ENTRY-LEVEL READING AND WRITING TESTS FOR CALIFORNIA LAW ENFORCEMENT: FINAL REPORT

STANDARDS RESEARCH PROJECT

Technical Report No. 4

Prepared by the
Standards and Evaluation Services Bureau

THE COMMISSION
ON PEACE OFFICER STANDARDS AND TRAINING

STATE OF CALIFORNIA

CALIFORNIA
COMMISSION ON PEACE OFFICER
STANDARDS AND TRAINING

ENTRY-LEVEL READING AND WRITING TESTS FOR CALIFORNIA LAW ENFORCEMENT: FINAL REPORT

TECHNICAL REPORT NO. 4

Richard A. Honey, M.A.
Research Specialist

John W. Kohls, Ph.D.
Chief of Standards and Evaluation Services Bureau

1981

This project was supported in part by Grant Number 78-DF-AX-0046, awarded by the Law Enforcement Assistance Administration, United States Department of Justice. Points of view or opinions stated in this publication do not necessarily represent the official position of the United States Department of Justice.

COMMISSION ON PEACE OFFICER STANDARDS AND TRAINING

Commissioners

Jacob J. Jackson, Chairman	Sergeant Sacramento Police Department
Robert A. Edmonds, Vice Chairman	Assistant Sheriff Los Angeles Sheriff's Department
Al Angele	General Manager California Organization of Police and Sheriffs (C.O.P.S)
William B. Kolender	Chief of Police San Diego Police Department
Richard F. Pacileo	Sheriff El Dorado County
Jay Rodriguez	Vice-President Corporate Information - West Coast KNBC-4, Los Angeles
Joseph Trejo	Chief of Police Sanger Police Department
Nathaniel Trives	Professor California State University, Los Angeles
John Van DeKamp	District Attorney County of Los Angeles
Robert L. Vernon	Assistant Chief Los Angeles Police Department
Joe Williams	Councilman City of Fresno
Rodney J. Blonien, ex-officio	Representative of Attorney General State of California
Norman C. Boehm	Executive Director

PREFACE

POST has long been committed to conducting research that will provide local law enforcement agencies with current, defensible, job-related selection procedures. Over the years, this commitment has resulted in such products as selection manuals on medical screening, background investigations, and physical performance testing as well as a statewide job analysis that included the participation of 219 law enforcement agencies in California. We believe POST to be among the leaders in such research.

In 1978 POST received an LEAA grant to continue its research into selection procedures. One of the purposes of that grant was the development of reading and writing tests to be used for the entry-level screening of law enforcement applicants.

So, continuing in its efforts to provide local agencies with the results of research into job-related selection standards, POST presents this report on the validation of reading and writing tests for the selection of entry-level law enforcement officers. This report details the progression of the research from the initial decision to develop reading and writing tests, through the project design and implementation, to the final product, the POST Entry-Level Law Enforcement Examination. Credit for this complex, multi-jursidictional research, must go to researchers Richard Honey and John Kohls of POST staff.

POST is proud to be able to offer to local agencies entry-level reading and writing tests based on solid research practices and to offer this report on the development of those tests. If you would like further information about the availability of the tests or have any questions about this report, please contact the POST Standards and Evaluation Services Bureau at (916) 322-3492.

Norman C. Boehm

NORMAN C. BOEHM
Executive Director

ACKNOWLEDGMENTS

The completion of this project would not have been possible without the extensive assistance of numerous individuals in the law enforcement field. POST would like to extend recognition to the following law enforcement personnel who contributed both their time and expertise to this project:

Captain John Anderson, Commander, California Highway Patrol Academy, Bryte,

Stan Anderson, Director of the Northern California Criminal Justice Training and Education System, S nt Rosa,

Captain W. P. Carlson, Commander, California Highway Patrol Academy, Bryte,

Sergeant Ed Doonan, Sacramento County Sheriff's Department Academy, Bryte,

Captain Thomas Hays, Academy Commander, Los Angeles Police Department, Los Angeles,

Captain Richard Klapp, Coordinator of Training, San Francisco Police Department,

Sergeant Dave Yancey, Academy Commander, Sacramento Law Enforcement Training Center, Bryte.

Project staff involved in the development of this publication included:

George Williams, Bureau Chief
John Berner, Ph.D., Staff Psychologist
Luella Luke, Staff Analyst
John Weiner, Psychometrician
Charlotte Luallin, Systems Analyst
Blanche-Marie Burch, Project Secretary
Sue Miguel, Word Processing Technician

TABLE OF CONTENTS

SCHEDULE OF TABLES

SCHEDULE OF APPENDICES

EXECUTIVE SUMMARY

This study was undertaken to develop and validate reading and writing tests to be used for the selection of entry-level law enforcement officers in California. In effect, this study was comprised of two research projects: one resulting in a reading test and one resulting in a writing test. There was however, considerable similarity in the logic and procedures by which these tests were validated.

The tests were validated using both the content and criterion-related strategies of validation. The tests were content validated using job skills as the criteria and empirically validated using academy academic performance as the criteria.

The basis for this research was the original 1979 POST job analysis (Kohls, Berner and Luke) of the entry-level law enforcement position. Additionally, supplemental job information was collected for both the reading and writing projects.

For the reading project, POST developed a "Source of Information Questionnaire" which was administered statewide to entry-level officers. The purpose of the questionnaire was to verify exactly how information was received by officers. The information generated by this questionnaire allowed POST to include in its analysis of reading requirements only those materials which were actually important to, and read by, officers. Readability analyses were undertaken and the reading requirements of the job were identified.

For the writing project, POST developed a "Survey of Writing Mechanics." The purpose of this survey was to identify the specific rules of grammar and punctuation which were important in an entry-level officer's written expression. This survey was administered statewide to a sample of officers and command level personnel.

From these studies, POST identified those language abilities prerequisite to the performance of job-related language tasks. POST then developed tests that measured those prerequisite abilities. By operationally defining the job skills in the job analysis, and by demonstrating logically that the POST tests measure those abilities prerequisite to the performance of job skills, POST has demonstrated the content validity of its tests.

In addition to content validating the tests, POST also demonstrated the empirical validity of the language ability tests. The empirical study, replicated in seven different academy classes, showed the predictor tests (the reading and writing tests) to be significantly correlated with academic performance in seven academy classes. The tests were also shown to be signficantly correlated with academy performance for 4 distinct ethnic groups.

The utility of the tests was demonstrated in two ways. The percent improvement in selection, above chance, that could be achieved using the predictor tests was established. Also, the precise relationship of predictor scores to criteria achievement was presented. Both demonstrated the tests to have significant utility.

In light of this research, POST feels the language ability tests are appropriate for use as entry-level selection procedures for all law enforcement agencies in the state of California.

INTRODUCTION

Origin of the Project

Basic language ability is necessary for most jobs. Most workers must be able to understand the meaning of common words, to read simple communications, to converse verbally with one another and to write down simple messages.

For the entry-level law enforcement position, language ability is critical. To do their jobs satisfactorily, officers must read difficult written material (e.g., the Penal Code), converse effectively with individuals of widely divergent backgrounds and daily write accurate and clear reports.

Society has traditionally depended upon the educational system to provide these abilities, and has accepted the high school diploma as verification that the abilities have been acquired. Unfortunately, employers of law enforcement officer candidates can no longer accept the high school diploma as proof of even a minimally acceptable level of basic language ability. In fact, during the course of the research described in this report, the researchers became aware of individuals with Bachelor's and Master's Degrees whose reading ability was below the eighth grade level. In addition, law enforcement academy and agency officials have become increasingly concerned about the steady rise in the number of recruits and officers who cannot read and write at a level necessary for satisfactory academy and job performance.

The California Commission on Peace Officer Standards and Training (POST) took action in 1976 to partially address this problem. On January 1, 1977, the Commission adopted the following regulation:

> 1002. Minimum Standards for Employment
>
> (a) Every peace officer employed by a department shall:
>
> (7) Be able to read at the level necessary to perform the job of a peace officer as determined by a "professionally developed" examination designed to test this skill. A professionally developed examination shall be job-related.

The Commission chose not to enforce the regulation until such time as acceptable job-related reading tests were available for use by the over 400 law enforcement agencies in California.

In 1979, POST obtained a grant from the Law Enforcement Assistance Administration to conduct research concerning entry-level employee selection standards. After a study of priorities and needs within the California law enforcement community, the decision was made that POST would begin its work on standards by developing and validating an original set of tests designed to evaluate reading and writing ability.

Goals of the Project

The project was designed to accomplish the following goals:

(a) analyze the reading and writing requirements of the entry-level law enforcement officer position,

(b) develop and validate tests which satisfy the requirements of the Federal Uniform Guidelines on Employee Selection Procedures,

(c) make the tests available to local law enforcement agencies for use in the entry-level screening process.

POST research staff was directed by the Commission to develop a test program which would provide local agencies with measurement tools which would satisfy local requirements. That is, rather than promulgating the tests and test cut-off scores as statewide standards, the Commission decided that use of the POST tests should be voluntary and that local agencies should be able to tailor the test use to local conditions (e.g. selection ratio, degree of reading and writing problems, and amount of adverse impact against groups protected by fair employment legislation).

The Role of POST

The California Commission on Peace Officer Standards and Training (established in 1959) is responsible for promulgating statewide employee selection and training standards. In the late 1960's POST began to fund or to assist in the funding of large scale research studies designed to further assist agencies in improving the abilities and qualifications of law enforcement personnel. Two of the more extensive and significant projects are System and Training Analyses Requirements of Criminal Justice Participants - Project STAR (1974-76), and the Basic Course Revision Project (Wilson, 1977). These projects were not necessarily intended to produce specific standards, but rather to produce new selection criteria and revised training content which local agencies and academies could use to improve the quality of the people being hired and the relevance and effectiveness of the training those people received. This preference for local assistance projects over standards promulgation has been maintained in subsequent POST supported research.

In 1975, POST decided that even better assistance could be provided to local agencies by conducting its own research (most of the previous research has been conducted with substantial outside assistance from contractors). A research unit was established in the fall of 1975. Before beginning the work described in this report, the research unit completed three major projects resulting in the following reports: (1) Medical Screening Manual for California Law Enforcement (Kohls, 1977); (2) Background Investigation Manual: Guidelines for the Investigator (Luke and Kohls, 1977); (3) California Entry-Level Law Enforcement Officer Job Analysis (Kohls, Berner and Luke, 1979). The research unit is now called the Standards and Evaluation Services Bureau. The Bureau designed and conducted the research described in this report.

Purpose of the Report

In documenting the procedures, results, and conduct of this type of research, researchers must fulfill a number of responsibilities:

(a) a sufficiently detailed report must be written so that other researchers, if they desire, can replicate the work,

(b) proper documentation, as outlined in the Federal Uniform Guidelines, must be provided to verify that the research was conducted, and the tests used, in a manner consistent with fair employment principles,

(c) the documentation required by the Standards for Educational and Psychological Tests (also known as the "American Psychological Association Standards") must be provided.

The goal of the Standards and Evaluation Services Bureau was to write a single report which satisfied all these requirements.

JOB-RELATEDNESS AND VALIDATION

Validity is an evaluation of the degree of relatedness between a test score and a criterion. In the specific case of personnel selection, validity is an expression of the extent to which test scores are related to, or predictive of, job performance. Different strategies of validation express this relationship differently. Content models establish the relationship conceptually, while empirical models establish it statistically. These strategies should not, however, be considered as being categorically different, but rather, as different aspects of the same thing. The common ground that unites them is that each is an expression of relatedness.

As different perspectives of the same thing, different validity strategies can be combined to achieve an even more complete understanding of the relationship being analyzed. This study utilized two validation models in establishing the test-criterion relationship. Both the content and criterion-related models were used, since each contributed unique and necessary information in defining the test-criterion relationships.

This synthesis of strategies is consistent with the position stated by the American Psychological Association in the latest Division 14 publication, "Principles for the Validation and Use of Personnel Selection Procedures" in which the interrelatedness of validation models is stressed. In the section on "Definition of Validity," in reference to the various methodologies, the authors stated:

> The three are really inseparable aspects of validity, not discreet types of validity. The Principles discuss these three validity strategies separately only to take advantage of traditional presentations. However, the reader is advised that in concept, and many times in methodology, the three cannot be logically separated.

The initial strategy of validation utilized in this study was content. Section 14C of the Uniform Guidelines on Employee Selection Procedures, defines the situations in which the content model is appropriate.

> A selection procedure can be supported by a content validity strategy to the extent that it is a representative sample of the content of the job. Selection procedures which purport to measure knowledges, skills, or abilities may in certain circumstances be justified by content validity, although they may not be representative samples, if the knowledge, skill, or ability measured by the selection procedure can be operationally defined as provided in Section 14C(4) below, and if that knowledge, skill or ability is a necessary prerequisite to successful job performance.

Thus, the first instance where content validity is appropriate is where the test can be shown to be a representative work sample, and the second is where the test samples knowledges, skills and abilities

prerequisite to successful job performance. The second application has the added requirement that the knowledges, skills, and abilities being measured must be operationally defined.

The tests developed by POST measure prerequisite abilities and can, therefore, be content validated under the conditions stated in Section 14C. Two tests were developed, a reading test and a writing test.[1] Both tests measure generic abilities prerequisite to performance of job-related, reading and writing tasks.

There is frequently the error, regarding the reading and writing tests, of considering them as being work sample tests. They are not. It would, in fact, be inappropriate if they were. Any work sample test is, in essence, a job knowledge test. By definition this is the case. It is clearly inappropriate, however, to test job knowledge at the entry level since no job knowledge can logically be assumed for those who will be trained. Officers frequently read legal codes, police reports, and various other law enforcement documents. It would be inappropriate, however, to test applicants on these materials. They contain numerous words and phrases which are uncommon in general reading, but crucial to the understanding of the law enforcement materials which contain them. Further, these technical terms are learned during academy training.

The same reasoning applies to the writing test. First-level officers spend approximately 25% of their time writing reports. It would be inappropriate, however, to test an applicant's ability to write police reports because applicants have no experience in proper format and style. Again, these specific abilities will be learned in the academy. What is appropriate, however, is to test those generic reading and writing abilities prerequisite to the performance of specific job tasks. The generic abilities cannot be easily learned on the job; in fact, research indicates that remediating reading and writing deficiencies in adults is extremely difficult and usually only partially successful.

While the content strategy is appropriate for establishing the job-relatedness of the tests, it does have limitations. These limitations are due primarily to the lack of external criteria in the content model. Without external criteria, regression analyses are impossible. In the absence of such statistical procedures many extremely important test-related issues could not be addressed (for example, the issues of the empirical evaluation of test fairness, differential validity, candidate ranking, and passpoint setting). To overcome these shortcomings of the content methodology, POST conducted a criterion-related study.

---

[1] In order to comply with the "Consideration of Alternative Selection Procedures" section of the Uniform Guidelines (Section 3B), POST actually developed a number of examinations: reading tests at different levels and writing tests using two different test formats. Ultimately, two tests, a reading test and a writing test, were selected to comprise the POST tests of language ability.

The rationale for justifying a test using the criterion-related strategy is set forth in the Uniform Guidelines, Overview of the 1978 Guidelines (VII).

> In criterion-related validity, a selection procedure is justified by a statistical relationship between scores on the test or other selection procedure and measures of job performance.

The Uniform Guidelines also establish the technical standards by which the adequacy of a criterion-related study is determined. These standards are presented in Section 14B(5) of the Technical Standards for Criterion-Related Validity.

> Generally, a selection procedure is considered related to the criterion, for the purpose of these guidelines, when the relationship between performance on the procedure and performance on the criterion measure is statistically significant at the 0.05 level of significance, which means that it is sufficiently high as to have a probability of no more than one (1) in twenty (20) to have occurred by chance.

In summary, the POST validation strategy incorporated two research models, content and criterion-related validity. To demonstrate content validity, POST operationally defined the work behaviors to be measured in terms of observable work tasks (job analysis). Once the behaviors were identified, POST developed tests to measure those abilities prerequisite to their performance. Additionally, POST conducted a criterion-related study so that important statistical analyses could be conducted.

CRITERION DEVELOPMENT

Once the selection of validation models was made, the next decision concerned the most appropriate criteria for each model. It was decided, for both the reading and writing tests, to use job performance for the content validity study and academy performance for the criterion-related study.

Criteria for the Content Model

The Uniform Guidelines stress that in demonstrating the content validity of a selection procedure one should establish a logical link between the content of the test and the content of the job. Specifically, it should be demonstrated that the behaviors (knowledges, skills, and abilities) being measured in the selection procedure are the same as those behaviors identified as being required for adequate job performance (Section 14C(4)).

A first step in this process consisted of defining the job. This was accomplished by conducting a job analysis. In this analysis, the job, its tasks, and the knowledges, skills and abilities required to perform those tasks, were thoroughly studied and defined.

POST conducted a job analysis that served as the foundation for this study (Kohls, Berner and Luke, 1979). The study sample was comprised of 1720 officers and 717 supervisory/command personnel from 219 police and sheriffs departments throughout the state. This sample was representative of all sizes of municipal police and county sheriffs departments in California.

The job data were collected and analyzed in two phases. In the first phase respondents evaluated 439 tasks in terms of the frequency of task occurrence, the importance of the tasks for the job, the relation of task performance to overall performance, and when the task performance needed to be mastered. In this manner the duties of the job were identified and ranked. These tasks were then grouped, using both statistical and rational techniques, into groups or clusters of tasks that required similar knowledges, skills, abilities and work behaviors. This process resulted in the identificaton of 33 homogeneous task groups.

In the second phase of the study, these 33 task groups were evaluated regarding the extent to which 29 identified personal attributes were required for successful job performance. This resulted in the identification of the specific behaviors which were required to adequately perform each of the task groups.

Thus, in phase one the work domain was identified and operationally defined (by identifying task groups), and in phase two the knowledges, skills, and abilities required to perform those tasks were identified (by relating personal attributes to task groups).

While this information served as the foundation for the present study, additional supplemental job analyses were conducted in the areas of reading and writing. These analyses are described, in depth, later in this report.

Criteria for the Empirical Model

For the empirical model, performance in the academy was identified as the best criterion. The academy curriculum has been demonstrated to be job related (Wilson, 1977) and all potential officers must successfully complete the academy curriculum.[2] This meets the technical standards for criteria established in the Uniform Guidelines, Section 14B(3), which state:

> Where performance in training is used as a criterion, success in training should be properly measured and the relevance of the training should be shown either through a comparison of the content of the training program with the critical or important work behavior(s) of the job(s), or through a demonstration of the relationship between measures of performance in training and measures of job performance. Measures of relative success in training include, but are not limited to, instructor evaluations, performance samples, or tests.

Two different assessments of success in the academy were identified for use as criteria. These were not combined, but used independently. One of these was a composite score of academic performance tests taken in the academy. This score was computed by adding all academy test scores on analytical subject matter. This was accomplished by converting all raw scores to T scores (mean equal to 50 and standard deviation equal to 10), adding all T scores together, and then dividing by the number of scores. Scores were standardized within each academy. A person's T score represents an aggregate assessment of all performance throughout the entire academy in all analytical work. Since each academy conducts its own assessment of students, the composite scores are on a different scale from academy to academy. Thus, a high correlation between the POST tests and this criterion would demonstrate a predictiveness of academy success regardless of the slightly different styles of presentation and assessment utilized by different academies.

The second criterion used in this study was the Training Proficiency Test administered by POST to all academy graduates throughout the state. In contrast to the other criterion (the composite of academy scores) which was unique for each academy, this criterion was consistent throughout the state. The Proficiency Test was designed to measure achievement on each of the twelve mandated functional areas covered by the academy curriculum. Thus, it is a final exam for all relevant knowledge taught in the academy.

By using these two independent criteria, one unique to each academy, and one common to all academies, staff felt assured that a reliable estimate of academy success would be achieved.

---

[2]It is, however, possible for individuals who have not gone through a California police academy to be hired into law enforcement. If they qualify (they must demonstrate that they received training equivalent to the POST Basic Course and be under consideration for hire) they may take the POST Basic Course Waiver Examination. If they pass that test, they can avoid retraining.

LANGUAGE USAGE

The purpose of this section is to briefly present a model of language usage and to identify those language processes which are the focus of the POST tests.

Language and Communication

The ease and frequency with which people communicate conceals the incredible complexity of the linguistic structure which makes communication possible. To communicate verbally requires language. In the absence of language there is no verbal communication.[3] Language, however, even though it has been the focus of study and speculation for thousands of years, has defied definitive explanation. The illusiveness of language to analysis and understanding is due largely to the fact that it exists in our minds. Language is the mental structure, the body of knowledge that underlies our ability to communicate. As such it cannot be directly addressed but only inferred from a study of overt language behavior.

Understanding the distinction between underlying structure and overt behavior is important because it has direct implications for the testing of language competence. Specifically, it impacts on what is to be subjected to measurement and how it should be measured. Slobin (1979) expresses the distinction between structure and behavior by contrasting the concepts of language and speech. The word "speech," he points out, has a corresponding verb form whereas the word "language" does not. Thus, to say that an individual speaks English is to say that the person makes meaningful sounds. Those sounds are meaningful because they can be systematically related to that thing called the English language. Speech is behavior, but a behavior that has meaning only within the context of the English language. The English language is not a behavior. Rather, it is the body of knowledge present in the minds of speakers of English that allows them to interpret English speech.

Observable linguistic behavior is comprised of the four language processes: reading and listening (which are receptive), and writing and speaking (which are productive). Each behavior represents a utilization of language for the purpose of communication. The speaker or writer uses language to go from thoughts to words; the reader or listener uses language to go from words to thought.

---

[3]There are many forms of communication which are non-verbal. Much information is conveyed through imagery, both visual (e.g., painting and dance) and auditory (e.g., music). Similarly, much can be communicated symbolically or by gesture. While there are many such forms of non-verbal communication, they are outside the scope of this study and were not addressed in this research. In this report POST is dealing only with verbal communication. Thus, whenever communication is referred to in this report, it is with reference to verbal communication only.

To be a user of the English language, (to read, listen, write or speak) one must know English grammar. It is important to note that there are two types of grammar. The type that first comes to mind is that which was encountered in school. This is prescriptive grammar and refers to "proper" speaking and writing. There is, however, another grammar that describes what one must know in order to use and understand a language. In this sense a grammar constitutes a system of rules, in effect a theory, of language usage. It is in this latter, theoretical sense with which POST is concerned here.

Daniel Slobin (1979) defines grammar in the following way:

> It is a theory which should be able to discriminate sentences from non-sentences, provide structural descriptions which relate meaning and sound, account for the meaning of sentences and combinations of sentences, and so forth.

Only a very few language users actually know the formal rules which are employed when making the determinations listed above. It is not important, however, that they do not. What is important is that they behave as if they knew the rules. That is, individuals must have internalized their knowledge of language (the rules) to the point where they simply "know" when a combination of words is meaningful and when it is not.

Thus, to be an effective language user does not require knowledge of the formal rules (grammar) of the language. Rather, it requires that individuals behave as if they knew the rules. Therefore, in its tests of language ability, POST focused on observable language behaviors, (reading and writing) and not on knowledge of formal language structure.

Language Theory and Test Construction

While the language user need not be aware of the formal rules of the language, it is important for the test developer to be aware of them. In the last quarter century linguistic theory has undergone significant evolution. The trend has been from relatively simple behaviorist models to increasingly complex cognitive, information processing models (Slobin, 1979). The significance of this evolution for the test developer is great. Different measurement procedures are sensitive to different types of variables. The interview, for example, is more sensitive to the measurement of interpersonal skills than is a true/false test. Similarly, different test formats are sensitive to different linguistic variables.

Format of the Tests

The next step became to identify the particular test formats to measure reading and writing abilities. Because of differences between these subject matters, the issue of format was addressed separately for each. In each case, however, the primary concern was to identify a test format which was sensitive to the particular linguistic variables identified as important by this research.

The next three sections will focus with greater detail on the two language processes being subjected to examination (reading and writing). Each process will be more fully defined and the link between process and test will be discussed.

READING ABILITY AND MEASUREMENT

This section describes specific models of reading, the procedures utilized in measuring reading ability, and the methods of estimating readability levels. The section begins with a general statement of the importance of reading in the law enforcement profession. This is followed by a short literature review addressing models or theories of reading. Next to be discussed are the procedures for measuring reading ability with emphasis placed on explaining the relationship between measurement techniques and the specific subject matter being measured. Last to be discussed is the concept of readability, what it is and how it is computed.

Reading and Law Enforcement

Ability to read is an essential skill for law enforcement officers in the State of California. On this point there is no disagreement. All current major studies which have included California law enforcement officers have supported this conclusion (Wollack, Clancy, and Beals, 1973; Rosenfeld and Thornton, 1976; Wollack, 1976; Honey, 1979; Kohls, Berner and Luke, 1979). The importance of reading for police work has also been officially recognized by the law enforcement community itself.

Law enforcement agencies, in response to a 1978 Needs Assessment Questionnaire (Kohls and Berner, 1978) designed to establish priorities for selection standards research, established reading as one of the selection areas that they felt to be of primary importance.

Thus, the problem was not to justify reading as an appropriate subject matter for examination, rather, it was one of determining how reading could best be measured in the selection process. To determinine how reading can best be assessed requires both an understanding of the various models of reading and an understanding of the techniques used in measuring reading ability.

Literature Review of Reading Models

A literature review was undertaken to explore research that related to models of reading, procedures for assessing reading, and methods for determining reading level. Based on this review, POST concluded: (1) that the "psycholinguistic" theory of reading best described the reading process, (2) that the Cloze testing procedure represented the most precise estimate of reading ability, and (3) that three readability formulas would be utilized in the assessment of reading difficulty (since no one readability equation emerged as clearly superior to all others, three equations were utilized so that a more stable readability index could be identified).

In reviewing the literature on reading models, it became immediately apparent that there is no generally accepted view of the nature of reading behavior. Rather, different reading models exist simultaneously, the consequence being that the same behavior is described and interpreted differently. This difference of interpretation is due largely to the different way in which reading behavior is perceived by different theorists. Some perceive reading to be the composite

of discrete components while others perceive it as a unity which cannot be subdivided. This difference in perception impacts greatly on the way in which reading is researched and also on the way in which reading comprehension is assessed.

The Components Model of Reading

Traditionally, reading behavior has been perceived as being the aggregate of several discrete cognitive sub-skills. A basic supposition of this model is that reading behavior is best described by enumerating its constituent skills and abilities. From this perspective, reading research tends to be directed at identifying the components that underlie reading behavior, and the measurement of reading ability is focused on quantifying proficiency on each of the constituent aspects of reading behavior.

Some of this research has utilized a rational methodology while other research has utilized statistical procedures. Neither approach, however, has resulted in a consensus regarding the number of components that comprise the act of reading and neither has resulted in the development of a comprehensive theory of reading behavior.

Some of the rational models stress large, global components such as vocabulary, reading speed, and reading comprehension. Others subdivide one or more of these larger components into smaller components. Comprehension, for example, can be divided into: (1) ability to follow the organization of a passage, (2) ability to select the main idea of a passage, and (3) ability to draw inferences about the passage, etc.

Another methodology used for identifying reading components is the statistical procedure of factor analysis. This procedure proved no more successful than the rational approach in the search for a comprehensive listing of the components of reading behavior.

The attempt to develop a theory of reading by identifying the components of reading behavior is not an approach without its critics. The basic theoretical objection to this procedure is that the results of such a process do not constitute a theory. Rather, they are simply lists of sub-skills. This is true regardless of the particular methodology (rational or statistical) used in deriving the list of sub-skills. And, while it might seem that the statistical approach is more precise and "scientific," there are actually additional problems inherent in this procedure.

Factor analytic studies have been criticized on both technical and theoretical grounds. The technical issue was raised by Raygor (1966) who questioned the validity of such studies. He noted that the integrity of the results based on any factor analytic study are largely a function of the adequacy of the test(s) being factored. This is due to the dependency of factor analytic results on the reliability and validity of the input tests. Thus, any weaknesses, inadequacies, or unreliability in the tests being factored must necessarily confound or distort conclusions based on those tests.

The primary theoretical criticism was made by Goodman (1966). Addressing the sub-strata theory of Holmes and Singer directly (but all factor studies by implication), Goodman contended that the factor analysts had not developed a theory of reading at all. Rather, they had simply defined the statistical dimensionality of the tests that they had analyzed. Goodman stresses the significant difference between a factor solution and a theory. All complex data can be summarized by using various mathematical procedures. To do so, however, is only to simplify and reduce the dimensionality of the original data.

Thus, the listing of reading sub-skills, whether the list is generated rationally or statistically, does not constitute an adequate model of reading behavior.

The Psycholinguistic Model of Reading

At the other end of the spectrum, theorists have attempted to define reading in psycholinguistic terms (Goodman, Smith, etc.). The essence of this model is that reading behavior is a process of the reader interacting with the information provided on the printed page. The meaning that a reader derives depends both on what is on the page and with what information the reader brings with him. Thus, reading is an integrated, information processing behavior which cannot be fragmented and its parts analyzed independently. The psycholinguistic model does not deny that reading requires a number of abilities, but it does assert that they cannot be divided and assessed individually. From this theoretical perspective, research tends to focus on the reading process as a unity.

The process of reading, according to Goodman (1973), is actually a sort of psycholinguistic guessing game. In this game, the reader forms hypotheses regarding meaning based on his knowledge of the grammatical structure of the language.

> He supplies semantic concepts to get meaning from structure. In turn, this sense of syntactic structure and meaning makes it possible to predict the graphic input to confirm his predictions. In reading, what the reader thinks he sees is partly what he sees, but largely what he expects to see.... The extent to which a reader can get meaning from written language depends upon how much related meaning he brings to it.

Further, this is done utilizing only a small portion of the clues (words) presented in the text.

A common example of this concept of reading behavior that each of us has experienced is the phenomenon of knowing the information contained at the end of a sentence before actually reading it. One hypothesizes, based on information obtained from what has already been read combined with a sensitivity to grammatical cues contained within the sentence, how the sentence will end. This is an interactive process with the passage continually providing more information and more cues and the reader supplying grammatical concepts and utilizing an ever-increasing knowledge about the subject matter being read.

As can be readily seen, the components and the psycholinguistic models are quite different in their definition of the reading process. These differences have implications for the subsequent testing of language competence. Since the nature of the variables to be measured largely determines the nature of the procedure most appropriate for their measurement, the two models require different measurement techniques to assess reading behavior. The variables identified as significant by the components models have traditionally been measured by the multiple-choice format, whereas those identified as significant by the psycholinguistic model are best measured by the Cloze procedure.

The term Cloze is derived from the Gestalt concept of closure - the tendency to complete that which is familiar but unfinished. "Just as there is an apparent tendency to 'see' a not quite complete circle as a whole circle by 'mentally closing the gap' and making the image conform to a familiar shape, a mutilated sentence is filled in by completing those words that make the finished language pattern conform to the intended or apparently intended meaning" (Potter, 1968). Thus, the Cloze test requires the test taker to identify words deleted from a passage in an effort to restore the passage to its original form.

Multiple-Choice Test Format

The multiple-choice format has been utilized in tests relying on the components model for basically two reasons. Since the various factors of the components model represents specific knowledges, they can be directly addressed by multiple-choice questions. A second reason, and probably the more compelling one, is that until the development of the Cloze procedure in 1953, there were really no alternatives to the multiple-choice format. Thus, the traditional reading test was usually comprised of a reading passage, followed by one or more multiple-choice items regarding that passage.

Even though the multiple-choice format is one of the most versatile of all testing procedures, there is a basic psychometric flaw when applying it to the reading context. The traditional reading test consists of a reading passage followed by one or more multiple-choice questions. The problem is that the reading difficulty of the passage, and the difficulty of the question regarding that passage, are independent of one another. Thus, there can be easy questions about hard-to-read passages and hard questions about easy-to-read passages. This being the case, getting an item correct does not indicate reading comprehension at the level of the passage, and getting an item wrong does not indicate lack of comprehension. Interpreting scores thus becomes troublesome because comprehension level is obscured by item difficulty.

Awareness of this problem is not recent. Lorge (1934) criticized multiple-choice reading tests citing the confounding influence of item difficulty on passage level. As long as the multiple-choice format is utilized in conjunction with a reading passage, the fundamental problem remains the same.

Thus, there are both theoretical and psychometric problems with the traditional methods of measuring reading ability. The components model is theoretically inadequate, as it is not really a theory at all, and the multiple-choice format when applied to reading has a fundamental and inexorable problem of score interpretation.

The Cloze Reading Test

The Cloze procedure suffers from neither of these problems. It is supported by a good theoretical base and it is psychometrically sound. Further, the variables measured by the Cloze procedure parallel those identified as being important by the psycholinguistic model. According to Taylor (1953), Cloze scores represent a measure of aggregate influences of all interacting factors which affect the understanding of a passage by a reader. This is entirely consistent with the psycholinguistic theory which asserts that reading is an integrated behavior which cannot be fragmented. Cloze makes no attempt to quantify discrete aspects of the reading process.

To complete a Cloze test, one must have knowledge of word meaning (semantics), words in combinations (syntax), and how these two interact. By systematically deleting words, Cloze measures an individual's ability to make whole again an incomplete passage. To complete the passage, the individual must hypothesize, based on the information present in the passage, combined with his or her knowledge of the language, what word best completes the blank space. This is a very similar behavior to the actual process of reading where the individual makes hypotheses regarding meaning based on what is presented on the printed page combined with his or her knowledge of the grammatical structure of the language. In both the reading process and in completing a Cloze passage, meaning is achieved as a consequence of the interaction of the individual with the printed material. Quoting Ryan and Semmel (1969), "...not all the information needed by the reader is on the printed page - nor are all the printed details needed by him." Thus, the Cloze procedure accurately replicates, and tests, the actual behavior required to read and comprehend written materials.

Readability

Whenever reading ability is the subject of analysis, the issue of reading level invariably comes up. While the notion of readability is conceptually clear, translating the concept into a numerical index is problematical. The actual difficulty stems from two conditions: the absence of any meaningful benchmark of difficulty and the small numbers of variables (usually 2), used to define readability. The former results in different readability equations generating different indices for the same reading sample, and the latter makes questionable any estimation of level because of variables not considered in the evaluation. A brief look at the readability equations explains why this is the case.

In a readability formula various language variables are counted. These counts are then utilized in an equation which yields an index of probable difficulty for readers (Klare, 1966). One problem inherent in this approach, however, is that there exists no stable norm group to which the index can be referenced.

Reading indicies are usually presented in terms of grade level. To predict, for example, that a particular document is written at the 10th grade level is to predict that students in the 10th grade can understand that document with at least a 70% comprehension rate. (A 12th grade level would require 70% comprehension by 12th graders.) There are, however, thousands of 10th grade classes, and the range of reading ability within these classes is great. It is, therefore, impossible for one index to be representative of this entire group.

A further weakness of readability equations is the limited number of variables that they include. Most traditional readability equations employ a two variable model: one semantic (dealing with words), one syntactic (dealing with sentences). When the two variables are analyzed independently, the semantic is consistently more predictive of reading difficulty than is the syntactic (Klare, 1966). The difference, however, is not great, and all popular readability equations utilize both variables together as each contributes unique information.

By utilizing only two variables to predict readability other relevant variables are ignored. Existing equations do not quantify conceptual complexity, are insensitive to the organization of written material, and do not consider the effect of physical factors such as format, type, and illustrations (Russell and Fea, 1951).

Conclusion

Based on the findings of the literature review, staff identified the psycholinguistic model as being superior to the components model as a theory of reading behavior. Staff also found the Cloze technique superior to the traditional multiple-choice format as a procedure for measuring reading ability. Relative to the issue of readability, staff identified the limitations of existing readability indices and concluded that three different indices would be utilized so that the deficiencies of any one of them would be minimized. A description of the readability equations utilized in this study is presented in Appendix A.

This section focuses on the procedures utilized to measure writing ability and the problems associated with each of the various alternatives. The section begins with a statement of the importance of writing skills for the entry-level job and then continues with the results of the literature review.

Writing and Law Enforcement

The 1979 POST job analysis (Kohls, Berner and Luke) examined the writing requirements of the entry-level job from various perspectives. In all instances, writing emerged as being extremely important for success on the job.

In one section of the job analysis questionnaire, respondents were asked to "Estimate the number of hours of your time as a radio car patrol officer/deputy that you spend during a typical week doing the following activities." Based on data collected from 1,720 incumbent officers, the average incumbent spends 10.7 hours per week writing and/or dictating reports. This is slightly over 25% of the officers'/deputies' entire time on the job. Thus, just from the standpoint of time spent, writing is a major part of the job.

The rating of writing tasks by supervisors also indicated that writing tasks are considered as being important. Writing tasks received a mean rating of 3.1 (statewide average) on a scale extending from one to five. This rating corresponds to a verbal scale label of "important."

Once the staff had established both the frequency and the importance of writing ability for performance on the job, the next step was to develop a strategy for the measurement of writing ability.

Literature Review

In reviewing the literature regarding the testing of writing ability, one finds a significant contrast between theory and practice. At the theoretical level, there is uniform agreement that writing ability cannot be properly evaluated with an objective test. At the practical level, however, almost all tests of writing ability utilize the multiple-choice format. The reason for this inconsistency is not difficult to identify. The fact is that it is extremely difficult and time consuming to use anything but a multiple-choice test to measure writing ability.

The obvious alternative to a multiple-choice test is to require an individual to demonstrate writing competence by actually writing a passage. There are, however, significant problems associated with this procedure. The initial problem lies in the development of criteria against which the essay should be evaluated. There are numerous dimensions that can potentially serve as criteria: style, logical development, syntactic correctness, spelling, etc. Thus, the first problem is one of identifying what aspect or aspects of

writing are to be rated. The next problem is finding scorers capable of making the determination if something is, given the criteria, acceptable or unacceptable. It is obvious that individuals can serve as reliable raters only for subject matter in which they themselves are competent. Finding individuals skilled in the various aspects of grammar and composition is extremely difficult. Finding a panel of such individuals in all the locations where a test might be administered is more difficult yet. Another problem is that of time. The amount of time required to score essays is excessive, especially when one considers that multiple raters are required if reliability of ratings is to be claimed.

In spite of the difficulties associated with the development of an essay test, POST decided to develop both a multiple-choice test and an essay test. This decision was based primarily on two factors. Since the literature stressed so strongly the superiority of the essay over the multiple-choice test as a measure of writing ability, staff felt it necessary to include such a test in the research. Also, by developing both types of exams, POST was complying with the alternative procedures section of the Uniform Guidelines (Section 3B). Once both were developed, POST could compare their relative reliability, validity, and adverse effect. With this information staff could then select the best test for inclusion in the final test battery.

This section will address the research undertaken for this study. Analyses will be discussed in the order that they occurred. The discussion of each analysis will include the reason it was undertaken, the specific procedures utilized, and the results obtained. Preliminary to addressing each analysis in detail, however, a general overview of the logic of the research will be presented.

General Overview of Research Logic

The purpose of this project was to establish the job-relatedness of the new reading and writing tests developed by POST. There was considerable similarity in the logic and procedures by which these tests were validated. As previously stated, both tests were validated using both the content and criterion-related validation strategies. Similarly, in both cases the tests were content validated using job skills as the criterion, and empirically validated using academy academic performance as the criterion.

POST came to the conclusion early in the project that it would utilize a combination of validation strategies in establishing the job-relatedness of its tests. This decision was based on the realization that a combination of approaches could establish a stronger, more complete link between test performance and job performance than could any one strategy alone.

To validate an examination under Section 14C of the Uniform Guidelines ("Technical Standards for Content Validity Studies") requires, if the test is not comprised of actual job samples, that the abilities to be tested be operationally defined and that they be demonstrated to be "prerequisite to successful job performance." In the specific case of this research, POST had to demonstrate that reading and writing abilities were prerequisites to successful job performance. POST accomplished this by: (1) documenting the reading and writing requirements of the job (by referring both to the original POST job analysis and to the supplemental job analyses conducted for this study); and (2) logically establishing a link between the tests and those reading and writing demands.

To validate an examination under Section 14B of the Uniform Guidelines, "Technical Standards for Criterion-Related Validity Studies," a quantitative link had to be established between the test and specific criteria. Relating this requirement to the current study, POST had to demonstrate a significant statistical correlation between the POST reading and writing tests and the criteria discussed earlier in this report.

The remainder of this section summarizes the research undertaken to establish the validity of the POST tests under the Guidelines sections cited above.

Identifying the Reading Demands of the Job

The Kohls, Berner and Luke (1979) job analysis study found the following reading tasks to be of sufficient frequency and importance to have implications for establishing the reading requirements of the job. A complete listing of the scales used to evaluate these tasks can be found in Appendix B.

Table 1: Reading variables that emerged as being important in the original POST job analysis.

Reading tasks
Review crime lab reports.
Read in-depth narrative reports containing complete sentences and paragraphs (e.g., investigative reports, supplemental/ follow-up reports).
Read reports consisting of several short descriptive phrases, sentence fragments, or very short sentences (e.g., incident reports).
Read reports consisting primarily of check-off boxes or fill-in blanks (e.g., vehicle impound reports).
Read street maps.
Read incoming correspondence.
Read interoffice memos.
Read departmental manuals.
Read weather forecasts and bulletins.
Read case law.
Read legal interpretations (e.g., California Attorney General's opinions, city attorney opinions).
Read legal transcripts.
Read teletype messages.
Read training bulletins.
Read and interpret coded material (e.g., NCIC printout, DMV drivers' records).
Read state, federal, and local statutes.
Read Alcoholic Beverage Control Act.
Read professional law enforcement publications (e.g., Police Chief, FBI Law Enforcement Bulletin).
Read Business and Professions Code, Administrative Code, Evidence Code, Vehicle Code, Civil Code, Government Code, Health and Safety Code, Penal Code, U.S. Code (e.g., regarding illegal aliens), U.S. Constitution, Welfare and Institutions Code, Municipal Code, County Ordinances, and Fish and Game Code.

The next step was to generate a more specific description of these tasks. To accomplish this, the reading variables were incorporated into the "Source of Information Questionnaire." A copy of this questionnaire can be found in Appendix C.

Source of Information Questionnaire

The purpose of this questionnaire was to verify exactly how information which originated in the written form is obtained by the entry-level officer. Specifically, staff was interested in determining if materials officers claimed to read were actually read by officers, or if the information contained in these materials was transmitted to the officer in some other form (e.g., orally during roll call). In the questionnaire, respondents were instructed to evaluate the reading tasks using the following response categories:

Does not apply: This material does not pertain to my job.

Orally: Information regarding this material is presented to me verbally (e.g., in lectures, over the radio, at roll call, etc.).

Written Summary: Information regarding this material is presented to me primarily in summarized written form. I read the summarized version but not the original document.

Original Written Form: Information regarding this material is presented to me primarily in its original written form (by "original form" we are referring to a non-summarized version of the original document itself).

Additionally, respondents were instructed to list, for the five codes included in the questionnaire, the specific sections (up to 10 sections) that they referred to most frequently. A listing of the specific code sections which were subjected to readability analysis is presented in Appendix D.

Questionnaires were completed by 508 officers in 41 departments throughout the State. The same sample stratification system (five size categories for police departments and three size categories for sheriff's departments) was used in this study as was developed for the original POST job analysis project. The sample is summarized in the Tables 2 and 3.

Table 2: Number of agencies and number of respondents within each size category to complete Source of Information Questionnaire.

	Size category	Number of entry-level officers in depts. in this category	Number of departments in this study in this category	Number of officers in this study in this category
Police Depts.	1	1 - 10	6	30
	2	11 - 25	6	42
	3	26 - 50	5	47
	4	51 - 150	5	71
	5	151+	6	115
Sheriff Depts.	6	1 - 40	3	28
	7	41 - 125	5	74
	8	126+	5	101
Total			41	508

Table 3: Size and geographic location of agencies completing Source of Information Questionnaire.

Size category	Geographic Location		
	North	Central	South
1	Fortuna Weed	Cotati Half Moon Bay	Ojai
2	Marysville So. Lake Tahoe	Davis	Coronado El Segundo Delano
3	Redding	Concord Petaluma	La Mesa Orange
4	---	Fremont Modesto Richmond	Glendale San Bernardino
5	---	Oakland Sacramento San Jose San Francisco	San Diego Los Angeles
6	Shasta	Madera	San Luis Obispo
7	Sonoma	Contra Costa Monterey	Kern Santa Barbara
8	---	Sacramento	Los Angeles Orange Riverside San Bernardino

Note. The empty cells in this table are due to the fact that there are no departments of that size in that part of the state.

[a]Size categories are defined in Table 2.

Each department was sent a contact letter requesting that a departmental coordinator be named for that department. This letter supplied the criteria for selecting the sample and the sample size. Copies of this correspondence are presented in Appendix E. The coordinator was to collect the questionnaires from the participants in his/her department, review them for completeness, and return them to POST. The coordinator was also to gather certain specified reading samples listed in a document entitled, "Request for Sample Reading," and send them to POST staff. This document is presented in Appendix F.

Frequencies were computed for the responses for each reading category and code in the questionnaire. To indicate how information was primarily received, the respondents were given a number of categorical choices, four choices for the first 14 reading categories (does not apply, orally, read summary, and read original) and five for the five legal codes (does not apply, orally, written summary, abridged original, unabridged original). For all reading categories and codes, the first choice indicated that the material did not pertain to the respondent's job, and the second that information regarding the category or code was received primarily orally. The selection of either of these choices indicated that the material being evaluated should not be included in the reading study. If the majority of respondents indicated that they actually read a particular material, it was retained in the study. Using these criteria, all reading materials in the questionnaire were retained for further analysis. Table 4 summarizes the responses to the Source of Information Ouestionnaire.

Table 4: Responses, expressed as the percentage of respondents that selected each categorical choice, to the Source of Information Questionnaire.

Reading Categories	Does Not Apply	Orally	Written Summary	Original Written Form	Percent of sample that indicated that reading was the source, in one form or another, of information regarding this variable.
General Reading Materials					
In-depth narrative	7.1	16.7	18.1	57.7	75.8
Short report	9.1	24.2	25.2	40.7	65.9
Check box report	11.8	15.0	15.7	55.9	71.6
Maps	7.7	4.9	10.8	76.2	87.0
Incoming correspondence	15.9	21.9	21.1	39.0	60.1
Memo	4.1	20.3	23.4	51.0	75.2
Vehicle bulletins	2.6	46.3	30.5	19.9	50.4
Departmental manuals	1.2	3.5	15.2	78.9	94.1
Case law	1.0	18.9	64.8	15.2	80.0
Legal interpretations	2.2	19.5	61.4	16.5	77.9
Teletype messages	3.9	29.3	18.7	47.2	65.9
Training bulletins	.8	9.4	31.5	57.7	89.2
Warrants	1.8	26.4	32.9	38.4	71.3
Coded material	3.0	25.2	23.0	48.6	71.6
Legal Codes					
Penal Code	.2	2.2	12.0	47.0 (38.6)[a]	97.6
Vehicle Code	0.0	2.0	25.4	20.7 (51.6)	97.7
Health & Safety	1.0	4.7	22.0	47.4 (24.4)	93.8
Welfare & Institutions	2.0	5.1	21.5	46.1 (24.8)	92.4
Municipal	1.6	5.5	27.0	29.9 (36.0)	92.9

Note. N for each variable = 508.

[a]The first number refers to codes read in abridged form, and the number in paranthesis refers to codes read in unabridged form.

Sample Reading Materials

A review of the reading materials showed that not all of the materials were conducive to readability analysis. To conduct a readability analysis requires narrative passages of at least 100 words which are comprised of complete sentences. Because of this requirement, the following reading categories were dropped from further consideration in this study: short reports, check box reports, maps, vehicle bulletins, teletype messages, warrants, and coded materials. Municipal codes were also omitted from further study since each municipal code is specific for each jurisdiction and, therefore, not appropriate for inclusion in a statewide study.

Six categories of materials were retained for further analysis. These are presented in Table 5.

Table 5: Reading materials that were retained for further analysis.

Reading materials
In-depth narrative
Memos
Manuals
Legal materials[a]
Training bulletins
Codes[b]

[a]Two variables, case law and legal interpretations were combined into one variable as, on closer analysis, the responses to these variables indicated that they were being perceived by respondents as being almost identical.

[b]Four of the five codes, Penal Code, Vehicle Code, Health and Safety Code, Welfare and Institutions Code, were combined into one category.

Readability

The sample reading materials were organized into the six categories described above. A random sample was selected from each of the six categories. The number of passages selected for each category ranged from a minimum of 26 to a maximum of 72. The goal was to generate a stable readability estimate for each category. The number of reading passages chosen for each category depended on the length of the passages being analyzed, the expected variability within a category, and the number of passages sent in by agencies.

Each specific passage was then entered into a computer file and the file was edited to be sure there were no typing errors. (For a discussion of computer generated readability indices, see Appendix G.) Once the file was verified, all passages were subjected to readability analysis. As was stated in a previous section of this report, the interpretation of readability indices is hazardous because of the absence of a benchmark for determining reading difficulty and because of the limited number of variables used in the computation of the indices. In an effort to limit the degree of hazard, and to make the information generated by the readability analysis as meaningful as possible, POST staff decided to present not only the final readability indices but also the intermediary values that were used to compute those indices. This includes measures of average sentence length, average number of syllables per word, number of syllables per 100 words, and number of polysyllables (words with 3 or more syllables) per 100 words. By knowing these values, direct comparisons can be made between different reading materials, independent of the notion of grade level. Thus, average sentence length, average syllables per word, etc., from different materials can be directly compared. Tables 6 and 7 summarize the results of the readability analysis.

Table 6: Reading difficulty of law enforcement material by reading category.

Category	Number of passages analyzed	Flesch index mean	Smog index mean	Fog index mean
Codes	27	17.7	17.7	26.6
Manuals	46	14.6	14.7	18.7
Training Bulletins	47	13.1	13.8	16.0
Legal	72	12.9	14.4	17.6
Memos	49	10.8	12.0	13.7
Reports	56	8.2	9.6	11.0
Overall	297	12.4	13.4	16.4

Table 7: Direct comparison of values used to compute readability indices.

Category	$\bar{x}$ sentence length	$\bar{x}$ syll. per wd.	$\bar{x}$ syll. per 100 wds.	$\bar{x}$ polysyll. per 100 wds.
Codes	50.4	1.6	160.7	16.8
Manuals	28.8	1.6	167.8	18.1
Training Bulletins	22.7	1.6	159.6	16.5
Legal	26.4	1.6	155.6	17.2
Memos	20.2	1.5	151.5	14.0
Reports	19.3	1.4	136.2	7.9
Overall Average	26.0	1.6	154.2	14.9

Tables 6 and 7 reveal some interesting characteristics of the readability estimates and the values (average sentence length, number of syllables per word, etc.) utilized to compute them. Regarding the ordering of reading materials in terms of difficulty, the tables demonstrate that the different equations were relatively consistent. This was expected.

In spite of the relative consistency in ordering, there were significant differences in the grade levels assigned to each category by the different readability equations. This outcome was also expected. The Flesch and the Smog were consistently similar in their results. The Fog was higher than either of the other two.

In comparing the values used to compute the indices, it is interesting to note only sentence length has significant variance. There is moderate variability in the number of polysyllables, but almost no variance in average syllables per word and mean number of syllables per 100 words.

When developing reading tests, staff concentrated on achieving values for the tests (number of polysyllables, average syllables per word, etc.) that corresponded to those generated in the analysis of work samples. This resulted in the reader levels of the tests corresponding approximately to the mean difficulty level for the job sample materials.

Throughout this report the reader level of the various tests are reported. This was done, however, to provide a traditional reference rather than to serve as a meaningful descriptor. The reader should focus on the constituent variables (mean sentence length, number of polysyllables, etc.) when comparing the tests to the job materials.

Table 8 shows the ranking of reading categories by each readability index and each intermediary values.

Table 8: Ordering of reading categories by different readability measures.

	Flesch	Smog	Fog	$\bar{x}$ sentence length	$\bar{x}$ syllables per word	$\bar{x}$ polysyllables per 100 words	$\bar{x}$ syllables per 100 words
1	Codes(17.7)	Codes(17.7)	Codes(26.6)	Codes(50.4)	Manuals(1.64)	Manuals(18.012)	Manuals(167.824)
2	Manuals(14.6)	Manuals(14.7)	Manuals(18.7)	Manuals(28.8)	Codes(1.62)	Legal(17.159)	Codes(160.745)
3	Training(13.1) Bulletins	Legal(14.4)	Legal(17.6)	Legal(26.4)	Training(1.60) Bulletins	Codes(16.799)	Training(159.560) Bulletins
4	Legal(12.9)	Training(13.8) Bulletins	Training(16.0) Bulletins	Training(22.7) Bulletins	Legal(1.55)	Training(16.538) Bulletins	Legal(155.611)
5	Memos(10.8)	Memos(12.0)	Memos(13.7)	Memos(20.2)	Memos(1.52)	Memos(13.955)	Memos(151.532)
6	Reports(8.2)	Reports(9.6)	Reports(11.0)	Reports(19.3)	Reports(1.37)	Reports(7.877)	Reports(136.228)

Review of Table 8 reveals that sentence length was the dominant factor in determining rank difficulty for this sample of reading materials. Codes were ranked as most difficult on only one of four values -- sentence length. Police manuals were regarded as most difficult by three of the four. Thus, a substantial difference on one factor had a greater impact on the readability indices than did the relatively small differences in the remaining three.

This review and analysis of the materials commonly read by entry-level officers reveals that the reading demands of the job are quite significant. Relative to many other entry-level positions, the reading level required by law enforcement positions can only be considered as being high. Further, since an officer must be able to read adequately from the first day on the job, ability to read is definitely a prerequisite to job performance.

Developing and Scoring a Content Valid Reading Test

Having established the range of reading difficulty of the job, the decision had to be made at which level to establish the test. Traditional wisdom relative to the content validating of reading tests indicates that the difficulty of the test should, as closely as possible, correspond to the difficulty of the job. This usually results in a test with passage difficulties spanning the range from the minimum to maximum reader level as indicated by the readability analysis, and with the mean difficulty again pegged to the job analysis. The issue, however, is not so simple. Preliminary to developing a content valid reading test, two related questions needed to be addressed:

1. Does one really want to test for the level of the job?

2. How does one know what level is actually being tested?

Regarding the first question it was concluded that testing for the span of difficulties was not appropriate. Rather, the test should be focused at the mean difficulty of the job materials. Regarding the second, it was decided that the Cloze procedure offered the most precise means of estimating the difficulty level being tested. These decisions are discussed below.

One complicating and frequently misleading factor in reading research, is the use of the term "level" when referring to the reading demands of the job. The problem with referring to the reading level of a job is the implication that there is a point, a specific reading difficulty "level," that represents the minimum competence level that an individual can possess and still be able to read job-related materials. It follows from this line of reasoning that only individuals who read at, or above, the level of the job can adequately read and comprehend job materials. The problem with this contention is that it incorrectly perceives the nature of an individual's reading competency. This position presupposes that an individual can read up to a level but not beyond it. This is obviously incorrect.

An individual's comprehension rate is, in fact, variable depending on the difficulty of the material being read. Comprehension increases as the reading difficulty decreases, and decreases as reading difficulty increases. A person does not have 100% comprehension up to a point and then zero comprehension beyond it. Thus, if a person is tested and found to read at the 10th reader level, that means the individual has achieved some criterion level of comprehension, usually 70%, at that level. Knowing this, one would expect that individual to read with higher levels of comprehension at lower than a 10th reader level and with lower levels of comprehension at higher than the 10th reader level. That person would not, however, under any circumstances, read up to the 10th level and be be incapable of comprehending beyond it.

In view of the above, POST decided not to test above the mean difficulty level of the job. As just stated, an individual can read with varying levels of competency along a continuum of levels. Staff felt that established competency around the mean level of difficulty indicated a sufficient reading ability to be able to handle the more difficult levels encountered on the job even though a substantial effort might be required on the individual's part.

In response to the second question regarding the identification of the level at which one is testing, POST decided to use the Cloze procedure. As was stated in the previous sections, it is difficult to ascertain reading level while utilizing multiple-choice tests. With the Cloze procedure, research has indicated that a score of approximately 44% correct on a Cloze test corresponds to a 70% comprehension rate at the level of the passage[4] (Bormuth, 1967).

Test Development

While there is extensive literature on the development of Cloze tests for reading assessment, it is focused entirely on the application of that procedure to the educational setting. Within that context, Cloze tests are utilized primarily to determine the reading level of the student, thus ensuring a proper match of student ability to curriculum difficulty.

In the personnel setting, Cloze tests will be utilized to determine suitability for employment. Within this setting there are many considerations which are not present in the educational area. A

---

[4]This conversion applies only when scoring is done on an exact replacement basis, allowing no synonyms. In the scoring of the POST tests synonyms were allowed, thus, changing the Cloze score that corresponds to 70% comprehension. While there is no research presently available that addresses the percent correct score that corresponds to 70% comprehension when allowing synonyms, POST research has found that Cloze test scores tend to improve by a constant when going from exact replacement to synonym scoring. This would tend to support the notion that a 70% comprehension rate would be achieved on a Cloze test using synonym scoring at the percent correct equal to 44 plus the percent improvement resulting from the acceptance of synonyms.

number of these are a function of the requirements imposed by the Uniform Guidelines. Included here are the issues of test fairness, differential validity, technical methodological requirements, pass-point setting and candidate ranking. Other considerations relate to the needs of the personnel process in general, such as standardizing scores so they can be combined with other components of the selection process, establishing expectancy charts to predict future job performance, and developing scoring procedures which make it administratively feasible to utilize the Cloze procedure on large candidate populations. These were all issues which the POST research had to address. The logic and analyses that produced the final test battery are summarized below.

The first issue to be addressed was whether to utilize one reading passage or multiple passages. POST decided to conduct research on numerous tests at different reader levels. By taking this approach, POST could both meet the Uniform Guidelines' requirement regarding the investigation of alternative tests and also identify that test, or combination of tests, that produced the greatest measurement precision.

Originally, 10 passages of approximately 400 words each were constructed. None of these passages were written from scratch; rather, existing materials were identified and modified. The passages were written to be essentially neutral in subject matter with perhaps some general relationship to law enforcement work. Also, the passages were written to be as free of jargon or cultural idiom as possible, and free of proper nouns, numbers, and words which could not be deduced from context.

Six of the 10 original passages were eliminated from further development because they either contained too many words which could not be deduced from context, were not sufficiently neutral, or they contained some wording which might potentially have some cultural bias. Four were retained for further analysis.

The issue to be addressed next was that of the deletion pattern. There are two dimensions of this issue: one relates to the frequency of deleted words (e.g., every 5th word, 6th word, etc.), and the other to which of the possible deletion patterns to utilize for testing purposes. Traditionally, Cloze tests use anywhere from every 5th to every 10th word deletions. The 5th word pattern results in a maximum number of words that can be deleted while retaining sufficient context to determine the deleted words. The 10th word pattern represents the opposite extreme, where deleting fewer words simply requires the generation of a longer text with no concomitant increase in measurement precision. When the 5th word pattern was utilized, respondents generally complained that there were too few words to adequately determine the content of the passage. The 10th word deletion pattern required tests which were too lengthy, thus, resulting in unacceptably long test administration times. Ultimately POST compromised and used a 7th word deletion. This has been totally acceptable, giving the test takers sufficient context to determine deleted words while resulting in a test administration length acceptable to user departments.

Regarding which of the seven possible deletion patterns to be utilized, staff again found themselves slightly at odds with existing research. Research indicates that all deletion patterns are equivalent in terms of difficulty. This is true in that each pattern will rank applicants in approximately the same way, but incorrect relative to the mean difficulty and variance of the different patterns. If the passages are extremely lengthy, then the assumption of equivalency of patterns is probably correct, but with passages of approximately 400 words, it definitely does not hold. Since one of POST's goals in testing was to maximize variance, staff selected those deletion patterns that resulted in this end.[5]

The final battery consisted of two tests, one at the 10th reader level, and one at the 12th reader level. The 5th and 8th level tests were eliminated from the battery because they were not sufficiently reliable or valid. A shortened version of the 5th level test was retained to serve as a warm-up for the test taker. It was assumed that the Cloze format would be new to most test takers. Subsequently, staff felt that a very easy example would aid the candidate in getting used to the new format. Table 9 summarizes the readability indices, and the values used to compute those indices, for the two tests in the test battery.

Table 9: Readability indices and the values used to compute those indices.

Test	Flesch index	Smog index	Fog index	$\bar{x}$ sen. length	$\bar{x}$ syll. per word	$\bar{x}$ syll. per 100 wds.	$\bar{x}$ polysyll. per 100 wds.
One	9.4	10.1	10.4	14	1.5	150	12.5
Two	11.5	11.7	12.7	16	1.6	162	16.5

The values presented in Table 9 correspond with those generated by the sample job materials. Average syllables per word and the number of polysyllables per 100 words are identical to the job sample. Similarly, the number of polysyllables per 100 words is in the middle of the range found on the job. The mean sentence lengths for the tests are slightly lower than those found in the sample materials. This is of no significance, however, because sentence length does not begin to markedly affect reading difficulty level until the average sentence length approaches 20 words per sentence. This is because the relationship of sentence length and reading difficulty is expressed by a curve. Up to 20 words per sentence, the curve is very flat indicating little effect of length on difficulty. Above 20 words per sentence, the curve becomes progressively steeper indicating a greater affect of length on difficulty.

[5]Staff selected randomly three of the seven possible deletion patterns for each test. These were then trial administered and scored. Staff then selected the deletion pattern for each that resulted in the maximum variance for the total test battery.

For the reading tests which comprised the final battery, the candidates were given the following instructions:

> This is a test of how well you can read. The test is completed by supplying missing words in sentences. Every place you see a blank line, you are to supply the correct word. All blank spaces, regardless of the length of the missing word, are the same size. Some of the missing words are long, while others are short.
>
> For example, a sentence in the test might read, *"The driver was injured when his car crashed 1_____ the tree."* You would complete the sentence by printing "into" in the blank space provided: *"The driver was injured when his car crashed 1 into the tree."* The actual test consists of three reading passages, each with about 50 missing words.

Test Scoring

The scoring of the Cloze test, has been and remains, one of the more controversial aspects of the entire procedure. The controversy focuses on two issues: one administrative, one technical. The administrative "problem" with the Cloze test is that it must be hand-scored. There is no doubt that hand scoring is less desirable than machine scoring. It is slower, more costly, and more prone to error than is automated scoring. On balance, however, staff felt that the psychometric advantages of the Cloze procedure far outweighed the cumbersome nature of its scoring.

The second issue of controversy is much more complex than is the first. It regards the rules for identifying acceptable words. A Cloze test can be scored in one of two ways: the exact word method, where synonyms are not accepted, and the correct synonym method, where exact word and correct synonyms are both accepted. POST chose the latter procedure. Four issues or questions were addressed prior to making this decision.

1. Is one method more valid than the other?

2. Does one procedure generate more adverse effect than the other?

3. On the individual level, are there many pass/fail reversals when changing from one method to the other?

4. Is one procedure more acceptable to the test users and test takers than is the other?

Relative Validity of the Two Scoring Methods

Research has been conducted that deals directly with this issue. Taylor (1953) conducted the original research on the question and found that counting grammatically correct synonyms in addition to

exact words did not produce results superior to those obtained by counting only exact words. Ranking (1957) and Ruddell (1963) concluded that variances were slightly increased when synonyms were allowed, but that overall results were not enhanced or made more precise as a consequence of allowing synonyms. Bormuth (1965) addressed the same issue and came up with similar results. Based on a multiple regression analysis he concluded 95% of the variance on a reading comprehension test could be explained by exact words alone. Bormuth concluded subsequently, that the most economical and objective scoring procedure was the exact word method. Thus, relative to the issue of validity, the research tends to support the equivalency of the procedures (equivalency in this sense means that both scoring methods are equally predictive).

To determine if the procedures were, in fact, equivalent for this study, staff scored all tests using both procedures and then correlated the two data sets with both criteria. This analysis demonstrated that the two procedures were in fact equivalent. For a complete presentation of all correlations, see Appendix H.

The two scoring procedures are also equivalent (equally valid) from a content validity standpoint. A Cloze test determines if an individual, given certain linguistic cues, can identify the deleted words thereby completing a passage. Whether an individual selects the exact word deleted, or another word which is also syntactically correct and semantically appropriate, makes no difference from a reading assessment perspective. The same abilities are required to identify an appropriate synonym as are required to identify the exact deleted word.

It should be noted that staff is not claiming that the synonym and exact word scoring methods are equivalent for all utilizations of the Cloze procedure. The intent of the POST research was only to demonstrate that from a content and criterion-related validity standpoint, there are no differences.

Group Level Differences Between Procedures

Two ways of comparing group performance on a test are by analyzing differences in mean score and analyzing differences in passing rates. The latter comparison is the one utilized in the Uniform Guidelines for making determinations of adverse effect. Analyzing mean scores is useful, however, as it does point out differences in group performance. When comparing means, POST found that the relative order of ethnic group means remained the same for both scoring procedures. For a complete presentation of all group means using both scoring methods, see Appendix I.

When analyzing the passing rates, a procedure must be developed that allows the passing rate generated by one scoring method to be directly compared to that generated by the other. This can be accomplished by designating an arbitrary passing rate for the entire group and then determining the passing rates for each of the different ethnic groups. Four passing rates were selected and group rates were computed for each. Scoring method was shown to have no impact on passing rate. Appendix J presents all group passing rates at the three designated overall rates.

Pass/Fail Reversals when Moving from One Method to the Other

In order to make a comparison between the two methods, staff identified, using both methods, the highest scores on the 10th and 12th level tests which did not violate the 80% rule (Uniform Guidelines Section 4D)[6].

Using these scores as passpoints staff compared the number of pass/fail reversals that resulted from the different methods. The total number of pass/fail reversals that occurred in a sample of 344, was only 16. This represented 4.7% of the total sample.

When addressing pass/fail reversals, one is focusing on but one point (the passpoint) on the entire distribution of scores. Having demonstrated that scoring method has no effect on the number of individuals passing at the passpoint, staff wondered if either procedure would tend to systematically affect any point in the distribution. In order to determine the effects of the scoring procedure across the entire distribution of scores, staff analyzed (for the entire distribution of scores) the score gain resulting from the acceptance of synonyms. The amount of gain was found to be relatively constant throughout the distribution. From this one can conclude that both scoring methods pass and fail the same people. Individuals are ranked in the same order regardless of the scoring method utilized. All analyses regarding pass/fail reversals are presented in Appendix K.

Candidate Acceptance

The last issue to be addressed relative to scoring was that of user and candidate acceptance. Staff perceived this to be of extreme importance, because acceptance of a test is to a large extent the primary determining factor of whether it ever gets administered. It was staff's observation that most individuals, when questioned on the issue, felt it "unfair" if synonyms were not allowed. A reiteration of the research just described, which demonstrated consistently that method of scoring made no difference on either validity, or group or individual level of performance, left most people unconvinced. The dominant attitude was, for example, that a "car" and an "auto" are the same thing regardless of what the statistics said. Since there was no measurement advantage with either method, staff decided to accept synonym scoring because that was the method that generated the highest level of acceptance.

---

[6]The scores expressed as percents are, for the 10th and 12th level tests respectively, 49 and 39 for the exact replacement method and 62 and 54 for the correct synonym method.

Disadvantages of Allowing Synonyms

It should be noted that, independent of the measurement issue, there are certain disadvantages to allowing synonyms. First among these was the problem of identifying acceptable synonyms. The POST procedure of identifying synonyms consisted of recording all responses by all candidates in the sample group to all questions on the test. Staff then produced a listing (by overall group and by individual ethnic group for each test) of all responses to each utilized item, and the frequency with which the responses were made. Staff assessed each word to determine if it could be considered an acceptable synonym. To be considered acceptable, the word had to be judged to be syntactically correct and semantically appropriate.

The primary problem with this approach is that it requires numerous trial administrations so that a complete list of synonyms can be generated before the test can be actually used as a selection device. This is, however, the only administratively feasible and technically acceptable procedure for generating a listing of correct synonyms.

Nevertheless, since the literature and POST's own research conclusively demonstrated that nothing was lost in terms of measurement precision by utilizing either of the two scoring methods, staff decided on the procedure that, although more difficult to develop, appeared more acceptable to the user agencies and candidates.

Identifying the Writing Demands of the Job

In the 1979 POST Job Analysis (Kohls, Berner and Luke), 23 different writing tasks were identified and found to be important to the job. The same rating scales were utilized to evaluate the writing tasks as were used to evaluate the reading tasks. (For a review of those scales see Appendix A.)

The writing tasks are presented in Table 10.

Table 10: Writing variables that emerged as being important in the original POST job analysis.

Writing tasks
Complete reports consisting primarily of check-off boxes or fill-in blanks (e.g., vehicle impound reports).
Take notes.
Write news releases.
Write interoffice memos.
Write letters or other correspondence as part of the job.
Prepare misdemeanor court complaint forms.
Prepare felony court complaint forms.
Prepare paperwork for process service.
Make entries in activity log, patrol log, daily report or departmental records.
Record and communicate descriptions of persons (e.g., suspects, missing persons).
Obtain search warrants.
Issue citations for non-traffic offenses.
Request that DMV re-administer driver's test to persons currently licensed.
Issue Vehicle Code citations.
Issue warning tickets (for equipment, moving, or parking violations).
Issue parking citations.
Summarize in writing statements of witnesses, complainants, etc.
Record formal confessions in writing.
Fill out surveys.
Write evaluations of training received.
Prepare lesson plans.
Write in-depth narrative reports containing complete sentences and paragraphs (e.g., investigative reports, supplemental/follow-up reports).
Write reports consisting of several short descriptive phrases, sentence fragments, or very short sentences (e.g., incident report).

As stated previously, the original POST job analysis found writing to be both a frequent and important aspect of the entry-level position. Having demonstrated this to be the case, the next step was to conduct a supplemental analysis to determine precisely what knowledges, skills, and abilities were required in order to write adequately in the entry-level job. This was accomplished by developing a "Survey of Writing Mechanics" and administering it to officers and supervisors throughout the State. A copy of this survey is presented in Appendix L.

Survey of Writing Mechanics

The purpose of this survey was to identify which specific rules of grammar and punctuation were crucial for the writing done by law enforcement officers. In the survey, respondents were told:

> Effective writing requires the knowledge and correct use of "writing mechanics" - rules and techniques which determine the content and structure of written material. This questionnaire is designed to identify the specific writing mechanics which are essential to effective writing by patrol officers. The information will be used in the design of a writing ability test which will be used to evaluate peace officer applicants.
>
> On the following pages are listed examples of a number of aspects of writing mechanics such as grammatical rules and rules of punctuation. Regarding each aspect, an example is provided in the form of a word or sentence which is written incorrectly (i.e., an example of an error), followed by one which is written correctly.
>
> We would like you to evaluate each aspect of writing mechanics listed on subsequent pages by means of the following process. Consider the first topic area on Page 3. Read the correct and incorrect examples. Next, decide which of the following three categories contains the most accurate statement regarding the type of writing error in the incorrect example:
>
> Category 1. Acceptable in a report...
>
> Category 2. Not acceptable, but person can be remediated...
>
> Category 3. Not acceptable and person should not be hired...

Surveys were completed by officers in a small sample of departments throughout the State (Los Angeles Police Department, Los Angeles County Sheriff's Department, Santa Rosa Police Department, Sacramento Police Department, Sacramento County Sheriff's Department, San Francisco Police Department).

Based on the analysis of these surveys, staff concluded that there were only three prerequisites of writing which could justifiably be subjected to examination: clarity, word usage, and spelling. Departments indicated that all other errors of grammar and punctuation were either acceptable or could be corrected by a reviewing officer. Thus, it was these subjects which would be the focus of the POST tests.

Developing a Content Valid Writing Test

Two writing tests were developed, an open-ended essay test, and a multiple-choice test. Both measured abilities prerequisite to the performance of job tasks. The open-ended essay measures one's ability to express in writing ideas that can be easily understood. This is a short test in which the candidate is required to write a 200-300 word passage on an assigned topic. The test takers were given the following instructions:

> In a 200-300 word passage, write about either: (a) the illustration which follows these instructions; or (b) the steps you took leading up to your making an application for a job in law enforcement. The topic you choose is not important. However, it is important that what you write is understandable.
>
> Remember, you will not be graded on what you write about but how clearly you express yourself. In other words, as long as the scorers can understand your ideas (whether they agree with them or not) you will pass this part of the screening process.

The objective test of writing mechanics was intended to measure three prerequisites of good writing: clarity, vocabulary, and spelling. Each of these subject matters was addressed by a fifteen item multiple-choice test segment. In writing the items for each segment, every effort was made to ensure that each item was both job related and psychometrically sound.

In the clarity section, the items are comprised of two sentences, one correct and one incorrect. The test taker is given the following instructions:

> In the following pairs of sentences, identify the sentence which is most clearly written. If sentence "A" is more clear than sentence "B," mark "A" on your answer sheet. If sentence "B" is more clear than sentence "A," mark "B" on your answer sheet.

Some of the sentences are neutral in terms of subject matter while others address law enforcement topics. In no instances, however, was a knowledge of either police procedures or law enforcement jargon required to correctly answer the test question. In each item the error in the incorrect sentence is one which makes the sentence difficult to interpret or ambiguous. Such an error in a police report could lead to an interpretation different from that which the writer had intended.

In the word usage and spelling segments, staff's primary concern was identifying specific words which were job-related, at the appropriate level of difficulty and would not be learned in the academy. To achieve this end, staff generated a word list comprised of words at various levels. Staff used the Living Word Vocabulary (Dale and O'Rourk, 1976) in developing this list. Once the list was completed, it was reviewed by various Law Enforcement Consultants on the POST staff with instructions to identify all words on the list which were job-related yet not taught in the academy. The words that they identified served as the pool from which staff developed test items. Specific words became word usage or spelling items based on the judgment of the POST research staff.

In the word usage section, test takers were given the following instructions:

> In each of the following sentences, choose the word that most nearly has the same meaning as the underlined word. Mark on your answer sheet the letter that identifies the correct choice.

In the spelling section they were given the following instructions:

> In the following sentences, choose the correct spelling of the missing words. Mark on your answer sheet the letter that identifies the correct choice.

After considerable research and analysis, the multiple-choice test was selected for the final test battery. The essay test was found to be neither as reliable nor as valid as the objective test. Appendix M presents the criteria used for the scoring of the essay test.

Thus, the POST test of writing mechanics was comprised of 45 items. The 15 clarity items utilized a two-choice format and the 15 word usage and 15 spelling items used a four-choice format.

Empirical Validation of Tests

The reading and writing tests were empirically validated so that the precise relationship of tests to criteria could be computed and so that all necessary quantitative analyses could be conducted. The tests were administered to seven different academy classes. Thus, the study was replicated seven times. The research sample is presented in Table 11.

Table 11: Sample by ethnicity and academy.

	Race					
Academy	Black	White	Asian	Spanish Surname	Total	Percent of total
Santa Rosa	0	17	1	3	21	6.1
Sacramento	0	27	4	2	33	9.6
San Francisco 1	5	23	6	1	35	10.2
San Francisco 2	3	17	6	3	29	8.4
San Francisco 3	4	18	6	1	29	8.4
Los Angeles PD	13	66	4	22	105	30.5
California Highway Patrol	11	61	2	18	92	26.7
Total	36	229	29	50	344	
Percent of total	10.5	66.6	8.4	14.2		

Note. The N's for the various analyses reported in this section are all smaller than 344. This is because not all individuals in the sample had complete data sets.

Tests were administered to individuals in the sample. The test norms resulting from these administrations, are presented in Table 12.

Table 12: Test norms for academy students.

Test	Number	Mean	Standard deviation (percent)	Standard error (percent)
Reading Level 10	343	74.8	10.4	.56
Reading Level 12	329	66.8	11.9	.69
Total Reading	329	71.0	10.0	.55
Writing: Clarity	252	70.4	16.1	1.02
Writing: Word Usage	237	80.7	12.6	.82
Writing: Spelling	225	77.2	17.2	1.15
Total Writing	225	76.5	11.1	.74

Note. Mean scores are presented as percent correct.

After looking at the test data, and considering the administrative problems associated with scoring, staff decided to combine the two reading tests to produce a total reading score and the three writing tests to produce a total writing score. In addition to simplifying scoring, combining tests in the above manner also increased test reliability. Utilizing the test-retest procedure, the reliability of the reading test was found to be .87 and, using the split-half procedure, the reliability of the writing test was found to be .86.[1]

Individual passpoints are set for both the total reading score and the total writing score. For those who pass, these two scores are combined to provide a single score for ranking purposes. By using this procedure agencies score the tests in the same way they were scored in the validation study. The procedure for combining scores, along with a general summary of the selection process, is presented in the Appendix N.

Regarding the criteria, reliability estimates were generated for the Proficiency Test but not for the composite of academy scores. The reliability of the Proficiency Test is .92. This is a stable estimate resulting from over 80 test administrations involving approximately 4,000 cadets. No reliability was computed for the composite scores because there was insufficient data available to do so. Composite scores were generated from academy records. Unfortunately, most academies maintained only total test scores as opposed to raw item data. Thus, reliabilities could not be computed. It does seem reasonable, however, to assume good reliability for the composite scores. The composites were comprised of numerous items, a condition which tends to increase reliability, and the correlations of both criteria with the predictors are very similar.

Validity coefficients were generated by individual academy class and by total sample (all classes combined into one group). Correlations between tests and criteria were also computed for each ethnic group for which there were sufficient representatives to produce stable statistics. These data are presented in Tables 13 and 14. As was previously the case, the N is not constant within either agency or ethnic group as some respondents did not have complete data sets.

---

[1]There are three basic categories or types of reliability: internal-consistency, alternate-form, and test-retest (Guilford and Fruchter, 1978). The method employed depends both on the meanining one wishes to attach to the index and to the kind of data being evaluated. This latter consideration was the primary determining factor in this research. The internal-consistency model was felt to be most appropriate for the multiple-choice test and the test-retest model most appropriate for the Cloze test.

Table 13: Validities for each academy class in the sample for both criteria.

Academy	Criteria	
	Proficiency Scores	Internal Scores
Santa Rosa		
Reading[a]	.58(N=20)**	.62(N=20)**
Writing[b]	.36(N=21)*	.53(N=20)**
Combined[c]	.58(N=20)**	.69(N=20)**
Sacramento		
Reading	.23(N=33)	.34(N=33)*
Writing	.36(N=33)*	.37(N=33)*
Combined	.35(N=33)*	.41(N=33)**
San Francisco 1		
Reading	.56(N=33)**	.41(N=31)**
Writing	-	-
Combined	-	-
San Francisco 2		
Reading	.67(N=29)**	.61(N=28)**
Writing	-	-
Combined	-	-
San Francisco 3		
Reading	.72(N=24)**	.50(N=24)**
Writing	.72(N=17)**	.60(N=17)**
Combined	.76(N=17)**	.63(N=17)**
Los Angeles PD		
Reading	.61(N=94)**	.50(N=96)**
Writing	.49(N=64)**	.55(N-62)**
Combined	.63(N=64)**	.62(N=62)**
California Highway Patrol		
Reading	.40(N=87)**	.49(N=88)**
Writing	.31(N=88)**	.42(N=89)**
Combined	.42(N=85)**	.53(N=86)**

Note. No writing test had been developed when the San Francisco 1 and 2 classes were tested.

[a] Reading = Reading Level 10 + Reading Level 12

[b] Writing = Clarity + Word Usage + Spelling

[c] Combined = Total Reading + Total Writing

* Significant at the .05 level

** Signficant at the .01 level

The validity coefficients for the total group and for each ethnic subgroup are presented in Table 14.

Table 14: Validities for total group and for all ethnic groups.

Group	Criteria	
	Proficiency scores	Internal Scores
Total group		
Reading[a]	.50(N=320)**	.47(N=320)**
Writing[b]	.40(N=223)**	.45(N=221)**
Combined[c]	.51(N=219)**	.53(N=218)**
White		
Reading	.45(N=215)**	.43(N=215)**
Writing	.35(N=158)**	.46(N=155)**
Combined	.46(N=154)**	.51(N=152)**
Black		
Reading	.44(N=33)**	.29(N=34)*
Writing	.51(N=19)**	.41(N=19)*
Combined	.62(N=19)**	.57(N=19)**
Asian		
Reading	.52(N=27)**	.38(N=24)*
Writing	.86(N=12)**	.61(N=12)**
Combined	.73(N=12)**	.51(N=12)*
Spanish Surname		
Reading	.52(N=45)**	.52(N=47)**
Writing	.38(N=34)*	.38(N=35)*
Combined	.49(N=34)**	.49(N=35)**

[a] Reading = Reading Level 10 + Reading Level 12
[b] Writing = Clarity + Word Usage + Spelling
[c] Combined = Total Reading + Total Writing
* Significant at the .05 level
** Signficant at the .01 level

Tables 13 and 14 report uncorrected Pearson product-moment correlations. Table 13 shows that the combined test scores are significantly correlated with both criteria in all seven academy classes. Even though both tests are highly correlated with both criteria, staff decided to keep both tests in the test battery. This decision was supported by the results of a multiple regression analysis which showed that each test was contributing unique variance for the prediction of the criteria. Thus, both tests were retained in the battery.

The magnitude of these correlations is quite impressive, particularly for this type of research. In his comprehensive review of validity studies, Ghiselli (1955) found average validities ranging in the .30s and low .40s. He found validities above .50 extremely rare. Of the ten combined validities reported in Table 13, one exceeded .70, four exceeded .60, two exceeded .50, two exceeded .40, and one exceeded .30. Thus, seven of the ten validities generated in this study exceeded what Ghiselli considered to be the upper limits for this type of research. The three lowest validities in this study fall in the range defined by Ghiselli as average. Clearly, the reading and writing tests are extremely predictive.

Table 14 demonstrates that the tests are also valid for all ethnic groups for both criteria. Again, the magnitude of the validities is impressive. Of the eight validities computed for the different ethnic groups, one exceeded .70, two exceeded .60, two exceeded .50, and three exceeded .40. No validity coefficient was below .46.

To determine if there was any differential prediction by race, a test of regression equations was computed (Cleary, 1968). Differential prediction was found, but in the range of usable test scores minorities are always overpredicted when the common regression line is used. This being the case, staff felt the common regression line appropriate.

Table 14 also presents the validities for the combined group. The validities for this large group with both criteria exceeded .50. The possibility of a correlation of such magnitude occurring randomly for a sample of 219 is 1 in 100,000.

Correlations were corrected for unreliability in the criteria but because of the high reliability of the criteria (.92 for the proficiency test), the increase was a negligible .02 for each correlation. Since the increase was so small, a table of corrected correlations was not presented.

Utility

The utility of a test can be defined as the, "...degree to which its use improves the quality of people being selected beyond what would have occurred had the device not been used" (Blum and Naylor, 1968). Utility can be expressed in different ways. Two different estimates of utility are presented in this report. Table 15 presents a direct measure of improvement in prediction that results from using the tests. Table 16 presents the gain in criterion performance, expressed in percentile terms, that results as scores on the predictor increase.

Table 15 presents the utility of the combined total reading and total writing scores with each of the criteria at various passpoints. The predictor scores are presented in standard score units (T Scores). Because two T Scores are added together to get the combined score, the mean of the combined scores equals 100. The standard deviation is 17.32.

Table 15: Utility of combined reading and writing test scores for predicting academy performance.

Passpoint on predictor	Proficiency scores			Internal scores		
	N	% of cadets above passpoint who were in top 50% of class	% of cadets who would be placed in top 50% of of class by chance	N	% of cadets above above passpoint who were top 50% of class	% of cadets who would be placed in top 50% of of class by chance
135 (+2 SD)[a]	2	--	--	2	--	--
117 (+1 SD)	22	72	50	33	79	50
100 (mean)	123	67	50	123	68	50
83 (-1 SD)	184	58	50	184	62	50
65 (-2 SD)	208	53	50	209	57	50
48 (-3 SD)	218	51	50	219	55	50

Note. The mean of the predictor is 100 and the standard deviation is 17.238. Therefore, each point in the predictor passpoint column is one standard deviation apart.

[a]An N of 2 was too small to produce any meaningful data.

Table 15 demonstrates a substantial gain in prediction resulting from test use across the range of scores. As would be expected, the relative utility of the tests tends to improve as test score increases. Considering that it costs between seven and nine thousand dollars to get an individual through the academy, and approximately two thousand dollars just to get an individual to the academy, there are substantial costs associated with errors in prediction by the selection procedure.

Another means of expressing utility is to show the gains in performance on the criteria that correspond to gains in performance on the predictor. Table 16 presents six different passpoints and the performance in the academy, expressed in percentiles, that correspond to these passpoints.

Table 16: Percentile rankings on the criteria that correspond to different predictor scores.

	Proficiency scores	Internal scores
Predictor score	Percentile ranking that corresponds to predictor score	Percentile ranking that corresponds to predictor score
135 (+2 SD)[a]	87	91
117 (+1 SD)	73	67
100 (mean)	50	47
83 (-1 SD)	33	31
65 (-2 SD)	18	19
48 (-3 SD)	9	9

Note. The mean of the predictor is 100 and the standard deviation is 17.238. Therefore, each point in the predictor passpoint column is one standard deviation apart.

Table 16 clearly demonstrates the relationship of performance on the criteria to performance on the predictor. It is important to note that the predictor scores are extremely accurate. The standard error of measurement for the test is 5.9. This means that 68% of the obtained scores on the test are within 5.9 scale points of the true scores that determined them (Guilford and Fruchter, 1978). Thus, for a true score on the test of 100, one would be 68% sure that the obtained score would fall between 94.1 and 105.9. One would be 95% sure that the obtained score would fall between 88.2 and 111.8. Considering that the test has a score range of 83 points (minimum of 52.8 and maximum of 136) a standard error of 5.9 is relatively small. A small standard error reflects an accurate test.

This relationship of test scores to academy performance has obvious implications for the setting of passpoints. It is important to remember that all of this data was generated from individuals who successfully completed the academy. Since all individuals in the study completed the academy, to set a passpoint which would eliminate anyone in the study is tantamount to raising the academic standards of the academy. If a passpoint was set at 83, for example, it would result in the elimination of the bottom 33% of all existing academy students. To set the passpoint at the mean (100) would result in the elimination of approximately 50% of existing students. In each case, the academic standards would be raised by a factor proportionate to the percentage of existing students eliminated. Thus, the expectancy charts can play a useful role in the rational setting of passpoints.

Summary

POST intended that its tests of language ability be valid for all law enforcement jurisdictions in the state. To realize that end, a multi-unit study was undertaken utilizing a representative sample of agencies from throughout California. From this study, POST identified those language abilities prerequisite to the performance of job-related language tasks. POST then developed tests that measured those prerequisite abilities. By operationally defining the job skills in the job analysis, and by demonstrating logically that the POST tests measure those abilities prerequisite to the performance of job skills, POST has demonstrated the content validity of its tests.

In addition, POST also demonstrated the empirical validity of the language ability tests. The empirical study, replicated in seven different academy classes, showed the predictor tests (the reading and writing tests) to be significantly correlated with academic performance in all seven academy classes. The tests were also shown to be significantly correlated with academy performance for all ethnic groups.

The utility of the tests was demonstrated in two ways. The percent improvement in selection, above chance, that could be achieved by using the predictor tests was established. Also, the precise relationship of predictor scores to criteria achievement was presented. Both demonstrated the tests to have significant utility.

In light of this research, POST feels the language ability tests are appropriate for use as entry-level selection procedures for all local law enforcement agencies in the State of California.

REFERENCES

APA Division of Industrial-Organizational Psychology. Principles for the validation and use of personnel selection procedures. Second edition, 1980.

Berner, J. G. & Kohls. J. W. Needs assessment questionnaire. Sacramento: Commission on Peace Officer Standards & Training, 1978.

Blum, M. L. & Naylor, J. C. Industrial psychology: Its theoretical & social foundations, New York: Harper & Row, 1968.

Bormuth, J. R. Comparable cloze & multiple-choice comprehension test scores. Journal of Reading, 1967, 10, 221.

Bormuth, J. R. Validities of grammatical & semantic classifications of cloze scores. International Reading Association Conference Proceedings, 1965, 10, 283.

Cleary, T. A. Test bias: Prediction of grade of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Dale, E. & O'Rourke, J. The living word vocabulary, Elgin, Il.: Field Enterprises Educational Corporation, 1976.

Farr, R. Reading: What can be measured? Newark: International Reading Association Research Fund, 1969.

Ghiselli, E. E. The Measurement of Occupational Aptitude, Berkeley: University of California Press, 1955.

Goodman, K.S. Reading: A psycholinguistic guessing game. Journal of the Reading Specialist, 1967, 4, 126-35.

Goodman, K. S. Analysis of oral reading miscues: Applied psycholinguistics. Reading Research Quarterly, 1969, 5, 9-30.

Goodman, K. S. ed., The psycholinguistic nature of the reading process. Detroit: Wayne State University Press, 1968.

Goodman, K. S. Miscues: windows on the reading process. In K. S. Goodman (Ed.) Miscue analysis: application to reading instruction. Champaigne, Urbana, Illinois: ERIC Clearinghouse on Reading and Communication NCTF, 1973.

Guilford, J. P. and Fruchter, B. Fundamental statistics in psychology and education, (6th ed.). New York: McGraw-Hill Book Company, 1978.

Holmes, J. A., & Singer, H. The substrata-factor theory: Substrata factor differences underlying reading ability in know groups at the high school level. (Final report covering contracts No. 538, SAE-8176 and 538A, SAE-8660) Washington, D.C.: U.S. Government Printing Office, 1966.

Holmes, J. A. & Singer, H. Theoretical models & trends toward more basic research in reading. Review of Educational Research, 1964, 34, 127-55.

Honey, R. A. Development and validation of an entry-level law enforcement examination for the State of California. Sacramento: Cooperative Personnel Services, 1979.

Kerlinger, F. N. & Pedhazer, E. J. Multiple regression in behavioral research, San Francisco: Holt, Rinehart & Winston, Inc. 1973.

Klare, R. K. Assessing readability. Reading Research Quarterly, 10 (6), 1974-75.

Klare, R. K. Comments on Bormuth's readability: A new approach. Reading Research Quarterly, 1 (4), Summer 1966.

Kohls, J. W. Medical Screening manual for California law enforcement. Sacramento: Commission on Peace Officer Standards and Training, 1976.

Kohls, J. W., Berner, J. G., & Luke, L. K. California entry-level law enforcement officer job analysis. Sacramento: Commission on Peace Officer Standard and Training, 1979.

Jones, J. P. & Guion, R. M. An examination of six fair selection models. Bowling Green State University.

Leonnon, R. T. What can be measured? The Reading Teacher, 1962, 15, 326-37.

Lorge, I. Predicting reading difficulty of selection for children. Elementary English Review, 1939, 16, 229-233.

Luke, L. K. and Kohls, J. W. Background investigation manual: Guidelines for the investigator. Sacramento: Commission on Peace Officer Standards and Training, 1977.

Potter, T. C. A taxonomy of cloze research. Southwest Regional Laboratory for Educational Research & Development. Part 1, June, 1968.

Project Star. Davis: Davis Publishing Company, 1974-76.

Rankin, E. G. An evaluation of the cloze procedure as a technique for measuring reading comprehension. Unpublished doctoral dissertation, University of Michigan, 1957.

Raygor, A. L. Problems in the substata-factor theory. Reading Research Quarterly, 1966. 1(4), 147-50.

Rosenfeld, M. & Thornton, R. F. The development and validation of multijurisdictional police examination. Educational Testing Service, Princeton, N. J. 1976.

Ruddell, R. B. The effect of oral & written patterns of language structure on reading comprehension. Unpublished doctoral dissertation, University of Indiana, 1963.

Russell, D. H. & Fea, H. R. Validity of six readability formulas as measures of juvenile action. Elementary School Journal, 1951, 52, 136-44.

Ryan, E. B. & Semmel, M. I. Reading as a constructive language process. Reading Research Quartery, 1969, 5, 59-83.

Slobin, D. A. Psycholinguistics (2nd Ed.). Palo Alto, California: Scott, Foresman and Company, 1979.

Taylor, W. L. Cloze procedure: A new tool for measuring readability. Journalism Quarterly, 1953.

Uniform employee selection guidelines-interpretation and clarification. Equal Employment Opportunity Commission. Washington, D. C., 1980.

Uniform Guidelines on employee selection procedures. Equal Employment Opportunity Commission, Civil Service Commission, Department of Justice and Department of Labor, Washington, D. C., 1978.

Wilson, B. Performance objectives for POST Basic Course. Sacramento: Commission on Peace Officer Standards and Training, 1977.

Wollack, S., Clancy, J. J., & Beals, S. The validation of entry-level law enforcement examinations in the states of California and Nevada. Sacramento: The Selection Consulting Center, 1973.

Wollack, Waibel and Gunther, Inc. The validation of entry-level police officer selection procedures. Fair Oaks, California: Author, 1976.

APPENDIX A

READABILITY FORMULAS

Three readability formulas were utilized in this study. The formula for each is presented below.

The Fog Index

Reading grade level = .4 (average sentence length + percentage of words of three of more syllables).

The Smog Index

Smog grade level = 3 + square root of polysyllable count.

A polysyllable is any word of 3 or more syllables.

Flesch index = 206.835 - 1.015 * average sentence length - 84.6 * average number of syllables.

This index is then utilized in the Flesch grading level equation.

Flesch grade level = If Flesch index

70 - (Flesch index - 150/10)

60 - (Flesch index - 110/15)

50 - (Flesch index - 93/3.33)

-50 - (Flesch index - 140/6.66)

-60 - (Flesch index - 150/10)

APPENDIX B

RATING SCALES

RATING SCALES

IMPORTANCE: When this task is done, how important is successful completion of this task to overall patrol officer/deputy job performance?

(1) Of little importance
(2) Of some importance
(3) Important
(4) Very important
(5) Critically important

WHEN LEARNED: To what extent is it necessary that officers/deputies learn to perform this task in the academy and prior to to any job assignment?

(1) Not necessary--can best be learned on the job.
(2) Some preparation in the academy is necessary but full competence can best be achieved on the job.
(3) Full competence must be achieved in the academy before any job assignment.

RELATION TO PERFORMANCE: To what extent do successful officers perform this task better than marginal or poor officers?

(1) In general, all officers perform this task about equally well.
(2) Some officers perform this task better than others, but they are not necessarily the better performers.
(3) Generally, successful officers perform this task better than marginal or poor officers.

Example:

If "transporting prisoners/inmates" is a very important task, if full task competence must be achieved in the academy before any job assignment, and if all officers generally perform this task equally well, your ratings would be:

	Never Performed	Importance	When Learned	Relation to Performance
Transport prisoner/inmate		4	3	1

If the task is never performed in your agency, you would simply put a check in the column labeled "NEVER PERFORMED" and go on to the next item. Do not skip any item. Make sure you provide the ratings of "IMPORTANCE", "WHEN LEARNED", and "RELATION TO PERFORMANCE" for each task performed in your agency.

Remember to go through the entire questionnaire three times--that is, rate all tasks for IMPORTANCE before using the WHEN LEARNED or RELATION TO PERFORMANCE scales.

APPENDIX C

SOURCE OF INFORMATION QUESTIONNAIRE

BACKGROUND INFORMATION

This job inventory or survey is not an exam or any type of position evaluation instrument. The information you provide is for research purposes only. We do request that you provide your name, but only for possible contact by the research staff in the unlikely occurrence of an unforeseen data processing problem.

PLEASE PRINT

1. Date _____

2. Name of Agency _____

3. County in Which Agency is Located _____

4. Your Name _____

5. Your Current Assignment _____

6. Office Telephone Number ___( )_____

Please respond to the following questions by indicating your answers in the boxes to the right.

7. Age: ☐

---

8. Sex: Male = 1 Female = 2 ☐

---

9. Ethnicity:

American Indian	= 1	Oriental/Asian	= 4
Black	= 2	Spanish Surname	= 5
Caucasian (white)	= 3	Filipino	= 6
		Other _____	= 7

☐

---

10. Education (indicate highest level completed):

High School or G.E.D.	= 1	College Senior	= 5
College Freshman	= 2	Bachelor's Degree	= 6
College Sophomore	= 3	Master's Degree	= 7
College Junior	= 4	Doctorate Degree	= 8

☐

---

11. Present Rank:

Officer/Deputy	= 1
Corporal	= 2
Sergeant	= 3

☐

12. How long have you been at your present rank with your present agency? (Please indicate months) ☐

13. Present shift:

Day	= 1	Night (graveyard)	= 3
Evening (swing)	= 2	Relief	= 4

☐

SOURCE OF INFORMATION QUESTIONNAIRE

Two of the primary ways an individual receives information are through listening and reading. We at POST are conducting this survey to determine the ways in which officers receive information regarding their job. Based on our findings we will develop an entry-level selection test of reading ability.

INSTRUCTIONS

On the following pages are listed various materials, such as narrative reports, memos, the Penal Code, etc., which are commonly encountered by the entry-level officer assigned to radio car patrol. Although these materials generally originate in written form, the information contained in them can be presented to you either verbally or in a written document.

You are asked to identify how information regarding these materials is presented to you. To assist you in making your decisions, four categories, each describing a different means of presenting information, have been identified.

For each type of material decide which of the following categories best describes the manner in which information is presented to you.

Does Not Apply: This material does not pertain to my job.

Orally: Information regarding this material is presented to me verbally (e.g., in lectures, over the radio, at roll call, etc.).

Written Summary: Information regarding this material is presented to me primarily in summarized written form. I read the summarized version but not the original document.

Original Written Form: Information regarding this material is presented to me primarily in its original written form. (By "original form" we are referring to a non-summarized version of the original, not the original document itself).

Place a check (✓) in the box which identifies the category that best describes how information is primarily presented to you for each type of material. Check only one category for each type of material.

Type of Material	Source of Information Category			
	Does Not Apply	Orally	Written Summary	Original Written Form
1. In-depth narrative reports containing complete sentences and paragraphs (e.g., investigative reports, supplemental/follow-up reports).				
2. Reports consisting of several short descriptive phrases, sentence fragments, or very short sentences (e.g., incident reports).				
3. Reports consisting primarily of check-off boxes or fill-in blanks (e.g., vehicle impound reports).				
4. Street maps.				
5. Incoming correspondence.				
6. Interoffice memorandums.				
7. Wanted vehicle bulletins.				
8. Departmental manuals.				
9. Case law.				
10. Legal interpretations (e.g., California Attorney General's opinions, city attorney opinions).				
11. Teletype messages.				
12. Training bulletins.				
13. Warrants.				
14. Coded material (e.g., NCIC printout, DMV drivers' records).				

For the following five codes, if you indicate that you read the original document, also indicate whether you read an abridged or unabridged version.

Type of Material	Source of Information Category				
	Does Not Apply	Orally	Written Summary (Ready Reference Card)	Original Written Form	
				Abridged	Unabridged
15. Penal Code.					
16. Vehicle Code.					
17. Health and Safety Code.					
18. Welfare and Institutions Code.					
19. Municipal Code/County Ordinances.					

SPECIFIC CODE SECTIONS

The following five codes were identified in a statewide study as being the most important and the most frequently used. For each of these codes, list the specific sections (up to ten sections) that you refer to most frequently. If your particular job does not require you to refer to one or more of these codes, leave the spaces corresponding to those codes blank.

Penal Code _____

Welfare and Institutions Code _____

Vehicle Code _____

Municipal Code/ County Ordinances _____

Health and Safety Code _____

APPENDIX D

CODE SECTIONS INCLUDED IN READABILITY ANALYSIS

Table 17: Code Sections included in readability analysis.

Penal Code	N
148	126
211	276
240	127
242	286
245	255
261	138
415	300
459	436
484	141
487	247
488	239
514	250
647F	186
Vehicle Code	
10851	194
22350	349
22450	248
23103	106
4000A	176
12951A	124
21453A	163
23102A	240
Health & Safety Code	
11350	269
11357	237
11358	152
11359	132
11377	191
11550	205
Welfare Code	
300	174
600	146
601	376
602	387
5150	237

Note. N equals the number of officers in sample of 508 that indicated this code section is among the most frequently referred to.

APPENDIX E

CONTACT LETTERS AND CRITERIA FOR INCLUSION IN READING SAMPLE

The POST Research and Evaluation Bureau is developing an entry-level selection test of reading ability. This activity is a continuation of POST's on-going research and development effort to provide entry-level selection procedures for the use of local law enforcement agencies. But we need your help. In return, we plan to produce a validated reading ability test which is to be made available to all agencies that participate in the POST Program. The purpose of this letter is to determine which agencies can assist us in providing reading job analysis data. We request that your agency participate.

In order for you to have an idea of what participation in the study will involve, I have enclosed sample copies of two documents that should you decide to participate, would call upon individuals in your department to perform certain tasks. The staff time that will be required is actually quite minimal. The Source of Information Questionnaire, which is to be filled out by entry-level radio car patrol officers, requires approximately 15 minutes to complete.

The number of officers from your department that we request complete this questionnaire can be determined by referring to the following chart. Note that we are using the number of patrol officers and not the total number of officers in your department to determine sample size.

Police Departments

Number of Radio Patrol Officers	Number of Radio Patrol Officers to Include in Sample
1-10	5
11-25	7
26-50	10
51-150	15
150+	20

Sheriffs' Departments

Number of Radio Patrol Deputies	Number of Radio Patrol Deputies to Include in Sample
1-40	10
41-125	15
126+	20

The specific individuals who complete the questionnaire are to be selected by you based on guidelines POST will supply if you desire to participate.

The Sample Reading Materials, we suggest, should be gathered by one individual. The time required should be no more than approximately two hours.

If you decide to participate in this project, we request that you designate one individual in your agency as coordinator. This individual would, for the purposes of this project, be responsible for coordinating all matters relative to your department and also serve as liaison between your department and the POST Research and Evaluation Bureau. Please do not attempt to complete and return the enclosed questionnaire, it is for your inspection only.

Please return the enclosed form which indicates whether you will participate or not. If you have any questions, please call me at (916) 322-3492.

Thank you.

We would like to begin by thanking you for your participation in the POST Entry-Level Law Enforcement Reading Project. There will be approximately 50 agencies statewide participating in this study. The final product of this project will be a series of validated reading tests which will be available to police and sheriffs departments throughout the state.

The selection process for the entry-level patrol officer/deputy is a complex one, comprised of numerous steps. Typically, these steps include a written examination, an oral, a physical performance test, a background investigation, a medical evaluation, and a psychological evaluation. The POST Research and Evaluation Bureau will conduct research into as many of these selection components as is possible. Our specific objective in this project, however, is to develop job related entry-level reading tests.

The notion of validity is central to this project. Validity is a technical term that refers to the extent to which a test measures what it is intended to measure. In the case of selection tests, validity is an assessment of the extent to which a test is predictive of success on the job. In short, it is the job relatedness of the examination. Preliminary to the development of any tests, however, one must first complete an extensive job analysis.

Much of the job analysis for the entry-level patrol officer/deputy job has already been completed. Your department has participated in that process.

The information requested at this time will supplement that collected earlier. Together, the information from these two sources will comprise the foundation on which the reading exam will rest. It is, therefore, crucially important that all information collected be accurate and complete.

Enclosed you will find the Source of Information Questionnaires and The Request for Samples of Reading Materials.

The Questionnaire is to be completed by officers/deputies assigned to radio car patrol. The departmental coordinator is to identify those persons who will complete questionnaires. The following guidelines should be utilized when selecting officers/deputies who will complete questionnaires.

- o include officers/deputies from all shifts.
- o include officers/deputies with at least one year of experience in patrol.
- o whenever possible, include women and minorities.

The following guide identifies the appropriate number of individuals to include. Note that we are using the number of patrol officers and not the total number of officers in your department to determine sample size.

Police Departments

Number of Radio Patrol Officers	Number of Radio Patrol Officers to Include in Sample
1-10	5
11-25	7
26-50	10
51-150	15
150+	20

Sheriffs' Departments

Number of Radio Patrol Deputies	Number of Radio Patrol Deputies to Include in Sample
1-40	10
41-125	15
126+	20

Once the specific officers/deputies who will complete the questionnaire have been identified, copies of the questionnaire should be distributed to them. All necessary instructions for completion of the questionnaire are presented on the questionnaire itself thus eliminating the need to bring the officers/deputies together at one place and time. Also, since the questionnaire requires only 10-15 minutes to complete, it should not be burdensome on the officers'/deputies' time.

The officers/deputies should be instructed to return the questionnaires to the departmental coordinator after they have been completed. The coordinator should then briefly review them to see that they have been properly completed.

The Sample of Reading Materials should be gathered by the departmental coordinator or someone of his/her choice. Again, complete directions are on the form. Allow about two hours to collect and label the materials.

These materials, along with all completed questionnaires, should be returned to POST by _____.

We realize that completing these documents may represent an inconvenience for you and your department. We hope, however, to compensate you by developing selection procedures for your use which identify only the best potential officers/deputies.

Thank you for you assistance.

APPENDIX F

SAMPLE READING MATERIALS INSTRUCTIONS

SAMPLE READING MATERIALS

In order to determine the level of reading difficulty of materials that officers encounter on the job, POST needs samples of those materials to subject to readability analysis. Please provide us with the materials listed below. When preparing the information for sending to POST, please take note of the following:

- The materials are requested by type of material (e.g., in-depth narrative reports, interoffice memorandums, department manuals, etc.). Label and clip together the examples of each type of material. This is extremely important as it ensures that all documents will be correctly classified for the purposes of analysis.

- Include only examples of materials that officers actually read. If the rating officers indicate on the questionnaire that they read primarily summaries of a particular category of material, send copies of the summaries - not of the material summarized. If they indicate that they read material in its original form - send those originals. If the officers indicate that a particular material does not pertain to their job, send nothing.

- When selecting materials, choose from the most recent.

- If you want any of the materials to be returned, make a note on those materials so specifying.

Type of material	Number of Examples
Narrative reports containing complete sentences and paragraphs (e.g., investigative reports, supplemental/ follow-up reports).	
Reports consisting of several short descriptive phrases, sentence fragments, or very short sentences (e.g., incident reports).	
Reports consisting primarily of check-off boxes or fill-in blanks (e.g., vehicle impound reports).	
Street maps.	
Incoming correspondence.	
Interoffice memorandums.	
Wanted vehicle bulletins.	
Departmental manuals.	
Case law.	
Legal interpretations (e.g., California Attorney General's Opinons, City Attorney Opinions).	
Teletype messages.	
Training bulletins.	
Warrants.	
Coded material (e.g., NCIC printout, DMV drivers' records).	

APPENDIX G

COMPUTER GENERATION OF READABILITY INDICES

While certainly a convenience, computer generation of readability indices is not without its problems. Primary among these is the difficulty a computer program has in actually "reading" a written passage. Readability equations require that the number of syllables, words, and sentences be counted and that these values then be entered into an equation. To accomplish this counting, the program must scan the written passage and, based on a system of semantic and syntactic rules, generate the necessary values. If our language was entirely logical and without exception or idiom, this would present no problem at all. English, however, is replete with exceptions. Thus, short of creating a program which contains all possible English words, and then matching each word in the readability passage against that file, any computerized analysis will contain a degree of error.

The program, for example, creates syllables by counting the number of vowel sets separated by consonants. Thus, the word "com pu ter" is read as a three syllable word and "pro gram" a two syllable word. "Actual", however, is not read as "act ual" but rather as "ac tu al". It is not sufficient simply to program contiguous vowels as representing two syllables for then the word "peo ple" becomes "pe op le." The point, simply, is that short of creating the program which contains all possible English words, any computerized analysis will contain errors. The important question then becomes, what is the extent of that error.

To determine the exact error rate of our program, staff selected a ten percent random sample from each reading category and conducted a manual count of syllables, words, and sentences. Staff then compared these with the machine generated count. The identified error rate was so small as to be totally irrelevant. Individual errors tended to be small, and more importantly, randomly distributed around the 0 value. Thus, if in one passage the computer read two syllables where there should be only one, in another it would read one where there should be two. There was thus, a self-canceling effect for each reading category. The specific errors, their mean, standard deviation, and distribution are presented in the Table 18.

Table 18: Frequency of errors resulting from computer analysis of readability.

	Training Bulletins		Memos		Reports		Legal		Manuals		Codes	
	Syllables	Poly Syll.	Syllables	Poly Syll.	Syllables	Poly Syll.	Syllables	Poly Syll.	Syllables	Poly Syll.	Syllables	Poly Syll.
*7												
*6												
*5												
*4												1
*3	1		1		2						2	3
*2	3	3	1	2		2				1	3	3
*1	1	4		4	4	2	2		3		2	6
0	3	7	1	1		7	4		3	7	9	6
-1	1		1	4	2	6	2		1	1	3	2
-2	5	1	3	2	3	2	4	2		1	3	4
-3		1	6		2	1	1	1	4	3	1	
-4	1	1	1			1		1				
-5	1			1			3	1	2		2	
-6	1	1					1					
-7					1							
Mean	-.83	-.28	-2.1	-.6	-.5	.67	-1.9	-1.2	-1.7	-.77	-.27	.81

*Note. Number counted in excess of those that actually existed.

-Note. Number missed that should have been counted.

APPENDIX H

COMPARISON OF VALIDITIES GENERATED USING EXACT WORD SCORING (EWS) AND SYNONYM SCORING (SS)

Table 19: Comparison of validities generated using Exact Word Scoring (EWS) and Synonym Scoring (SS).

Predictors	Criteria			
	Proficiency Scores		Internal Scores	
	(EWS)	(SS)	(EWS)	(SS)
Total Group				
Reading 10	.46 (N=334)**	.45 (N=334)**	.46 (N=333)**	.43 (N=333)**
Reading 12	.46 (N=320)**	.45 (N=320)**	.47 (N=320)**	.43 (N=320)**
Total Reading[a]	.51 (N=320)**	.51 (N=320)**	.49 (N=320)**	.46 (N=320)**
Black				
Reading 10	.41 (N=35)**	.36 (N=35)**	.32 (N=36)	.28 (N=36)
Reading 12	.49 (N=33)**	.45 (N=33)**	.26 (N=34)	.34 (N=34)*
Total Reading	.49 (N=33)**	.41 (N=33)**	.22 (N=34)	.25 (N=34)
White				
Reading 10	.40 (N=222)**	.43 (N=222)**	.45 (N=221)**	.42 (N=221)**
Reading 12	.39 (N=215)**	.39 (N=215)**	.44 (N=215)**	.37 (N=215)**
Total Reading	.45 (N=215)**	.47 (N=215)**	.47 (N=215)**	.43 (N=215)**
Asian				
Reading 10	.45 (N=29)**	.44 (N=29)**	.32 (N=26)	.33 (N=26)*
Reading 12	.57 (N=27)**	.50 (N=27)**	.43 (N=24)*	.33 (N=24)
Total Reading	.55 (N=27)**	.55 (N=27)**	.41 (N=24)*	.39 (N=24)*
Hispanic				
Reading 10	.53 (N=48)**	.47 (N=48)**	.37 (N=50)**	.41 (N=50)**
Reading 12	.46 (N=45)**	.50 (N=45)**	.56 (N=47)**	.55 (N=47)**
Total Reading	.62 (N=45)**	.57 (N=45)**	.51 (N=47)**	.54 (N=47)**

[a]Total Reading = Reading Level 8 + Reading Level 10 + Reading Level 12.
* Significant at the .05 level
** Significant at the .01 level

APPENDIX I

COMPARISON OF MEAN SCORES GENERATED USING EXACT WORD SCORING (EWS) AND SYNONYM SCORING (SS)

Table 20: Comparison on mean scores generated using Exact Word Scoring (EWS) and Synonym Scoring (SS).

Group	Reading 10				Reading 12			
	N	EWS	SS	Diff	N	EWS	SS	Diff
Total	343	61.3	74.8	13.5	329	47.1	66.8	19.7
Black	36	57.0	70.4	13.4	34	44.2	63.2	19.0
White	228	63.0	76.5	13.5	221	48.8	69.1	20.3
Spanish	50	57.6	70.7	13.1	47	43.0	62.3	19.3

APPENDIX J

COMPARISON OF RELATIVE PASSING RATES OF TOTAL GROUP AND ETHNIC GROUPS USING EXACT WORD SCORING (EWS) AND SYNONYM SCORING (SS)

TABLE 21: Comparison of relative passing rates of total group and ethnic groups using Exact Word Scoring (EWS) and Synonym Scoring (SS).

		% of Total Group Passing	% of Blacks Passing	% of Whites Passing	% of Asians Passing	% of Hispanics Passing	80 % of Whites Pass Rate
	EWS	90.0	76.9	93.7	85.3	85.1	75.0
	SS	90.0	83.2	94.2	86.1	77.9	75.4
	EWS	85.0	67.2	90.7	80.8	75.4	72.6
10th Reader Level Test	SS	85.0	68.4	90.3	84.9	72.9	72.2
	EWS	80.0	59.9	86.6	81.5	63.8	69.3
	SS	80.0	60.3	86.2	78.6	66.2	69.0
	EWS	75.0	51.4	82.6	67.4	61.7	66.1
	SS	75.0	56.6	80.8	73.6	62.1	64.6
	EWS	90.0	83.9	95.4	75.4	78.0	76.3
	SS	90.0	91.2	93.7	74.1	80.9	75.0
	EWS	85.0	79.3	89.9	66.6	76.5	71.9
12th Reader Level Test	SS	85.0	84.3	89.9	65.9	73.6	71.9
	EWS	80.0	72.0	87.2	60.5	63.0	69.8
	SS	80.0	81.5	85.0	62.3	65.3	68.0
	EWS	75.0	66.8	81.8	57.4	59.3	65.4
	SS	75.0	72.2	81.4	53.1	58.7	65.1

APPENDIX K

ANALYSIS OF PASS/FAIL REVERSALS RESULTING FROM DIFFERENT SCORING METHODS (EXACT WORD REPLACEMENT AND SYNONYM SCORING) AND ANALYSIS OF MEAN SCORE IMPROVEMENT RESULTING FROM THE ACCEPTANCE OF SYNONYMS IN SCORING

Table 22: Number of pass/fail reversals resulting from different scoring methods (exact word replacement and synonyn scoring).

Scoring method	N
Failed using both methods	47
Failed exact word scoring but not synonym scoring	11
Failed synonym but not exact word scoring	5
Total failed using exact word scoring	58
Total failed using synonym scoring	52

Table 23: Mean score gain resulting from the acceptance of synonyms in scoring. (Mean gain was computed for each quartile in the range of scores)

Group	10th level test				Overall mean gain	12th level test				Overall mean gain
	Q_1	Q_2	Q_3	Q_4		Q_1	Q_2	Q_3	Q_4	
Total	6.2	6.4	6.5	6.1	6.3	8.1	9.3	9.4	9.1	9.0
Black	5.9	6.0	7.2	6.1	6.3	9.1	8.9	7.8	8.9	8.7
White	6.4	6.4	6.5	5.8	6.3	8.1	9.5	9.6	8.6	9.0
Asian	5.8	6.6	7.0	6.3	6.4	6.6	6.7	10.3	8.9	8.1
Hispanic	5.9	5.8	6.5	5.9	6.0	8.0	9.2	9.5	8.8	8.9

It is important to note that while there is no difference in gain scores within tests, there is a difference in gain scores between tests. The overall mean gain on the 10th level test is 6.3, while the mean gain on the 12th level is 9. It appears that mean gain is a function of available synonyms, with the greater number of synonyms the greater the mean gain. There is a total of 45 items on the 10th level test and 44 on the 12th. In addition to the 45 exact words, there are 25 additional synonyms making a total of 70 possible words for the 10th level test. On the 12th level test there are 44 exact words and 54 additional synonyms for a total of 98 possible words. Thus, the difference between tests seems due to the greater number of possible words in one test as opposed to the other.

APPENDIX L

SURVEY OF WRITING MECHANICS

INSTRUCTIONS

Effective writing requires the knowledge and correct use of "writing mechanics" - rules and techniques which determine the content and structure of written material. This questionnaire is designed to identify the specific writing mechanics which are essential to effective writing by patrol officers. The information will be used in the design of a writing ability test which will be used to evaluate peace officer applicants.

On the following pages are listed examples of a number of aspects of writing mechanics such as grammatical rules and rules of punctuation. Regarding each aspect, an example is provided in the form of a word or sentence which is written incorrectly (i.e., an example of an error), followed by one which is written correctly.

We would like you to evaluate each aspect of writing mechanics listed on subsequent pages by means of the following process. Consider the first topic area on page 3. Read the correct and incorrect examples. Next, decide which of the following three categories contains the most accurate statement regarding the type of writing error in the incorrect example:

Category 1. Acceptable in a report. Put an "x" on the line corresponding to this label if you either typically accept the exemplified type of error in an officer's reports or if someone other than the officer routinely corrects the error (e.g., a secretary or typist makes the correction). This category refers to errors which do not require any editing or rewriting on the part of the officer.

Category 2. Not acceptable, but person can be remediated. Put an "x" on the line corresponding to this label if you feel that the person who tends to make the type of error which is exemplified can be remediated during academy training, field training and/or inservice training. This category refers to errors which would require editing or rewriting on the part of the officer.

Category 3. Not acceptable and person should not be hired. Put an "x" on the line corresponding to this label if you feel that a person who tends to make this type of error should not be hired. Choosing this category means that you feel that your agency and/or academy training program cannot devote time to teaching recruits to avoid this type of error.

Topic: Conciseness.

Incorrect: The victim obviously suffering from the severe trauma and distress of the draining experience of the attack heretofore could not adequately describe the specific physical characteristics as the alleged perpetrator.

Correct: The victim was upset and could not give a complete description of the suspect.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Vague antecedent.

Incorrect: This took place near dawn at approximately 5:00 a.m.

Correct: This robbery took place near dawn at approximately 5:00 a.m.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Fragment.

Incorrect: Officers responding to the call.

Correct: Officers responding to the call found no one.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Parallel construction.

Incorrect: When they heard the call, the officers acknowledged the dispatcher, drove to the proper location, witnesses were interviewed, and took a report.

Correct: When they heard the call, the officers acknowledged the dispatcher, drove to the proper location, interviewed the witnesses, and took a report.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Vague wording.

Incorrect: The youths hassled the shopkeeper.

Correct: The youths demanded free food from the shopkeeper.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Erudite wording.

Incorrect: The bar patrons were creating a tumultuous din.

Correct: The bar patrons were noisy and disorderly.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Incomplete expression.

Incorrect: The officer distrusted the witness as much as the judge.

Correct: The officer distrusted the witness as much as the judge did.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Run-on sentence.

Incorrect: Officers responded to a barking dog complaint neighbors had called the police earlier.

Correct: Officers responded to a barking dog complaint. Neighbors had called the police earlier. (or, could use a semicolon)

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Sentence structure.

Incorrect: Mr. Fenton recognized the woman who had robbed him from her picture in the paper.

Correct: Mr. Fenton recognized from her picture in the paper the woman who had robbed him.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Poor word order.

Incorrect: The reported accident earlier has been cleared.

Correct: The accident reported earlier has been cleared.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Adverbs (modifying verbs).

Incorrect: The suspect was considerable taller than the victim described.

Correct: The suspect was considerably taller than the victim described.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Adverb (modifying verb).

Incorrect: He slid the wallet out slow.

Correct: He slid the wallet out slowly.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Subject verb agreement.

Incorrect: Neither of the boys want to testify.

Correct: Neither of the boys wants to testify.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Subject verb agreement.

Incorrect: The victims was bound and gagged.

Correct: The victim was bound and gagged.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Double Negative.

Incorrect: The captain did not tell us nothing.

Correct: The captain told us nothing. or, The captain did not tell us anything.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Choice of the right verb.

Incorrect: The witness said he picked up the gun and sat it on the table.

Correct: The witness said he picked up the gun and set it on the table.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Proper verb tense.

Incorrect: The witness has testified yesterday.

Correct: The witness testified yesterday.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Irregular verb.

Incorrect: The citizen brung in the missing bicycle.

Correct: The citizen brought in the missing bicycle.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Unclear reference of pronouns.

Incorrect: The officer was highly critical of the witness because he was a conservative.

Correct: The officer was highly critical of the witness because the witness was a conservative.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Objective pronoun.

Incorrect: Them are the best fingerprints that we could find.

Correct: Those are the best fingerprints that we could find.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Pronoun whose antecedent is singular indefinite pronoun.

Incorrect: Neither of the boys would admit that they was the offender.

Correct: Neither of the boys would admit that he was the offender.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Nominative/Objective case of pronouns.

Incorrect: Officer Brown and me responded as backup.

Correct: Officer Brown and I responded as backup.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Plural, singular pronouns.

Incorrect: When everybody had arrived, the officer told him the good news.

Correct: When everybody had arrived, the officer told them the good news.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Misplaced modifier.

Incorrect: The victim stated that a ring was taken by the burglar with three diamonds.

Correct: The victim stated that a ring with three diamonds was taken by the burglar.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Use of comma for clarity.

Incorrect: After that training was begun promptly at 5:30 a.m.

Correct: After that, training was begun promptly at 5:30 a.m.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Use of comma in figures.

Incorrect: The department recovered $35,00 in cash.

Correct: The department recovered $3,500 in cash.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Commas between series.

Incorrect: The thief stole silver goblets and crystal.

Correct: The thief stole silver, goblets, and crystal.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Punctuation (commas).

Incorrect: The accident happened on Monday October 26, 1979.

Correct: The accident happened on Monday, October 26, 1979.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Proper use of Quotation Marks.

Incorrect: The suspect confessed I did it while under custody.

Correct: The suspect confessed, "I did it," while under custody.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Capitalization.

Incorrect: Mr. Wilkins told the officers his chevrolet had been stolen.

Correct: Mr. Wilkins told the officers his Chevrolet had been stolen.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Capitalization.

Incorrect: The search covered an area from franklin street to walnut avenue and included the banks of the sacramento river.

Correct: The search covered an area from Franklin Street to Walnut Avenue and included the banks of the Sacramento River.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Punctuation.

Incorrect: Testimony of FBI agents sent the counterfeiters to prison.

Correct: Testimony of F.B.I. agents sent the counterfeiters to prison.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Punctuation.

Incorrect: The neighbor said, "I do not believe he is guilty, no one who knows him could believe it."

Correct: The neighbor said, "I do not believe he is guilty; no one who knows him could believe it."

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Conjunctions.

Incorrect: The victim said he kept his revolver on the table and in the desk drawer.

Correct: The victim said he kept his revolver on the table or in the desk drawer.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Prepositions.

Incorrect: The victim can't remember of having heard any noise.

Correct: The victim can't remember having heard any noise.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Unnecessary Prepositions.

Incorrect: A witness saw the two boys divide up the money.

Correct: A witness saw the two boys divide the money.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Possessives (singular).

Incorrect: The suspects name is Coleman.

Correct: The suspect's name is Coleman.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Possessives (plural).

Incorrect: The thieve's fingerprints were found on the door.

Correct: The thieves' fingerprints were found on the door.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Possessives (noun).

Incorrect: The garage's door was unlocked.

Correct: The door of the garage was unlocked.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Spelling.

Incorrect: The feirce dog attacked the child at approximately 11:30 a.m.

Correct: The fierce dog attacked the child at approximately 11:30 a.m.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic: Spelling.

Incorrect: The officers patrolled the dessert.

Correct: The officers patrolled the desert.

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

Topic:

Incorrect:

Correct:

____ Acceptable in a report; would not be corrected.
____ Not acceptable, but person can be remediated.
____ Not acceptable and person should not be hired.

APPENDIX M

SCORING CRITERIA FOR THE OPEN-ENDED ESSAY TEST

SCORING THE POST ENTRY-LEVEL LAW ENFORCEMENT TEST OF WRITTEN EXPRESSION

The POST Test of Written Expression is designed to measure basic factors which affect the clarity of written expression. Scoring the test requires the following simple steps:

First, the end of each sentence should be marked with a slash mark (/), and the total number of sentences should be counted.

Second, a minus sign (-) should be placed at the end of each sentence containing one or more of the following errors which are called "basic errors":

> (a) a misspelled word;* (b) a word not used properly;** (c) an unclear pronoun reference, or a misplaced modifying word or phrase;(d) an improper sentence (e.g., sentence fragment); or (e) wording which is difficult to understand or illegible.

Third, the number of sentences which were not assigned a minus should be counted. This includes sentences containing no errors and sentences containing writing mistakes other than the basic errors listed above (unless such errors make the writing unclear).

Fourth, the total number of sentences in the essay should be divided into the number which resulted from the previous step, and then multiplied by 100. The percentage which results from this step is a person's score on the test.

This scoring procedure is based upon the assumption that an appropriate characteristic to assess in law enforcement applicants is the basic ability to clearly express ideas in writing (as opposed to knowledge of the more esoteric rules of written expression).

---

* Count as basic errors only the first misspelling or misusage of words which are misspelled or misused more than once.

** The meaning of the word does not fit its use in the sentence (e.g., using the word "extant" when the correct word is "extinct").

SPECIFIC SCORING INSTRUCTIONS

The following scoring instructions represent a more detailed explanation of the four steps described on the previous page.

Step 1. Mark the end of each sentence with a slash mark (/) and count the sentences. Record the total number of sentences at the top of the first page of the essay.

As you read the essay for the first time, you should make a slash mark at the end of each sentence.

Consider two independent clauses separated by a semicolon as two separate sentences; however, consider as one sentence two independent clauses separated by a conjunction. An independent clause is defined the same as a sentence (i.e., a group of words complete in themselves, containing a subject and a predicate, expressed or understood). For a full list of definitions mentioned in these instructions, refer to the glossary.

On occasion, a test taker will indicate the intended end of a complete thought by incorrect punctuation (e.g., a colon), or alternatives to punctuation (e.g., starting a new line for each sentence, leaving a space between sentences, capitalizing the first word in a new sentence without the preceding period). If you, the scorer, judge that the test taker intended to complete a sentence at a particular point in the essay, put a slash mark at that point and count it as a complete sentence. If, however, there is a point in the essay which you feel should be the end of a sentence but the test taker provides no such indication, do not put a slash mark at that point.

Step 2. Place a minus sign (-) at the end of each sentence containing one or more basic errors.

Go back and review each sentence. Place a minus sign (-) next to the slash mark at the end of a sentence if the sentence contains one of the following basic errors:

- A misspelled word.
- A word not used properly.
- An unclear pronoun reference, or misplaced modifying word or phrase.
- The group of words designated as a sentence by the test taker is not a proper sentence; instead, it is an incomplete sentence or a run-on sentence.
- The meaning of the sentence is unclear; it is difficult to understand what the writer intended. This could be due to omitted words, illegible handwriting, awkward or ungrammatical construction, or other factors. However, it

is the lack of clarity which is the basic error and not the writing mistake. If a writing mistake such as an omitted word or ungrammatical construction does not obscure the meaning of a sentence, a basic error has not been committed and the sentence should not be assigned a minus.

Indicate, as near to the point of the error as possible, the type of error made by the test taker. For example, write "misspelled" or "incorrect word" directly above the word in question. Also circle each misspelled and misused word.

Place a check mark (✓) at the end of each sentence which contains no mistakes or which contains writing mistakes other than the five types of basic errors previously listed (unless such mistakes make the writing unclear, in which case the sentence should be assigned a minus).

Step 3. Count the sentences which were assigned a check mark.

At the completion of the previous step, each sentence should have been assigned either a check mark or a minus. Count the number of sentences which were assigned a check mark (i.e., the sentences which contain no basic errors).

Step 4. Compute the percentage of sentences in a person's essay which were assigned a check mark.

Using the following formula, compute the percentage of sentences which were assigned a check mark (i.e., the percentage of sentences which contain no basic errors). Record the percentage on the top of the first page of the essay.

$$\text{Percentage of Sentences* Assigned a Check Mark} = \frac{\text{Number of Sentences Assigned a Check Mark}}{\text{Total Number of Sentences}} \times 100$$

Special Scoring Considerations

A) Assign a minus only to those basic writing errors listed above. Do not assign a minus for any other violations of the commonly held rules of correct writing (i.e., rules of grammar, punctuation, paragraph construction, etc.), unless the mistakes cause the meaning of the sentence to be unclear. For example, do not assign a minus to a sentence containing the following types of writing mistakes:

- o lack of conciseness**
- o lack of parallel construction
- o use of adjective when an adverb is proper
- o lack of subject-verb agreement

---

* Round percentage off to the nearest whole number.

** For definitions of these non-basic, writing mistakes, refer to the glossary.

- failure to use articles
- double negatives
- improper verb form, such as wrong tense (unless the meaning is obscured)
- incorrect use of objective or subjective pronoun
- use of incorrect punctuation, including apostrophe
- incorrect capitalization

B) If a word is misspelled or misused more than once, consider only the first misspelling or misuse of the word as an error. In the evaluation of succeeding sentences, do not score as basic errors such repeated spelling or word usage errors.

C) Consider a misspelled or misused word to be a basic error even if it is spelled or used correctly somewhere else in the essay.

D) If a common, easy-to-spell or easy-to-understand word is misspelled or misused (e.g., "to" instead of "too", or "there" instead of "their", or "an" instead of "and", etc.), score the word as misspelled or misused.

E) Some action words can take many different forms depending on number (singular versus plural), tense (e.g., past, present, future), and grammatical usage (e.g., verb, infinitive, participle). For example, consider some possible alternative forms of the word "throw": threw, thrown, throws, and throwing. If the test taker uses the wrong word form, but the meaning of the sentence is not obscured, do not assign a minus to the sentence.

F) A sentence is assigned a minus when the sentence contains a basic error. Additional basic errors in the same sentence should be identified and labeled, but they do not affect the scoring in any way.

Additional Scoring Aids

To assist the scorer, an example of a scored essay and a glossary of terms have been provided. The glossary is divided into sections containing: (a) definitions of general terms, (b) examples of several types of basic errors, and (c) examples of non-basic writing mistakes.

EXAMPLE

misspelled word
I had been (thing) about becoming a police officer for about Six
year./ I consider the good misspelled word (an) bad points about the job./ After
getting out of the service I decide not to try to become a police
misspelled word
(office) until I try a few other thing I had been considering also./
Well after about a year of trying other job I decide that I want,
for sure to become a police officer./ So I began watching and
waiting for the different department to start accepting people./
After about 8 month the field open up for employment./ I put in
misspelled word and run-on
three (appilication) the CHP was the first to respond with a test
misspelled word misspelled word
day./ i took all the test and (enter view) require to (quilify) for
misspelled word misspelled word
coming to the (Acedemy)./ Nine month later I made it to the (Acedemy)./
misspelled word misspelled word and sentence fragment
(Thoe) the work is hard (an) not what I expected./ I will do the work
because I want the job of being a cHP OFFICER./

5* / 11** 45%

---

* Number of sentences assigned a check mark.
** Total number of sentences.

GLOSSARY

General Terms

SUBJECT - The word or group of words which names the thing, person, place or idea about which a statement is made in the sentence.

Example: The helicopter landed.

In this example "helicopter" is the subject.

PREDICATE - That part of a sentence or clause that expresses something about the subject. It regularly consists of a verb and may include objects, modifiers, or complements of the verb.

Example: The application is being processed.

In this example "is being processed" is the predicate.

INDEPENDENT CLAUSE - A clause containing a subject, a verb, and sometimes an object and modifiers, capable of standing alone as a complete sentence.

Example: The back-up car arrived within a short time; the ambulance arrived a few moments later.

In this example, two independent clauses are separated by a semicolon. Each clause is capable of standing alone as a complete sentence.

ADJECTIVE - A word that modifies a noun. An adjective may tell the size, shape, color, or quality, or it may point out, or tell the number.

Example: A large, vacant lot adjoined the house.

In the example, "large" and "vacant" are adjectives modifying the noun "lot".

ADVERB - A word, phrase, or clause that modifies a verb, adjective, or other adverb. It answers such questions as where, how, how much, when.

Example: The alarm was periodically set off by the wind.

The adverb "periodically" modifies the verb "was set" and answers the question when.

ARTICLE - "a", "an", and "the" used as adjectives.

Example: A minute can seem like an hour at the dentist's office.

CONJUNCTION - A word used to connect words, phrases, or clauses.

Example: A date was chosen for the conference, and the announcements were mailed.

In this example, the conjunction "and" connects two independent clauses.

UNCLEAR PRONOUN REFERENCE - The antecedent of the pronoun (the word or group of words to which a pronoun refers or for which it stands) is vague.

Example: The icing should not be put on the cake until it is cold.

What "it" refers to, in the example, is unclear.

MISPLACED MODIFYING WORD OR PHRASE - A modifying word or phrase so placed that there is doubt as to which word it modifies.

Example: The witness identified the suspect with calm certainty.

It is unclear whether the witness or suspect possessed calm certainty.

INCOMPLETE SENTENCE (phrase or dependent clause) - A group of related words with or without a subject and a predicate that cannot stand alone as a complete sentence.

Example: Into the crowd. (phrase)

Example: Before the car could be found. (dependent clause)

Neither of these examples can stand alone as a complete sentence even though the second example contains both a subject and a predicate.

RUN-ON SENTENCE - Two independent clauses combined without using a comma and a conjunction, or a semicolon between them.

Example: The caller was persistent he telephoned every half hour.

This example is a basic error because there is no semicolon or comma and conjunction between "persistent" and "he".

Non-Basic Writing Mistakes

(Sentences with these types of errors are not assigned a minus)

LACK OF CONCISENESS - Lack of clarity and compactness. Use of unnecessary words.

Example: The suspects, of whom there were two in number seemed to decide that the most advantageous place of hiding was a tool shed under whose shelter they might escape the notice of their pursuers.

A more concise way to state this thought might be, "The two suspects hid in a tool shed."

PARALLEL CONSTRUCTION - When two or more elements in a sentence are joined by coordinating conjunctions, they should be expressed with the same grammatical construction.

Example: John's duties include meeting with community representatives, making recommendations and supervising three employees.

Example: John's duties include meeting with community representatives, making recommendations and to supervise three employees.

SUBJECT-VERB AGREEMENT - A subject and its verb must agree in number (singular or plural) and person (first person, second person, or third person).

Example: Richard arrives early.

Example: They arrives early.

In the first example both the subject and verb are singular. In the second example "they" is plural while the verb form is singular.

(Sentences with these types of errors are not assigned a minus)

DOUBLE NEGATIVES - The use of more than one negative (e.g., "no", "not", "n't", "none", "nothing", "nobody", "no one", "neither", "never", and "nowhere") in a sentence to express a negative idea.

Example: Nobody did nothing.

OBJECTIVE-SUBJECTIVE CASE PRONOUNS - A pronoun used as the subject or as part of a compound subject of a verb is in the subjective (or nominative) case. Some examples of subjective case pronouns include "I", "we", "he", "she", "who". A pronoun used as the direct or indirect object of a verb or as a part of a compound object or compound indirect object of a verb is in the objective case. A pronoun used as the object of a preposition or as part of a compound object of a preposition is also in the objective case. Some examples of objective case pronouns include "me", "us", "him", "her", "whom".

Example: He and I entered the building behind them.

Example: Him and me entered the building behind they.

In the second example, "him" and "me" are pronouns forming the compound subject of the sentence. However, subjects require pronouns in the subjective case and "him" and "me" are in the objective case. Therefore, the subjective case "he" and "I" in the first sentence are correct. Similarly, the subjective case "they" in the second sentence is incorrectly used as the object of the preposition which requires the objective case.

IMPROPER VERB FORM - Wrong tense, lack of agreement with subject in person or number, improper verb choice based on meaning.

Example: They will took him to the hospital. (wrong tense)

Example: She jog three miles a day. (lack of subject-verb agreement)

Example: He set down on the chair. (improper verb choice based on meaning)

APPENDIX N

THE SELECTION PROCESS

This appendix addresses the issue of the entry-level selection process and the components which comprise it. The entry-level job is dimensionally complex, thus requiring a selection procedure comprised of multiple components. The range of behavioral requirements for police work is much broader than in most other professions. In the POST job analysis, 29 specific job behaviors were identified. For the sake of brevity these 29 were collapsed into seven categories: cognitive ability, communication skill, special skills, interpersonal skills, personality characteristics, worker characteristics, and physical characteristics. Numerous testing procedures, of which the POST tests are but one, are required to accurately assess applicants across the spectrum of behaviors listed above. In order for the POST language ability tests to be properly utilized, it is important that the entire selection process be understood. To accomplish this one must be aware of all possible components of police testing, how these components are ordered, and how they are scored and weighted.

The actual test process is comprised of from three to eight individual test components. Three represents the minimum number of components as there are three legally mandated selection procedures or requirements (a background investigation, a medical examination, and a minimum education requirement of high school graduation or its equivalent). For those departments in the POST program there is a minimum of four procedures; for in addition to the three legally mandated requirements, there is one additional POST requirement, the interview. Beyond these four, there are four additional steps that may be incorporated into the selection process: a reading test, a writing test, other written tests (i.e., analytical skills, memory, etc.) and a psychological test.

Assuming that each test component is reliable and valid, it is only reasonable to conclude that the accuracy of prediction of the process will be a direct function of the inclusiveness of its scope. Thus, the more test components utilized, the greater the subsequent precision.

How the components are ordered, scored, and weighted, also impacts on the accuracy of the final ranking of candidates. Some components are intrinsically pass/fail. These include meeting the high school requirement, the background investigation, the medical examination, and the psychological test. The physical performance test is also usually scored pass/fail although it may be used as a ranking device. The reading, writing, other written tests, and the interview may be used either pass/fail or as a ranking procedure, depending upon the method of validation. POST makes no recommendation regarding the exact ordering of components. POST does stress, however, that components be used consistent with the results of the job analysis and validation procedures, and that they not be used solely or primarily as screen-out mechanisms.

A more complicated issue is that of test weighting. Once a decision to rank has been made, the question then becomes how to combine the various component scores into a final ranking based on some predetermined test weighting plan. Raw scores obviously cannot be used because differences in test length and test variances will hopelessly confound the ultimate ranking. Percent and percentile scores are an improvement, but they are ordinal values, thus leaving the distance between scores undefined. This being the case, to combine and weight scores creates a distortion in rank which cannot be accurately determined. Before accurate component weighting can be accomplished, test scores must be converted to scales with comparable means and standard deviations and with equal distances between scale points. It is only when these conditions are met that the achieved weighting of components is also the intended weighting. This can be accomplished by standardizing test scores. As with test ordering, POST makes no specific recommendation regarding the weighting of components. Again, however, it is stressed that components be used in a manner consistent with job analysis and the validation procedures.

As was stated initially, the tests of language ability are but one component among many. Regarding these tests specifically, however, POST does have recommendations. There are two tests: one reading and one writing. Because POST does not feel that these skills are compensatory, POST requires a separate passpoint on each. The scores on the three writing sub-tests are added together to form a total writing score and the scores on the two reading tests are combined to form a total reading score. A passpoint is set for each total score. The two sections (reading and writing) are weighted 50-50 to generate a final score for language competency. This is accomplished by converting both total section scores to scores with a mean of 50 and standard deviation of 10. These two standard scores are then added together to form the final score of the test. Regarding ranking, POST recommends agencies use the tests as a ranking device. Agencies may use the tests pass/fail if they wish, but this results in a considerable loss of useful information regarding the candidates because of the extremely high validity of the tests.