

An Evaluation of the Validity of California's Driving Performance Evaluation Road Test

By
Patricia A. Romanowicz and Robert A. Hagge

August 1995

Research and Development Section
Division of Program and Policy Administration
California Department of Motor Vehicles
RSS-95-154

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE August 1995		3. REPORT TYPE AND DATES COVERED
4. TITLE AND SUBTITLE An Evaluation of the Validity of California's Driving Performance Evaluation Road Test			5. FUNDING NUMBERS	
6. AUTHOR(S) Patricia A. Romanowicz and Robert A. Hagge				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) California Department of Motor Vehicles Research and Development Section P. O. Box 932382, F-126 Sacramento, CA 94232-3820			8. PERFORMING ORGANIZATION REPORT NUMBER RSS-95-154	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This report presents findings of an evaluation of the validity of the Driving Performance Evaluation (DPE) road test that was piloted in 30 California Department of Motor Vehicles field offices. The study represents the fourth stage in a four-stage project to develop an improved competency-based drive test for possible statewide implementation. The DPE was found to have construct validity as demonstrated by experienced good drivers having had significantly lower fail rates and mean point scores than did inexperienced drivers and drivers with physical or mental disabilities that affected their driving. The evaluation also found the DPE to be more difficult than the current drive test, with fail rates of 45.6% and 26.2% for the two tests, respectively. The DPE was also found to take 11 minutes longer to administer than did the current drive test. The impact on test validity of several modifications to shorten the DPE test time was also evaluated.				
14. SUBJECT TERMS drive test, driver competency, fail rates, disqualifications, test time, modifications			15. NUMBER OF PAGES 37	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT	

PREFACE

This report presents findings of a validity evaluation of a prototype drive test that was piloted in 30 California Department of Motor Vehicles field offices. The study represents the fourth stage in a four-stage project to develop an improved competency-based drive test for possible statewide implementation. The present report is being issued as an internal monograph of the Department of Motor Vehicles' Research and Development Section rather than an official report of the State of California. The findings and options may therefore not represent the views and policies of the State of California.

ACKNOWLEDGMENTS

This report presents results of a validity evaluation of a prototype drive test that was piloted in 30 California Department of Motor Vehicles field offices. The study was conducted under the general direction of Raymond C. Peck, Research Chief. Mark Harling, Manager III of Field Operations Division's Program Support Staff, acted as a liaison between Research and Development and the study field offices.

Special thanks go to Eric Berube, Research Analyst II, for doing the analysis of covariance (ANCOVA) and to Debbie McKenzie, Staff Services Analyst, for producing the tables and graphs, monitoring data collection, and coding data. Thanks are also given to David Hironaka, Information Systems Technician of Headquarters Operations' Data Entry, for writing the data entry procedures.

The authors also wish to thank all of the Licensing Registration Examiners Supervising Motor Vehicle Representatives, Driver License Supervisors, and Field Office Managers in the 30 Driving Performance Evaluation offices for their efforts in the data collection during the pilot, especially staff in the Fullerton, West Covina, and Westminster field offices. Thanks are also given to LeeAnn Richardson, Hearing Officer in the Driver Safety Review Unit, for locating and notifying three-accident reexamination subjects for the validity study.

EXECUTIVE SUMMARY

Introduction

- In 1990 the California Department of Motor Vehicles (DMV) established a program to increase the level of competency of California drivers. One of the components of

this program is the development and implementation of an improved drive test, the Driving Performance Evaluation (DPE).

- The purpose of this study was to evaluate the validity of Stage 4 of the DPE and its effect on drive test fail rate and average administration time.
- The DPE differs from the current drive test in several important ways. For example, it is around 10 minutes longer, includes an off road skill test and freeway driving, and uses more objective or standardized criteria in which defined maneuvers are observed and scored only at preplanned locations and times during the test.

Methods

- Six groups of subjects—novice originals, out-of-state originals, congratulations letter renewals, written-test renewals, three-accident reexamination referrals, and limited-term renewals—were administered the DPE by two Licensing Registration Examiners (LREs) in each of three DMV field offices from September 1994 through January 1995. Additional data on DPE performance and process measures were collected 1 week each month during the same time period for all applicants appearing for a drive test in one of the 30 DPE offices.
- The validity of the DPE was assessed by comparing the six groups on fail rate and mean point score, by correlating group status with DPE pass/fail result and point score, and by comparing the DPE performance of accident-free and accident-involved drivers.
- Several possible changes to the DPE were evaluated to determine their effects on validity as measured by correlation of group status with DPE score.
- DPE fail rate and average test time were compared with the fail rate and average test time for the current drive test in the 30 DPE offices.

Results

- The novice original and limited-term renewal groups had significantly higher fail rates than did the out-of-state original, congratulations renewal, and written-test renewal groups, and the out-of-state original group had a significantly higher fail rate than did the congratulations renewal group.
- The novice original, three-accident reexamination, and limited-term renewal groups had significantly higher mean point scores than did the out-of-state original, congratulations renewal, and written-test renewal groups.

- Validity was also found at the individual-examiner level; all LREs rated the DPE performance of the novice original group as being significantly worse than that of the congratulations and written-test renewal groups.
- Group status (novice original vs. congratulations renewal) was significantly correlated with DPE pass/fail result and point score.
- Although the difference was not statistically significant, accident-involved subjects had a higher DPE fail rate than did accident-free subjects.
- Fail rate increased by 19.4 percentage points, from 26.2% for the current drive test to 45.6% for the DPE.
- Average test time increased by 11.2 minutes, from 13.6 minutes for the current drive test to 24.8 minutes for the DPE.
- All simulated modifications to the DPE reduced validity, except eliminating street parking which slightly increased validity. Eliminating freeway driving reduced validity the most, causing the correlation of point score with group status to decrease from -.30 to -.25.

Conclusions

- The overall results of the analyses provide strong evidence that the DPE is a valid test.
- The DPE is more difficult than the current drive test, substantially increasing test fail rate.
- The DPE takes more than 10 minutes longer to administer than does the current drive test.
- If the DPE must be modified to reduce test length, freeway driving should not be eliminated.

TABLE OF CONTENTS

	<u>PAGE</u>
PREFACE.....	i
ACKNOWLEDGMENTS.....	i
EXECUTIVE SUMMARY.....	i
Introduction.....	i
Methods.....	ii
Results.....	ii
Conclusions.....	iii
INTRODUCTION.....	1
METHODS.....	1
Subjects.....	1
Data Collection.....	3
Data Analysis.....	4
DPE validity.....	4
Evaluation of DPE modifications	5
RESULTS.....	5
DPE validity	5
Construct validity.....	5
Ultimate criterion validity.....	13
General Process Measures.....	14
DPE difficulty	14
DPE test time.....	18
Evaluation of DPE Modifications.....	19
DISCUSSION AND RECOMMENDATIONS	21
REFERENCES	23

APPENDICES

<u>NUMBER</u>		
A	Driving Performance Evaluation Score Sheet	24
B	Daily DPE Activity Form	25
C	Table C-1. Correlation of Item Score and Group Status (Novice Original vs. Congratulations Renewal) for First-Attempt Subjects Completing the DPE (DQs Excluded) in Fullerton, West Covina, and Westminster (N = 1,563).....	26

TABLE OF CONTENTS (Continued)

APPENDICES (Continued)

<u>NUMBER</u>	<u>PAGE</u>
<u>Table C-2.</u> Number of Subjects (<i>n</i>) and Fail Rate for the DPE and Current Drive Tests by Field Office	29
<u>Table C-3.</u> Number of Subjects (<i>n</i>) and Turn-and-Stop DQ Rate by Field Office.....	30
<u>Table C-4.</u> Number of Subjects (<i>n</i>) and Mean Test Time for the DPE and Current Drive Tests by Field Office	31

LIST OF TABLES

1	Number of Subjects in Each Office by First-Attempt Applicant Group.....	6
2	Number of Subjects (<i>n</i>), Fail Rate, and Mean Score by First-Attempt Applicant Group	6
3	Fail Rate (%) and Number of Subjects (in Parentheses) by Office, LRE, and First-Attempt Applicant Group	8
4	Fail Rate ANOVA Results.....	9
5	Mean Score and Number of Subjects (in Parentheses) by Office, LRE, and First-Attempt Applicant Group (DQs Excluded).....	10
6	Mean Score ANOVA Results.....	10
7	Number of Subjects (<i>n</i>) and Correlation (<i>r</i>) of Pass/Fail Result and Total Point Score with Applicant Group by LRE.....	12
8	Number of Subjects (<i>n</i>), Fail Rate, and Mean Score by First-Attempt Applicant Group (30 Offices).....	13
9	Number of Subjects (<i>n</i>), Mean Age, Percentage of Women, and Observed and Statistically-Adjusted DPE Fail Rates for Accident-Free and Accident-Involved Renewal and 3-Accident Reexamination Subjects	13

TABLE OF CONTENTS (Continued)

LIST OF TABLES (Continued)

<u>NUMBER</u>		<u>PAGE</u>
10	Number (<i>n</i>) and Percentage of Applicants with Usable Data Reported on the Daily DPE Activity Form by DPE Test Status (30 Offices).....	14
11	DPE Results for All Applicants (30 Offices)	15
12	DPE Results for First-Attempt Novice Originals (30 Offices).....	16
13	Fail Rate (%) and Number of Subjects (in Parentheses) by Group and Attempt Number for All Applicants Tested (30 Offices).....	17
14	Total Fails, On-Road DQs, Turn-and-Stop DQs, and Point-Score Fails as Percentages of all Novice Original First-Attempt Applicants by Month for September 1994 through February 1995.....	17
15	Number of Subjects (<i>n</i>) and Correlation of Total Point Score with Group Status by LRE for Various DPE Modifications (DQs Excluded).....	20

LIST OF FIGURES

1	DPE fail rate by first-attempt applicant group	7
2	DPE mean score by first-attempt applicant group (DQs excluded)	7
3	DPE fail rate by office, LRE, and first-attempt applicant group.....	9
4	DPE mean score by office, LRE, and applicant group (DQs excluded).....	11
5	Total fails, on-road DQs, turn-and-stop DQs, and point-score fails as percentage of all novice original first-attempt applicants by month for September 1994 through February 1995.	18
6	DPE mean test time by month for September 1994 through February 1995.....	19

INTRODUCTION

This study evaluated the validity of the Driving Performance Evaluation (DPE) drive test piloted in 30 California Department of Motor Vehicles (DMV) field offices in Los Angeles, Orange, and San Diego counties. The evaluation is the final step (Stage 4) in a four-stage project for developing and evaluating a new drive test for possible statewide implementation in California. The Stage 1 study (Shumaker, 1994) evaluated the reliability of the department's current drive test in six field offices that were selected in a pre-Stage 1 study (Williams & Shumaker, 1994) to be a representation of field offices statewide. A prototype of the DPE was piloted in Bellflower, Laguna Hills, Sacramento, and South Sacramento in Stage 2. The Stage 3 pilot (Hagge, 1994) evaluated the reliability of the DPE in the six field offices used for the Stage 1 study and found it to be much more reliable than the current drive test. It also provided information for further improving the DPE prior to the pilot for the current study.

The DPE is based on the driver performance model applied to commercial road tests described in a report by Mackie et al. (1989). A brief review of the relevant drive test literature is included in the Stage 1 report.

METHODS

This study assessed the validity of the DPE in three field offices, and provides general descriptive statistics on the DPE testing process in the 30 DPE pilot offices. Validity is defined here as the ability of the test to accurately measure driving competency. The higher the validity of the test, the greater the likelihood that the correct licensing decision will be made based on DPE performance. The selection of subjects and the data collection and analysis procedures are described below.

Subjects

Subjects in the validity study were drawn from applicants administered the DPE in the Fullerton, West Covina, and Westminster field offices. These offices were also involved, along with three other offices, in the Stage 1 and Stage 3 evaluations mentioned above. The subject pool consisted of:

- (1) novice original driver license applicants taking a California drive test for the first time;
- (2) original license applicants previously licensed in another state and taking a California drive test for the first time;
- (3) reentrant license renewal applicants who received a good-driver "congratulations" letter and, therefore, were not required to take an 18-item written test;
- (4) renewal license applicants who were required to take an 18-item written test;
- (5) licensed drivers involved in three or more traffic accidents in 1 year;

- (6) limited-term license renewal applicants taking a drive test for the first time on the current license renewal application.

Only drivers taking the DPE for the first time were selected as subjects. Subjects who failed the DPE were not included in the study again upon taking a DPE retest. Participation in the study was represented as mandatory and applicants were not told they were subjects in a study. If subjects asked whether the results of the test would count against them, they were told that their performance on the DPE would not be used to make a licensing decision. Subjects who performed poorly on the DPE may have been asked to take additional tests of driver competency that could be used as a basis for a licensing action.

Applicants who were selected as subjects, but refused to take the DPE, were to be sent to a Driver License (DL) Supervisor for further discussion. Subjects who strongly objected to the drive test were to be informed by the DL Supervisor that DMV has legal authority to give a drive test to a license applicant at any time. Subjects who still refused to cooperate were to be referred to the Field Office Manager for additional discussion and possible release from the study.

Renewal subjects who were willing to be tested, but indicated that they were unable to take the drive test at that time, were to be rescheduled for a DPE to be administered within 2 weeks, and were not to be allowed to complete the licensing transaction until they returned to take the DPE in the same office. Novice originals who indicated that they were not prepared to drive on the freeway were not used as subjects on that day. These applicants were asked to practice freeway driving and to later reschedule the DPE when they felt they were ready.

Three-accident reexamination subjects—those with three or more accidents in a year—were identified by Driver Safety Review Unit and instructed to phone a designated contact person for a DPE appointment in one of the three validation study offices. The contact person was to schedule the DPE early enough for the subject to be given the drive test within 2 weeks.

Because the number of limited-term renewals and three-accident subjects was expected to be small, these applicants were to be tested only by the study Licensing Registration Examiners (LREs) and given the highest priority in testing throughout the study. When none of these subjects were available, each study LRE was to try to alternate testing between the other four groups in an attempt to roughly balance the number of subjects in each group. In an effort to minimize subject-selection bias, each study LRE was to choose the next person in line as the next subject to be tested in the specified group.

Subjects selected for the development of general DPE performance and process measures consisted of all applicants who appeared for a drive test in one of the 30 DPE offices during the general data collection periods specified below.

Data Collection

All LREs in the three validation study offices administered the DPE, but only two designated study LREs in each office administered DPEs for the DPE validity assessment. The validation study used the same LREs, DL Supervisors, and Field Office Managers who participated in the Stage 3 DPE study, except for two LRE replacements. Data were collected from September 19, 1994 through January 27, 1995 for all subjects except the three-accident reexamination group. Data collection was extended through February for the latter group in order to obtain a larger sample size for that group.

In addition to the information normally required on the DPE score sheet (see Appendix A), each study LRE was to record on the sheet the subject's study group, the DPE start and end times, and, if applicable, the reason why the subject was not scored on the DPE freeway maneuvers (e.g., freeway closure or subject's refusal).

Each study LRE also recorded data on a separate Daily DPE Activity form (see Appendix B) for each subject they tested that day. The information collected included test start and end times, applicant status, drive test attempt number, whether the applicant was tested on the freeway, whether the applicant was restricted to no freeway, and drive test result. The form also provided space for the validation study LREs to explain why an applicant selected as a subject may not have been tested on the DPE (e.g., they refused to participate or did not have auto insurance).

At the end of each day, the DL Supervisor was to collect and send to the Research and Development Section (R&D) all DPE score sheets and Daily DPE Activity forms completed by the study LREs.

Data for the descriptive analysis were collected from all 30 DPE offices for 1 week each month during September 1994 through June 1995. The data were collected using the Daily DPE Activity form mentioned above. The form was to be completed by each LRE. Each DL window technician was also to complete the form for applicants who appeared for a drive test but for some reason (e.g., no auto insurance) were not referred to an LRE to take the DPE. Data were not collected for phone-in cancellations and no-shows. The completed Daily DPE Activity forms were submitted to R&D at the end of each week of data collection.

Score sheets for the department's current drive test (DL 179) were also collected from the 30 offices. LREs were instructed to record test start and end times on all score sheets completed in May 1994, just before the DPE was implemented. A random sample of score sheets was selected for each office to reduce the number of documents needing to be key entered. Sampling ratios were used that were expected to yield within-office sample sizes of 75 or larger. Average drive test time and fail rate for the current drive test, normalized to reflect the same proportional representation of DPE score sheets collected from each office, served as baseline measures for assessing the operational impact of the DPE.

All DPE score sheets and Daily DPE Activity forms were screened by R&D and key entered by Data Entry.

Data Analysis

DPE validity. The validity of the DPE was evaluated by comparing the six applicant groups on average DPE fail rate and mean DPE point score. If performance on the DPE were a valid measure of driving competency, experienced drivers (e.g., renewals and California originals previously licensed in another state) would be expected to perform better than novice drivers and limited-term renewals drivers with worsening physical or mental conditions that affect their driving. The groups were also compared within each office and LRE to determine whether the validation results varied as a function of who did the scoring. The analysis of variance (ANOVA) statistical procedure was used to test whether group differences on the DPE performance measures were statistically significant (i.e., not likely to have occurred by chance).

Two additional DPE validity measures were calculated by correlating subject status (novice original vs. congratulations renewal) with DPE pass/fail result and DPE point score. Item validity measures were also computed by correlating subject status with item scores for subjects who completed the DPE. Subjects who were automatically disqualified (DQued) due to making a serious error that caused the examiner to terminate the test were excluded from the computation. Correlation as measured here refers to the strength of the relationship between subject status and DPE performance. The correlation coefficient theoretically can range from -1 to 1. A large positive or negative coefficient indicates the two variables are strongly related, whereas a 0 coefficient indicates the two variables are completely unrelated. A positive value means that the variables tend to increase or decrease together, and a negative value means that one variable tends to increase as the other decreases, or visa versa.

"Ultimate" criterion validity—the relationship between performance on the DPE and accident risk—was assessed using a contrasted-groups design. Driver records for the 3-year period prior to testing were obtained from the department's automated DL masterfile for renewal subjects (groups 3 and 4) and three-accident reexamination subjects (group 5). These subjects were divided into two groups: accident-free drivers and drivers with two or more accidents. If the DPE has criterion-related validity, the accident-free group would be expected to have a significantly lower fail rate than would the accident-involved group. Analysis of covariance (ANCOVA) was used to test whether the differences in group fail rates were statistically significant. This statistical technique enabled the group fail rates to be statistically adjusted to partial out the effects of other variables (or covariates), in this case age, sex, and LRE. The adjusted means reflect what the relative performance of the two groups would be expected to be had all subjects been of the same age and sex and been tested by the same LRE. Because of the small number of accident-repeaters and other limitations in the use of accidents as a criterion measure, the results of the analysis were not expected to be definitive.

A future assessment of the DPE's criterion-related validity will also be made by comparing the driver records subsequent to drive testing for original applicants given the DPE with those for a sample of original applicants given the current drive test in matched northern California offices. The former group will be identified from the Daily

DPE Activity forms collected from the 30 DPE offices. Drivers in the latter group will be selected either from the DL masterfile or from score sheets for the current drive test collected from northern offices before Stage 1. The results of this analysis will be presented in a follow-up report to be completed by September 1998.

Evaluation of DPE modifications. Several possible changes to the DPE were evaluated to determine the effect of each modification on test validity. The validity measure used for this purpose was the correlation of groups status (novice original vs. congratulations renewal) with DPE total score for subjects who completed the DPE (DQs excluded). Total test score under each DPE-change scenario was simulated by counting only errors made on maneuvers that would remain on the test. It was not possible to compute validity coefficients based on pass/fail results, which would have included all subjects, because data indicating at what point in the DPE a DQ occurred were not available.

The following changes to the DPE were assessed:

- (1) eliminate the turn-and-stop skill test;
- (2) eliminate street parking;
- (3) eliminate two intersections (#7 and #8 on score sheet);
- (4) eliminate one left turn (#4) and one right turn (#4);
- (5) eliminate two intersections (#7 and #8), one left turn (#4), and one right turn (#4);
- (6) eliminate straight business and residential items;
- (7) eliminate freeway driving.

The effect of changing the DPE on total fail rate could be estimated only for the turn-and-stop and freeway options due to the inability to match DQs with specific DPE maneuvers on the score sheet. DQ data for the turn-and-stop skill test were available from the DPE score sheets and the Daily DPE Activity forms. The occurrence of DQs on the freeway were determined from LRE comments made on DPE score sheets collected for a future R&D study over 2 weeks during October 1994 in all 30 DPE offices.

RESULTS

DPE Validity

Construct validity. Table 1 presents the number of subjects tested in each validation study office by applicant group. (The counts do not reflect the proportional representation of each group that would be expected for normal operations—i.e., they reflect the LREs' selection of subjects rather than actual workload volumes.) As expected, the number of subjects in the three-accident reexamination and limited-term renewal groups was small relative to the number of subjects in the other groups, especially in West Covina where only one three-accident reexamination subject was tested. Novice originals far outnumbered other subjects, comprising over 2/3 of the total subject count.

Table 1
Number of Subjects in Each Office by First-Attempt Applicant Group

Group	Fullerton	West Covina	Westminster	Total
Novice original	535	720	1,095	2,350
Out-of-state original	68	53	85	206
Congratulations renewal	107	129	107	343
Written-test renewal	104	167	142	449
3-accident reexamination	16	1	12	29
Limited-term renewal	44	17	67	128
Total	910	1,087	1,508	3,505

The number of subjects who were selected for the study but were not tested due to their refusing to participate or wanting to be rescheduled for testing on another day is unknown. The study procedures called for LREs to select study subjects and record information on any refusals or rescheduled subjects on the Daily DPE Activity form. However, post-study comments made by employees in the study offices indicated that this procedure was not followed. It appears that the selection of subjects was made by technicians at the DL window, who then referred willing subjects to the LRE. Therefore, refusals and reschedules could not be accounted for.

Table 2 presents the number of subjects, DPE fail rate, and mean DPE point score (number of errors) for each group in the three offices combined. The performance measures are shown graphically in Figures 1 and 2, respectively. Only subjects who completed the DPE (i.e., who were not DQed) were included in the computation of mean DPE score. The 0.2% of subjects who DQed on the pre-drive check list were excluded from both analyses.

Table 2
Number of Subjects (*n*), Fail Rate, and Mean Score
by First-Attempt Applicant Group

Group	<i>n</i>	Fail rate (%)	Mean score
Novice original	2,350	52.6	9.9
Out-of-state original	206	31.6	8.3
Congratulations renewal	343	14.9	7.5
Written-test renewal	449	24.1	8.3
3-accident reexamination	29	31.0	12.1
Limited-term renewal	128	48.4	10.4
Total	3,505	43.7	9.2

Note. Groups differed significantly on fail rate ($F = 59.97, p < .001$, two-tailed) and point score ($F = 21.35, p < .001$, two-tailed).

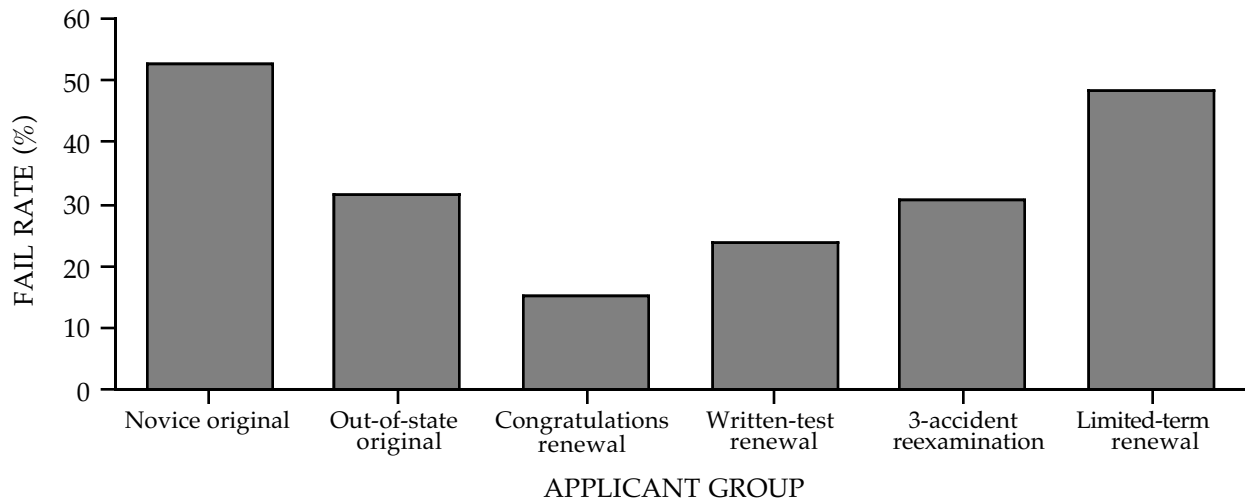


Figure 1. DPE fail rate by first-attempt applicant group.

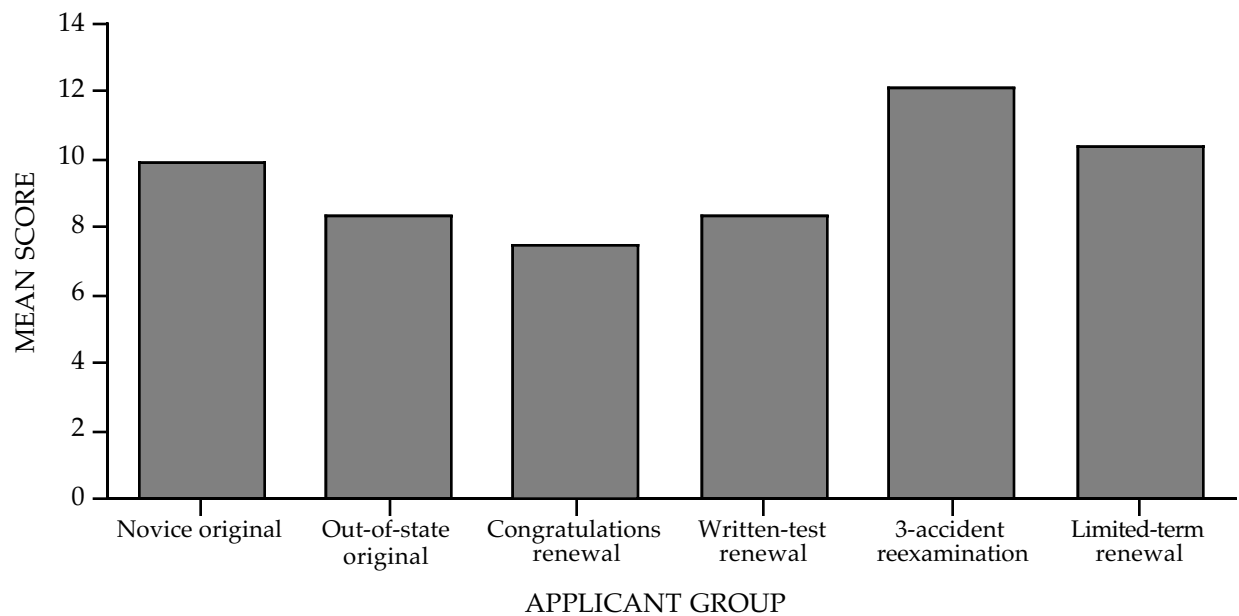


Figure 2. DPE mean score by first-attempt applicant group (DQs excluded).

ANOVA results indicate that the group differences on fail rate were statistically significant ($p < .001$). Results of Tukey post-hoc significance tests (familywise alpha = .05) indicate that the novice original and limited-term renewal groups had significantly higher fail rates than did the out-of-state original, congratulations renewal, and written-test renewal groups, and that the out-of-state original group had a significantly higher fail rate than did the congratulations renewal group.

Results of the ANOVA conducted on point scores indicate the groups also differed significantly on average score ($p < .001$). Tukey post-hoc tests found mean scores for the novice original, three-accident reexamination renewal, and limited-term renewal groups to be significantly higher than the mean scores for the out-of-state original, congratulations renewal, and written-test renewal groups. This pattern of differences generally parallels the results from the fail rate analysis reported above.

Two supplemental three-way factorial ANOVAs were conducted to determine whether the pattern of group differences on fail rate and mean point score was the same for different study LREs in different offices. The three-accident reexamination group and the limited-term renewal group were excluded from these analyses due to the small number of these subjects tested by each LRE. The factors or independent variables in each ANOVA were applicant group, office, and. (Only the main effect of group, and the interaction effects that included group as a factor, are of interest here because the other effects represented in the factorial design—the main effects of office and LRE within office—do not directly address the question of the DPE's validity. In addition, effects averaged across applicant groups are not meaningful because the proportional representation of each applicant group was different for each office and LRE.)

Table 3 presents DPE fail rate and number of subjects by office, LRE, and applicant group. The nested factorial ANOVA results are presented in Table 4. The main effect of group was statistically significant ($p < .001$), indicating significant differences in group fail rates averaged across offices and LREs (which essentially duplicates results from the one-way ANOVA presented above). The Group $\times$ Office interaction was not significant ($p = .20$). The Group $\times$ LRE interaction within office was significant ($p < .05$), indicating that at least one pair of LREs (within office) differed in their patterns of group fail rates. However, the LREs' general pattern of scoring, as reflected in Figure 3, clearly supports the validity of the DPE. In particular, all LREs indicated that the novice original group performed worse on the DPE than did both renewal groups.

Table 3
Fail Rate (%) and Number of Subjects (in Parentheses) by Office,
LRE, and First-Attempt Applicant Group

Group	Fullerton		West Covina		Westminster	
	LRE 1	LRE 2	LRE 1	LRE 2	LRE 1	LRE 2
Novice original	69.9 (286)	53.4 (249)	52.6 (352)	71.2 (368)	33.8 (515)	48.5 (580)
Out-of-state original	38.7 (31)	18.9 (37)	35.0 (20)	42.4 (33)	10.0 (10)	32.0 (75)
Congratulations renewal	15.6 (64)	23.3 (43)	8.4 (83)	30.4 (46)	8.2 (61)	10.9 (46)
Written-test renewal	39.2 (51)	28.1 (89)	24.1 (87)	27.5 (80)	19.2 (47)	11.6 (95)

Table 4
Fail Rate ANOVA Results

Source of variation	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Group	3	22.72	106.59	.000
Office	2	2.59	12.16	.000
LRE within office	3	1.18	5.54	.001
Group x Office	6	0.30	1.43	.200
Group x LRE within office	9	0.45	2.11	.025
Error within cells	3324	0.21		

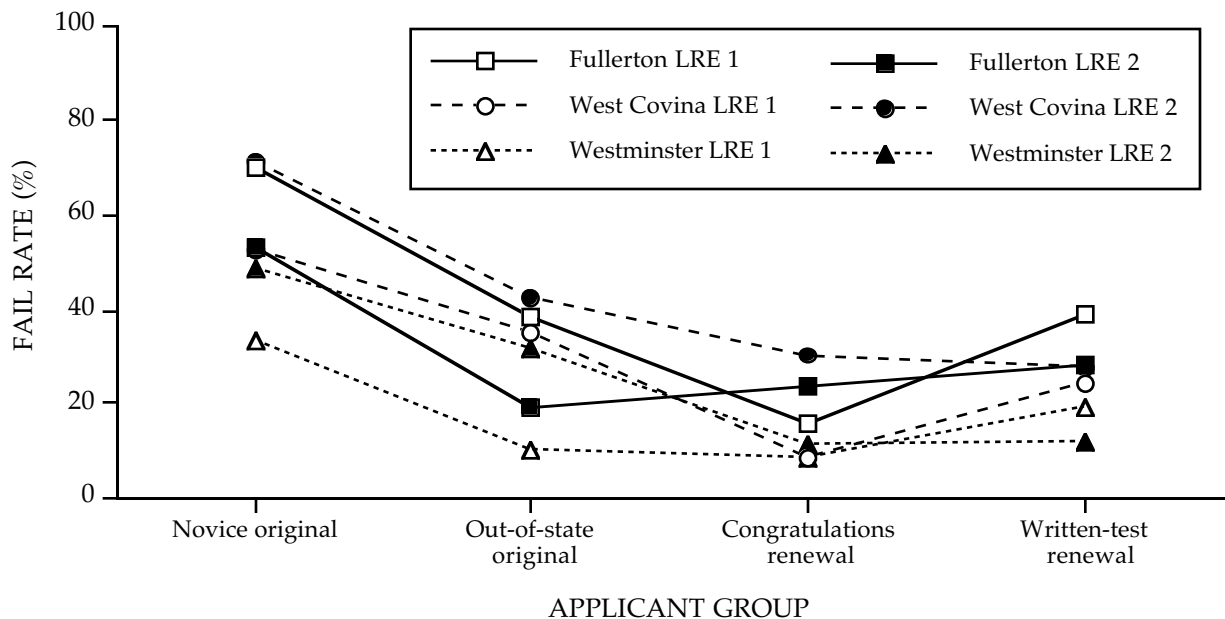


Figure 3. DPE fail rate by office, LRE, and first-attempt applicant group.

Table 5 presents mean DPE point score and number of subjects by office, LRE, and applicant group for subjects in each group who completed the DPE (DQs excluded). Table 6 presents the results of the nested factorial ANOVA conducted on point scores. The main effect of group on point score was statistically significant ($p < .001$). However, the significant Group x LRE interaction effect within office ($p < .05$) indicates that the pattern of differences between group means underlying the main effect of group does not represent the scoring of every LRE individually. This interaction effect is evidenced

in Figure 4, which shows the group mean for each LRE. However, as was the case for fail rates, their general pattern of scoring clearly supports the validity of the DPE. In particular, all LRE's mean scores indicated that the novice original group performed worse on the DPE than did both renewal groups.

Table 5

Mean Score and Number of Subjects (in Parentheses) by Office, LRE, and First-Attempt Applicant Group (DQs Excluded)

Group	Fullerton		West Covina		Westminster	
	LRE 1	LRE 2	LRE 1	LRE 2	LRE 1	LRE 2
Novice original	12.8 (122)	12.3 (154)	9.8 (188)	12.1 (107)	7.6 (349)	9.4 (331)
Out-of-state original	10.9 (22)	10.1 (34)	3.6 (13)	9.3 (21)	6.1 (9)	7.2 (51)
Congratulations renewal	11.4 (61)	10.0 (41)	4.5 (77)	7.8 (33)	4.4 (56)	8.6 (42)
Written-test renewal	11.2 (35)	11.0 (81)	5.6 (67)	9.1 (58)	6.1 (38)	7.2 (84)

Table 6

Mean Score ANOVA Results

Source of variation	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Group	3	974.56	58.79	.000
Office	2	1666.91	100.55	.000
LRE within office	3	460.77	27.80	.000
Group x Office	6	91.46	5.52	.000
Group x LRE within office	9	31.31	1.89	.049
Error within cells	2052	16.58		

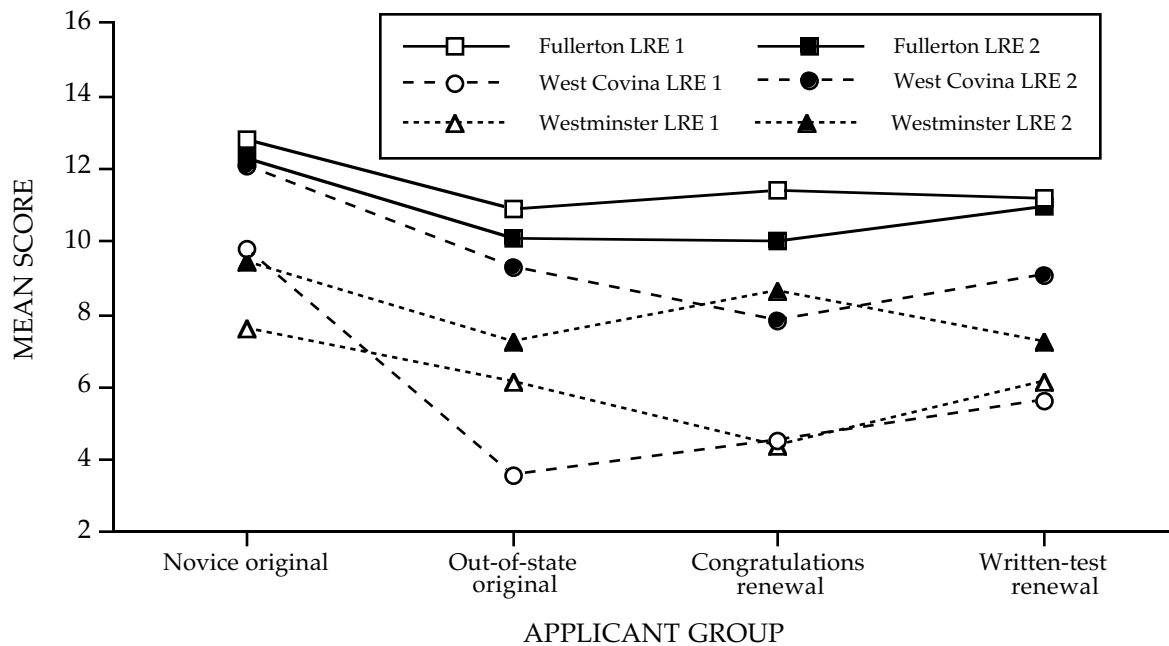


Figure 4. DPE mean score by office, LRE, and applicant group (DQs excluded).

Table 7 presents the correlations of DPE pass/fail result and DPE point score with applicant status for each validation study LRE. DQs were excluded from the computation involving point scores. Applicant status was coded 0 for novice originals and 1 for congratulations renewals. Correlations involving pass/fail result ranged from -.17 to -.43, and averaged -.27. All of these coefficients were statistically significant ($p < .01$). The point score correlations fluctuated more widely, from -.06 to -.52, and averaged -.30. All but two of these correlations were significant ($p < .05$). These results indicate that novice originals tended to fail the test more often and make more errors than did congratulations renewals, and thus validate the DPE.

The correlations of item scores with group status for first-attempt subjects completing the test (DQs excluded) are presented in Table C-1 in the Appendix. A positive correlation indicates that novice originals tended to perform better on the item than did congratulations renewals. Items without a correlation coefficient entered in the table did not have enough variance (i.e., were failed by too few applicants) to enable this measure to be computed.

Table 7
Number of Subjects (*n*) and Correlation (*r*) of Pass/Fail Result and
Total Point Score with Applicant Group by LRE

Office LRE	Pass/fail		Point score	
	<i>n</i>	<i>r</i>	<i>n</i>	<i>r</i>
Fullerton				
LRE 1	350	-.43**	183	-.13
LRE 2	292	-.21**	195	-.18*
West Covina				
LRE 1	435	-.35**	265	-.51**
LRE 2	414	-.27**	142	-.52**
Westminster				
LRE 1	576	-.17**	405	-.37**
LRE 2	626	-.20**	373	-.06
Total (avg.)	2,693	-.27**	1,563	-.30**

Note. Total (average) correlations are not weighted to reflect differences in within-LRE sample sizes.

* $p < .05$. ** $p < .01$.

Additional support for the validity of the DPE is provided by data reported on the Daily DPE Activity forms collected from the 30 DPE offices during September 1994 through February 1995, excluding data collected from the six study LREs. (In September and October only 29 offices reported, in November 27 offices reported, in December and January all offices reported, and in February 28 offices reported.) Table 8 presents the number of subjects, fail rate, and mean DPE score (DQs excluded) for each first-attempt applicant group identified on the forms. The physical/mental group consisted of applicants who were identified as having a driving-related physical or mental condition that warranted their being given a regular (rather than special) drive test. Drivers in the written test failure group were renewals who had to take a drive test because they failed the written test three times.

ANOVA results indicate the differences in fail rates to be statistically significant ($p < .001$). Tukey post-hoc tests found that the renewal group, which consisted mostly of California drivers with expired licenses, had a significantly lower fail rate than did all the other groups. The post-hoc tests also found that the out-of-state original group had a significantly lower fail rate than did the novice original or limited-term renewal groups, and that the novice original and written test failure renewal groups had significantly lower fail rates than did the limited-term renewal group.

Table 8
Number of Applicants (*n*), Fail Rate, and Mean Score by
First-Attempt Applicant Group (30 Offices)

Group	<i>n</i>	Fail rate (%)	Mean score
Novice original	18,942	45.9	10.8
Out-of-state original	900	36.2	9.1
Limited-term renewal	717	57.9	12.1
Physical/mental renewal	151	49.0	12.4
Written-test failure renewal	352	41.5	11.6
License-expired renewal	169	23.1	9.2

Note. Groups differed significantly on fail rate ($F = 22.88$, $p < .001$, two-tailed) and point score ($F = 24.06$, $p < .001$, two-tailed).

The groups also had significantly different mean scores ($p < .001$). Tukey post-hoc tests found that the renewal and out-of-state original groups had significantly lower mean scores than did all other groups, and that the novice original group had a significantly lower mean score than did the limited-term and physical/mental renewal groups ($p < .05$).

Ultimate criterion validity. Table 9 presents demographic information and the observed and statistically-adjusted DPE fail rates for accident-free and accident-involved renewal and three-accident reexamination subjects. The two groups were not significantly different on mean age ($p = .08$), proportion of women ($p = .78$), or fail rate ($p = .31$).

The ANCOVA results showed age and LRE to be significant covariates ($p < .01$), but not sex ($p = .12$). Homogeneity of slopes tests showed that the contrasted groups did not have different slopes on any of the covariates, therefore a common-slopes ANCOVA model was selected. The adjusted DPE fail rates were farther apart in the direction of validating the DPE than were the unadjusted means, but the difference was still nonsignificant ($p = .17$). Although the results fail to substantiate the validity of the test in discriminating accident-repeaters from accident-free applicants, the very small sample size for the accident group greatly limited the statistical power for detecting a significant effect. The size and direction of the differences are encouraging.

Table 9
Number of Subjects (*n*), Mean Age, Percentage of Women, and
Observed and Statistically-Adjusted DPE Fail Rates for Accident-Free and
Accident-Involved Renewal and 3-Accident Reexamination Subjects

Group	<i>n</i>	Mean age	% women	Unadjusted fail rate (%)	Adjusted fail rate (%)
Accident-free	654	46.6	45.1	19.7	19.6
Accident-involved	42	41.9	42.9	26.2	28.2

Note. The accident-involved group consisted of drivers with two or more accidents during the 3 years prior to DPE testing. Covariates used to adjust DPE fail rate were age, sex, and LRE. None of the differences were statistically significant ($p > .05$, two-tailed).

General Process Measures

The Daily DPE Activity forms collected from the 30 offices contained data on 36,812 applicants scheduled for a drive test, excluding those tested by the six study LREs. Data were not usable for 2,165 cases because the information was not sufficient to determine whether the DPE was administered. Of the remaining 34,647 applicants for which data were usable, 3,175 (9.2%) had the DPE postponed or rescheduled. Table 10 lists the reported reasons for not giving the DPE on the scheduled day and the number and percentage of subjects associated with each one. Nearly two-thirds of the cases were due to the applicant's not having auto insurance or to vehicle mechanical failure. The table also indicates that, of the 31,472 applicants who did take the test, 208 (less than 1%) were not tested on the freeway. Only six of the freeway-waiver cases were due to inability to test on the freeway due to a traffic accident, congestion, or freeway closure; the remainder were due to the applicant's stating that they never drive on the freeway and consequently may have been given a no-freeway license restriction. (An additional 105 applicants given the DPE may not have been tested on the freeway, but this could not be determined with certainty because of ambiguity of data reported for these cases.)

Table 10

Number (*n*) and Percentage of Applicants with Usable Data
Reported on the Daily DPE Activity Form by DPE Test Status (30 Offices)

Test status	<i>n</i>	% of total (<i>N</i> = 34,647)
DPE postponed / rescheduled	<u>3,175</u>	9.2
no insurance	1,059	3.1
vehicle mechanical failure	884	2.6
not ready for freeway	251	0.7
no driver training	119	0.3
no registration	102	0.3
no accompanying driver	54	0.2
other	139	0.4
DPE administered	<u>31,472</u>	90.8
not tested on freeway	<u>208</u>	0.6
no-freeway restriction	202	0.6
accident / congestion / closure	6	0.0

Note. Data were provided for a total of 36,812 driver license applicants, excluding those tested by the six study LREs. However, data were usable for only 34,647 of these cases because information provided for the other 2,165 did not indicate whether the DPE was administered.

DPE difficulty. Table 11 presents the DPE results for all 31,159 applicants tested in the 30 DPE offices during the general data collection periods from September 1994 through February 1995 (excluding applicants tested by the six validation study LREs). Data were collected from September 1994 through June 1995, however only data for September

1994 through February 1995 were used to compute the descriptive measures due to key entry delays and time constraints on data analysis. The overall fail rate was 45.5%. The vast majority of the fails were due to DQs, which accounted for 38.8 percentage points of the total fail rate. A total of 11.6% of all applicants were DQed on the turn-and-stop skill test and, consequently, were not tested on the road. Only 6.7% of all applicants failed because of high point (error) scores.

Table 11

DPE Results for All Applicants (30 Offices)

DPE result	Number of applicants	% of total (N = 31,159)
Total fails	<u>14,182</u>	<u>45.5</u>
Point-score fails	2,085	6.7
DQ fails	<u>12,097</u>	<u>38.8</u>
DQ pre-drive	70	0.2
DQ turn-and-stop	3,609	11.6
DQ on road	8,118	26.1
DQ unidentified	300	1.0

Note. Applicants tested by the six designated LREs involved in the validation study are not represented.

The percentage of applicants who failed the DPE in each of the 30 offices is presented in Table C-2 in the Appendix. The table also shows the fail rate in each office for the current drive test. The latter estimates are based on test results recorded on the current drive test score sheets that were collected in May 1994 (just before DPE implementation) to capture baseline test time data. Within-office fail rate ranged from 32.0% to 53.3% for the DPE and from 17.7% to 36.3% for the current drive test. To compare the fail rates for the DPE and current drive test across offices, a normalized fail rate for the current test was computed based on the within-office frequencies for the DPE. The normalized fail rate is the average fail rate that would be expected if the same number of current drive tests as the number of DPEs were administered in each office. The San Clemente office was excluded from the comparison because it did not submit current drive test score sheets to R&D. DPE fail rate for the 29 reporting offices was 45.6%. The current drive test's normalized fail rate was 26.2%. These results represent a fail rate increase of 19.4 percentage points following implementation of the DPE.

The percentage of total applicants who were DQed on the turn-and-stop skill test in each of the 30 offices is presented in Appendix Table C-3. The estimates ranged from 3.5% to 18.3%. (Laguna Hills [0.9%] substituted a 3-point turn for the turn-and-stop skill test.)

Table 12 presents DPE results for novice original first-attempt applicants only. These figures better reflect the inherent difficulty of the DPE because they reflect how well inexperienced drivers performed on their first try. The results are very similar to those presented in Table 11. The total fail rate was 45.8%, with 38.9 of the percentage points being due to DQs. The rate of DQs on the turn-and-stop skill test was 13.2%.

Table 12
DPE Results for First-Attempt Novice Originals (30 Offices)

DPE result	Number of applicants	% of total (N = 18,522)
Total fails	<u>8,481</u>	<u>45.8</u>
Point-score fails	1,278	6.9
DQ fails	<u>7,203</u>	<u>38.9</u>
DQ pre-drive	35	0.2
DQ turn-and-stop	2,451	13.2
DQ on road	4,561	24.6
DQ unidentified	156	0.8

Note. Applicants tested by the six designated LREs involved in the validation study are not represented.

Table 13 presents the fail rate and number of subjects by applicant group and test attempt number for the 30 offices (excluding applicants whose applicant group or attempt number could not be identified). Novice originals had about the same fail rate for all drive test attempts. However, the fail rates for other groups tended to increase on successive attempts. (The third-attempt fail rates for physical/mental and renewal applicants should be interpreted with caution because of the small sample sizes for these groups.) It should be noted that some of the second and third drive test attempts may actually have been the first attempt on the DPE. This should not have substantially biased the results because the DPE had been implemented for several weeks before the data collection started and, consequently, the number of drive tests attempts that were not DPEs was expected to be small. To determine the extent to which the results may have been biased by the inclusion of nonDPE drive tests, the fail rates for the three test attempts made by novice originals were calculated for February 1995 only, more than 6 months after DPE testing had begun. The results were nearly identical to those for all months combined.

Table 13

Fail Rate (%) and Number of Applicants (in Parentheses)
by Group and Attempt Number (30 Offices)

Applicant group	Test attempt		
	1st	2nd	3rd
Novice original	45.9 (18,724)	44.8 (6,224)	46.3 (1,909)
Out-of-state original	36.2 (900)	41.9 (301)	50.0 (90)
Limited-term renewal	57.9 (717)	58.8 (369)	64.1 (142)
Physical/mental renewal	49.0 (151)	48.1 (54)	56.3 (16)
Written test failure renewal	41.5 (352)	45.6 (114)	46.9 (49)
License-expired renewal	23.1 (169)	44.4 (27)	54.5 (11)

Table 14 and Figure 5 present the DPE results for novice original first-attempt applicants by month. The general trend appears to be a decrease in total fail rate over time. This decrease parallels the decrease that occurred in the turn-and-stop DQ rates. On-road DQ rate appears to have remained reasonably steady.

Table 14

Total Fails, On-Road DQs, Turn-and-Stop DQs, and Point-Score Fails
as Percentages of all Novice Original First-Attempt Applicants
by Month for September 1994 through February 1995

Month	Total fail rate	On-road DQ rate	Turn-and-stop DQ rate	Point-score fail rate
September	49.1	25.4	15.3	7.6
October	46.2	24.6	14.0	6.3
November	44.6	23.7	12.4	7.7
December	45.5	24.6	13.3	6.1
January	44.7	23.9	13.4	6.5
February	44.4	25.2	11.4	7.0

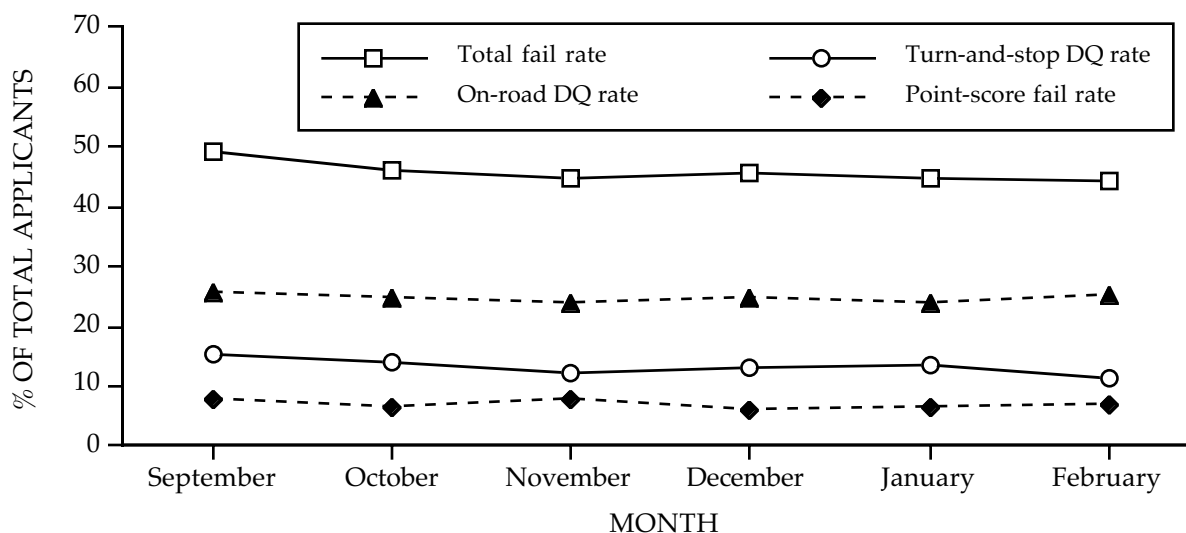


Figure 5. Total fails, on-road DQs, turn-and-stop DQs, and point-score fails as percentage of all novice original first-attempt applicants by month for September 1994 through February 1995.

DPE test time. The average reported test time was 24.7 minutes for all DPEs given in the 30 offices (excluding DPEs administered by the six LREs in the validation study). The average time reported for completed DPEs (DQs excluded) was 28.4 minutes. As would be expected, the average test time for completed DPEs was greater than that for DPEs with DQ rates excluded because DQs frequently occur early in the test.

Table C-4 in the Appendix presents the mean reported test times for the DPE and the department's current drive test for each of the 30 DPE field offices. Within-office mean test time ranged from 20.5 to 28.6 minutes for the DPE and from 10.6 to 18.1 minutes for the current drive test.

To compare the mean test time for the DPE and current drive test across offices, a normalized mean time for the current test was computed based on the within-office test frequencies for the DPE. The San Clemente field office was excluded from the analysis because it did not report any data on current drive test time. The mean DPE test time for the 29 reporting offices was 24.8 minutes. The normalized mean time computed for the current drive test was 13.6 minutes. These results represent an average increase of 11.2 minutes per drive test.

Figure 6 shows reported mean DPE test time by month. Mean test time increased by almost 1 minute over the 6 months. The reduction in turn-and-stop DQ rate over the

same time period (See Figure 5) may have contributed to this reduction in total test time, since the former trend would have been associated with an increase in the percentage of applicants tested on the road.

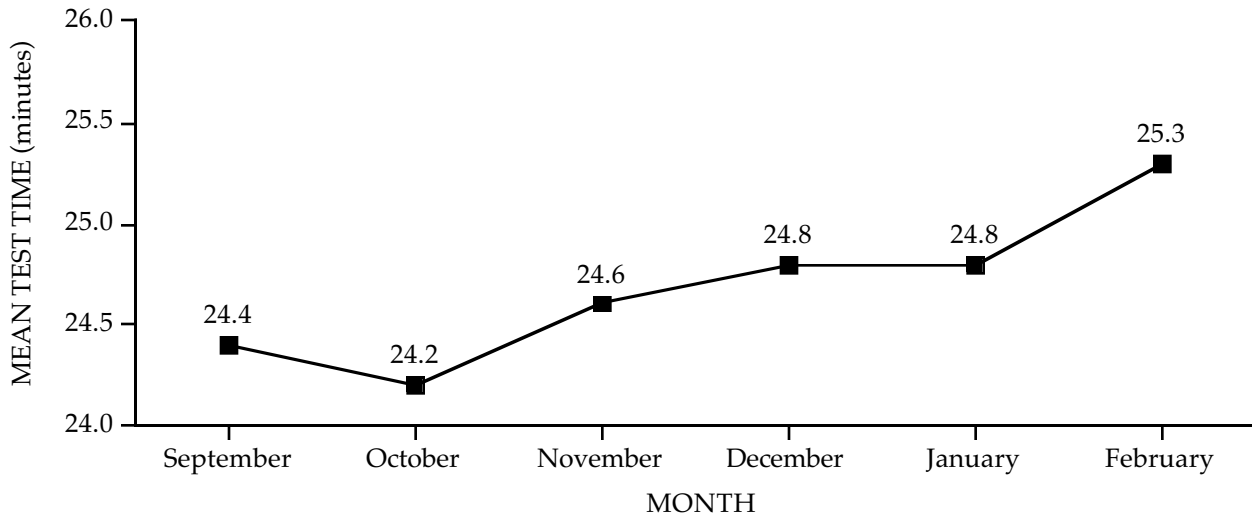


Figure 6. DPE mean test time by month for September 1994 through February 1995.

Evaluation of DPE Modifications

Table 15 presents the correlation of group status with DPE point score for each DPE-modification option (including making no changes) by LRE. Eliminating freeway driving from the DPE had the largest impact on validity, significantly reducing the average coefficient for all LREs from $-.30$ to $-.25$ ($p < .001$). The loss in validity caused by dropping the freeway maneuvers tended to occur regardless of which LRE did the testing. The effect on the validity of the other DPE modifications was much smaller but still statistically significant ($p < .05$). The one exception was the removing of two intersections, which did not have a significant effect on validity ($p > .10$). One of the options—eliminating street parking—slightly increased the average validity correlation coefficient, from $-.30$ to $-.32$. Had DQs been included in the simulations, the decrease in validity produced by each change option would have been greater.

Table 15

Number of Subjects (*n*) and Correlation of Total Point Score
with Group Status by LRE for Various DPE Modifications (DQs Excluded)

Office LRE	<i>n</i>	DPE modification (what to eliminate)							
		No change	Turn- and- stop	Street park	2 inter- sections	1 L turn 1 R turn	1 L turn 1 R turn 2 inter- sections	Bus./ resid.	Free- way
Fullerton									
LRE 1	183	-.13	-.08	-.19**	-.14	-.11	-.12	-.12	-.09
LRE 2	195	-.18*	-.18*	-.21**	-.16*	-.20**	-.17*	-.17*	-.14
West Covina									
LRE 1	265	-.51**	-.51**	-.51**	-.48**	-.48**	-.46**	-.51**	-.48**
LRE 2	142	-.52**	-.51**	-.54**	-.54**	-.48**	-.50**	-.52**	-.45**
Westminster									
LRE 1	405	-.37**	-.37**	-.39**	-.36**	-.36**	-.36**	-.34**	-.32**
LRE 2	373	-.06	-.04	-.07	-.05	-.07	-.07	-.05	-.02
Total (avg.)	1,563	-.30**	-.28**	-.32**	-.29**	-.28**	-.28**	-.28**	-.25**

Note. Total (average) correlations are not weighted to reflect differences in the number of subjects tested by each LRE.

* $p < .05$. ** $p < .01$.

Eliminating the turn-and-stop skill test would be expected to reduce the DPE fail rate by a maximum of 12 percentage points, which represents the percentage of all applicants tested in the 30 offices who were identified on the Daily DPE Activity forms as having been DQed on one of the turn-and-stop maneuvers. The reduction in fail rate that would actually result from this DPE modification, however, would likely be much smaller than this because applicants who were DQed on the turn-and-stop would probably have made a DQ error somewhere else on the test, or failed based on point score, if they had been permitted to continue testing on the road.

Eliminating freeway driving from the DPE would be expected to reduce fail rate by at least five percentage points. This represents the percentage of applicants who were identified from LRE comments recorded on the score sheet as having been DQed due to an error made on the freeway. (This is considered a lower bound estimate because some freeway DQs were probably not explicitly identified as such on the DPE score sheet.)

DISCUSSION AND RECOMMENDATIONS

The results provide strong evidence that the DPE is a valid test. Inexperienced drivers and drivers with mental or physical conditions that affect their driving tended to perform worse on the DPE than did experienced drivers with no known driving-related debility. The results show that the novice original and limited-term renewal groups had significantly higher fail rates than did the out-of-state original, congratulations renewal, and written-test renewal groups. Comparisons of the point scores (number of errors) yielded similar results. Novice original, limited-term renewal, and three-accident reexamination driver groups had significantly higher mean scores than did the out-of-state original, congratulations renewal, and written-test renewal groups. In addition, the finding that all LREs consistently scored novice originals as being worse drivers than both renewal groups evidences the DPE's validity even at the individual examiner level.

The correlational analyses also confirmed the validity of the DPE. The significant correlation of applicant status (novice original vs. congratulations renewal) with pass/fail result and with point score indicate that novice originals failed the test more often and made more errors than did congratulations renewals. Although there was a wide range in the within-LRE correlation coefficients, the LREs were consistent in scoring the congratulations renewal group better than the novice original group.

Analyses of data collected from the 30 DPE offices yielded further evidence of the validity of the DPE. The results show that license-expired renewals had significantly lower fail rates than did all other applicant groups, out-of-state originals had a significantly lower fail rate than did novice originals and limited-term renewals, and novice originals and written test failure renewals had lower fail rates than did limited-term renewals. The same pattern of ranking was found for mean scores, but with fewer distinctions: The license-expired renewal and out-of-state groups had significantly lower mean scores than did all other groups, and the novice original group had a significantly lower mean score than did the limited-term renewal and physical/mental renewal groups. It should be noted that the validation results may be biased due to LRE's having known the status of subjects prior to testing, which may have influenced their scoring. Any such criterion contamination would be expected to make the DPE look more valid than it actually is.

The ultimate criterion analysis did not find a significant difference in either the unadjusted or adjusted fail rates for accident-free and accident-involved renewals and 3-accident reexamination subjects. However, the direction of the difference in group means was consistent with the DPE being a valid test. The failure to detect a significant difference was not surprising considering that statistical power (the probability of finding a significant effect) was only 17% due to the small sample size in the accident-

involved group. In addition, accident involvement per se is not a good indicator of level of driving ability, which is what the DPE is intended to measure. The likelihood of having an accident not only depends upon level of driving competency, but also exposure variables such as the amount and conditions of driving. Had the accident measure used to establish the contrasted groups been adjusted for these exposure variables, the probability of finding a significant difference between the contrasted groups would have been increased.

Information collected from the 30 offices revealed the operational impact of the DPE. Implementation of the DPE substantially increased both the length and the difficulty of the drive test. Fail rate increased from 26.2% for the current drive test to 45.6% for the DPE. This 19.4 percentage-point gain in fail rate represents over a 70% increase in the total proportion of applicants who failed. Average test time increased by 11.2 minutes, from 13.6 minutes for the current test to 24.8 minutes for the DPE. These increments in fail rate and test time translate into a large increase in the cost of administering the drive test.

The finding that DPE fail rate for original applicants decreased slightly during the first 6 months of the pilot (from 49.1% in September 1994 to 44.4% in February 1995) suggests that, over time, applicants became better prepared for the DPE. It is uncertain whether this decline will continue.

The results of the assessment of possible DPE modifications that would reduce test length showed that eliminating freeway driving was the only option that substantially impacted test validity, reducing the correlation of point score with group status from $-.30$ to $-.25$. All the other changes slightly decreased validity, except eliminating street parking, which slightly increased the correlation coefficient. These results should be interpreted with caution because the most incompetent subjects (DQs) were not included in the simulations and it is uncertain whether the findings can be generalized to them. It is also possible that the simulation results do not accurately represent what would occur had shorter versions of the DPE actually been administered (all subjects in the simulations having completed the full DPE). In spite of these limitations, the relatively large drop in the validity coefficient produced by eliminating freeway driving provides strong evidence that not testing drivers on the freeway is the worst of all of the options considered. The fact that dropping freeway driving would greatly reduce the DPE's face validity also argues against that option.

If the DPE had to be shortened to reduce administration costs, consideration should be given to the expected time savings that would be produced by each modification, which was not assessed in this study. Dropping the turn-and-stop skill test would almost certainly save more time than any of the other options (excluding dropping freeway driving). The time that would be saved by eliminating intersections and turns would probably be small, although it would vary according to the route. Dropping the

straight business and residential items may save a little more time than would dropping turns and intersections, but the savings would probably still be marginal.

Another factor to be considered before changing the DPE is the effect of the modification on test fail rate, which would impact the cost of retesting applicants. Eliminating the skill test would reduce fail rate by at most 12 percentage points (the percentage of applicants DQed on the skill test), however the actual net increase would probably be much lower than this because many of those who DQ on the turn-and-stop would probably also have failed the on-road portion of the test. Not testing on the freeway would reduce fail rate by a maximum of five percentage points, the percentage of applicants identified from LRE comments on the DPE score sheets to have been DQed on the freeway. None of the other change options could be evaluated for their effect on fail rate due to the inability to match DQs with specific maneuvers on the DPE route.

REFERENCES

- Hagge, R. A. (1994). *The California driver performance evaluation project: An evaluation of a new driver licensing road test* (Report No. 150). Sacramento: California Department of Motor Vehicles.
- Mackie, R. R., Wylie, C. D., Shultz, T., Engel, R., Townsend, M., Lammlein, S. E., & Johnson, S. (1989). *Development of a recommended testing program for commercial motor vehicle operators (the CDL system)*. Lincoln, NE: Nebraska Department of Motor Vehicles and Washington DC: American Association of Motor Vehicle Administrators.
- Shumaker, N. C. (1994). *The California driver performance evaluation project: An evaluation of the current driver licensing road test* (Report No. 151). Sacramento: California Department of Motor Vehicles.
- Williams, R. L., & Shumaker, N. C. (1994). *Class C drive test baseline study: Preliminary report*. Sacramento: California Department of Motor Vehicles.

Appendix A

SUPPLEMENTAL/AREA DRIVING PERFORMANCE EVALUATION SCORE SHEET

[illegible]

Appendix B

DAILY DPE ACTIVITY

Scheduled or Walk-in Drive Test Applicants

Field office number: _____

Today's date (mmddyy) _____

LRE number _____

DL window number _____

[illegible]

Appendix C

Table C-1

Correlation of Item Score and Group Status (Novice Original vs. Congratulations Renewal) for First-Attempt Subjects Completing the DPE (DQs Excluded) in Fullerton, West Covina, and Westminster ($N = 1,563$)

TURN AND STOP				E	X				
Traffic check				-.11*	-.11*				
Speed				-.02	-.02				
Braking				-.01	-.01				
Vehicle position				.00	-.04				
PARKING LOT DRIVING				1	2				
Traffic check				-.15*	-.09*				
Speed				-.00	-.03				
STREET PARK				E	X				
Traffic check				.04	.05*				
Signal				.07*	-.01				
Speed				-.04	.01				
Parking				-.02	.01				
Parallel				-.10*					
INTERSECTIONS	1	2	3	4	5	6	7	8	
Through									
Traffic check	-.01	-.10*	-.08*	-.10*	.02	-.03	.01	.04	
Speed	-.06*	.07*	.05	-.01	.01	.03	-.03	.03	
Unnecessary stop	.05*	-.02	-.02	-.02	-	-	-.02	-.01	
Stop									
Traffic check	-.03	-	-.04	-.05	-.07*	-.03	-.04	-.03	
Speed	-.03	-.01	.05*	.05*	-.03	.03	-.02	-.01	
Full stop	-.05*	-.01	-.06*	-.03	-.02	-	.01	-.02	
Gap or limit line	.05	.02	.03	.07*	-.07*	.03	-.03	-.04	
Start									
Traffic check	-.07*	-.06*	-.04	-.03	-.03	-.03	-.06*	-.10*	
Yield	.00	-	-.02	-.05	-.06*	-	-	-.01	
Speed	-.01	-	-.01	-.01	.03	-	-	-.02	

(Continued on next page)

Table C-1 (Continued)

TURNS		LEFT				RIGHT			
	1	2	3	4	1	2	3	4	
Approach									
Traffic check	-.04	-.01	-.05*	-.02	-.02	.05	-.05	-.07*	
Signal	.05*	-.02	-.04	-.01	.01	-.01	.01	-.02	
Speed	.01	-.02	-.01	.02	-.02	-.03	.01	.06	
Lane	-.03	-.01	-.02	-.03	-.01	-.04	-.03	-.03	
Unnecessary stop	-.09*	-.01	-.04	-.03	.01	-.07*	-.04	-.07*	
Stop									
Traffic check	-.02	-.02	-.02	-.02	-.01	-.01	-.02	-.02	
Speed	-	-	-.01	-	-.01	-.01	-	-.01	
Full stop	-.01	.01	.02	-.02	-.04	-.04	-.05*	-.03	
Gap or limit line	.08*	.04	.02	.01	.05	-.03	.03	.03	
Wheels straight	-.06*	-.03	-.05	-.04					
Turn/Complete									
Traffic check	-.05*	-.03	-.06*	-.07*	-.05*	-.02	-.06*	-.03	
Steering control	-.04	-.03	-.02	-.05*	-.02	-.03	-.02	-.02	
Too wide/short	-.03	-.01	-.05*	-.07*	-.01	.01	-.04	.03	
Correct lane	-.02	-.02	-.02	-.02	-	-.04	.01	.07*	
Speed	-.04	-.03	-.03	-.04	-.02	-.03	-.03	-.03	
Signal	-	.05*	-	-	-	-	-	-.01	
STRAIGHT BUSINESS AND RESIDENTIAL									
	B	R							
Traffic checks	-.12*	-.13*							
Speed	-.06*	.02							
Spacing	-.04	.01							
Lane position	-.03	.03							
LANE CHANGE									
	L	R							
Traffic check	.14*	.06*							
Signal	-.01	-.01							
Speed	-.05*	-.02							
Spacing	-.04	-.04							
Steering control	-.02	-.04							

(Continued on next page)

Table C-1 (Continued)

FREEWAY	
Entering	
Traffic check	-.15*
Signal	-.02
Speed	-.06*
Spacing	-.01
Lane position	-.02
Merge	
Traffic check	.01
Signal	-.04
Speed	-.15*
Spacing	-.05*
Lane position	-.08*
Steering control	-.06*
Lane Use	
Traffic check	-.10*
Speed	-.02
Spacing	-.05*
Lane position	.03
Exiting	
Traffic check	-.09*
Signal	-.04
Speed	-.06*
Spacing	-.03
Lane position	-.05*
Steering control	.01
CURVE	
Entering speed	-.02
Through speed	-.04
Lane position	-.03

Table C-2
 Number of Subjects (*n*) and Fail Rate for the DPE and
 Current Drive Tests by Field Office

Office	Current drive test		DPE	
	<i>n</i>	Fail rate (%)	<i>n</i>	Fail rate (%)
Arleta	89	18.0	1,655	40.7
Bell Gardens	97	26.8	1,998	43.9
Bellflower	91	28.6	1,654	51.9
Chula Vista	83	21.7	958	52.5
Compton	82	22.0	1,082	41.1
Culver City	76	38.2	666	52.0
Escondido	86	29.1	717	33.9
Fullerton	72	31.9	704	47.6
Glendale	99	20.2	1,780	46.9
Hawthorne	84	22.6	860	33.7
Hollywood	101	34.7	529	44.8
Inglewood	96	17.7	626	32.0
Laguna Hills	86	19.8	1,032	51.6
Lincoln Park	101	29.7	1,166	39.5
Montebello	99	21.2	1,039	44.4
Oceanside	97	24.7	1,339	43.7
Pasadena	170	34.1	1,749	49.3
Placentia	112	21.4	901	47.8
Pomona	92	23.9	738	50.8
Poway	79	22.8	737	48.2
San Clemente	*	*	507	35.1
San Diego	97	26.8	645	45.6
San Diego Clairemont	98	24.5	1,444	49.7
San Pedro	87	21.8	913	44.8
San Ysidro	79	32.9	720	53.3
Torrance	99	30.3	1,033	35.9
Van Nuys	70	24.3	624	52.2
West Covina	92	29.4	1,277	48.6
Westminster	102	36.3	1,250	48.2
Winnetka	75	25.3	957	47.3

*not reported

Table C-3

Number of Subjects (*n*) and Turn-and-Stop
DQ Rate by Field Office

Office	<i>n</i>	DQ rate (%)
Arleta	1,655	8.4
Bell Gardens	1,998	14.3
Bellflower	1,654	20.1
Chula Vista	958	12.3
Compton	1,082	11.6
Culver City	666	3.9
Escondido	717	16.9
Fullerton	704	12.4
Glendale	1,780	13.0
Hawthorne	860	8.0
Hollywood	529	15.3
Inglewood	626	4.2
Laguna Hills*	1,032	0.9
Lincoln Park	1,166	7.2
Montebello	1,039	14.8
Oceanside	1,339	3.5
Pasadena	1,749	12.8
Placentia	901	13.3
Pomona	738	18.3
Poway	737	12.8
San Clemente	737	7.9
San Diego	645	14.3
San Diego Clairemont	1,444	5.6
San Pedro	913	16.3
San Ysidro	720	6.4
Torrance	1,033	9.5
Van Nuys	624	9.3
West Covina	1,277	16.5
Westminster	1,250	20.0
Winnetka	957	6.7

*Laguna Hills adopted a 3-point turn in lieu of the turn-and-stop skill test.

Table C-4
Number of Subjects (*n*) and Mean Test Time (Minutes) for
the DPE and Current Drive Tests by Field Office

Office	Current drive test		DPE	
	<i>n</i>	Mean time	<i>n</i>	Mean time
Arleta	80	13.2	1,605	25.1
Bell Gardens	94	13.4	1,985	24.9
Bellflower	86	12.5	1,619	20.9
Chula Vista	82	14.6	948	26.7
Compton	80	12.0	1,068	23.1
Culver City	70	13.5	663	23.2
Escondido	84	16.5	708	23.5
Fullerton	56	14.6	704	24.1
Glendale	97	13.0	1,768	24.4
Hawthorne	80	12.9	831	27.4
Hollywood	91	11.2	524	25.5
Inglewood	94	18.1	612	28.4
Laguna Hills	79	14.3	1,048	20.5
Lincoln Park	98	11.9	1,140	23.6
Montebello	98	13.3	1,031	25.6
Oceanside	86	13.9	1,322	24.1
Pasadena	163	12.0	1,724	25.7
Placentia	108	13.3	897	26.5
Pomona	85	12.5	720	24.7
Poway	77	14.0	726	22.7
San Clemente	*	*	507	28.6
San Diego	91	13.8	634	24.0
San Diego Clairemont	92	13.0	1,420	22.4
San Pedro	85	10.6	903	24.3
San Ysidro	79	12.3	713	25.4
Torrance	87	16.6	1,024	28.3
Van Nuys	70	15.1	619	28.4
West Covina	91	14.3	1,269	23.9
Westminster	94	15.5	1,240	24.5
Winnetka	67	13.6	936	27.3

*not reported