POST Entry-Level Dispatcher

Selection Test Battery

USER'S MANUAL

1996

POST Entry-Level Dispatcher

Selection Test Battery

USER'S MANUAL

John A. Weiner
Sr. Personnel Selection Consultant

February 1996

CALIFORNIA COMMISSION ON PEACE OFFICER STANDARDS AND TRAINING

Published 1996
Printed February 1996

For information about copies of this publication contact:

POST Media Distribution Center
1601 Alhambra Boulevard
Sacramento, CA 95816
(916) 227-4856

COMMISSIONERS

Devallis Rutledge Chairman	Deputy District Attorney Orange County District Attorney's Office
Manuel Ortega Vice-Chairman	Chief of Police Placentia Police Department
Sherman Block	Los Angeles County Sheriff's Department
Collene Campbell	Public Member
Jody Hall-Esser	Chief Administrative Officer City of Culver City
George W. Kennedy	District Attorney Santa Clara County
William B. Kolender	Sheriff San Diego County Sheriff's Department
Marcel L. Leduc	Sergeant San Joaquin County Sheriff's Department
Ronald Lowenberg	Chief of Police Huntington Beach Police Department
Daniel E. Lungren Ex Officio Member	Attorney General
Raquel Montenegro, Ph.D	Professor of Education
Lou Silva	Officer Oakland Police Department
Dale Stockton	Detective Sergeant Carlsbad Police Department
Rick Terborch	Chief of Police Arroyo Grande Police Department
Norman C. Boehm	Executive Director

PREFACE

This manual provides information regarding a battery of cognitive ability tests developed by the Commission on Peace Officer Standards and Training (POST) for use in selecting candidates for entry-level employment in the public safety dispatcher occupation. The manual is intended to provide guidance to human resource specialists, managers, and others in using Test Battery scores to make employment decisions. Information is presented at a level which assumes that the reader possesses a fundamental understanding of statistics and personnel testing and measurement principles.

The manual describes the test protocols, abilities measured, scoring procedures, and guidelines for interpreting and using test scores. A summary of technical information pertaining to test development, validity, utility and fairness is also presented. Further information regarding the development and validation of the tests is given in a separate technical report. Instructions for administering the tests are contained in an *Administrator's Guide* and an *Examinee Guide* provides an introductory overview of the tests for job candidates.

Inquiries regarding the Test Battery should be directed to the Dispatcher Testing Program Coordinator at the Commission on POST, 1601 Alhambra Boulevard, Sacramento CA 95816; telephone (916) 227-4834; or FAX (916) 227-3895.

NORMAN C. BOEHM
Executive Director

ACKNOWLEDGEMENTS

The development and validation of the Test Battery was completed with the invaluable assistance of many individuals and agencies in the law enforcement community. The following individuals played a key role in the development of one or more tests in the Battery: Sergeant Greg Kast, Oakland PD Communications Division; Gerald Verwolf, Communications Operations Manager, City of Costa Mesa; Deborah Davis, Operations Manager, SHASCOM (formerly with San Jose PD); RoxAnn Brown, Deputy Director, Stanislaus County Emergency Dispatch; Lieutenant Nick Bercuta, Los Angeles County SD Communications Center; Beverly Kent, Communications Supervisor and Instructor, Los Angeles County SD Communications Center; and Greg Miraglia, Communications Center Manager, Fairfield DPS. Special recognition is also owed to the following agencies for their assistance in providing staff and resources to help develop the Battery: San Jose PD Communications, Stanislaus County Emergency Dispatch, and the California Highway Patrol audio/visual unit.

The test validation research was completed with the assistance of the following individuals who participated in the development of job performance criterion measures: Commander Jim Sida and Sergeant Pat Lantz, Kern County SD; Lisa Vasquez, Training Manager, San Jose PD; Lori Thompson, Communications Manager, Riverside County SD; Cherry Teter, Communications Supervisor, Los Angeles PD (Sergeant Kast and Mr.Verwolf also participated in this phase of the project).

POST staff Donna Lively and Vickie Pruden, Office Technicians, were instrumental in the project, administering experimental versions of the Test Battery, helping to produce and implement criterion data collection forms, key-entering data, and assisting in the production of the final Test Battery and related materials. The test development and validation research was completed under the general direction of Dr. John Berner, Chief of the POST Standards and Evaluation Service Bureau.

The validation research could not have been completed without the cooperation of the following academies: Allan Hancock College, California Highway Patrol Academy, Evergreen Valley College CJTC, Fresno City College/State Center RTC, Golden West College CJTC, Modesto CJTC, Napa Valley College, Oakland PD, Rio Hondo RTC, Riverside Community College/Academy of Justice, Sacramento County SD, Sacramento Public Safety Center, San Diego LETC, and Ventura College/County CJTC.

Finally, POST expresses its thanks to the over 100 agencies that participated in the validation research, including:

ALAMEDA COUNTY SD
ALHAMBRA PD
ATWATER PD
BALDWIN PARK PD
BANNING PD
BERKELEY PD
BEVERLY HILLS PD
BISHOP PD
CA DEPT OF PARKS & REC
CAL POLY POMONA PD
CALEXICO PD
CALIFORNIA CITY PD
CAPITOLA PD
CERES PD
CHINO PD
(continued)

CHP
CHULA VISTA PD
CLOVIS PD
COALINGA PD
COMPTON PD
CONTRA COSTA COUNTY SD
CORCORAN PD
CORONA PD
COVINA PD
CSU HAYWARD PD
CSU HUMBOLDT PD
CSU SAN DIEGO PD
CYPRESS PD
DALY CITY PD
DAVIS PD
EL CAJON PD
EMERYVILLE PD
FAIRFIELD PD
FARMERSVILLE PD
FREMONT PD
FRESNO COUNTY SD
FRESNO PD
GARDEN GROVE PD
GLENDALE PD
GLENDORA PD
HANFORD PD
HAYWARD PD
HEMET PD
HUNTINGTON PARK PD
INYO COUNTY SD
KERN COUNTY SD
LODI PD
LONG BEACH PD
LOS ALAMITOS PD
MADERA COUNTY SD
MANTECA PD
MARIN COUNTY SD
MARYSVILLE PD
MENDOCINO COUNTY SD
MILLBRAE PD
MORGAN HILL PD
MOUNT SHASTA PD
MOUNTAIN VIEW PD
NEWPORT BEACH PD
NOVATO PD
OAKLAND PD
OCEANSIDE PD
ONTARIO PD
ORANGE COUNTY MARSHAL
ORANGE COUNTY SD
ORANGE PD
OXNARD PD
PACIFICA PD
PASADENA PD
PLACENTIA PD
PLACER COUNTY SD
PLEASANTON PD
PLUMAS COUNTY SD
RANCHO SANTIAGO COLLEGE SAFETY DEPT.
REDLANDS PD
RIALTO PD
RICHMOND PD
RIPON PD
RIVERSIDE COUNTY SD
RIVERSIDE PD
ROCKLIN PD
SACRAMENTO COUNTY SD
SACRAMENTO PD
SAINT HELENA PD
SAN BERNARDINO COUNTY SD
SAN BERNARDINO PD
SAN BERNARDINO VALLEY COLLEGE PD
SAN BRUNO PD
SAN DIEGO COMM COLLEGE DIST
SAN DIEGO COUNTY SD
SAN DIEGO PD
SAN FRANCISCO INTL AIRPORT COMM.
SAN JACINTO PD
SAN RAFAEL PD
SANGER PD
SANTA ANA PD
SANTA BARBARA COUNTY SD
SANTA MONICA COLLEGE PD
SANTA MONICA PD
SCOTTS VALLEY PD
SEAL BEACH PD
SEBASTOPOL PD
SHAFTER PD
SIMI VALLEY PD
SOLANO COUNTY SD
SONOMA COUNTY SD
SONOMA PD
SOUTH BAY REGIONAL PUB COMM AUTH
SOUTH LAKE TAHOE PD
STANISLAUS CO. EMERGENCY DISPATCH
STOCKTON PD
SUTTER COUNTY SD
TULARE COUNTY SD
TUSTIN PD
UNIV OF CA-LOS ANGELES PD
UNIV OF SOUTHERN CA PSD
VENTURA COUNTY SD
VERNON PD
WALNUT CREEK PD
WHITTIER PD
YOLO COUNTY SD
YUBA CITY PD
YUBA COUNTY SD

TABLE OF CONTENTS

COMMISSIONERS i

PREFACE iii

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vii

LIST OF TABLES ix

LIST OF FIGURES xi

OVERVIEW OF THE TEST BATTERY 1

- Measures of Cognitive Ability 1
- General Features of the Tests 2
- Test Protocols 2
- Summary of Technical Information 5

TEST SCORING PROCEDURES 7

- Overview 7
- Total Battery Score: To be Used for Employment Decision Making 8
- Ability and Single Test Scores Provided for Examinee Feedback 8
- Sample Test Score Reports 10

USING TEST SCORES TO MAKE EMPLOYMENT DECISIONS 15

- Norms 15
- Utility 21
- Establishing a Minimum Passing Score 27
- Combining Test Battery Scores With Other Assessments 31
- Score Banding: An Alterative Approach to Using Test Scores 32

OVERVIEW OF TEST DEVELOPMENT 33

- Job Analysis Foundation 33
- Test Design 37
- Test Construction 38

PSYCHOMETRIC CHARACTERISTICS 42

Summary of Test Properties 42
Test Intercorrelations 42

VALIDITY 47

The Concept of Validity 47
Criterion-Related Validity Evidence for the Test Battery 47
Comparison of Findings with Other Research 51

FAIRNESS 53

The Concept of Fairness 53
Differential Prediction Analysis and Results for the Test Battery 54

ADA CONSIDERATIONS 57

REFERENCES 59

APPENDIX 63

Example Computational Procedure for Combining Test Battery Scores with Other Assessments A-1

LIST OF TABLES

		Page
Table 1	Abilities Measured by POST Dispatcher Selection Tests	1
Table 2	POST Entry-Level Dispatcher Selection Tests	3-4
Table 3	Technical Information Summary	6
Table 4	Test Battery Norms: Job Applicants and Non-affiliated Students	16
Table 5	Test Norms: Job Applicants and Non-affiliated Students	17
Table 6	Test Battery Norms: Incumbent Dispatchers	19
Table 7	Test Norms: Incumbent Dispatchers	20
Table 8	Expected Gain in Basic Course Student Performance Associated with Test Battery Performance Level	22
Table 9	Expected Gain in Job Performance Associated with Test Battery Performance Level	24
Table 10	Expected Gain in Probation Success Rate Associated with Test Battery Performance Level	25
Table 11	Test Battery Scores Corresponding to Minimum Acceptable Academy Performance and Job Performance	29
Table 12	Applicant Passing Rates Expected to Result from Alternative Passing Scores	30
Table 13	Dispatcher Job Duties	34
Table 14	Essential Dispatcher Abilities	35
Table 15	Summary of Test Characteristics	43
Table 16	Test Intercorrelations	44
Table 17	Factor Structure of the Tests	46
Table 18	Validation Sample Characteristics	48-49
Table 19	Performance Criterion Measures	50

Page

Table 20 Summary of Validity Evidence for the POST Entry-Level Dispatcher Selection Test Battery.....52

Table 21 Differential Prediction Analysis Results for the POST Entry-Level Dispatcher Selection Test Battery.....55

LIST OF FIGURES

Page

Figure 1 Sample Examinee Roster 11

Figure 2 Sample Test Score Report 13

Figure 3 Sample Test Score Summary Report 14

Figure 4 Gain in Probation Success Rate Associated with Test Battery Performance Level 26

Figure 5 Job Duty Linkages: Abilities Measured by the POST Dispatcher Selection Test Battery 36

OVERVIEW OF THE TEST BATTERY

Measures of Cognitive Ability

The POST Entry-Level Dispatcher Selection Test Battery was designed to measure abilities that are both essential for successful performance of dispatcher duties and necessary for job candidates to possess before hire. These abilities fall into four primary areas:

Verbal ability (V): This ability includes *Written and Oral Comprehension*: the ability to read passages and listen to orally imparted information and retrieve facts, draw conclusions, and derive meaning; and *Written Expression*: the ability to use language to convey information clearly in writing.

Reasoning ability (R): This ability includes *Deductive Reasoning*: the ability to apply general rules to specific problems to attain logical answers; and *Information Ordering*: the ability to correctly follow a given rule or set of rules to arrange things or actions in a certain order.

Memory ability (M): This ability includes the capacity to store and retrieve facts, details, and other information.

Perceptual ability (P): This ability includes *Speed and Accuracy*: the ability to quickly and accurately compare letters and numbers presented orally and in written form; and *Time Sharing*: the ability to shift back and forth between two or more sources of information, both written and orally imparted, in performing a task or set of tasks.

The Battery is comprised of eleven (11) tests designed to measure one or more facets of the four primary abilities. The tests are listed below along with the abilities that they measure.[1]

Table 1
Abilities Measured by POST Dispatcher Selection Tests

Tests	V	R	M	P
1. Public Safety Bulletin		❍		
2. Assigning Field Units	❍	●	❍	
3. Evaluating Facts	●			
4. Setting Priorities		●		
5. Reading Comprehension	●	❍		
6. Clarity	●			
7. Recalling Facts & Details			●	
8. Call-Taking	❍	●		❍
9. Oral Directions	❍	●	❍	
10. Checking Coded Information				●
11. Checking & Listening		❍		●

●=primary measure; ❍=secondary measure

[1]Primary measures have a factor loading ≥.50 and secondary measures have a factor loading ≥.30, as derived in a Principal Components analysis with Varimax rotation

General Features of the Tests

The tests are presented through a combination of audiotape and paper-pencil media. Six of the tests use a traditional written multiple-choice format, while the remaining five tests incorporate an audiotape format where information is presented orally and examinees answer questions or perform tasks that require a written multiple-choice response. Instructions for all tests are provided on audiotape, including timed start and stop points. The tests are separately timed, ranging from 5 minutes to 17 minutes each. The Total Battery requires approximately 2-1/2 hours to administer, including instructions; actual time will vary depending upon the size of the examinee group. All of the tests employ an optically scan able answer sheet so that the Battery may be administered in a large group setting.

An important feature of the Test Battery is that it is designed to measure *aptitude* for performing public safety dispatcher work. The tests are not intended to measure job-specific knowledge or skills that would be expected to be learned quickly in training or on the job.

Test Protocols

Although the tests were designed to measure general cognitive abilities (verbal, reasoning, memory, and perceptual abilities), the test formats utilize job relevant tasks and are presented in a public safety context. An overview of each test in the Battery is given in Table 2, including descriptions of the tasks to be performed by the examinee, test format, test type (power, speeded), number of item response options, number of items, and test administration time (excluding examinee instructions).

Tests which are designated as *power* tests were designed to measure abilities that are not heavily dependent upon speed of response; i.e., the time limits were established such that the vast majority of examinees complete the test in the allotted time. Conversely, tests designated as *speeded* are intended to measure abilities where speed of response is an integral part of the ability being measured; thus, these tests are purposely administered under highly restrictive time limits such that virtually no examinee can respond to all items. The tests designated as *hybrid speed/power* tests present information to examinees under somewhat speeded conditions, while examinees are given sufficient time to respond to questions regarding the information presented.

Table 2
POST Entry-Level Dispatcher Selection Tests[2]

1. Public Safety Bulletin (r M): The examinee is given 3 minutes to review a one-page simulated "shift bulletin" containing several single-paragraph descriptions, each one pertaining to a different event. After the study period, the examinee answers multiple-choice questions regarding facts and details about the events based solely upon memory. *(paper-pencil format; power test; 4 response options; 15 items; 6 min.)*

2. Assigning Field Units (v R m): The examinee reads a novel set of rules for assigning field units and then determines which of five field units should be assigned to various "incidents." The incidents occur in different geographic regions and are designated emergency or non-emergency. The examinee uses a multiple-choice response format to designate no, one, or more units to be dispatched to each incident. Each item response alternative is scored. This test was designed to measure the deductive facet of Reasoning ability, primarily. *(paper-pencil format; power test; 32 response options; 20 items; 5 min.)*

3. Evaluating Facts (V): The examinee reads a set of public safety-related facts and then determines whether statements that follow are true, false, or cannot be determined on the basis of the facts. The answer is marked on a multiple-choice response sheet. *(paper-pencil format; power test; 3 response options; 15 items; 5 min.)*

4. Setting Priorities (R): The examinee is given a novel set of rules to read and follow in order assign priority codes to events. The events are presented in sets of three. A multiple-choice response format is used to designate the priority of events in each triad as 1st, 2nd, and 3rd priority. This test was designed to measure Information Ordering ability, a facet of Reasoning ability. *(paper-pencil format; power test; 3 response options; 15 sets/45 items; 10 min.)*

5. Reading Comprehension (V r): The examinee reads a brief passage and then answers multiple-choice questions regarding facts and details contained in the passage, as well as the meaning of the information, how it may be interpreted, and conclusions that may be drawn. While primarily designed to measure Written Comprehension ability, this test also measures Reasoning ability. *(paper-pencil format; power test; 4 response options; 20 items; 15 min.)*

6. Clarity (V): The examinee compares two versions of the same sentence and identifies the one that is more clearly written. The answer is marked on a multiple-choice response sheet. This test measures Written Expression, a facet of Verbal ability. *(paper-pencil format; power test; 2 response options; 15 items; 5 min.)*

7. Recalling Facts & Details (M): The examinee listens to a tape recording of a simulated call for law enforcement service received by a public safety dispatcher. The examinee is not allowed to take notes. The examinee then answers multiple-choice questions regarding various facts and details contained in the call, based solely upon memory. *(audiotape format; hybrid speed/power test; 4 response options; 18 items; 9 min.)*

(continued)

[2]The abilities measured by each test are shown in parentheses: V=Verbal, R=Reasoning, M=Memory, and P=Perceptual ability, where upper case denotes primary measure of ability (factor loading $\geq$.50) and lower case denotes secondary measure of ability (factor loading $\geq$.30). Test times do not include examinee instructions.

Table 2 (continued)
POST Entry-Level Dispatcher Selection Tests

8. Call-Taking (v R p): The examinee listens to a tape recording of three simulated calls for law enforcement service received by a dispatcher. The examinee is allowed to take notes during the calls and is given a brief period to complete the notes after all calls have been presented. The examinee is allowed to use the notes to answer a series of multiple-choice questions regarding facts and details pertaining the calls, as well as interpretations and conclusions regarding the meaning of each call. While primarily a measure of Reasoning ability, this test was also measures Oral Comprehension (Verbal) and Perceptual Speed and Accuracy abilities. *(audiotape format; hybrid speed/ power test; 4 response options; 25 items; 17 min.)*

9. Oral Directions (v R m): The examinee listens to a tape recording of a simulated radio call from a patrol officer to a dispatcher. The examinee is allowed to take notes during the call and is given a brief period to complete the notes after the call is presented. The examinee is allowed to use the notes to answer multiple-choice questions regarding various tasks to be performed, the order in which they are to be performed, various details contained in the call such as names, times, locations, etc., and interpretations and conclusions that may be drawn. While primarily a measure of Reasoning ability (Information Ordering), this test also measures Oral Comprehension (Verbal) and Memory abilities. *(audiotape format; hybrid speed/power test; 4 response options; 17 items; 14 min.)*

10. Checking Coded Information (P): The examinee listens to a tape recording of a narrator presenting a series of random letter-number codes. The codes range from two to four alphanumeric characters. As each code is presented orally, the examinee refers to a "Code Sheet," and identifies and marks the corresponding code among 5 written alternatives. The information is presented slowly at first, increasing in speed until the task becomes very difficult. After the information is presented, the examinee the marks his/her answers onto a scan able answer sheet. This test was designed to measure Perceptual Speed and Accuracy. *(audiotape format; speeded test; 5 response options; 60 items; 9 min.)*

11. Checking & Listening (r P): The examinee performs two tasks at the same time: (1) compare a list of names, addresses, and license numbers with a "HOT SHEET" containing similar information, and identify as many matches as possible; and (2) listen to a tape recording of simulated radio broadcasts from several field units and keep track of their status. The examinee records the unit status transmissions on a "RADIO LOG." After the simulated radio broadcasts have ended, the examinee is instructed to stop the comparison task and answer a series of multiple-choice questions regarding the various status changes of each unit. This test was designed to measure Perceptual Time Sharing ability, primarily. *(audiotape format; speeded part I, hybrid speed/ power part II; 2 response options part I; 5 response options part II; 105 items; 13 min.)*

Summary of Technical Information

Statistical characteristics of the Test Battery are summarized in Table 3, including information regarding the distribution of applicant scores (mean, SD, skewness, and kurtosis), the precision with which test scores estimate an examinees' true abilities (reliability and standard error of measurement), and the extent to which Test Battery scores are predictive of performance in basic training and on the job (correlation coefficients). The normative and validation samples are described later in the manual.

Total Battery score yields a reasonably precise estimate of an examinee's ability. The internal consistency reliability of Total Battery scores (i.e., the linear combination of alpha coefficients and Spearman-Brown estimates for the 11 tests) is .94, which translates into a standard error of measurement (SEM) of approximately 2.5 points (note that scores are reported on a T scale, where mean=50 and SD=10). In very broad terms, this means that an examinee's "true" ability is likely to fall within 2.5 points of his or her obtained score (68% probability) and is very likely to fall within approximately 4 points (90% probability).

The validity evidence indicates that Test Battery scores are significantly predictive ($p<.01$, one-tailed) of overall success or failure in completing basic training, rated performance in basic training, rated performance of dispatcher job duties, and employee retention (overall success or failure in completing probation). The uncorrected validity coefficients, ranging from .21 to .35, are of sufficient magnitude to offer substantial utility to test users (see section entitled "Utility"). Further details regarding the validation research are provided later in this manual.

Table 3
Technical Information Summary
POST Entry-Level Dispatcher Selection Test Battery

Psychometric Characteristics[a]	Descriptive Statistics
Total Score:	
Mean	50.0
SD	10.0
Skewness	- 0.503
Kurtosis	0.004
Reliability	.938
SEM	2.490
Validity Evidence	Zero-order correlations
Basic Academy:	
Pass/Fail[b]	.21***
Total Performance[c]	.35***
Job Performance	
Supervisor Ratings (Total Performance)[d]	.28**
Self-Ratings (Total Performance)[e]	.24**
Employee Retention (pass/fail probation)[f]	.30***

***p<.0001; **p<.01 (one-tailed)

[a]N=1,036. Job applicants and non-affiliated academy students.

[b]N=627. Completed Dispatcher's Basic Course=1; failed to complete for any reason=0.

[c]N=629. Mean of standardized mean knowledge/skill rating and standardized class rank (approximately 15% of the students were ranked within class on the basis of academy curriculum test scores only; instructor ratings were not available).

[d]N=148. Mean of: (a) global effectiveness and relative performance ratings; (b) mean of 20 job duty ratings; (c) mean of 18 KSA ratings; and (d) mean of 8 work behavior ratings related to conscientiousness and performance under stress.

[e]N=134. Mean of: (a) global effectiveness and relative performance ratings; (b) mean of 20 job duty ratings; (c) mean of 18 KSA ratings; and (d) mean of 8 work behavior ratings related to conscientiousness and performance under stress.

[f]N=215. Completed probation=1; resigned or terminated while job performance was unsatisfactory due to inadequate job knowledge, skills, or abilities=0.

TEST SCORING PROCEDURES

Overview

The tests are scored in a multi-step procedure that yields a Total Battery score (T-Score) which reflects test performance relative to other job applicants throughout the state. The eleven tests are scored such that they contribute equally to an examinee's Total Battery score. Because each test measures one or more of the four primary abilities (Verbal, Reasoning, Memory, Perceptual), these abilities are implicitly weighted in Total Battery scores as follows:

Verbal:	29%
Reasoning:	36%
Memory:	19%
Perceptual:	16%

The above percentage weights for Verbal, Reasoning, Memory, and Perceptual abilities are consistent with the relative importance of these abilities as identified in a statewide job analysis of the dispatcher position conducted by POST (Weiner, 1991).[3]

The scoring procedure entails: (1) computing total number right score on each test; (2) adjusting the scores for guessing; (3) standardizing the scores so that the means and standard deviations for each test are equal; (4) averaging the standard scores to compute a battery score; and (5) rescaling the battery score to a T-score. The scoring procedures are outlined below. Further details regarding the scoring formulas are given in the technical report.

Correction for Guessing

Examinees' total number right scores are adjusted by subtracting a fractional value for each wrong answer.[4] The purpose of this adjustment is to take into account the possibility of guessing the correct answer by randomly selecting among multiple-choice item response alternatives. This adjustment is particularly important for speeded tests and difficult power tests in which examinees do not respond to all items. The adjustment has no net effect upon tests for which all examinees respond to all items.

[3]The overall importance of each ability measured by the Battery was indicated in the job analysis results as follows: dispatch supervisors' mean ratings of ability importance were multiplied by the number of tasks for which the ability was identified as essential by a majority of subject matter experts, and the products were rescaled to relative percentages of the sum of products.

[4]Scores on Test #2, *Assigning Field Units*, are not adjusted for guessing in view of the complex nature of the test items (i.e., each item alternative is scored and a non-response may be the correct answer in some instances) and the low probability of guessing correctly (.03125).

Adjusted test scores are computed using the following formula (Guilford & Fruchter, 1973, pp. 442-444):

$$S = R - (W / (k-1))$$

where S=raw score adjusted for guessing, R=number of right answers, W=number of wrong answers, and k=number of item response alternatives.

Standard Scores

The guessing-adjusted scores are standardized so that the tests receive equal weight in contributing to Total Battery Score. The standard scores are expressed in T-score units (mean=50, SD=10). The formula used to compute standard test scores is:

$$T = (S - M) / SD * 10 + 50$$

where T=standardized test score, S=raw score adjusted for guessing, M=mean of examinees' adjusted raw scores, SD=standard deviation of examinees' adjusted raw scores. The mean and SD are based upon a normative calibration sample comprised of job applicants and non-affiliated students (see *Norms* section below for further details regarding the sample).

Total Battery Score:
To be used for Employment Decision Making

Total Battery scores are reported to local hiring authorities for use in making personnel decisions.[5] Total score is computed by averaging the 11 standardized test scores and then rescaling the average to a T-score based upon the above-referenced normative sample mean and SD.

Ability and Single Test Scores Provided for Examinee Feedback

Ability composite scores and individual test scores are reported for feedback purposes only; that is, to provide candidates with information regarding the portions of the Battery in which their performance was relatively strongest. The ability scores are computed by averaging the standardized adjusted scores for tests that are primary or secondary measures of each ability. Primary measures receive full weight while secondary measures receive 50% weight in computing an ability score. The tests that contribute to each ability score are shown below.

[5]Test users should rely on Total Battery scores for several reasons: Each of the 11 tests measures a unique facet of an essential ability by incorporating oral and/or written media; the tests were intentionally designed to be very brief so that it would be feasible to incorporate a variety of formats; all of the tests are significantly predictive of job performance; and the reliability (precision) of measurement increases with the number of tests used.

Ability Score	Contributing Tests
Verbal:	100%: Reading Comprehension, Clarity, Evaluating Facts 50%: Assigning Field Units, Call Taking, Oral Directions
Reasoning:	100%: Assigning Field Units, Setting Priorities, Call Taking, Oral Directions 50%: Reading Comprehension, Checking & Listening, Public Safety Bulletin
Memory:	100%: Public Safety Bulletin, Recalling Facts & Details 50%: Oral Directions, Assigning Field Units
Perceptual:	100%: Checking Coded Information, Checking & Listening 50%: Call Taking

Sample Test Score Reports

At present, all test scoring and reporting is performed by POST. Three types of reports are generated for each test administration, including an *Examinee Roster*, a *Test Score Report*, and a *Summary Report*. Each of these reports is briefly described below.[6]

Examinee Roster. This report lists the examinees in descending order of performance on the Total Battery. A sample *Examinee Roster* is shown in Figure 1. As seen in the example, three indices of performance are reported for each candidate:

> ***Rank:*** The examinees are ranked according to their Total Battery T-score (rounded to the nearest whole number), where rank 1 is assigned to the highest scoring person, rank 2 is assigned to the second highest person, and so on.[7] This information is provided to assist users in making practical decisions (e.g., when it is feasible to interview or further assess a limited number of candidates)**.**
>
> ***Total T-score:*** This is the resulting Total Battery score described earlier (see page 8). Users should exercise caution in comparing examinees with very small differences between test scores (e.g., avoid cut scores that separate examinees by only 0.1 point).
>
> ***Statewide Percentile***: This value represents the examinee's standing on the test relative to other examinees in the state. For example, an examinee who receives a percentile score of 90 has performed better than 90% of all examinees in the statewide comparison sample, while an examinee who receives a percentile score of 50 has scored better than 50% of those in the statewide comparison sample (further information regarding the normative sample is given in the next section).

Summary statistical information based on normative samples of job applicants and incumbent dispatchers is reported at the bottom of the *Examinee Roster* for comparative purposes (e.g., mean, standard deviation, and standard error of measurement). However, see the following section for more comprehensive interpretive information.

[6]While the report formats may change periodically, their content will remain essentially the same.

[7]In the case of ties, examinees are assigned the same rank, with the rank value representing the number of examinees scoring above that level, plus one.

Figure 1
Sample Examinee Roster

Examinee Roster

POST Entry-Level Dispatcher Selection Test Battery

Agency: River City PD
Test Date: January 2, 1996
Test Form: 9501
No. Examinees: 20

Rank	Examinee Name	SSN	Total T-Score	Statewide Percentile
1	SING JENNIFER M	163-49-5098	72.0	99
2	KAULKNER JOAN C	257-87-6212	66.6	97
3	JONES RYAN M	355-31-4367	66.5	97
4	SMITHE MARGO R	463-75-8887	65.3	95
5	MINDENOAS BARBARA	568-83-3450	61.6	90
6	CARLYLE LOIS G	460-73-0986	61.5	90
7	JAMES MICHAEL M	347-25-7657	59.1	80
8	WILFFORD VALERIE A	246-45-4444	57.4	75
9	JASPER CHARMAYNE B	146-08-6648	55.5	70
10	BOSSED CHRISTINA M	265-89-6129	54.5	65
11	ARNNOLD ELLEN H	355-82-3160	53.5	60
12	DENTON MARIA C	465-88-1323	52.5	55
13	KLANTZ CHRISTOPHER S	572-65-9988	50.7	50
14	JEFFRIES KEN A	548-51-8755	48.5	45
15	KURBIN BRIAN K	466-02-6922	47.5	40
16	ADAMS DONNA R	361-35-0488	45.5	35
17	GAILON JAYSON M	273-39-3139	44.8	30
18	ASCOTT TRENA L	168-41-1092	43.5	25
19	LORENTZ LORRIE A	148-81-0703	41.7	20
20	ROBERTS LISA L	262-11-7045	40.1	15

Note: Percentile based on norm sample (applicants & academy students)
Norm sample mean=50, SD=10, N=1,036, standard error=2.5
Incumbent Dispatcher mean=57.8, SD=7.3, N=283

Test Score Report. This report provides detailed information regarding each examinee's performance on various components of the Battery. As seen in the example shown in Figure 2, the report includes T-scores and percentiles (centiles) for the Total Battery, four ability scores, and eleven tests which comprise the Battery. This information is provided primarily for feedback purposes rather than decision-making.

For example, a candidate who is not hired may seek information regarding his or her strengths and weaknesses on the Battery. Accordingly, the employer may respond by indicating that the candidate demonstrated average, above average, and/or below average ability in certain areas (say, for example, that average verbal and reasoning ability centile scores of 50 and 60 were obtained, but below average memory and perceptual ability centile scores of 20 and 30 were obtained). Further diagnostic information may be provided by identifying specific tests among the eleven on which the candidate performed particularly well (e.g., a centile of 70 or higher) or poorly (e.g., a centile of 30 or lower).

When providing feedback regarding specific test performance or otherwise interpreting such results, test users are advised to consider the precision of the scores (SEM) and avoid drawing conclusions on the basis of small differences in scores. For example, the SEM for the *Public Safety Bulletin* test is 5.39 points (see Table 5). Thus, it would be inadvisable to classify an examinee who obtains a score of 47 as "below average" since there is considerable likelihood that his/her "true" score might fall at or above the median score (50); i.e., the 90% confidence interval for an obtained score of 47 ranges from 38 to 56.

The information in the *Test Score Report* is not intended for purposes of comparing the profiles of two candidates in order to determine which one to hire. As stated earlier, the Total Battery score is regarded as the best overall appraisal of a candidate's aptitude to perform dispatcher work.

Summary Report. This report summarizes the total performance of all examinees in the test administration. As seen in the example report in Figure 3, the report contains the mean, standard deviation, minimum and maximum of scores obtained by all examinees, as well as by race/ethnicity and by gender. The report also contains frequency distributions of Total Battery scores for the total sample and by subgroup (not shown in example). This information is reported to gauge the overall caliber of the candidate group and to help users evaluate the impact of alternative minimum passing scores upon various subgroups.

Figure 2
Sample Test Score Report

Test Score Report

POST Entry-Level Dispatcher Selection Test Battery

Agency: Bay City PD
Test Date: January 3, 1996
Test Form: 9501

		Ability Scores:				Subtests:										
Examinee	**Total**	**V**	**R**	**M**	**P**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
123-45-67-8910																
T-score:	67.4	65.7	61.4	66.5	70.9	64.2	66.1	60.9	67.3	65.4	62.8	63.7	61.4	69.9	68.2	67.3
Centile:	97	97	85	97	99	97	95	90	95	90	97	90	95	90	99	97
223-45-67-8910																
T-score:	53.1	55.7	48.5	52.2	50.2	54.2	50.3	45.1	52.2	48.3	53.9	51.8	49.4	57.2	52.2	50.8
Centile:	55	70	40	60	45	60	50	30	50	40	60	50	40	60	50	40
323-45-67-8910																
T-score:	56.2	55.3	58.3	57.0	58.4	54.0	47.3	52.0	58.0	55.2	63.2	58.7	52.7	61.1	57.6	59.4
Centile:	70	65	75	70	75	60	40	50	70	60	80	70	50	80	70	80
423-45-67-8910																
T-score:	54.9	57.0	53.9	55.3	56.2	62.2	58.5	54.4	59.2	65.4	60.5	55.0	58.0	57.7	60.1	63.3
Centile:	65	70	60	65	70	80	70	60	80	90	70	60	70	70	80	90
523-45-67-8910																
T-score:	61.1	64.1	59.2	62.0	66.3	64.2	62.2	57.9	61.0	60.1	63.0	63.7	61.4	63.3	65.4	59.0
Centile:	85	90	80	85	95	97	80	70	90	80	90	90	95	90	95	80
623-45-67-8910																
T-score:	45.2	43.3	49.1	45.9	48.3	46.1	46.7	44.0	47.4	52.6	47.7	46.9	53.5	45.0	47.1	47.3
Centile:	30	20	40	30	35	30	40	20	30	50	40	30	50	20	30	30

Note: V = Verbal, R = Reasoning, M = Memory, P = Perceptual ability scores,
1 = Public Safety Bulletin, 2 = Assigning Field Units, 3 = Evaluating Facts, 4 = Setting Priorities, 5 = Reading Comprehension, 6 = Clarity, 7 = Recalling Facts & Details,
8 = Call-Taking, 9 = Oral Directions, 10 = Checking Coded Information, 11 = Checking & Listening

Figure 3
Sample Test Score Summary Report

Test Score Summary Report
POST Entry-Level Dispatcher Selection Test Battery

Agency: Foothill SD
Test Date: January 5, 1996
Test Form: 9501

Total Battery Means and SDs

All Examinees

N	Mean	Std Dev	Minimum	Maximum
262	51.91	7.91	21.60	66.30

by Race/Ethnicity

RACE	N	Mean	Std Dev	Minimum	Maximum
American Ind.	2	54.41	2.40	50.41	58.41
Asian	5	55.00	4.82	49.20	65.30
Black	27	50.19	6.78	25.10	60.20
Filipino	2	52.65	3.92	50.15	55.15
Hispanic	85	50.47	6.78	32.30	62.70
White	137	53.52	6.16	21.60	66.30
Other	3	50.64	3.22	49.70	66.10
Not reported	1	50.20	.	50.20	50.20

by Gender

SEX	N	Mean	Std Dev	Minimum	Maximum
Female	190	52.67	6.22	29.60	66.30
Male	72	52.59	6.78	21.60	65.90

USING TEST SCORES TO MAKE EMPLOYMENT DECISIONS

A number of factors must be taken into consideration in order to interpret and best use examinees' Test Battery scores. Two principal considerations are test norms and utility. Thought must also be given to how the test scores will be used. For example, will the Battery be used as a pass/fail hurdle, or will examinees be ranked on the basis of their scores? How will Test Battery scores be combined with scores resulting from other employee selection procedures, such as oral interviews and practical exams? These issues are briefly addressed below.

Norms

The tests are "norm-referenced." This means that an examinee's score on the Battery is interpreted by comparing it to the scores obtained by other examinees in the relevant population. Given that the Battery is intended for use in selecting entry-level dispatcher candidates, the principal normative group for interpreting test scores is the population of *job applicants and students interested in pursuing a career as a public safety dispatcher.* An additional, secondary point of reference for interpreting test scores is provided by incumbent dispatchers' performance on the Battery. While the Battery is not intended for the assessment of job incumbents, it is useful to consider their performance on the tests to interpret job applicants' performance in the light of "normal expectations of the work force."

The following norm tables are based upon examinees tested between 1993 and 1995. The norms will be updated periodically as new data become available.

Job Applicant/Non-affiliated Student Norms

Test Battery norms for over 1,000 job applicants and non-affiliated Basic Course students are presented in Table 4. The table shows Total Battery scores and separate ability scores (Verbal, Reasoning, Memory, and Perceptual) corresponding to percentiles of the distribution of scores obtained by the normative sample.[8] Norms for the eleven individual tests are shown in Table 5.

The norm sample was comprised of a diverse sample of examinees which included a substantial percentage of racial/ethnic minorities (25% Black, 14% Hispanic, and 6.8% other non-White race/ethnicity) and females (71%). Approximately 89% of the examinees reported having no previous dispatching experience.[9]

[8]A percentile is the score at which a certain percentage of examinees fall below. For example, a Total Battery score of 62 corresponds to the 90th percentile; i.e., an examinee who obtains a score of 62 has performed better than 90% of the examinees in the norm sample.

[9]Examinee age was not available.

Table 4: Test Battery Norms
Job Applicants and Non-affiliated Students

PERCENTILE	Total Battery[a]	Verbal[b]	Reasoning[c]	Memory[d]	Perceptual[e]
99	68	67	67	69	70
97	66	66	66	67	67
95	65	64	65	65	65
90	62	62	62	63	62
85	60	61	61	61	60
80	58	59	59	59	59
75	57	58	58	57	57
70	56	56	56	56	56
65	55	55	55	55	54
60	54	54	54	53	53
55	52	52	53	52	52
50	51	51	51	51	51
45	49	50	50	49	50
40	48	48	48	48	49
35	46	47	47	47	47
30	45	45	45	45	46
25	44	43	43	43	44
20	42	41	42	41	42
15	39	38	39	39	40
10	36	36	36	37	37
5	31	32	31	33	31
3	29	30	29	30	28
1	24	26	26	24	23
SEM	2.49	2.95	2.45	3.71	3.46

Note: Mean=50, SD=10 for all tests; N=1,036 to 1,048. Test scores are adjusted for guessing.

Total Battery score=mean of 11 standardized test scores; rescaled to mean=50 and SD=10.

Verbal ability score=mean of T scores: Reading Comprehension + Clarity + Evaluating Facts + (Assigning Field Units/2) + (Call-Taking/2) + (Oral Directions/2); rescaled to mean=50, SD=10.

Reasoning ability score=mean of T scores: Assigning Field Units + Setting Priorities + Call Taking + Oral Directions + (Public Safety Bulletin/2) + (Reading Comprehension/2) + (Checking & Listening/2); rescaled to mean=50, SD=10.

Memory ability score=mean of T scores: Public Safety Bulletin + Recalling Facts & Details + (Assigning Field Units/2) + (Oral Directions/2); rescaled to mean=50;SD=10.

Table 5: Test Norms
Job Applicants and Non-affiliated Students

PERCENTILE	1. Public Safety Bulletin	2. Assigning Field Units	3. Evaluating Facts	4. Setting Priorities	5. Reading Comprehension	6. Clarity	7. Recalling Facts/Details	8. Call-Taking	9. Oral Directions	10. Checking Coded Info.	11. Checking & Listening
99	-	69	67	66	-	-	-	66	-	69	71
97	68	68	-	-	65	-	67	64	-	66	67
95	-	67	64	63	-	-	-	-	65	65	65
90	65	65	-	-	62	63	63	63	-	62	62
85	-	63	-	60	-	-	-	-	62	61	60
80	61	60	60	59	59	-	59	60	59	59	58
70	57	56	57	57	57	59	56	57	-	56	55
60	53	52	53	55	54	54	-	54	56	53	53
50	50	49	-	52	51	-	52	51	52	51	51
40	47	46	50	50	48	50	48	48	49	48	49
30	44	43	45	46	45	45	45	45	46	46	46
20	42	40	41	41	-	-	-	42	40	43	42
10	37	38	36	34	36	36	37	36	36	37	37
5	34	36	32	30	31	32	34	30	32	32	32
3	31	34	29	29	28	27	30	27	30	29	28
1	27	32	25	24	23	23	19	22	27	22	22
SEM	5.39	3.46	5.20	4.47	4.80	5.92	6.08	5.39	5.10	5.00	4.58

Note: Mean=50, SD=10 for all tests; N=1,039 to 1,092. All scores (except Test #2) are adjusted for guessing prior to standardization.

Incumbent Dispatcher Norms

Table 6 contains Total Battery score and ability score norms based upon a sample of nearly 300 incumbent dispatchers who successfully completed probation.[10] Norms for each of the eleven tests are shown in Table 7. These tables show T-scores corresponding to percentiles of the dispatcher norm sample, along with means and standard deviations of Total Battery and ability scores. The T-scores are scaled relative to the job applicant/student means and standard deviations.

Clearly, incumbent dispatchers perform substantially higher on the tests: the Total Battery mean is 57.75 for incumbents versus a mean of 50.0 for applicants and students. This result is to be expected since, presumably, individuals who possess more of the abilities measured by the tests are more likely to be selected and remain on the job.

The incumbent dispatcher norm sample was also comprised of substantial percentages of racial/ethnic minorities (12% Black, 16% Hispanic, and 6.8% other non-White race/ethnicity) and females (73%). The mean age of the norm sample was 31.5 years and ranged from 20 to 61. The examinees had approximately 2 years of dispatching experience, on average, with the range of experience extending from 10 months to over 18 years.

[10]These individuals were subjects in a test validation study who completed the test battery under experimental conditions, while attending the Dispatchers' Basic Course. For a small percentage of these subjects (for whom probation outcome was unknown), job tenure was used as a proxy; i.e., currently employed as a public safety dispatcher with at least 12 months experience.

Table 6: Test Battery Norms
Incumbent Dispatchers

PERCENTILE	Total Battery	Verbal	Reasoning	Memory	Perceptual
99	71	69	69	71	72
97	69	68	68	70	69
95	68	67	67	69	68
90	66	66	66	67	65
85	-	65	65	66	64
80	65	64	64	64	63
70	62	61	62	63	61
60	60	60	61	60	60
50	58	59	59	58	57
40	57	56	58	55	56
30	54	53	56	52	53
20	52	50	54	50	51
15	50	48	52	48	49
10	48	46	50	46	47
5	45	43	47	42	43
3	42	41	44	40	41
1	35	28	35	35	36
Mean	57.75	56.75	58.46	57.10	56.75
SD	7.28	7.85	6.43	8.15	7.40
SEM	1.80	2.51	1.97	3.66	2.75
N	283	295	283	288	283

Note: Norms based on tenured dispatchers (completed probation). Scores are standardized relative to job applicant/student means and SDs, and are adjusted for guessing.

Total Battery score=mean of 11 standardized test scores; rescaled to mean=50 and SD=10.

Verbal ability score=mean of T scores: Reading Comprehension + Clarity + Evaluating Facts + (Assigning Field Units/2) + (Call-Taking/2) + (Oral Directions/2); rescaled to mean=50, SD=10.

Reasoning ability score=mean of T scores: Assigning Field Units + Setting Priorities + Call Taking + Oral Directions + (Public Safety Bulletin/2) + (Reading Comprehension/2) + (Checking & Listening/2); rescaled to mean=50, SD=10.

Memory ability score=mean of T scores: Public Safety Bulletin + Recalling Facts & Details + (Assigning Field Units/2) + (Oral Directions/2); rescaled to mean=50; SD=10.

Perceptual ability score=mean of T scores: Checking Coded Information + Checking & Listening + (Call Taking/2); rescaled to mean=50, SD=10.

Table 7: Test Norms
Incumbent Dispatchers

PERCENTILE	1. Public Safety Bulletin	2. Assigning Field Units	3. Evaluating Facts	4. Setting Priorities	5. Reading Comprehension	6. Clarity	7. Recalling Facts/Details	8. Call-Taking	9. Oral Directions	10. Checking Coded Info.	11. Checking & Listening
99	-	-	-	-	-	-	-	-	-	71	72
97	-	-	-	66	-	-	-	-	-	68	70
95	-	69	67	65	-	-	67	-	-	67	69
90	68	67	-	63	65	-	64	66	65	64	65
85	-	-	64	62	-	-	-	-	-	63	64
80	65	66	-	-	-	63	63	63	-	62	62
70	61	64	60	60	62	-	-	-	62	59	60
60	-	62	57	-	59	59	59	-	-	58	58
50	57	60	-	58	57	-	56	60	59	56	56
40	53	56	53	55	-	54	-	-	-	53	54
30	51	53	50	53	54	-	52	57	56	51	53
20	46	49	44	50	51	50	48	52	54	48	50
10	42	45	37	46	45	41	41	51	47	43	46
5	39	41	32	40	43	36	38	46	43	41	41
3	37	40	29	37	37	32	37	45	41	39	39
1	33	37	25	32	28	18	29	33	36	33	24
Mean	55.43	57.51	52.83	55.42	55.63	53.83	54.27	57.62	56.69	54.74	55.79
SD	9.21	8.64	10.04	7.12	7.65	8.98	8.63	6.24	6.77	8.25	8.41
SEM	5.04	3.35	5.02	3.77	4.19	5.24	5.59	4.37	4.17	4.44	4.61
N	296	296	296	296	296	296	288	295	295	284	283

Note: Norms based on tenured dispatchers (completed probation). Scores are standardized relative to job applicant/student means and SDs, and all (except Test #2) are adjusted for guessing.

Utility

Test norms are useful in placing an examinee's test performance in context relative to other examinees. However, they do not convey information regarding the extent to which an examinee scoring at a given level is likely to be a successful dispatcher. To obtain this perspective, several "utility" tables are presented which allow test users to consider potential gains in *basic academy performance*, *job performance*, and *employee retention* that would be expected when using the Battery with alternative cut scores. These tables are based upon the empirical validation results summarized later in this manual and represent gains over and above existing selection procedures (i.e., subjects in the validation research were selected for employment on the basis of procedures other than the POST Battery); and to the extent that these procedures are correlated with the Battery, the utility estimates presented below are underestimates. Overall, the tables illustrate how employers may realize dramatic gains in employee performance and retention through the use of the Battery.

Gains in Basic Academy Performance

Table 8 shows the expected gain in basic academy student performance that would be realized by using the Battery with alternative passing scores ranging from the 10th to the 90th percentile (centile). For each centile score level, the corresponding Total Battery T-score is reported, along with the percentage of students who were rated above average and who scored at or above that level (achieve cut score), the percentage of above average students who scored below that level (below cut score), and the gain in student performance that would be realized if the cut score were implemented (% Gain vs. base rate).[11] Note that gains in performance represent the percentage of above-average-rated students achieving the cut score relative to the percentage of students that would be rated above average if the test were not used (i.e., the base rate).

For example, at the 50th percentile (a Total T-score of 51.4), 59% of students who achieved this level of test performance were rated above average in basic academy performance, a gain of 16% relative to the base rate of 50.9%. Only 35% of students scoring below this level on the Battery were rated above average (a significantly lower success rate).

As seen in the table, gains in student performance increase steadily with Total Battery score, ranging from 2.4% (at the 10th percentile) to 41.8% at the 90th percentile.

[11]The median academy performance composite rating was used as a "cut point" to classify students as above and below average.

Table 8
Expected Gain in Basic Course Student Performance
Associated with Test Battery Performance Level

Total Battery Performance[a]		Percent Students Rated Above Average[b]		% Gain vs. base rate[c]
Centile	T Score	Achieve cut score	Below cut score	
90	62.3	72.2	46.1***	41.8
80	58.5	67.3	42.1***	32.2
70	55.9	66.8	37.1***	31.2
60	53.8	62.8	35.7***	23.4
50	51.4	59.1	34.5***	16.1
40	47.8	57.3	29.9***	12.6
30	45.2	55.5	26.3***	9.0
20	41.7	54.1	21.3***	6.3
10	36.4	52.1	26.7**	2.4
Base rate = 50.9% (N=629)				

Note: Significant differences (chi-square or Fisher's exact test) between percent achieve cut score vs. percent below cut score denoted as follows: ***p<.001; **p<.01 (one-tailed).

[a]Centiles based on job applicant/student norms (N=1,036).

[b]Percentage of students performing at or above the median (49.2) as measured by Total Academy Performance Composite (mean of mean knowledge/skill ratings [T] and standardized class rank [T]).

[c]Percent gain=((% achieve cut score/base rate)-1)*100.

Gains in Job Performance

Table 9-A summarizes expected gains in job performance as measured by *supervisor ratings* of entry-level dispatchers' performance of important job duties.[12,13] The largest gains in job performance are expected at the highest test score levels: 31% gain at the 90th percentile and 46% gain at the 95th percentile. Expected gains in performance are somewhat lower in the 60 - 80 percentile range of scores (9% to 11% gain).

Table 9-B describes expected gains in job performance as measured by entry-level dispatchers' *self-ratings*.[14] Consistent with the results for supervisor ratings, significant gains in the numbers of satisfactory performers are associated with Battery scores ranging from the 60th to 95th percentiles. The greatest gains are expected at the 85 - 95 percentile range (33% to 34% gain), with smaller gains expected at the 60 - 80 percentile range of test scores (9% to 22% gain).

Gains in Employee Retention (Probation Success Rate)

Successful completion of probation (employee retention) is perhaps the most important criterion measure of dispatcher job success. As seen in Table 10, significant gains in probation success rate are associated with cut scores throughout the percentile range, increasing monotonically with test score percentile level.[15] Gains in probation success rate are expected to range from 1.7% (at the 20th percentile), up to 24.8% (at the 95th percentile). When gains are viewed in terms of the percentage reduction in employee turnover, the results are far more dramatic, ranging from 5.8% to 86.5%.

These results are illustrated graphically in Figure 4. Clearly, employers who select examinees scoring in the 80 - 90 percentile range are expected to realize substantially greater reductions in turnover than those who select examinees scoring at or below the 50th percentile (49%-87% vs. 6%-11%).

[12]Total composite of supervisor ratings on scales representing performance of job duties and demonstrated knowledge, skills, abilities, and work behaviors. The median composite rating was used as a "cut point" to classify dispatchers as above and below average.

[13]A higher range of Test Battery percentiles were selected for the table (30 - 95) because the validation sample score distribution was substantially skewed relative to the normative sample of job applicants and students (mean Battery Score for entry-level dispatchers was 58.9; the SD of scores was 6.3).

[14]Total composite of self-ratings on the same scales as used in the supervisory evaluation, again using the median composite rating to classify dispatchers as above and below average.

[15]Note that the mean Battery score was 57.8 and the SD was 7.0 for the validation sample, and none of the examinees in the sample scored as low as the 10th percentile.

Table 9: Expected Gain in Job Performance
Associated with Test Battery Performance Level

A. Supervisor Ratings

Total Battery Performance[a]		Percent Students Rated Above Average[b]		% Gain vs. base rate[c]
Centile	T Score	Achieve cut score	Below cut score	
95	65.1	72.0	44.7*	46.0
90	62.3	64.7	41.2**	31.2
85	60.4	58.8	41.3*	19.3
80	58.5	53.7	43.9	8.9
70	55.9	54.5	38.3*	10.5
60	53.8	54.9	31.4*	11.4
50	51.4	51.5	33.3	4.5
40	47.8	51.1	22.2	3.7
30	45.2	50.3	0	2.0
Base rate = 49.3% (N=148)				

B. Self-Ratings

Total Battery Performance[a]		Percent Students Rated Above Average[b]		% Gain vs. base rate[c]
Centile	T Score	Achieve cut score	Below cut score	
95	65.1	68.0	46.8*	33.9
90	62.3	67.4	41.2**	32.7
85	60.4	68.3	35.2***	34.4
80	58.5	61.8	36.2**	21.7
70	65.9	59.1	31.7**	16.3
60	53.8	55.2	34.5*	8.7
50	51.4	52.1	40.0	2.6
40	47.8	52.0	28.6	2.4
30	45.2	-	-	-
Base rate = 50.8% (N=134)				

Note: Significant differences (chi-square or Fisher's exact test) between percent achieve cut score vs. percent below cut score denoted as follows: ***p<.001; **p<.01; *p<.05 (one-tailed).

[a]Centiles based on job applicant/student norms (N=1,036).

[b]Percentage of dispatchers with Total Composite Rating greater than or equal to the median for the validation sample (supervisor rating ≥3.37; self-rating ≥3.72).

Table 10
Expected Gain in Probation Success Rate
Associated with Test Battery Performance Level

Test Battery[a]		Percent Completing Probation[b]		% Gain vs. base rate[c]	% Reduction in Turnover[d]
Centile	T Score	Achieve cut score	Below cut score		
95	65.1	97.0	74.2**	24.8	86.5
90	62.3	95.3	70.2***	22.7	78.9
85	60.4	89.5	69.8***	15.2	52.9
80	58.5	88.7	67.0***	12.4	49.3
75	57.2	86.5	65.2***	11.3	39.5
70	55.9	83.6	66.7**	7.6	26.5
65	55.8	83.2	65.1**	7.1	24.7
60	53.8	82.4	64.3**	6.0	21.1
50	51.4	80.1	64.7*	3.1	10.8
40	47.8	80.1	52.6*	3.1	10.8
30	45.2	79.5	40.0**	2.3	8.1
20	41.7	79.0	33.3**	1.7	5.8
10	36.4	-	-	-	-
Base rate = 77.7% (N= 215); maximum possible gain=28.7%					

Note: Significant differences (chi-square or Fisher's exact test) between percent achieve cut score vs. percent below cut score denoted as follows: ***p<.001; **p<.01; *p<.05 (one-tailed).

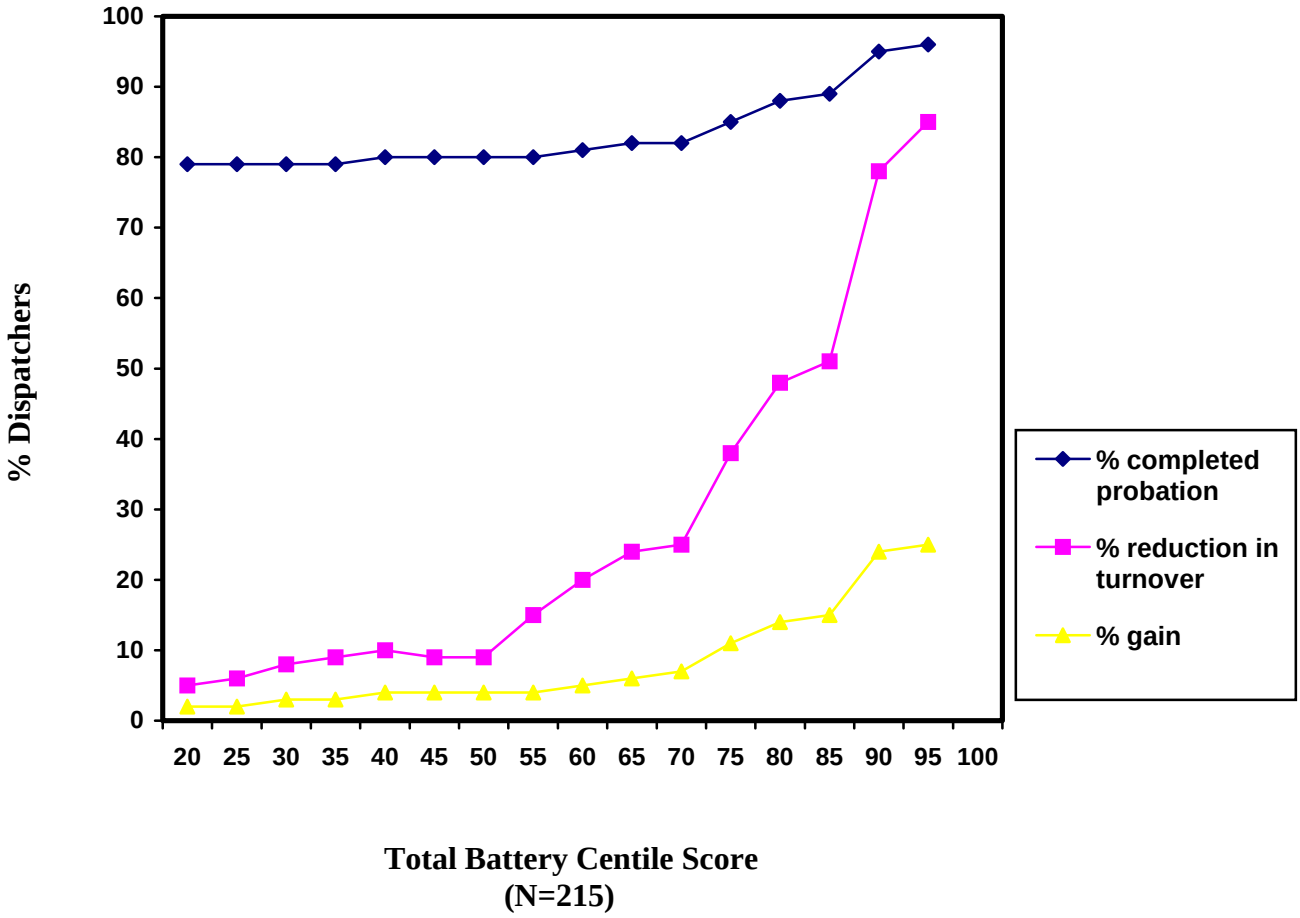
[a]Centiles based on job applicant/non-affiliated student norms (N=1,036).

[b]Completed probation=1; resigned or terminated while performance was unsatisfactory due to inadequate job knowledge, skills, or abilities=0.

[c]Percent gain=((% achieve cut score/base rate)-1)*100.

[d]Turnover base rate=100-base rate. % Reduction in Turnover=(Turnover base rate-(100-% Completing Probation and Achieve cut score))/Turnover base rate*100.

Figure 4
Gain in Probation Success Rate
Associated with Test Battery Performance Level



Establishing a Minimum Passing Score

Generally speaking, the Test Battery will provide the most benefit to employers when candidates are ranked on the basis of their total scores and selected "top-down," rather than using the Battery as an initial pass/fail "hurdle" with no further consideration being given to candidates' relative performance. This principle applies to most empirically validated ability tests (e.g., see SIOP Principles, 1987, pp. 32-33). By the same token, even with top-down selection there comes a point in the distribution of test scores where the employer must "draw a line" above which candidates are considered to be at least minimally qualified for the job, and below which candidates are considered unqualified.

At present, there is no single recommended minimum passing score for the Test Battery. However, ranges of scores have been found to reflect minimum competency in several different performance contexts. Test users are advised to consider the below alternative scores and rationales when establishing a minimum passing standard for the Test Battery.

Recommended Minimum Passing Score Range:
48 – 57

Rationale for Alternative Passing Scores

Alternative passing scores were identified on the basis of two different approaches, one *norm-referenced* and the other *criterion-related* (see Cascio & Barrett, 1988). Each approach is aimed at identifying the Total Battery score that corresponds to "minimally acceptable performance."

Norm-Referenced Approach. The *norm-referenced* approach is founded upon the base rate for unsatisfactory job performance among job incumbents. Under this rationale, alternative cut scores were established by first identifying the percentage of entry-level dispatchers who generally perform at an unsatisfactory level, as reflected by different performance indices. Then, the Test Battery scores were identified that would screen out the same percentages of entry-level dispatchers.

Total Battery scores of 47.5 and 49.5 were identified as points in the test score distribution that would screen entry-level dispatchers in the same proportions as the base rates for unsatisfactory job performance among entry-level dispatchers. The former score reflects unsatisfactory performance with regard to demonstrated knowledge, skills and abilities (base rate= 3.9%), while the latter represents total job performance (base rate=9.2%).[16]

[16]Based on a validation sample of 150 dispatchers with 12 months or less experience. Unsatisfactory performance was defined as a supervisor composite rating of 2.5 or lower. Total job performance rating composite was comprised of 20 job duty effectiveness ratings, 18 KSA ratings, 8 conscientiousness/performance-related work behavior ratings, and two global ratings of overall job performance.

This strategy for establishing passing scores is consistent with the federal *Uniform Guidelines on Employee Selection Procedures* (EEOC, et al, 1978, Sec. 5.H.) which state:

> "Where cutoff scores are used, they should normally be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force."

Criterion-Related Approach. The *criterion-related* approach is based upon the empirical relationship between test scores and academy/job performance measures. A statistical procedure (linear regression analysis) was used to derive conversion formulas to "translate" Test Battery scores into corresponding academy and job performance criterion scores so that the Battery score corresponding to *minimally acceptable* performance could be identified.

The resulting *criterion-related* passing scores are shown in Table 11. They range from 48.6 (the mid-point between the predicted mean Battery scores for academy success/failure groups) to 56.7 (corresponding to a supervisor Total Performance composite rating of 2.6). The mean of the derived criterion-related passing scores is 53.6. The grand mean of the seven norm-referenced and criterion-related passing scores is 52.2.

Passing Rates Resulting from Alternative Minimum Passing Scores

The various alternative passing scores were applied to the normative sample of job applicants and students, and the resulting passing rates were tabulated for the total sample, as well as by racial/ethnic and gender groups. Passing rates were also computed for selected percentile points, ranging from 20 to 99. The results are presented in Table 12.

As seen in the table, the grand mean of norm-referenced and criterion-referenced passing scores (52.2) would be expected to result in an overall passing rate of 46.1%; the lowest derived passing score, 47.5 (Demonstrated KSA base rate, norm-referenced) would result in an overall passing rate of 61.3%; and the highest derived passing score, 56.7 (Supervisor Rating: Total Job Performance=2.6) would result in an overall passing rate of 27.0%.

The passing rates for males and females were found to be comparable, where the female/male passing rate ratios are 80% or higher in most instances. The relative passing rates within racial/ethnic subgroups follow traditional patterns, where the rates for minority group members tend to be lower than that for the majority group (whites).[17] It is encouraging, however, that even at the higher percentile ranges, minority groups are represented by substantial percentages of examinees.

[17] For example, at the grand mean of norm-referenced and criterion-related scores (52.2), 31.2% of Blacks passed, 35.0% of Hispanics passed, and 55.6% of Whites passed (Black:White ratio=56%; Hispanic:White ratio=63%). Of the reported alternative passing scores, it is only at the 20th percentile that the Black:White and Hispanic:White passing score ratios meet the "4/5ths rule of thumb" espoused by the federal EEOC in judging the comparability of subgroup passing rates (EEOC, et al., 1978).

Table 11
Test Battery Scores Corresponding to Minimum Acceptable
Academy Performance and Job Performance

Performance Index	Total Battery T-Score
Norm-Referenced Score:	
Total Job Performance[a]	49.5
Demonstrated Job Knowledge, Skills, Abilities[b]	47.5
Criterion-Related Score:	
Basic Academy Success/Failure[c]	48.6
Basic Academy Performance[d]	51.2
Supervisor Rating[e]	56.7
Self-Rating[f]	55.2
Probation Success/Failure[g]	56.5
Mean of criterion-related scores:	53.6

[a]Test Battery score that would screen out entry-level dispatchers in accordance with the base rate for unsatisfactory Total Job Performance (9.2%; N=150) as measured by a composite of supervisor ratings, including: Mean of: (a) mean of global effectiveness and relative performance ratings; (b) mean of 20 job duty effectiveness ratings; (c) mean of 18 KSA ratings; and (d) mean of 8 work behavior ratings related to conscientiousness and performance under stress.

[b]Test Battery score that would screen out entry-level dispatchers in accordance with the base rate for unsatisfactory demonstrated job knowledge, skills, and abilities (3.9%; N=150) as measured by the mean of supervisor ratings on 18 KSA scales.

[c]Test Battery score corresponding to mid-point between mean predicted Battery scores for successful students vs. unsuccessful students, as derived via regression of Battery scores onto academy success/failure index; N=627. This procedure is equivalent to the "contrasting groups" method, wherein the mid-point between the means of unsuccessful and successful examinees is computed.

[d]Test Battery score corresponding to basic academy knowledge/skill rating of 2.6, as derived via regression of Battery scores onto academy ratings; N=526.

[e]Test Battery score corresponding to Total Job Performance composite supervisor rating of 2.6, as derived via regression of Battery scores onto ratings; N=148.

[f]Test Battery score corresponding to Total Job Performance composite self-rating of 2.6, as derived via regression of Battery scores onto ratings; N=134.

[g]Test Battery score corresponding to mid-point between mean predicted Battery scores for dispatchers who successfully completed probation vs. those who failed probation due to inadequate performance related to job knowledge, skills, or abilities, as derived via regression of Battery scores onto success/failure index; N=215. This procedure is equivalent to the "contrasting groups" method.

Table 12
Applicant Passing Rates Expected to Result from
Alternative Passing Scores

Total T-Score [a]	Rationale	Passing Rate (%)					
		Total	Black	Hispanic	White	Female	Male
47.5	Demonstrated KSA base rate (norm-referenced score)	61.3	44.9	46.2	72.2	60.5	63.4
48.6	Basic academy success/failure (mid-point between means)	58.1	41.4	44.8	68.6	57.7	59.4
49.5	Total job performance base rate (norm-referenced score)	54.5	38.8	42.0	64.2	53.6	57.0
51.2	Basic academy performance (knowledge/skill rating=2.6)	50.9	36.1	37.8	60.2	49.7	54.0
52.2	Grand mean of norm-referenced and criterion-related scores	46.1	31.2	35.0	55.6	45.3	48.3
53.6	Mean criterion-related score	41.2	28.9	28.7	49.1	40.3	43.6
55.2	Self-Rating (Total Performance=2.6)	32.5	22.4	24.5	40.9	32.2	38.6
56.7	Supervisor Rating (Total Performance=2.6)	27.0	17.1	17.5	33.0	29.4	31.9
41.7	20th percentile	80.1	*70.7+*	*69.9+*	86.9	79.8	81.2
45.2	30th percentile	70.2	57.8	56.6	79.4	69.6	71.8
47.8	40th percentile	60.2	43.3	45.5	71.1	59.8	61.4
51.4	50th percentile	50.1	35.7	47.8	59.1	49.1	52.7
53.8	60th percentile	40.4	29.5	27.3	48.6	39.5	43.0
55.9	70th percentile	30.0	20.6	21.0	36.9	(28.4)	36.6
57.2	75th percentile	25.0	15.6	16.8	31.0	24.5	29.2
58.5	80th percentile	20.0	9.9	15.4	24.6	(18.5)	23.8
60.4	85th percentile	15.0	7.2	13.3	18.3	(13.8)	18.8
62.3	90th percentile	10.0	5.7	7.7	11.8	10.2	9.7
65.1	95th percentile	5.0	2.7	3.5	6.1	(4.5)	6.4
68.3	99th percentile	1.0	0.8	0.7	1.3	1.4	0.7
	No. Examinees	1,036	263	143	558	737	298

Note: Passing rates are for job applicants and non-affiliated basic academy students (N=1,036). Racial/ethnic minority group passing rates denoted by "+" are at least 80% of the passing rate for whites. All passing rates for females are at least 80% of the rates for males, except those in parentheses.

[a]Total battery T-score normed to applicant/student sample (N=1,036); test scores adjusted for guessing.

Combining Test Battery Scores with Other Assessments

When multiple assessment procedures are used in the employee selection process, there are a number of alternative approaches to combining or otherwise using the scores yielded by each evaluation component. Under one approach, all assessment procedures may be given to all examinees and their scores combined before ranking candidates and/or setting a minimum passing score. This is a *compensatory approach* in that examinees are given the opportunity to compensate for relatively low performance on initial assessments by scoring relatively higher on subsequent assessments. On the other hand, practical limitations (e.g., time, staff, cost), and measurement concerns (e.g., ensuring minimum competence in essential areas) often lead employers to use a *noncompensatory* or "multiple hurdles" approach in which each successive assessment procedure is scored using a minimum passing score and only those examinees passing the first procedure are admitted to the next one.

Whatever strategy is chosen, test users are encouraged to take full advantage of the potential benefits offered by the Test Battery by combining examinees' Total scores with scores obtained on other pre-employment assessment procedures, such as practical exams, oral interviews, etc. An example procedure for combining Battery scores with those yielded by other procedures is outlined below (corresponding computational formulas are given in the Appendix).

Example Procedure for Combining Test Battery Scores with Other Assessments:

1. Standardize Test Battery scores and scores resulting from other procedures so that the means and SDs are equal;

2. Determine the proportional weight that each component of the assessment process should receive (e.g., Test Battery 50%, Oral Examination 50%). In assigning the weights, consideration should be given to the relative importance and scope of abilities measured by each procedure, as well as the validity and reliability of the assessment scores;

3. Compute a weighted total assessment score (sum the products of the proportional weights and the standard scores);

4. Rank-order candidates on the basis of their total assessment scores. Optimal utility will be obtained if candidates are selected according to overall rank.

Note: "Selection" in this case means making a *conditional offer* of employment, subject to passing a medical examination and background investigation, as required by POST Regulation 1018(c).

Score Banding: An Alterative Approach to Using Test Scores

Score banding is a procedure that has been suggested as a means of using test scores in light of the measurement error that is inherent in all tests; i.e., the imprecision with which test scores represent examinees' true abilities.

While there are a variety of approaches to banding (Cascio, et al, 1991; Sackett and Roth, 1991) the basic premise is that examinees are grouped within ranges of test scores, or bands. The width of the score band (range of highest to lowest scores in the group) is based upon a statistical index of measurement error (SEM). The score bands may be fixed or may be recomputed after each applicant is selected (sliding bands). Examinees within a score band are considered to be equivalent and are selected through a variety of methods (e.g., strict top-down selection, within-group percentiles, random selection, nonrandom selection).

Proponents of score banding (Cascio, et al., 1991) cite the advantage of increased work force diversity (reduced adverse impact upon protected examinee groups). Critics of banding procedures (Schmidt, 1991) point out a fundamental inconsistency between banding methods and classical selection theory which holds that, given a linear relationship between test scores and job performance, groups of higher scoring applicants will be expected to perform better than lower scoring groups in the long run, even if differences between adjacent scoring individuals' test scores are not statistically significant.

The test score reports presently generated for the POST Test Battery do not incorporate score bands. However, information regarding the degree of measurement error in test scores (SEM) is reported which enables users to apply banding methods.

Test users with an interest in applying score banding methods to the results of a test administration may contact POST for assistance.

OVERVIEW OF TEST DEVELOPMENT

The Test Battery was developed through an extensive three-phase process that entailed: (1) job analysis, (2) test design, and (3) test construction. The developmental process is summarized below; further details are provided in the technical validation report.

Job Analysis Foundation

The Battery is founded upon the results of a statewide job analysis of the public safety dispatcher occupation (Weiner, 1991). The job analysis was truly a statewide effort in which over 1,000 dispatchers and supervisors participated, representing over 160 agencies. Job information was collected through a variety of modes, including site visits, group interviews/workshops, and survey instruments. The job information included: (a) important duties performed by a majority of dispatchers throughout California; (b) knowledge, skills, abilities and traits (KSATs) essential for successful performance of dispatcher duties which are suitable for development in training or necessary for entry-level candidates to possess before hire; and (c) KSAT-duty linkages.

The identified essential dispatcher abilities were taken from a well-established taxonomy of human performance (Fleishman & Quaintance, 1984). The Fleishman taxonomy consists of a wide spectrum of cognitive, psychomotor, sensorimotor, and physical abilities (52 in total) that were derived through extensive factor analytic research spanning several decades (Guilford, 1967; French, Ekstrom & Price, 1963; Fleishman, 1964). The cognitive abilities in the taxonomy are consistent with other summaries of abilities measurement (Nunnally, 1978; Carroll, 1993).[18]

Tables 13 and 14 summarize the identified dispatcher duties and essential abilities. Figure 5 illustrates the extent to which abilities measured by the Test Battery were rated by subject matter experts as essential for the successful performance of dispatcher duties.

---

[18] An *ability* as measured by the POST Battery is operationally defined as "*... a general, underlying capacity which enables or limits the performance of a wide variety of tasks."* This definition was adapted from Fleishman and Quaintance (1984) and is consistent with the more recent work of Carroll (1993), who devotes the entire first chapter of his book to defining and explicating the term *ability*. An example of an ability is *Reasoning*: the ability to apply general rules to specific problems to come up with logical answers; the ability to combine separate pieces of information or specific answers to problems, to form general rules or conclusions; and the ability to correctly follow a given rule or set of rules to arrange things or actions in a certain order. Reasoning ability underlies the performance of a broad range of dispatching tasks, such as following rules and procedures to handle 9-1-1 calls; evaluating, classifying, and "prioritizing" complaints and incidents; and assigning the appropriate types and numbers of field units to calls for service.

Abilities defined in this context are distinct from *skills* in that the latter are much more task-specific in nature and are, thus, more likely to be amenable to training. Examples of dispatcher skills include typing or using a computer keyboard, and broadcasting information over the radio. Abilities are also distinct from *traits*, as defined in the POST research, in that traits do not reflect capabilities so much as they reflect styles of behavior (i.e., traits reflect more of what an individual *will* do than what he or she *can* do). An example of an essential dispatcher trait is *Emotional Control*.

Table 13
Dispatcher Job Duties

- Receiving Complaints and Requests for Service
- Communicating with Difficult Callers
- Obtaining Complaint-Dispatching Information
- Evaluating and Summarizing Complaint-Dispatching Information
- Processing Complaint-Dispatching Information
- Advising the Public
- Providing Information to the Public and Other Agencies
- Monitoring and Responding to Radio and Emergency Systems
- Keeping Track of Field Units and Complaints/Incidents
- Dispatching Field Units
- Contacting Other Agencies
- Providing Information to Other Agencies
- Querying Data Bases
- Recordkeeping
- Maintaining Resource Materials
- Testifying in Court
- Performing Office Duties
- Providing Training

Note: Job duties represent 121 "core" tasks; i.e., important tasks that are performed by a majority of dispatchers in California. On average, 85% of incumbent dispatchers reported performing each task; the mean task frequency rating was 6.0 (performed more than once per week); 86% of supervisors rated each task as part of the dispatcher job; and the mean task importance rating was 3.8 (between Important and Very Important); see 1991 POST Dispatcher Job Analysis, Component 1: Job Task Analysis, for further information.

Table 14
Essential Dispatcher Abilities

COGNITIVE:

VERBAL

*Oral Comprehension
*Written Comprehension
Oral Expression
*Written Expression
Fluency of Ideas

REASONING

*Deductive Reasoning
*Inductive Reasoning
*Information Ordering

MEMORY

*Memorization

PERCEPTUAL

*Perceptual Speed
*Time Sharing
Selective Attention
Speed of Closure

PSYCHOMOTOR:

MANUAL DEXTERITY/SPEED

Multilimb Coordination
Finger Dexterity
Response Orientation
Reaction Time

SENSORIMOTOR:

VISION

Near Vision

HEARING

General Hearing
Auditory Attention
Speech Hearing

SPEECH

Speech Clarity

*Targeted for measurement by POST Entry-Level Dispatcher Selection Test Battery.

Note: On average, 99% of supervisors rated each ability as relevant to dispatcher work; 71% rated each ability as necessary before hire; and the mean importance rating was 4.2 (Very Important); see 1991 POST Dispatcher Job Analysis, Component 2: Analysis of Job Requirements.

Figure 5
Job Duty Linkages:
Abilities Measured by the POST Dispatcher Selection Test Battery

Ability	JOB DUTY																	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Oral Comprehension	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X
Written Comprehension		X		X	X			X	X			X	X	X	X	X	X	X
Written Expression	X			X	X				X					X	X			X
Deductive Reasoning	X	X	X	X		X	X			X								
Inductive Reasoning	X	X	X	X		X				X								
Information Ordering	X	X	X	X	X				X	X		X		X				X
Memorization	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X
Perceptual Speed/ Accuracy	X		X	X					X	X		X	X					
Time Sharing	X			X				X	X	X	X	X						

Note: "X" denotes essential ability for performance of at least one "core" task within a job duty, as rated by a majority of subject matter experts (see 1991 POST Dispatcher Job Analysis, Component 3: KSAT Linkage Analysis.

Job Duties:

1. Receiving Complaints and Requests for Service
2. Communicating with Difficult Callers
3. Obtaining Complaint-Dispatching Information
4. Evaluating and Summarizing Complaint-Dispatching Information
5. Processing Complaint-Dispatching Information
6. Advising the Public
7. Providing Information to the Public and Other Agencies
8. Monitoring and Responding to Radio and Emergency Systems
9. Keeping Track of Field Units and Complaints/Incidents
10. Dispatching Field Units
11. Contacting Other Agencies
12. Providing Information to Other Agencies
13. Querying Data Bases
14. Recordkeeping
15. Maintaining Resource Materials
16. Testifying in Court
17. Performing Office Duties
18. Providing Training

Test Design

The tests were developed according to detailed design specifications. The test specifications incorporated results of the statewide job analysis along with a review of published research literature and tests and a supplemental analysis of dispatcher reading and writing demands.

Review of Literature and Published Instruments

A review was made of cognitive ability testing and measurement literature, published tests and supporting validation evidence, commercial test publisher catalogs, local agency testing procedures, and previous POST research on reading and writing ability testing for entry-level peace officers. Major contributing sources included: The *Manual for Kit of Factor-Referenced Cognitive Tests* (Ekstrom, French & Harman, 1976); the *Mental Measurements Yearbook* (published by Buros, 1992, 1990, 1989, 1972); *Tests in Print* (Buros, Vol 2, 1974); and the *Handbook of Human Abilities* (Fleishman & Reilly, 1992).

Supplemental Analyses of Reading and Writing Demands

Additional job information was collected with regard to dispatcher reading and writing demands to support the development of tests of these abilities.

Readability Analysis. A readability analysis was performed on a sample of passages extracted from manuals, memos, bulletins and sections from legal codes commonly used by dispatchers. Two different indices were used to gauge the reading level of the passages. One index estimates reading level as a function of the number of polysyllable words in a passage (the SMOG index, McLaughlin, 1960). The other index is a function of average sentence length and word syllables (Flesch, 1951). The overall average reading grade levels identified by the two respective indices were 13.2 and 12.3.[19]

Analysis of Dispatcher Writing Demands. A panel of subject matter experts (supervisors and managers) reviewed dispatcher-writing tasks identified in the 1991 POST Dispatcher Job Analysis, sample documents written by dispatchers, and a survey of potential writing problems.[20] As a result of the review, it was concluded that the central concern in assessing writing ability for entry-level dispatcher candidates should be with sentence construction errors that lead to ambiguity in communication, rather than mechanical problems (e.g., spelling, punctuation) or with problems pertaining to writing lengthy documents (e.g., organization and analysis of information).

---

[19]Analyses of entry-level peace officer resource materials yielded similar results: overall mean readability levels of 13.4 and 12.4 were obtained with the SMOG and Flesch indices, respectively (Honey & Kohls, 1981, p. 32).

[20]The writing problems were based on the above 1981 POST entry-level peace officer reading and writing test research. They generally involve ambiguity in sentence construction resulting from unclear references to nouns and objects (i.e., indefinite, ambiguous, and vague reference; incomplete comparison; incomplete sentence; number shift), and confusing use of modifying clauses (misplaced, dangling, and squinting modifier; split construction).

Test Specifications

Design specifications were developed for tests representing those abilities that were feasible to be both assessed in a group setting and objectively scored (a practical requirement of the Battery).[21] The specifications incorporated the literature review and job analysis results, along with guidelines for test design given by Millman and Greene (1989).

Test features that were specified included: (a) the principal ability to be measured; (b) the types of information to be presented to the examinee, i.e., the stimulus; (c) how the information will be presented, i.e., written or oral media; (d) the tasks that the examinee is to perform in order to give the correct response; (e) how the examinee's performance will be scored; (f) any factors explicitly intended to affect performance on the test, such as speed and complexity of information presented, speed and complexity of tasks to be performed, complexity of questions asked, and reading level (calibrated for the *Reading Comprehension* test); and (g) alternative item formats.

Test Construction

The tests were constructed in a series of steps that included item writing and protocol development, focus group review, revision and further development of experimental test forms, pilot test administration, psychometric analyses, and assembly of final experimental test forms for empirical validation research.

Item Writing

POST staff developed test protocols, scenarios, and questions with the assistance of an ad hoc test development committee.[22] Target numbers of items to be developed were established for each test in consideration of the minimum number that would be needed to build a reliable scale, the number of items expected to be discarded for psychometric reasons after experimental testing, and practical concerns regarding testing time. The draft test items were assembled into mock test booklets, examinee instructions were drafted for each test, and initial draft recordings were made for the audio-based tests.

[21]A number of important dispatcher abilities were identified that would necessitate individualized assessment and/or subjective scoring procedures and thus, were not targeted for measurement by the Battery. These include Oral Expression, Fluency of Ideas, Manual Speed and Dexterity (Response Orientation, Reaction Time, Multilimb Coordination, Finger Dexterity), and the various sensorimotor abilities (Speech, Vision and Hearing). These abilities are recommended for assessment through other means.

[22]Managers and supervisors who had previous experience in dispatcher selection and evaluation.

Focus Group Review

The experimental tests were administered to small groups of newly hired dispatchers, experienced dispatchers and supervisors in focus group sessions. In each group session, participants were first briefed on the purpose of the tests and were given general instructions for completing the tests and documenting questions or comments. Participants then completed each test, as it would be given operationally. After each test, participants were asked to discuss their strategies for completing the tests along with various issues pertaining to test quality.[23] The focus group review sessions yielded valuable information, which led to substantial modifications of the tests.

Experimental Test Forms

Following the focus group reviews, POST staff edited the test items, scenarios, and instructions, and assembled experimental test booklets and response sheets. Test items were selected for the experimental test forms on the basis of: (a) content representation, i.e., type of item, passage, task, problem, etc., (b) focus group feedback, (c) practical concerns regarding test administration time, i.e., keeping testing time to a minimum, and (d) psychometric concerns, e.g., including sufficient items to obtain a reliable measure of ability.

Pilot Test Administration

The experimental tests were administered in two separate batteries: one comprised of the audio tape-based tests, the other comprised of the paper-and-pencil tests. The audio-based tests were administered on an experimental basis to approximately 450 examinees, including students of the Public Safety Dispatcher's Basic Course and entry-level dispatcher job applicants.[24] The written test battery was administered experimentally to approximately 160 examinees under similar conditions.

---

[23]The following questions were addressed: Are the instructions clear and concise? Are the tasks to be performed straightforward? Are the test questions written clearly and at an appropriate level of difficulty? Is the keyed answer correct and are the distracters incorrect, yet plausible? How much time should be given to examinees? Are the tests packaged in such a way that examinees are able to complete the tests without undue complexity? Are the audio taped components clear and understandable? Any other comments, concerns, or suggestions for improvement?

[24]The academy students were briefed regarding the purpose of the test and were encouraged to exert maximal effort. The job applicant data were obtained under operational, "real-stakes" conditions in one testing session, and the remaining applicant data were obtained under experimental conditions in a voluntary testing session that was held prior to the "real" test. Examinees in the latter session were given instructions similar to those for the academy students and as an incentive to exert maximal effort, they were offered a short feedback report indicating their relative performance on the battery (many in this group completed both the experimental audio tests and written tests on the same day).

Psychometric Analyses

Examinees' responses to the experimental test items were analyzed using classical item analysis procedures (e.g., see Millman & Greene, 1989; Allen & Yen, 1979). The analyses of the power and hybrid test items focused on the following three psychometric properties, while the speeded test items were analyzed only with respect to the first:[25]

1) Difficulty: the percentage of examinees responding correctly to an item (p);

2) Discrimination: the item-total score point-biserial correlation (r), adjusted for spuriousness due to inclusion of the item in the total score (Henrysson, 1971);[26]

3) Distracter Effectiveness: the item response-total score point-biserial correlation (r_d) and percentage of examinees choosing the alternative (p_d).

Item Bias Analysis. The written tests were also analyzed to identify items that were potentially "biased" against racial/ethnic minorities (Blacks and Hispanics) using a statistical procedure known as Angoff's Transformed Item Difficulty (Delta Plot) Method (Angoff & Ford, 1973).[27] These tests load heavily upon Verbal and Reasoning abilities, and tests measuring these abilities have traditionally yielded substantial racial/ethnic group performance differences (e.g., see Jensen, 1980, Ch. 4). In this analysis, "bias" is suggested by instances where an item is relatively more difficult for a minority group relative to other items of the same type, as indicated by a statistical index (*d*).[28] Thus, the analysis goes beyond examining simple group differences in item difficulty by taking into account the relative functioning of the items (see Angoff, 1982 for an overview of this procedure and computational formulas).[29]

[25]The *Setting Priorities Test* items are answered in triads, where only one A, one B, and one C answer are possible for each triad. To remove the effects of item interdependence, the items are scored by triad, with each triad being treated as a 3-point item.

[26]The hybrid tests items were scored with omitted responses treated as missing (rather than incorrect) to control for any effect of speediness upon the point-biserials.

[27]This procedure is not appropriate for speeded tests.

[28]Technically speaking, items that fall a relatively large distance from the major axis of the ellipse formed by the bivariate plot of transformed item difficulties (e.g., Hispanic vs. White) are regarded as contributing to the item-by-group-interaction. These items are especially more difficult for one group than another group relative to other items and are likely to represent a different psychological meaning to the members of each group.

[29]The analysis of item bias is a highly technical procedure and there is considerable controversy among statistical experts with regard to the most effective procedures and experimental designs for detecting "true" bias, and there are many alternative approaches (Holland & Wainer, 1993; Berk, 1982). The purpose of the analysis in the present study was to detect bias in a relative rather than definitive sense, using a practical and well tried method in an attempt to minimize the potential for any bias of Verbal and Reasoning test items.

Item Selection

Final experimental test forms were cast by systematically selecting test items on the basis of psychometric, content, and practical considerations. The overall goal was to select the best functioning test items in order to reduce testing time without sacrificing psychometric quality. The principal aim in selecting items for the power and hybrid tests was to construct maximally reliable scales of moderate difficulty which yield wide and approximately normal distributions of test scores, with minimal truncation at the extremes (i.e., avoiding ceiling and floor effects).[30] The speeded tests were constructed with the aim of assembling homogeneous sets of items representing simple tasks (as reflected by high p values) which, when administered under restrictive time limits, create a moderately difficult test with a wide score distribution.

[30]Items were selected for the power and hybrid tests using the following general approach:

a) Rank items in descending order of r;
b) Flag items to be excluded from further consideration which meet any of the following conditions: negative or statistically nonsignificant r value; extreme p value; positive r_d in conjunction with substantial p_d; and for the written tests only, extreme Angoff *d* values or, in some instances substantially low minority group p values relative to those of the majority group.
c) Classify items by type, passage, scenario, or other significant content designation;
d) Select items representing important content areas, giving priority to items with highest r and moderate p values; some items with extreme (high and low) p values were selected.
e) Compute test-level statistics for the set of selected items, including the mean, SD, coefficient alpha, and total score distribution. In most instances, shorter alternative tests were assembled and compared to longer versions to assess the potential loss in measurement precision. The mean, SD and distribution of total scores were also examined for alternative selected sets of items in an effort to maximize the dispersion of examinees on the scale, while minimizing any ceiling effect. The decision to use a selected set of items was made on the balance of testing time vs. reliability of measurement.

PSYCHOMETRIC CHARACTERISTICS

Summary of Test Properties

Table 15 presents characteristics of the tests based upon all available data obtained in experimental administrations of the Battery to job applicants and basic academy students. The sample of approximately 1,500 examinees is comprised largely of applicants and non-affiliated students (N=1,039 to 1,092); the remaining examinees were student/dispatchers with varying amounts of prior experience. The summary information includes test means, SDs and reliabilities,[31] as well as item difficulty (p) and point-biserial (r, unadjusted) means, SDs, and ranges for each test.

The test means are fairly well centered in the middle to upper-middle range of possible scores and the magnitudes of the SDs indicate that examinees vary substantially in their performance on the tests. The item difficulties fall in a fairly wide range, from somewhat difficult to very easy; and the point-biserials for the power tests are all positive and significant ($p<.01$), with mean values ranging from .353 (*Call-Taking*) to .577 (*Assigning Field Units*) indicating that the items distinguish well between examinees of varying ability levels.

In general, the tests exhibit adequate reliability. The power test reliabilities range from .652 (*Clarity*) to .884 (*Assigning Field Units*), while estimates for the speeded and hybrid tests range from .633 (*Recalling Facts and Details*) to .785 (*Checking and Listening*). Guilford and Fruchter (1973, p. 436) suggest that reliabilities of .60 are sufficient for component tests in a battery, and that not much will be gained by achieving single test reliabilities of greater magnitude when the battery is scored as a composite.

Test Intercorrelations

Table 16 displays the correlations between the tests. Scores on all of the tests are shown to be positively and significantly correlated, ranging from .30 (*Checking Coded Information* with *Setting Priorities*) to .60 (*Call-Taking* with *Oral Directions*).

[31] Reliabilities are alpha coefficients, with the following exceptions: Spearman-Brown estimates are reported for *Checking Coded Information* (based on the intercorrelation between parts 1-3) and *Checking and Listening-Part 1* (based on the correlation with scores on the similar perceptual speed test, *Checking Coded Information*). See Guilford and Fruchter (1973) for further information regarding these reliability estimation procedures.

Table 15
Summary of Test Characteristics

Test	No. Items (max score)	Raw Score			Item Difficulty			Item point-biserial				
		Mean	SD	N	Mean	SD	Range	Mean	SD	Range	Reliability[a]	SEM
1. Public Safety Bulletin	15	10.21	2.91	1534	.681	.163	.33-.96	.438	.087	.26-.56	.706	1.58
2. Assigning Field Units	20 (100)	72.24	16.40	1535	.722	.146	.45-.95	.577	.155	.20-.74	.844	5.58
3. Evaluating Facts	15	10.15	2.94	1533	.677	.196	.22-.93	.458	.120	.28-.65	.731	1.52
4. Setting Priorities	15 (45)	34.09	7.42	1533	.758	.158	.27-.96	.512	.118	.27-.66	.801	3.31
5. Reading Comprehension	20	15.02	3.64	1534	.751	.107	.56-.90	.435	.072	.31-.56	.773	1.74
6. Clarity	15	12.16	2.30	1534	.811	.128	.44-.95	.408	.082	.23-.51	.652	1.36
7. Recalling Facts & Details	18	13.75	2.76	1482	.764	.104	.54-.50	.373	.061	.25-.48	.633	1.67
8. Call-Taking	25	20.30	3.31	1504	.812	.112	.62-.98	.353	.076	.16-.51	.707	1.79
9. Oral Directions	17	12.71	3.10	1502	.748	.131	.44-.92	.441	.060	.30-.54	.739	1.59
10. Checking Coded Info	60	37.47	8.84	1468	.625	.230	.13-.97		N/A		.747[b]	4.45
11. Checking & Listening												4.81
Part 1	80	22.03	8.54	1467	.275	.351	.01-.98		N/A		.683[c]	1.58
Part 2	25	20.87	4.52	1467	.835	.101	.57-.96	.501	.099	.34-.68	.878	5.06
Total	105	42.90	10.91	1467	.409	.310	.01-.96		N/A		.785[d]	

Note: Sample includes job applicants and non-affiliated academy students (N=1,039 to 1,092), as well as student/dispatchers with prior dispatching experience (N=428 to 443).

[a]Coefficient alpha, unless noted otherwise.

[b]Spearman-Brown estimate based on correlation between 3 separately timed parts.

[c]Spearman-Brown estimate based on correlation with test #6 (Checking Coded Information).

[d]Linear composite reliability estimate (cf. Nunnally, 1978, p. 248).

Table 16
Test Intercorrelations

Test	1	2	3	4	5	6	7	8	9	10	11
1. Public Safety Bulletin	--										
2. Assigning Field Units	.47	--									
3. Evaluating Facts	.36	.40	--								
4. Setting Priorities	.36	.49	.33	--							
5. Reading Comprehension	.48	.56	.50	.49	--						
6. Clarity	.35	.42	.43	.39	.55	--					
7. Recalling Facts & Details	.44	.40	.31	.35	.41	.35	--				
8. Call-Taking	.47	.51	.40	.47	.55	.42	.43	--			
9. Oral Directions	.52	.57	.39	.49	.57	.43	.48	.60	--		
10. Checking Coded Info.	.31	.37	.31	.30	.41	.34	.35	.45	.44	--	
11. Checking & Listening	.40	.47	.35	.41	.49	.37	.38	.51	.52	.57	--

Note: N=1,442 examinees, including job applicants and academy students (both non-affiliates and student/dispatchers). All correlations are significant ($p<.0001$, two-tailed).

Factor Structure. The test intercorrelations were subjected to a statistical procedure called "factor analysis" in an effort to assess the degree of commonality among the tests in distinguishing among examinees.[32] The results of the analysis indicated that examinees' performance on the 11 tests could be reasonably accounted for by four underlying ability factors consistent with the original test design specifications.

As seen in Table 17, the *Public Safety Bulletin* and *Recalling Facts and Details* tests were both found to load substantially on a single factor entitled **Memory**, while the loadings for these same tests were marginal on the remaining factors. The *Reading Comprehension, Clarity,* and *Evaluating Facts* tests loaded highest on a single factor, entitled **Verbal**.[33] The *Setting Priorities, Assigning Field Units, Call-Taking*, and *Oral Directions* tests loaded highest on a factor entitled **Reasoning**.[34] Finally, the *Checking Coded Information* and *Checking & Listening* tests loaded highest on a single **Perceptual** factor.

[32]A Principal Components analysis with Varimax rotation was performed, resulting in a 4-factor solution, which accounted for approximately two-thirds of the total variance in test scores. Four experimental tests were included in the analysis that was eventually dropped from the battery. These included a *Word Usage* test and a *Cloze* test, both of which loaded >.70 on the Verbal factor. Also included were two perceptual tests: *Associating Information With Incidents* and *Selective Listening*; these loaded in the .40-.50 range on both the Perceptual and Memory ability factors.

[33]The *Evaluating Facts* test was originally designed to measure Reasoning ability. It is hypothesized that the inductive ability required to answer these questions is primarily related to the analysis of relationships between verbal conditions, as opposed to abstract or mathematical conditions; hence, the high loading on the Verbal factor.

[34]The *Call-Taking* and *Following Oral Directions* tests were designed to measure Verbal ability (Oral Comprehension) and the tests load substantially on this factor (.35 and .30, respectively). However, performance on the tests was found to be a function of a combination of Verbal, Reasoning, Memory and Perceptual abilities, as evidenced by substantial loadings on each of these factors.

Table 17
Factor Structure of the Tests

Test	Verbal	Reasoning	Memory	Perceptual
1. Public Safety Bulletin	.25	.33*	.69**	.09
2. Assigning Field Units	.33*	.58**	.32*	.20
3. Evaluating Facts	.70**	.03	.26	.17
4. Setting Priorities	.23	.76**	.13	.13
5. Reading Comprehension	.62**	.45*	.24	.22
6. Clarity	.73**	.21	.11	.17
7. Recalling Facts & Details	.20	.23	.71**	.15
8. Call-Taking	.35*	.53**	.29	.34*
9. Oral Directions	.30*	.56**	.41*	.28
10. Checking Coded Info.	.24	.18	.07	.81**
11. Checking & Listening	.19	.42*	.15	.69**

**Largest factor loading.
* Factor loading ≥.30.

Note: N=1,442, including job applicants and students of the Public Safety Dispatcher's Basic Course (both non-affiliates and student/dispatchers). Values are factor loadings yielded by a Principal Components analysis with Varimax rotation. Four experimental tests were included in the analysis that were eventually dropped from the batter, including: a *Word Usage* test and a *Cloze* test, both of which loaded >.70 on the Verbal factor; and two perceptual tests, *Associating Information With Incidents* and *Selective Listening*, which loaded in the .40-.50 range on both the Perceptual and Memory ability factors.

VALIDITY

The Concept of Validity

Validity is an essential characteristic of all measurement instruments. The validity of a test refers to the appropriateness, meaningfulness, and usefulness of specific inferences made from test scores (AERA, APA & NCME, 1985, p. 9). It is the particular use of test scores that is supported by validity evidence, not the test itself.

Test validation is the process by which evidence is gathered to support the use of test scores. Three professionally and legally recognized approaches to establishing validity are the *content-related, criterion-related*, and *construct-related* strategies. In very broad terms, content-related validity is demonstrated to the extent that a test is shown to be comprised of items which are representative of a well-defined domain of content; criterion-related validity is demonstrated when test scores are systematically related with one or more criterion measures of performance; and construct-related validity evidence is demonstrated through research establishing that a test measures a theoretical psychological characteristic.

Despite the differences in alternative validation strategies, validity is regarded as a "unitary concept" and "...always refers to the degree to which evidence supports the inferences that are made from the scores" (AERA, et al, 1985; also, see Messick, 1995). Further detailed discussion of the concept of validity is provided in professional standards (APA, et al) and principles (SIOP, 1987), as well as federal legal guidelines (EEOC, et al, 1978).

Criterion-Related Validity Evidence for the Test Battery

A predictive criterion-related validation study of the POST Test Battery was conducted during 1993-1995. The Test Battery was administered to several hundred students of the POST Public Safety Dispatcher's Basic Course at 13 training institutions. Specially developed measures of the students' overall performance in the course were then collected and, for those students who were employed as dispatchers, measures of their subsequent job performance were also collected. Statistical analyses were then conducted examining empirical relationships between subjects' test scores and the measures of academy and job performance.

Table 18 presents descriptive characteristics of the validation research sample, including breakdowns by race/ethnicity, gender, education, assignment, shift worked, age, experience, and time between testing and criterion data collection.

The performance criterion measures developed by POST expressly for the validation research are summarized in Table 19. Further details regarding the criterion measures are provided in the technical report.

Table 18
Validation Sample Characteristics

	Academy Performance Criterion Data		Job Performance Criterion Data					
			Supervisor Ratings		Self-Ratings		Probation Success/Failure	
	Freq	Pct	Freq	Pct	Freq	Pct	Freq	Pct
Race/Ethnicity[a]								
American Indian	12	1.8%	5	3.2%	3	2.2%	5	2.1%
Asian	11	1.6%	4	2.6%	4	2.9%	6	2.6%
Black	74	10.9%	18	11.6%	13	9.5%	25	10.7%
Hispanic	114	16.7%	27	17.4%	30	21.9%	33	14.2%
Filipino	5	0.7%	-	-	-	-	3	1.3%
White	453	66.4%	98	63.2%	86	62.8%	156	67.0%
Other	13	1.9%	3	1.9%	1	0.7%	5	2.2%
Gender[b]								
Male	201	29.5%	38	24.5%	34	24.6%	69	29.6%
Female	481	70.5%	117	75.5%	104	75.4%	164	70.4%
Education[c]	N/A							
High School/GED			32	21.5%	28	20.3%	48	22.5%
1 year college			35	23.5%	34	24.6%	53	24.9%
AA/2 yr college			53	35.6%	49	35.5%	75	35.2%
BA/BS			26	17.4%	24	17.4%	33	15.5%
MA/MS			1	0.7%	1	0.7%	1	0.5%
Other			2	1.3%	2	1.4%	3	1.4%
Assignment[d]	N/A							
Call-Taking			20	13.3%	45	32.4%	87	37.3%
Dispatching			3	2.0%	64	46.0%	92	39.5%
Combined			47	31.3%	61	43.9%	84	36.1%
Rotate/Both			81	53.6%	8	5.8%	9	3.9%
Shift(s) Worked[e]	N/A							
Day			49	31.6%	45	32.4%	87	37.3%
Evening			66	42.6%	64	46.0%	92	39.5%
Night			69	44.5%	61	43.9%	84	36.1%
Relief			9	5.8%	8	5.8%	9	3.9%
Total	682		155		139		233	

Table 18 (continued)

[a]Not reported for 2 self-ratings.

[b]Not reported for 1 self-rating.

[c]Not available for academy performance sample; not reported for 6 supervisor ratings, 1 self-rating, 20 probation outcomes.

[d]Not available for academy performance sample; not reported for 4 supervisor ratings, 2 self-ratings, and 12 probation outcomes.

[e]Not available for academy performance sample; more than one category may be reported for an individual.

Validation Sample Characteristics

	Job Performance Criterion Data					
	Supervisor Ratings		Self-Ratings		Probation Success/Failure	
	Mean	SD	Mean	SD	Mean	SD
Age[a]	30.5	7.4	30.1	7.2	31.3	7.9
Dispatching experience when tested (months)	3.9	4.2	3.8	4.2	12.7	25.5
Months between testing and criterion data[b]	9.2	3.2	9.2	3.3	8.2	3.3

[a]Not reported for 5 supervisor ratings, 1 self-rating, 24 probation outcomes.

[b]Not reported for 2 supervisor ratings, 12 self-ratings, and 71 probation outcomes.

Table 19
Performance Criterion Measures

Basic Academy Performance

Academy Instructor Ratings. Students were rated at the conclusion of basic training with respect to: (a) overall levels of demonstrated knowledge and skill, and (b) overall class rank.[35] An overall academy success index was also constructed (successful completion=1, failure to complete the course=0). These data were collected for 682 students.

Job Performance

Supervisor Ratings. Students who were employed as dispatchers at the time they were tested in the academy were later rated by their immediate supervisors at the end of probation using POST-developed scales which covered: (a) effectiveness in performing important job duties, as identified in the 1991 POST Dispatcher Job Analysis; and (b) performance outcomes, including instances of commendable performance, complaints regarding poor performance, and inability to perform a critical job duty. Supervisor ratings were obtained for 155 entry-level dispatchers from over 100 agencies.

Self-Ratings. Dispatchers who were rated by their supervisors were asked to rate themselves using the same job effectiveness scales. The self-ratings were completed under conditions of strict confidentiality and were obtained for 139 entry-level dispatchers.

Probation Success/Failure (Employee Retention). Dispatchers were tracked throughout their probationary period and their overall success or failure in completing probation was coded. General reasons for failure were obtained and used to identify students who performed poorly for reasons that would be expected to be relevant to the Battery (e.g., inadequate job knowledge, skills or abilities). A quantitative index of success/failure was constructed by scoring probation outcomes as a dichotomy; i.e., successfully completed probation=1; resigned or terminated while performing unsatisfactorily for relevant reasons=0. Retention data were obtained for 233 dispatchers.

[35]Approximately 15% of the academy students were ranked within class on the basis of academy-specific curriculum test scores (instructor ratings were not available).

Summary of Empirical Validity Evidence

Table 20 summarizes the criterion-related validity evidence for the Test Battery. Total Battery scores were found to be significantly predictive of *basic academy performance*, subsequent *job performance*, and *employee retention*. Validities of .21 and .35 were obtained for basic academy completion/failure and total performance level, respectively. Correlations between Total Battery scores and total job performance were comparable for supervisor ratings and self-ratings, .28 and .24. A validity coefficient of .30 was obtained with overall probation success/failure.

Scores on each of the eleven tests, which comprise the Test Battery, were found to predict *academy performance* and subsequent *job performance level*, and eight of the tests were found to predict *employee retention*. The obtained significant validities range from .12 (*Clarity* test scores predicting academy completion/failure) to .33 (*Reading Comprehension* test scores predicting total performance in basic training). See Table 20.

Comparison of Findings with Other Research

The validity findings in the present study of the POST Dispatcher Test Battery are consistent with cumulative validity research which indicates that cognitive ability tests are valid predictors of performance in a wide range of occupations, as described below.

In one of the earliest published summaries of the predictive validity of employee aptitude tests in personnel selection (Ghiselli, 1973), the results of numerous studies conducted between 1920 and 1971 were aggregated for cognitive tests in predicting training success and job performance for a variety of occupational classifications, including: managerial, clerical, sales, protective, service, trades and crafts, industrial, vehicle operator, and sales clerk. Overall, cognitive ability tests were found to be predictive of both training success and job performance for all occupations; the overall mean validities were .39 and .22, respectively.[36]

Subsequent validity generalization studies have further demonstrated the wide-ranging validity of cognitive ability tests in predicting training success and job performance in thousands of jobs. (e.g., Hunter, 1980; Pearlman, Schmidt & Hunter, 1980; Hunter & Hunter, 1984; see Hunter, 1989, for a concise summary of these and other studies). Additional important findings of this research are that: (a) variations in the magnitudes of validities across studies of the same occupation may be attributed largely to statistical artifacts (e.g., sampling error, predictor range-restriction, criterion unreliability; see Schmidt & Hunter, 1977) and when adjustments are made for these artifacts, validity estimates increase markedly; (b) the validity of cognitive ability tests increases as job complexity increases; (c) cognitive ability tests are generally better predictors of job performance than alternative measures, such as biodata, personality, and interviews (Hunter & Hunter, 1984); and (d) such tests are fair (do not understate job performance) for minority applicants.

[36]Training success data were not available for sales occupations.

Table 20
Summary of Validity Evidence for the
POST Entry-Level Dispatcher Selection Test Battery

Tests[a]	Basic Academy		Job Performance					
			Supervisor Ratings		Self-Ratings		Probation	
	Pass/ Fail[b]	Total Perf[c]	Total Perf[d]	KSAs[e]	Unable to perf[f]	Total Perf[g]	Global Rating[h]	Pass/ Fail[i]
TOTAL BATTERY	.21***	.35***	.28**	.32***	-16*	.24**	.26**	.30***
1. Public Safety Bulletin (r M)	.15***	.30***	.18*	.14*	-14*	.18*	.17*	.16**
2. Assigning Field Units (v R m)	.13**	.29***	.18*	.20**	-.06	.22**	.16*	.20**
3. Evaluating Facts (V)	.14***	.23***	.13	.22**	-.04	.07	.08	.09
4. Setting Priorities (R)	.14**	.18***	.06	.06	-.16*	.01	.02	.14*
5. Reading Comprehension (V r)	.18***	.33***	.22**	.21**	-.21**	.15*	.17*	.13*
6. Clarity (V)	.12**	.24***	.12	.18*	-.11	.15*	.17*	.08
7. Recalling Facts & Details (M)	.15***	.21***	.12	.09	-.19*	.14*	.15*	.13*
8. Call-Taking (v R p)	.16***	.30***	.21**	.20**	-.11	.09	.09	.28***
9. Oral Directions (v R m)	.15***	.28***	.15*	.16*	-.11	.15*	.18*	.25***
10. Checking Coded Information (P)	.17***	.22***	.18*	.17*	-.01	.13	.17*	.28**
11. Checking & Listening (r P)	.14**	.25***	.25**	.32***	-.10	.19*	.20**	.26***

***p<.0001; **p<.01; *p<.05 (one-tailed).

[a]Abilities measured are shown in parentheses; V=verbal, R=reasoning, M=memory, P=perceptual; uppercase denotes major factor loading (>=.50); lower case denotes minor loading (>=.30).

[b]N=627 to 680 non-affiliated students and student/dispatchers. Completed Dispatcher's Basic Course=1; failed to complete for any reason=0.

[c]N=629 to 682 non-affiliated students and student/dispatchers. Mean of standardized mean knowledge/skill rating and standardized class rank (15% of students ranked within class on the basis of academy curriculum test scores; no ratings available).

[d]N=148 to 153 entry-level dispatchers. Mean of: (a) mean of global effectiveness and relative performance ratings; (b) mean of 20 job duty ratings; (c) mean of 18 KSA ratings; and (d) mean of 8 work behavior ratings related to conscientiousness and performance under stress.

[e]N=150 to 155 entry-level dispatchers. Mean of 18 KSA ratings.

[f]N=144 to 149 entry-level dispatchers. Note: negative correlation is desired direction: Any instances where the dispatcher was unable to perform a critical job duty due to inadequate knowledge, skills, abilities, or other characteristics, over the last month (1=yes, 0=no).

[g]N=134 to 139 entry-level dispatchers. Mean of: (a) mean of global effectiveness and relative performance ratings; (b) mean of 20 job duty ratings; (c) mean of 18 KSA ratings; and (d) mean of 8 work behavior ratings related to conscientiousness and performance under stress.

[h]N=134 to 139 entry-level dispatchers. Mean of global effectiveness and relative performance ratings.

[i]N=215 to 221 entry-level probationary dispatchers. Completed=1; Resigned or terminated while job performance was unsatisfactory due to inadequate job knowledge, skills, or abilities=0.

FAIRNESS

The Concept of Fairness

"Fairness" is not a test characteristic per se. Rather, it is a condition that may result from a combination of factors related to the test, the job, the employee population, and how the test scores are used. In its broadest sense, fairness is a social concept whose definition depends upon what one considers being fair (SIOP, 1985, p.18).

A more narrow conception of fairness is espoused by federal guidelines pertaining to employment testing; namely, the *Uniform Guidelines on Employee Selection Procedures* (EEOC, et al, 1978). These guidelines are given great deference by the courts in litigation involving employment test discrimination. *The Guidelines* define test fairness as follows:

> "When members of one race, sex, or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences in scores are not reflected in differences in a measure of job performance, use of the selection procedure may unfairly deny opportunities to members of the group that obtains the lower scores" [Sec. 14.B (8)].

Thus, according to this definition, a test is not necessarily unfair simply because one group of people systematically obtains lower test scores than another group. The fairness or unfairness of the test depends upon the joint relationship between test scores and some external measure of job performance. Also, from this definition it is possible for a test to be fair in some situations and not in others, depending upon the criterion measure of performance that is being considered.

This model of test fairness fits within the framework of "differential prediction" described in the *Standards for Educational and Psychological Testing* (AERA, et al, 1985, p. 12). According to the *Standards*, differential prediction exits when a single linear regression line is inadequate to describe test score predictions of a criterion measure for two different examinee groups. The adequacy of a single regression line is assessed by comparing regression parameters computed separately for each group (i.e., standard error of estimate, slope and intercept).

It should be noted that the literature for cognitive tests indicates that differential prediction is not supported for major racial/ethnic groups and there is no compelling research or theory to warrant the use of cognitive ability tests differently for different groups (SIOP, 1985, p. 18). Jensen (1980) provides an extensive review of issues, research methodologies, and results pertaining to test bias. On the basis of a review of major research pertaining to predictive bias of cognitive ability tests for racial/ethnic minorities, Jensen concludes that such tests are not biased (minorities' criterion performance is generally *over* predicted) and further, that the relatively few predictive studies that have identified test bias were most likely the result of sampling error.

Differential Prediction Analysis and Results for the Test Battery

POST Dispatcher Test Battery scores were examined with respect to fairness in the differential prediction framework outlined in the *Standards* and recognized by the *Guidelines*. A series of analyses were performed comparing test-criterion prediction equations yielded for racial/ethnic minority vs. majority groups and by gender. The analyses were limited to those test-criterion combinations for which significant validities were obtained for the total sample.

Procedure

A 3-step procedure was conducted comparing racial/ethnic minority vs. majority groups and males vs. females, as follows: (1) within-group **error variances** resulting from the overall regression of academy/job performance onto test scores were compared; (2) if no significant differences were detected, then the **slopes** of separate test-academy/job performance regression lines were compared; and (3), if significant slope differences were not detected, then the **intercepts** of separate test-academy/job performance regression lines were compared. The analytic procedure was based upon a multiple regression model described by Cohen and Cohen (1975) and is further described in the technical report.

The net effect of any group differences in regression parameters (error variances, slopes and intercepts) was examined by computing the mean difference between minority groups' actual academy/job performance and that predicted on the basis of the regression equation derived for the minority and majority groups combined; that is, a "residuals" analysis was performed. A positive value for the difference between actual academy/job performance and that predicted on the basis of Test Battery score (the residual) is indicative of *under* prediction; that is, the examinee's actual work capability is underrepresented by his/her performance on the test. A negative residual indicates the opposite; i.e., *over* prediction by Test Battery score.

Results

The results are summarized in Table 21. The table includes within-group descriptive statistics (mean and SD) for the Test Battery and criterion measures, mean predicted and residual criterion scores, within-group validities (r), and results of the analysis of minority vs. majority regression parameters (F-test of error variance differences, t-test of slope differences, and t-test of intercept differences).

Differential Prediction. Overall, *no instances of systematic under prediction* by Total Battery scores were found for racial/ethnic minorities (Blacks and Hispanics) or females. Similarly, males' academy performance, supervisor ratings, and probation success were not systematically under predicted by Battery scores. Males' *self-ratings* of job performance were *under* predicted, although the extent of under prediction was of little practical significance as their mean rating of 3.87 and their predicted mean rating of 3.70 both fall in the range of effective-to-very effective job performance.

Analyses of the individual tests are fully described in the technical report. The results indicated that the tests are not unfair to racial/ethnic minorities (Blacks, Hispanics) or females in predicting their academy performance or job performance. While significant differences in regression parameters were detected in a number of instances, there were *no instances of significant under prediction of racial/ethnic minorities' or females' academy or job performance* (Hispanics' academy performance was actually *over* predicted in two instances; males' self-ratings of job performance were *under* predicted in one instance).

Within-Group Validity. As seen in Table 21, Total Battery scores were found to be significantly predictive of *basic academy performance* and *job performance level* for all subgroups studied (Blacks, Hispanics, Whites, females, and males); significant validities range from .15 (predicting males' academy completion/failure) to .51 (predicting males' probation success/failure). *Employee retention* (probation success/failure) was predicted for males, females and Whites. The relatively low and non-significant correlations obtained for Blacks (.05) and Hispanics (.06) are inconclusive due to relatively small sample sizes (N=25 and 32).[37]

Within-group results for the individual tests in the Battery are reported in the technical report. Each test was found to significantly predict *basic academy performance* for all racial/ethnic and gender groups studied with one exception (the *Evaluating Facts* test was positively correlated with Hispanics' academy performance, however the correlations were not statistically significant).

With regard to *job performance* and *employee retention*, within-group test score predictions were usually positive, however, few were found to be statistically significant. In many instances, the nonsignificant within-group validities were of comparable or greater magnitude than significant validities yielded for the total sample. In addition to the obvious problem of statistical power, the fluxuations in within-group validities serve as a reminder that the correlation coefficient is subject to sampling error.

[37]A correlation of .05 based on a sample of 25 subjects has a 95% confidence interval ranging from -.342 to +. 397. Furthermore, when statistical significance is detected with a small sample, the result is considered conclusive in view of the small probability of obtaining such a finding by chance (e.g., .05). However, if significance is not detected, the result is considered equivocal due to the relatively high probability of making a Type II error (failing to reject the null hypothesis when it is true). For example, when N=30, there is only a 50% probability of detecting a significant correlation when the true value for the population r=. 30 and a .05 (one-tailed) significance level are used (see Cohen, 1988, p.101).

Table 21
Differential Prediction Analysis Results for the
POST Entry-Level Dispatcher Selection Test Battery

Criterion Measure	N	Test Score Mean	Test Score SD	Criterion Score SD	Criterion Score Mean Observed	Criterion Score Mean Predicted	Criterion Score Mean Residual[a]	r	Test of Regression Parameters F SE	Test of Regression Parameters t Slope	Test of Regression Parameters t Intercept
Basic Academy											
Pass/Fail[b] Black											
Hispanic	67	50.19	6.04	.122	.985	.958	.028	.40**	1.93***		
White	105	50.47	6.78	.214	.952	.959	-.007	.24**	1.75***		
Female	417	53.52	6.17	.160	.974	.976	-.003	.20***	N/A		
Male	442	52.67	6.22	.163	.973	.971	.001	.24***	1.24*		
	185	52.59	6.78	.178	.968	.971	-.0003	.15*	N/A		
Total Performance[c]											
Black											
Hispanic	67	50.16	8.72	9.83	49.62	48.89	0.73	3.37**	1.02	0.12	-0.43
White	107	50.84	9.91	9.88	48.19	49.15	-0.96	6.30**	1.04	0.91	1.24
Female	417	54.96	8.90	9.89	51.00	50.70	0.30	0.36***	N/A		
Male	444	53.79	8.99	9.68	49.80	50.26	-0.45	5.38***	1.27		
	185	53.61	9.77	10.55	51.28	50.19	1.09	9.29***	N/A		
Job Performance											
Supervisor Ratings[d]											
Black											
Hispanic	17	53.87	5.31	0.83	3.14	3.28	-0.14	.46*	1.78*		
White	25	54.33	4.30	0.48	3.45	3.30	0.15	.34*	1.57	-0.09	-1.18
Female	96	56.77	3.94	0.58	3.39	3.39	0.00	.24*	N/A		
Male	112	55.87	4.52	0.62	3.34	3.35	-0.01	.27*	1.41	-0.28	-0.38
	36	56.44	3.87	0.53	3.41	3.38	0.03	.32*	N/A		
Self-Ratings[d]											
Black	13	53.90	5.68	0.44	3.65	3.64	0.00	.34	1.06	-0.24	-0.13
Hispanic	23	54.47	4.39	0.41	3.74	3.66	0.08	.43*	1.09	-0.91	-1.10
White	87	56.92	3.91	0.41	3.70	3.71	-0.01	.20*	N/A		
Female	102	56.10	4.54	0.38	3.64	3.69	-0.05	.24***	1.44	-0.24	-2.80**
Male	32	56.59	3.88	0.45	3.87	3.70	0.17	.21	N/A		
Probation Pass/Fail[e]											
Black											
Hispanic	25	52.76	5.05	.476	.680	.665	.015	.06	1.47	1.78	-0.49
White	32	54.16	4.56	.440	.750	.706	.044	.05	1.26	1.94	-0.84
Female	153	55.87	4.77	.441	.739	.755	-.017	.42***	N/A		
Male	161	55.44	4.82	.433	.752	.743	.009	.23**	1.07	-2.07	
	66	54.59	5.00	.463	.697	.718	-.021	.51***	N/A		

***p<.001; **p<.01; *p<.05 (one-tailed for r; two-tailed for regression parameters). Significance tests are F-ratio for SE^2 and t-test for slopes, intercepts and residuals.

[a]Residual=observed minus predicted criterion score based on common regression line.

[b]Successful completion of basic course=1; failure to complete basic course for any reason (except administrative reassignment, transfer, college student withdrawal, etc.)=0.

[c]Total academy performance index: mean of standardized mean knowledge/skill rating (T) and standardized class rank (T).

[d]Total rating composite.

[e]Completed probation=1; Resigned or terminated while performance was unsatisfactory due to any reason, or performance level unknown=0 (this index of probation success was used to maximize the sample size).

ADA CONSIDERATIONS

Test users should be well aware of the Americans with Disabilities Act (ADA) and its impact upon employee selection. In general, the ADA protects qualified individuals with physical, medical, or mental disabilities from discrimination in any aspect of employment, including: application, **testing**, hiring, training, assignment, evaluation, disciplinary action, promotion, termination, compensation, and benefits. To be considered lawful, a selection procedure that screens out disabled individuals must be *job-related* and consistent with *business necessity*. During the pre-employment phase, employers are permitted to assess abilities, skills, competencies, etc.; however, there are strict prohibitions on the types of inquiries that can be made regarding an individual's disability.

While a full review of ADA pre-employment proscriptions and requirements is beyond the scope of this manual, certain key issues are noted below which pertain to the use of the POST Entry-Level Dispatcher Selection Test Battery.[38]

Reasonable Accommodation. A significant requirement of the ADA is for employers to provide reasonable accommodation during pre-employment testing when requested by disabled applicants who are otherwise qualified for the job. The choice of an accommodation is generally made on a case-by-case basis (POST does not offer a set of specific prescribed accommodations for the Test Battery). ***Test users should exercise caution in modifying any test formats and protocols*** as there is a potential to substantively alter their psychometric properties and thereby make the "altered" test scores incompatible with established test norms and validity results. Agencies are encouraged to contact POST to discuss accommodation options and their potential impact on the Battery.

Notification of Applicants. The test user/agency is responsible for informing applicants of its willingness to provide testing accommodation. Notification of applicants should begin early in the recruitment process and should include a brief description of the testing procedures and conditions, along with a request to be notified in advance of the test administration if an accommodation is needed.

Notification of POST When Accommodations are Made. In the event that an agency does make an accommodation for an individual when administering the Test Battery, ***it is important that POST be notified*** prior to scoring the examination so that test norms and other program evaluations are not unduly affected.

[38]Those involved in law enforcement personnel administration are strongly urged to consult the EEOC Enforcement Guidance (Notice 915.002, May 19, 1994), as well as the POST publication: *The Americans with Disabilities Act: Questions and Answers* (Spilberg, May 1995).

REFERENCES

Allen, M., & Yen, W. (1979). Introduction to measurement theory. Monterey: Brooks-Cole.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee (1985). Standards for educational and psychological testing. Washington, D.C.: American Psychological Association, Inc.

Angoff, W. (1982). Use of difficulty and discrimination indices for detecting item bias. (In Handbook of methods for detecting item bias, edited by R.A. Berk). Baltimore: The Johns-Hopkins University Press.

Angoff, W. & Ford, S. (1973). Item-race interaction on a test of scholastic aptitude. Journal of Educational Measurement, 10, 95-105.

Berk, R. (Ed.) (1982). Handbook of methods for detecting item bias. Baltimore: The Johns-Hopkins University Press.

Bloom, B.S. (Ed.). (1984). Taxonomy of educational objectives: Book 1: Cognitive domain. New York: Longman, Inc.

Buros, O.K. (Ed.). (1974). Tests in print II. Highland Park, New Jersey: The Gryphon Press.

Buros, O.K. (Ed.). (1972). The seventh mental measurements yearbook, Volumes I & II. Highland Park, New Jersey: The Gryphon Press.

Carroll, J.B. (1993). Human cognitive abilities: A survey of factor-analytic studies. New York: Cambridge University Press.

Cascio, W.F. (1982). Costing human resources: The financial impact of behavior in organizations. Boston: Kent Publishing.

Cascio, W.F., & Barrett, G.V. (1988). Setting cutoff scores: Legal, psychometric, and professional issues and guidelines. Personnel Psychology, 57, 1-4.

Cascio, W.F., Outtz, J., Zedeck, S., & Goldstein, I.L. (1991). Statistical implications of six methods of test score use in personnel selection. Human Performance, 4, 233-264.

Cohen, J., & Cohen, P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences. New York: Lawrence Erlbaum Associates.

Cohen, J. (1988). Statistical Power for the Behavioral Sciences. Erlbaum, p. 101.

Conoley, J. & Kramer, J. (Ed.). (1989). The tenth mental measurements yearbook. The University of Nebraska Press.

Conoley, J. & Kramer, J. (Ed.). (1990). The Supplement to the tenth mental measurements yearbook. The University of Nebraska Press.

Ekstrom, R., French, J., & Harman, H. (1976). Manual for kit of factor-referenced tests. Princeton, NJ: ETS.

Fleishman, E., & Quaintance, M. (1984). Taxonomies of human performance: The description of human tasks. Orlando: Academic Press, Inc.

Fleishman, E., & Reilly, M. (1992). Handbook of human abilities. Palo Alto: Consulting Psychologists Press, Inc.

Flesch, R. (1951). How to test readability. New York: Harper-Row.

French, J., Ekstrom, R., & Price, L. (1963). Kit of reference tests for cognitive factors. Princeton, N.J.: Educational Testing Service.

Ghiselli, E.E. (1973). The validity of aptitude tests in personnel selection. Personnel Psychology, 26, 461-477.

Guilford, J.P. (1967). The nature of human intelligence. New York: McGraw-Hill.

Henrysson, S. (1971). Gathering, analyzing, and using data on test items. (In Educational Measurement, R.L. Thorndike, editor). Washington, D.C.: American Council on Education.

Holland, P. & Wainer, H. (1993). Differential Item Functioning. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Honey, R.A., & Kohls, J.W. (1981). Entry-level reading and writing tests for California law enforcement: Final report. Sacramento: Commission on Peace Officer Standards and Training.

Hunter, J.E. (1980). Test validation for 12,000 jobs: An application of synthetic validity and validity generalization to the General Aptitude Test Battery. Washington D.C.: U.S. Employment Service.

Hunter, J.E, & Hunter, R.F. (1984). Validity and utility of alternate predictors of job performance. Psychological Bulletin, 96, 72-98.

Hunter, J.E. (1989). The Wonderlic Personnel Test as a predictor of training success and job performance. Northfield, IL: E.F. Wonderlic Personnel Test, Inc.

Jensen, A. (1980). Bias in mental testing. New York: The Free Press.

Kramer, J.J., & Conoley, J.C. (Eds.). (1992). The eleventh mental measurements yearbook. Lincoln, NE: The Buros Institute of Mental Measurements of The University of Nebraska-Lincoln.

Kramer, J.J., & Conoley, J.C. (Eds.). (1994). The supplement to the eleventh mental measurements yearbook. Lincoln, NE: The Buros Institute of Mental Measurements of The University of Nebraska-Lincoln.

McLaughlin, H. (1969). SMOG grading--A new readability formula. Journal of Reading, 639-646.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50 (9), 741-749.

Millman, J., & Greene, J. (1989). The specification and development of tests of achievement and ability. In Educational Measurement, Third Ed. (Edited by R.L. Linn). New York: Macmillan.

Miner, M. G., & Miner, J. B. (1979). Analysis of uniform guidelines on employee selection procedures. Washington, D.C.: The Bureau of National Affairs, Inc.

Nunnally, J. (1978). Psychometric theory. NY: McGraw-Hill.

Pearlman, K., Schmidt, F.L., & Hunter, J.E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. Journal of Applied Psychology, 65, 373-406.

Sackett, P.R., & Roth, L. (1991). A Monte Carlo examination of banding and rank order selection methods of test score use in personnel selection. Human Performance, 4, 279-296.

Schmidt, F.L. (1991). Why all banding procedures in personnel selection are logically flawed. Human Performance, 4, 265-278.

Schmidt, F.L., & Hunter, J.E. (1977). Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 62 (5), 529-540.

Society for Industrial and Organizational Psychology, Inc. (1987). Principles for the validation and use of personnel selection procedures. (Third edition) College Park, MD: Author.

Weiner, J. (1991). Public Safety Dispatcher Job Analysis. Sacramento: Commission on POST.

APPENDIX

EXAMPLE COMPUTATIONAL PROCEDURE FOR

COMBINING TEST BATTERY SCORES WITH OTHER ASSESSMENTS

Appendix: Example Computational Procedure for Combining Test Battery Scores with Other Assessments

1. Standardize Test Battery scores and scores resulting from other procedures so that the means and standard deviations are equal:

   a) Compute standardized Test Battery score (Z_t):

   $$Z_t = (X_t - M_t) / SD_t$$

   where: X_t=Total Battery score (Total T-score reported by POST)
   M_t=mean of Total Battery scores for the group of examinees
   SD_t=standard deviation of Total Battery scores for the group of examinees

   b) Compute standardized score(s) on other assessment procedures (Z_o):

   $$Z_o = (X_o - M_o) / SD_o$$

   where: X_o=score on other procedure
   M_o=mean of scores on other procedure for the group of examinees
   SD_o=standard deviation of scores on other procedure for the group of examinees

2. Assign proportional weights (W) to Test Battery score and score(s) on other procedure(s). The weights must total 1.0. In assigning the weights, consideration should be given to the relative importance of the abilities measured by each procedure, as well as the validity and reliability of the assessment scores.

   For example, to assign 50% weight to Test Battery scores and 50% weight to an oral examination, weights would be assigned as follows:

   Example: W_t=. 50 (Test Battery score weight)
   W_o=. 50 (oral examination score weight).

   If scores on three procedures were to be equally weighted, the weights would be as follows: W_1=. 33, W_2=. 33, W_3=. 33.

3. Compute weighted total assessment score (Z_{wt}):

   $$Z_{wt}=(W_t*Z_t) + (W_o*Z_o)$$

   The scores can be scaled to have any mean and SD; e.g., to report scores on a scale with mean=50 and SD=10, multiply all scores by 10 and add 50 to each score.

4. Rank candidates on the basis of total assessment score after rounding scores to the nearest integer. Consideration should be given to the reliability of scores (standard error of measurement) in making distinctions between examinees at different score levels.