.9**	2.8	96.7	62.9	11.3	2.7	32.4
	c	2.4	**1.1**	86.9	47.2	64.4	**0.0**	16.3	65.3	**1.6**	5.8
12	a	2.3	80.0	7.4	90.8	63.3	**1.7**	**0.0**	61.8	71.1	16.2
	b	3.5	4.2	89.4	**2.0**	9.4	**0.0**	98.0	5.6	14.8	81.9
	c	94.2	15.8	3.2	7.1	27.3	98.3	**2.0**	32.6	14.1	**1.9**
13	a	2.7	13.6	92.9	11.8	27.5	6.2	4.9	36.2	94.6	31.2
	b	5.4	71.6	7.1	82.7	6.6	**1.7**	11.1	**32.1**	4.3	8.8
	c	91.9	14.8	**0.0**	5.5	66.0	92.1	84.0	**31.7**	**1.2**	60.0
14	a	97.3	4.2	4.6	68.2	**2.0**	15.3	24.2	**1.1**	**0.4**	4.2
	b	**1.5**	2.7	91.4	20.9	93.5	78.9	46.4	98.5	**1.6**	95.4
	c	**1.2**	93.2	3.9	10.9	4.5	5.8	29.4	**0.4**	98.1	**0.4**
15	a	98.5	11.8	13.8	95.3	63.6	13.7	**0.0**	35.2	9.4	86.5
	b	**1.2**	11.5	7.8	**0.0**	23.0	58.9	**0.8**	**0.0**	80.6	8.1
	c	**0.4**	76.7	78.4	4.7	13.4	27.4	99.2	64.8	9.8	5.4
16	a	3.5	6.5	13.5	3.1	7.4	**0.8**	**1.6**	95.9	6.7	4.2
	b	**0.8**	85.9	85.5	9.0	83.6	**65.1**	97.6	**1.5**	80.4	11.9
	c	95.8	7.6	**1.1**	87.8	9.0	34.0	**0.8**	2.6	12.9	83.8
17	a	**0.0**	87.8	**1.8**	6.6	**1.2**	5.8	26.7	74.3	3.5	24.9
	b	98.5	7.6	**1.1**	80.1	**1.2**	5.4	17.7	19.3	**1.2**	17.9
	c	**1.5**	4.6	97.2	13.3	97.6	88.8	55.6	6.3	95.3	57.2
18	a	82.2	2.7	45.7	61.5	10.6	**0.8**	40.6	6.4	70.0	12.8
	b	6.6	97.3	29.5	30.6	83.7	98.8	57.0	87.6	6.6	51.2
	c	11.2	**0.0**	24.8	7.9	5.7	**0.4**	2.5	6.0	23.3	36.0

Note. Underlining of a percentage indicates that the answer choice was the correct response according to the official answer key. Shading indicates that the item needs to be reviewed and possibly revised due to the item pass rate being too low or too high. A boldface percentage indicates that the distractor selection rate is too low or too high.

Table B2

Item-Total Correlation for Each Item on Each Form of the English DL 5 (Rev. 10/98) Test for First-Attempt Renewal Applicants

Item	Form 1 (*n* = 259)	Form 2 (*n* = 263)	Form 3 (*n* = 283)	Form 4 (*n* = 258)	Form 5 (*n* = 247)	Form 6 (*n* = 242)	Form 7 (*n* = 248)	Form 8 (*n* = 271)	Form 9 (*n* = 258)	Form 10 (*n* = 261)
1	.22	.11	.22	.23	-.03	.03	.20	.18	.11	.22
2	.21	.20	.27	-.03	.29	.10	.15	.11	.08	.23
3	.08	.10	.26	.22	.20	.20	.28	.21	.39	.21
4	.33	.28	.39	.09	.15	.14	.25	.02	.15	.12
5	-.00	.11	.09	.18	.01	.05	-.06	.05	.19	.11
6	.20	.22	.28	.31	.27	.20	.35	.19	.05	.28
7	.09	.10	.27	.30	.21	.11	.05	.20	.21	.20
8	.25	.03	.09	.20	.36	.05	.15	.21	.08	.22
9	.09	.17	.27	.15	.15	.27	.06	.06	.23	.15
10	.14	.07	.26	.08	.12	-.01	.14	.27	.20	.08
11	.26	.27	.17	.39	.20	.01	.20	.20	.39	.32
12	.28	.14	.09	.15	.26	.07	.12	.32	.18	.18
13	.10	.26	.15	.19	.30	.19	.22	.20	.37	.21
14	.08	.14	.34	.21	.04	.17	.05	.10	.31	.20
15	.29	.12	.27	.19	.21	.14	-.08	.21	.20	.12
16	.20	.18	.33	.18	.24	-.13	.16	.14	.16	.12
17	.26	.13	.18	.29	.13	.21	.13	.19	.37	.18
18	.30	.13	.11	.21	.22	.15	.23	.28	.19	.31

Note. Shading indicates that an item needs to be reviewed and revised or replaced because the item-total correlation is negative or less than .10.

Table B3

Percentage of First-Attempt Renewal Applicants Who Would Pass if Different Cut-Points Were Used for Each Form of the English DL 5 (Rev. 10/98)

Number Missed	Form 1 (*n* = 259)	Form 2 (*n* = 263)	Form 3 (*n* = 283)	Form 4 (*n* = 258)	Form 5 (*n* = 247)	Form 6 (*n* = 242)	Form 7 (*n* = 248)	Form 8 (*n* = 271)	Form 9 (*n* = 258)	Form 10 (*n* = 261)	Total (*N* = 2,590)
0	12.7	16.3	4.6	4.7	5.3	2.1	0.8	6.6	5.8	3.8	6.3
1	29.7	37.3	16.6	15.1	11.3	14.0	8.9	20.7	23.3	11.5	19.0
2	58.7	54.8	31.1	29.1	25.1	37.6	18.5	37.6	45.7	26.4	36.6
3	76.8	75.3	48.1	43.8	37.7	55.4	35.5	59.8	67.4	41.8	54.3
4	84.9	87.1	65.7	63.2	55.1	76.4	55.6	73.4	82.9	54.4	69.9
5	90.7	93.2	77.4	74.8	70.0	87.6	68.1	86.3	91.5	70.9	81.1
6	96.1	97.3	85.9	82.9	81.4	95.5	79.0	94.1	95.0	80.8	88.8
7	98.8	98.1	92.9	91.5	90.3	97.5	90.3	97.0	96.5	89.7	94.3
8	99.2	99.6	96.8	95.3	96.0	99.2	95.6	98.9	96.9	95.0	97.3
9	100.0	100.0	97.9	97.7	98.4	99.6	98.4	99.6	98.1	97.3	98.7
10+	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. The shaded line represents the pass rate at the current passing standard, which allows up to three errors.

Table B4

Summary of Problem Items on Each Form of the English DL 5 (Rev. 10/98) Test for First-Attempt Renewal Applicants

Problem indicator	Form 1 items	Form 2 items	Form 3 items	Form 4 items	Form 5 items	Form 6 items	Form 7 items	Form 8 items	Form 9 items	Form 10 items
Item-total correlation too low or negative[a]	03, 05, 07, 09, 14	08, 11	05, 08, 12	02, 04, 10	01, 05, 14	01, 05, 08, 10, 11, 12, 16	05, 07, 09, 14, 15	04, 05, 09	02, 06, 08	10
Pass rate too high[b]	09, 10, 11, 14, 15, 16, 17	04, 07, 08, 18	05, 08, 17	05, 15	01, 04, 05, 09, 17	03, 11, 12, 18	08, 12, 15, 16	03, 05, 06, 08, 14, 16	03, 04, 10, 11, 13, 14, 17	07, 08, 09, 14
Pass rate too low[c]	05	none	01, 10, 18	09, 11	03, 07	07, 15, 16	02, 14, 17, 18	13	08	01, 03, 13, 17, 18
Pass rate too high or low and item-total correlation too low or negative	05, 09, 14	08	05, 08	none	01, 05	11, 12, 16	14, 15	05	08	none
Distractor selected too often[d]	none	none	01	11	03, 07	16	none	13	none	none
Distractor selected too infrequently[e]	02, 04, 07, 08, 09, 10, 11, 14, 15, 16, 17	03, 04, 07, 08, 09, 11, 18	03, 05, 08, 13, 16, 17	01, 02, 05, 07, 10, 12, 15	01, 04, 05, 09, 14, 17	03, 05, 06, 09, 10, 11, 12, 13, 16, 18	01, 06, 08, 10, 12, 15, 16	01, 03, 05, 06, 08, 14, 15, 16	02, 03, 04, 06, 10, 11, 13, 14, 17	06, 07, 08, 09, 10, 12, 14

[a]The item-total correlation was negative or less than .10. [b]More than 95% of applicants answered the item correctly. [c]Less than 60% of applicants answered the item correctly. [d]A distractor was chosen more, or almost, as often as the correct answer to the item. [e]A distractor was selected by 2% or fewer applicants.

Appendix C

Item Statistics for the English DL 5T (Rev. 8/98) for First-Attempt Provisional Applicants

Table C1

Percentage of Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5T (Rev. 8/98) Test for First-Attempt Provisional Applicants

Item	Answer choice	Form 1 (*n* = 98)	Form 2 (*n* = 102)	Form 3 (*n* = 88)	Form 4 (*n* = 91)	Form 5 (*n* = 95)
1	a	80.6	**0.0**	93.2	5.5	6.3
	b	4.1	**2.0**	6.8	9.9	16.8
	c	15.3	98.0	**0.0**	84.6	76.8
2	a	94.8	17.6	86.2	95.6	36.8
	b	5.2	75.5	3.4	3.3	2.1
	c	**0.0**	6.9	10.3	**1.1**	61.1
3	a	8.2	76.5	64.8	8.9	14.7
	b	**0.0**	**1.0**	15.9	78.9	76.8
	c	91.8	22.5	19.3	12.2	8.4
4	a	40.2	24.5	8.0	8.8	95.8
	b	**1.0**	**1.0**	67.8	79.1	2.1
	c	58.8	**74.5**	24.1	12.1	2.1
5	a	**2.0**	**2.0**	10.3	16.5	**0.0**
	b	71.4	94.1	19.5	78.0	3.2
	c	26.5	3.9	70.1	5.5	96.8
6	a	30.9	17.0	3.4	**0.0**	**0.0**
	b	20.6	8.0	19.5	94.5	**1.1**
	c	48.5	75.0	77.0	5.5	98.9
7	a	2.1	**0.0**	6.8	**0.0**	**53.2**
	b	87.6	9.8	**1.1**	**66.7**	42.6
	c	10.3	90.2	92.0	33.3	4.3
8	a	17.3	98.0	**0.0**	2.2	22.1
	b	78.6	**2.0**	95.4	14.3	69.5
	c	4.1	**0.0**	4.6	83.5	8.4
9	a	15.3	**1.0**	97.7	82.2	83.2
	b	72.4	99.0	**1.1**	6.7	4.2
	c	12.2	**0.0**	**1.1**	11.1	12.6
10	a	19.1	9.8	64.8	6.7	**53.7**
	b	44.7	6.9	4.5	93.3	35.8
	c	**36.2**	83.3	30.7	**0.0**	10.5
11	a	85.7	7.8	8.0	76.4	10.5
	b	**2.0**	70.6	84.1	**1.1**	15.8
	c	12.2	21.6	8.0	22.5	73.7
12	a	96.9	88.1	77.3	82.4	**0.0**
	b	**2.0**	**2.0**	5.7	14.3	**0.0**
	c	**1.0**	9.9	17.0	3.3	100.0
13	a	70.4	**1.0**	**0.0**	**0.0**	92.6
	b	17.3	96.1	94.3	82.2	7.4
	c	12.2	3.0	5.7	17.8	**0.0**
14	a	88.8	**1.0**	**0.0**	93.4	**0.0**
	b	3.1	7.8	96.6	4.4	94.7
	c	8.2	91.2	3.4	2.2	5.3
15	a	67.3	**49.0**	2.3	**0.0**	96.8
	b	19.4	**2.0**	12.5	**0.0**	**1.1**
	c	13.3	49.0	85.2	100.0	2.1
16	a	**1.0**	68.6	94.3	28.6	96.8
	b	6.1	3.9	3.4	**45.1**	**1.1**
	c	92.9	27.5	2.3	**26.4**	2.1
17	a	10.2	74.5	88.5	4.4	74.7
	b	72.4	10.8	10.3	56.7	13.7
	c	17.3	14.7	**1.1**	38.9	11.6
18	a	**0.0**	94.1	20.5	4.4	**1.1**
	b	100.0	**1.0**	73.9	85.6	32.6
	c	**0.0**	4.9	5.7	10.0	66.3

Table C1 (continued)

Item	Answer choice	Form 1 (*n* = 98)	Form 2 (*n* = 102)	Form 3 (*n* = 88)	Form 4 (*n* = 91)	Form 5 (*n* = 95)
19	a	2.1	3.9	15.9	2.2	11.6
	b	**0.0**	4.9	73.9	96.7	**0.0**
	c	97.9	91.2	10.2	**1.1**	88.4
20	a	98.0	**2.0**	2.3	7.7	37.9
	b	**0.0**	**2.0**	94.3	92.3	**40.0**
	c	**2.0**	96.1	3.4	**0.0**	22.1
21	a	14.4	16.7	78.4	9.9	25.3
	b	79.4	15.7	13.6	84.6	73.7
	c	6.2	67.6	8.0	5.5	**1.1**
22	a	92.9	**1.0**	93.2	2.2	24.2
	b	**1.0**	**2.0**	5.7	16.5	68.4
	c	6.1	97.1	**1.1**	81.3	7.4
23	a	17.5	85.3	**1.1**	12.2	86.3
	b	68.0	9.8	93.2	67.8	13.7
	c	14.4	4.9	5.7	20.0	**0.0**
24	a	19.4	99.0	5.7	18.7	10.6
	b	69.4	**1.0**	10.2	3.3	63.8
	c	11.2	**0.0**	84.1	78.0	25.5
25	a	84.7	99.0	9.2	**0.0**	8.4
	b	**2.0**	**0.0**	90.8	3.3	86.3
	c	13.3	**1.0**	**0.0**	96.7	5.3
26	a	6.1	14.9	94.3	16.7	12.8
	b	22.4	77.2	4.5	16.7	29.8
	c	71.4	7.9	**1.1**	66.7	57.4
27	a	**0.0**	13.7	19.3	70.0	12.6
	b	**0.0**	84.3	13.6	26.7	85.3
	c	100.0	**2.0**	67.0	3.3	2.1
28	a	21.4	8.8	**1.1**	**0.0**	60.0
	b	71.4	13.7	96.6	**0.0**	31.6
	c	7.1	77.5	2.3	100.0	8.4
29	a	**0.0**	**37.0**	4.6	91.2	6.3
	b	100.0	37.0	10.3	7.7	3.2
	c	**0.0**	26.0	85.1	**1.1**	90.5
30	a	**0.0**	**0.0**	**0.0**	9.9	73.7
	b	100.0	5.0	100.0	76.9	16.8
	c	**0.0**	95.0	**0.0**	13.2	9.5
31	a	92.9	55.4	97.7	78.9	6.5
	b	**0.0**	39.6	2.3	14.4	14.0
	c	7.1	5.0	**0.0**	6.7	79.6
32	a	66.3	83.3	53.4	**0.0**	2.1
	b	5.1	15.7	14.8	**0.0**	92.6
	c	28.6	**1.0**	31.8	100.0	5.3
33	a	18.8	93.1	71.3	**1.1**	**0.0**
	b	60.4	**2.0**	**1.1**	98.9	**1.1**
	c	20.8	4.9	27.6	**0.0**	98.9
34	a	19.8	40.2	**58.6**	3.3	3.2
	b	19.8	**0.0**	29.9	96.7	55.3
	c	60.4	59.8	11.5	**0.0**	41.5
35	a	**1.0**	4.9	12.6	89.0	21.1
	b	99.0	**1.0**	54.0	**0.0**	72.6
	c	**0.0**	94.1	33.3	11.0	6.3
36	a	**0.0**	**1.0**	4.5	3.3	11.6
	b	**0.0**	**1.0**	89.8	3.3	56.8
	c	100.0	98.0	5.7	93.4	31.6
37	a	17.5	15.8	34.1	3.3	9.5
	b	12.4	**1.0**	14.8	5.6	16.8
	c	70.1	83.2	**51.1**	91.1	73.7

Table C1 (continued)

Item	Answer choice	Form 1 (n = 98)	Form 2 (n = 102)	Form 3 (n = 88)	Form 4 (n = 91)	Form 5 (n = 95)
38	a	4.1	12.9	8.1	20.9	92.6
	b	38.1	13.9	86.0	69.2	**1.1**
	c	57.7	73.3	5.8	9.9	6.3
39	a	38.8	91.0	92.0	20.9	11.7
	b	4.1	4.0	2.3	71.4	7.4
	c	57.1	5.0	5.7	7.7	80.9
40	a	37.5	15.7	10.2	89.0	11.6
	b	2.1	76.5	**0.0**	3.3	88.4
	c	60.4	7.8	89.8	7.7	**0.0**
41	a	10.3	5.9	**1.1**	91.2	95.8
	b	78.4	93.1	59.1	**1.1**	3.2
	c	11.3	**1.0**	39.8	7.7	**1.1**
42	a	9.3	8.8	**50.6**	23.1	2.1
	b	17.5	**2.0**	5.7	34.1	97.9
	c	73.2	89.2	43.7	**42.9**	**0.0**
43	a	**1.0**	**37.3**	82.8	**1.1**	6.3
	b	99.0	25.5	14.9	**0.0**	71.6
	c	**0.0**	37.3	2.3	98.9	22.1
44	a	**1.0**	80.2	21.6	4.4	34.8
	b	93.9	18.8	72.7	64.8	16.3
	c	5.1	**1.0**	5.7	30.8	48.9
45	a	4.1	61.8	92.0	**1.1**	2.1
	b	14.4	17.6	3.4	97.8	83.2
	c	81.4	20.6	4.5	**1.1**	14.7
46	a	7.1	7.9	11.4	8.8	10.6
	b	83.7	90.1	**0.0**	**0.0**	50.0
	c	9.2	**2.0**	88.6	91.2	39.4

Note. Underlining of a percentage indicates that the answer choice was the correct response according to the official answer key. Shading indicates that the item needs to be reviewed and possibly revised due to the item pass rate being too low or too high. A boldface percentage indicates that the distractor selection rate is too low or too high.

Table C2

Item-Total Correlation for Each Item on Each Form of the English DL 5T (Rev. 8/98) Test for First-Attempt Provisional Applicants

Item	Form 1 (n = 98)	Form 2 (n = 102)	Form 3 (n = 88)	Form 4 (n = 91)	Form 5 (n = 95)
1	.34	.39	.17	.13	.27
2	.14	.35	.27	.11	.14
3	.23	.25	.15	.31	.33
4	.01	.31	.27	.29	.11
5	.20	.16	.23	.25	-.02
6	.21	.37	.05	.22	.18
7	.12	.20	.25	.08	.38
8	.19	.06	.31	.26	.17
9	.10	.18	.30	.12	.24
10	.38	.13	.22	.41	.34
11	.11	.29	.31	.35	.27
12	.23	.15	.17	.09	*
13	.17	.31	.32	.25	.18
14	.11	.36	.24	.13	.23
15	.35	.21	.06	*	.06
16	.30	.47	.09	.19	.16
17	.19	.17	.07	.12	.28
18	*	.45	.24	.26	.32
19	.25	.23	.02	.22	.10
20	.01	.12	.23	.30	.18
21	.40	.20	.35	.00	.21
22	.06	.34	.27	.36	.36
23	.39	.09	.15	.11	.16
24	.24	-.01	.19	.17	.32
25	.38	-.00	.16	-.09	.14
26	.41	.56	-.16	.27	.29
27	-.14	.05	.18	.12	.11
28	.35	.04	.35	*	.21
29	*	.11	.16	.41	.29
30	*	.20	*	.31	.31
31	.14	.29	.12	.22	.28
32	.15	.16	.25	*	-.05
33	.30	.32	.00	-.09	.18
34	.26	.20	.34	.22	.34
35	.19	.32	.04	.05	.33
36	*	.52	.39	.29	.14
37	.26	.25	.20	.35	.00
38	.19	.11	.22	.16	.07
39	.33	.06	.15	.21	.29
40	.23	.19	.13	.25	.06
41	.44	.22	.24	.03	.29
42	.30	.03	.29	.21	-.11
43	.29	.13	-.03	.16	.35
44	.16	.26	.23	.35	.18
45	.30	.05	.16	.22	.16
46	.13	-.06	.07	-.06	.03

Note. Shading indicates that an item needs to be reviewed and revised or replaced because the item-total correlation is negative or less than .10.
*Item-total correlation could not be calculated because the item had no variance.

Table C3

Percentage of First-Attempt Provisional Applicants Who Would Pass if Different Cut-Points Were Used for Each Form of the English DL 5T (Rev. 8/98)

Number Missed	Form 1 (n = 98)	Form 2 (n = 102)	Form 3 (n = 88)	Form 4 (n = 91)	Form 5 (n = 95)	Total (N = 474)
0	2.0	2.0	0.0	1.1	1.1	1.3
1	3.1	2.0	1.1	3.3	2.1	2.3
2	6.1	3.9	1.1	5.5	3.2	4.0
3	8.2	6.9	4.5	9.9	6.3	7.2
4	13.3	8.8	12.5	14.3	9.5	11.6
5	21.4	22.5	18.2	22.0	15.8	20.0
6	24.5	29.4	30.7	35.2	21.1	28.1
7	34.7	37.3	44.3	44.0	24.2	36.7
8	41.8	51.0	52.3	57.1	32.6	46.8
9	51.0	59.8	55.7	67.0	44.2	55.5
10	61.2	70.6	63.6	76.9	51.6	64.8
11	67.3	77.5	69.3	81.3	62.1	71.5
12	74.5	84.3	77.3	83.5	69.5	77.9
13	78.6	86.3	84.1	86.8	73.7	81.9
14	82.7	90.2	89.8	87.9	80.0	86.1
15	88.8	92.2	92.0	92.3	83.2	89.7
16+	100.0	100.0	100.0	100.0	100.0	100.0

Note. The shaded line represents the pass rate at the current passing standard, which allows up to five errors.

Table C4

Summary of Problem Items on Each Form of the English DL 5T (Rev. 8/98) Test for First-Attempt Provisional Applicants

Problem indicator	Form 1 items	Form 2 items	Form 3 items	Form 4 items	Form 5 items
Item-total correlation too low or negative[a]	04, 18, 20, 22, 27, 29, 30, 36	08, 23, 24, 25, 27, 28, 39, 42, 45, 46	06, 15, 16, 17, 19, 26, 30, 33, 35, 43, 46	07, 12, 15, 21, 25, 28, 32, 33, 35, 41, 46	05, 12, 15, 32, 37, 38, 40, 42, 46
Pass rate too high[b]	02, 12, 18, 19, 20, 27, 29, 30, 35, 36, 43	01, 08, 09, 13, 20, 22, 24, 25, 30, 36	08, 09, 14, 28, 30, 31	02, 15, 19, 25, 28, 32, 33, 34, 43, 45	04, 05, 06, 12, 14, 15, 16, 33, 41, 42
Pass rate too low[c]	04, 10, 33, 34, 38, 39, 40	04, 15, 29, 31, 34, 43	32, 34, 35, 37, 41, 42	07, 16, 17, 42	07, 10, 20, 26, 28, 34, 36, 44, 46
Pass rate too high or low and item-total correlation too low or negative	04, 18, 20, 27 29, 30, 36	08, 24, 25	30, 35	07, 15, 25, 28, 32, 33	05, 12, 15, 42, 46
Distractor selected too often[d]	10	04, 15, 29, 43	34, 37, 42	07, 16, 42	07, 10, 20
Distractor selected too infrequently[e]	02, 04, 05, 11, 12, 16, 18, 19, 20, 22, 25, 27, 29, 30, 31, 35, 36, 43, 44	01, 04, 05, 07, 08, 09, 12, 13, 14, 15, 18, 20, 22, 24, 25, 27, 30, 32, 33, 34, 35, 36, 37, 41, 42, 44, 46	01, 07, 08, 09, 13, 14, 17, 22, 23, 25, 26, 28, 30, 31, 33, 40, 41, 46	02, 06, 07, 10, 11, 13, 15, 19, 20, 25, 28, 29, 32, 33, 34, 35, 41, 43, 45, 46	05, 06, 12, 13, 14, 15, 16, 18, 19, 21, 23, 33, 38, 40, 41, 42

[a]The item-total correlation was negative or less than .10. [b]More than 95% of applicants answered the item correctly. [c]Less than 60% of applicants answered the item correctly. [d]A distractor was chosen more, or almost, as often as the correct answer to the item. [e]A distractor was selected by 2% or fewer applicants.

Appendix D

Table of Fail Rates by Field Office

Table D1

Number of Tests and Fail Rates for English DL 5 (Rev. 10/98)
Original and Renewal Applicants (Over All Attempts) by Field Office

RPU	Name	Data comments	Originals		Renewals	
			n	Fail rate	*n*	Fail rate
501	Sacramento		35	74.3	45	55.6
502	Los Angeles	WTR	0	*	0	*
503	San Francisco		131	65.7	74	43.2
504	Oakland		58	67.2	73	54.8
505	Fresno		15	66.7	9	88.9
506	San Diego	N	0	*	0	*
507	Long Beach		53	69.8	81	34.6
508	Hollywood		115	59.1	46	60.9
509	Pasadena		67	68.7	106	50.9
510	Glendale	N	0	*	0	*
511	Montebello		48	62.5	64	48.4
512	San Bernardino		1	100.0	32	50.0
513	Truckee	WTR	0	*	0	*
514	Culver City	N	0	*	0	*
515	Van Nuys		45	71.1	31	35.5
516	San Jose		48	62.5	53	56.6
517	Stockton		42	66.7	47	59.6
518	Mountain View	N	0	*	0	*
519	San Diego-Clairmont		75	61.3	84	41.7
520	Chico	N	0	*	0	*
521	Jackson	WTR	0	*	0	*
522	Oroville		3	66.7	25	64.0
523	Concord		34	67.7	44	40.9
524	Crescent City		0	*	1	100.0
525	Placerville		9	55.6	30	36.7
526	Eureka		16	62.5	21	33.3
527	El Centro		14	78.6	7	71.4
528	Blythe		5	40.0	2	100.0
529	Bakersfield		25	80.0	17	35.3
530	Lakeport		7	71.4	18	33.3
531	Susanville		1	0.0	5	40.0
532	Pomona		40	55.0	49	44.9
533	Madera		9	77.8	12	58.3
534	Corte Madera		33	51.5	37	45.9
535	Ukiah		2	100.0	7	85.7
536	Merced		13	61.5	22	45.4
537	Alturas	N	0	*	0	*
538	South Lake Tahoe		2	0.0	4	25.0
539	Salinas		24	66.7	21	42.9
540	Napa		23	56.5	22	22.7
541	Grass Valley		4	50.0	7	42.9
542	Santa Ana		29	34.5	48	27.1
543	Roseville		24	70.8	37	59.5

Table D1 (continued)

RPU	Name	Data comments	Originals		Renewals	
			n	Fail rate	*n*	Fail rate
544	Quincy	N	0	*	0	*
545	Riverside		27	66.7	29	51.7
546	Hollister		5	60.0	10	60.0
547	San Luis Obispo		4	100.0	32	56.2
548	Redwood City	N	0	*	0	*
549	Santa Barbara		21	42.9	25	44.0
550	Capitola		14	57.1	45	37.8
551	Redding		13	38.5	23	34.8
552	Yreka		1	100.0	4	25.0
553	Tulelake	N	0	*	0	*
554	Vallejo		24	83.3	30	56.7
555	Santa Rosa		27	59.3	45	51.1
556	El Cerrito		45	55.6	46	39.1
557	Modesto	WTR	0	*	0	*
558	Red Bluff		5	40.0	11	45.4
559	Visalia		16	68.7	23	52.2
560	Ventura		21	61.9	38	31.6
561	Woodland	N	0	*	0	*
562	Yuba City		15	66.7	22	59.1
563	Santa Maria		14	50.0	23	73.9
564	Colusa		1	0.0	4	50.0
565	Hanford		18	38.9	21	47.6
566	Mariposa		4	75.0	7	28.6
567	Seaside		11	27.3	14	50.0
568	San Andreas		3	66.7	8	37.5
569	Sonora		7	71.4	18	44.4
570	Auburn		6	50.0	23	34.8
571	Willows		0	*	1	0.0
572	Weaverville	N	0	*	0	*
573	Porterville		5	40.0	9	11.1
574	Paso Robles		8	62.5	13	46.1
575	Taft		1	100.0	6	0.0
576	Bell Gardens	WTR	0	*	0	*
577	Ridgecrest		6	16.7	12	33.3
578	Indio		13	61.5	42	50.0
579	Hayward	WTR	0	*	0	*
580	Clovis		4	50.0	25	40.0
581	Compton	N	0	*	0	*
582	Barstow		8	50.0	12	16.7
583	Watsonville		4	75.0	6	66.7
584	Needles		1	0.0	3	66.7
585	Bishop		0	*	3	66.7
586	Norco	N	0	*	0	*
587	Arleta		48	62.5	21	52.4
588	Vacaville		12	58.3	23	52.2
589	Lompoc		11	45.4	16	37.5
590	Fort Bragg		0	*	4	25.0

Table D1 (continued)

RPU-Name		Data comments	Originals		Renewals	
			n	Fail rate	*n*	Fail rate
591	Whittier	N	0	*	0	*
592	Pittsburg		20	75.0	40	60.0
593	San Mateo		32	40.6	42	40.5
594	Tulare		12	75.0	14	50.0
595	Lancaster		24	54.2	35	45.7
596	Oceanside		59	54.2	64	42.2
597	Brawley		2	100.0	6	66.7
598	Davis		5	20.0	12	50.0
599	Daly City	N	0	*	0	*
601	Paradise	N	0	*	0	*
602	Sacramento-South		34	67.7	46	45.6
603	Coalinga		6	83.3	3	33.3
604	Oakland Coliseum		40	77.5	45	57.8
605	Laguna Hills		47	46.8	65	29.2
606	Bellflower		62	75.8	77	51.9
607	Fullerton	N	0	*	0	*
608	Torrance	WTR	0	*	0	*
609	Hawthorne		49	79.6	54	68.5
610	Inglewood	N	0	*	0	*
611	Westminister		55	63.6	53	49.1
612	Rancho Cucamonga		64	84.4	38	28.9
613	Chula Vista		70	68.6	72	52.8
615	Delano	N	0	*	0	*
616	Santa Monica	N	0	*	0	*
617	Lincoln Park		6	83.3	2	50.0
618	West Covina		95	67.4	106	49.1
619	San Pedro		0	*	1	100.0
620	Escondido		36	61.1	47	36.2
621	Fairfield		18	83.3	12	50.0
622	Lodi		17	64.7	40	52.5
623	Gilroy		8	50.0	23	52.2
624	Walnut Creek		38	63.2	49	36.7
625	Carmichael		47	68.1	71	62.0
626	Redlands		30	63.3	40	35.0
627	Garberville	N	0	*	0	*
628	Costa Mesa		36	52.8	69	42.0
629	Victorville		32	65.6	57	52.6
630	Santa Paula		9	77.8	22	50.0
631	Pleasanton		39	46.1	38	39.5
632	Santa Clara		65	60.0	61	29.5
633	Reedley		19	68.4	15	46.7
634	Petaluma		7	42.9	23	43.5
635	Hemet		16	56.2	42	38.1
636	Oxnard		30	76.7	31	38.7
637	Winnetka		55	65.4	59	50.8
638	Twentynine Palms		14	71.4	13	38.5
639	Mount Shasta		0	*	2	100.0

Table D1 (continued)

RPU	Name	Data comments	Originals		Renewals	
			n	Fail rate	*n*	Fail rate
640	Los Gatos		32	68.7	38	39.5
641	Banning	N	0	*	0	*
642	Tracy		8	75.0	19	47.4
643	Fall River Mills	WTR	0	*	0	*
644	Fremont		35	62.8	30	53.3
645	Orland		2	50.0	5	20.0
646	Fresno-North	N	0	*	0	*
647	King City		3	100.0	2	50.0
648	San Clemente		23	43.5	37	37.8
649	Turlock		20	65.0	23	26.1
650	Los Banos		1	0.0	4	75.0
655	Folsom		19	47.4	32	31.2
656	Riverside-East		30	70.0	38	57.9
657	Fontana		45	73.3	34	41.2
658	Manteca		7	71.4	26	61.5
659	Palm Springs		29	70.0	42	47.6
660	Shafter		7	85.7	2	50.0
661	Arvin		0	*	1	0.0
662	Newhall	N	0	*	0	*
663	Thousand Oaks		36	58.3	22	40.9
668	Santa Teresa		26	73.1	21	38.1
669	El Cajon	N	0	*	0	*
670	Goleta		15	80.0	28	50.0
672	Temecula		18	72.2	47	51.1
673	Rocklin		1	100.0	15	53.3
676	Poway		36	58.3	54	48.1
677	San Ysidro		25	72.0	14	64.3
679	Bakersfield-Southwest		0	*	1	100.0
680	Simi Valley		11	45.4	22	40.9
686	Novato	N	0	*	0	*
687	Lake Isabella	N	0	*	0	*
690	Palmdale		12	58.3	20	44.4
	Total	35	3,112	63.8	3,944	46.3

Note. Office fail rates that are based on fewer than 20 test forms are highly unstable estimates, and should *not* be interpreted as accurate estimates. Fail rates are not presented by office for Spanish DL 5 originals and renewals, or for English DL 5T provisionals, because too few forms were collected to compute accurate estimates for the majority of the offices. *The fail rate could not be computed because no test forms were received of that type. N = No test forms were received. WTR = No current revision test forms were received.

[a]The figures presented for total fail rates are weighted averages.

Appendix E

Test Forms used in This Evaluation

Due to security considerations the test forms are not distributed outside the department