
3A

Action

Professional Services Committee

Passing Standard Review for the California Subject Matter Examinations for Teachers (CSET): American Sign Language Examination

Executive Summary: This report provides the Commission with recommendations relevant to the passing standard for the CSET: American Sign Language Examination.

Recommended Action: That the Commission adopt the recommendations of the 2009 Passing Standard Review Advisory Panel concerning the CSET: American Sign Language Examination.

Presenter: Phyllis Jacobson, Ed.D.,
Administrator, Professional Services Division

Strategic Plan Goal: 1

Promote educational excellence through the preparation and certification of professional educators

- ◆ Sustain high quality standards for the preparation and performance of professional educators and for the accreditation of credential programs

Passing Standard Review for the California Subject Matter Examinations for Teachers (CSET): American Sign Language Examination

Introduction

This report describes the 2009 standard setting study review for the California Subject Examinations for Teachers (CSET): American Sign Language (ASL) Examination, and provides advisory panel recommendations concerning the passing standard for this examination.

The first administration of the CSET: ASL examination in November 2005 included 15 examinees, and of these 20% passed the examination. The original 2005 agenda item presenting the passing standard adopted by the Commission indicated that “a subsequent passing standard activity will be held to review the passing standards in light of the increased number of examinees. . . . Following further review, recommendations for any change in the standards will be presented to the Commission for consideration and adoption.”

Since 2005, 84 examinees have taken the CSET: ASL Examination, with a cumulative passing rate of 27.4% for the period through November 2008. Because the passing rate remains low, and because additional data are available from the increased number of examinees since the initial 2005 administration, a subsequent standard setting review was conducted in January 2009 by a new standard setting advisory panel. Given the agenda deadline for the January 2009 Commission meeting relative to the panel’s January 15-16, 2009 meeting date, the panel’s recommendations will be provided as an infolder item.

Background

The development of the CSET: ASL Examination began in 2004. That spring, the Commission’s Executive Director appointed a subject matter advisory panel for the CSET: ASL to advise Commission staff on the development of new subject matter program standards and an examination in this subject area. A Single Subject Teaching Credential for ASL was not previously available. National Evaluation Systems, Inc., the Commission’s CSET testing contractor which later became NCS Pearson, and Commission staff worked with the panel to facilitate this work during 2004-05. The original subject matter advisory panel consisted of a diverse group of college and university faculty, classroom teachers, ASL interpreters, and members of relevant professional organizations and committees, all of whom had expertise in ASL.

Test structures were developed by the subject matter advisory panel, and the multiple-choice and constructed-response items were drafted, reviewed, and revised as needed by both the Bias Review Committee and the subject matter advisory panel. Once these items were field tested, a panel consisting of some members of the subject matter advisory panel and some new individuals with appropriate backgrounds in ASL selected marker responses (i.e., candidate responses that exemplify the range of possible score points) and scored the constructed-

responses from the field test. Additionally, a test guide including the subject matter requirements, test structures, and sample questions was developed to assist candidates in preparing to take the CSET: ASL.

On November 5, 2005, the first test administration of the CSET: ASL Examination was conducted. On December 6-7, 2005, the original standard setting study for this examination was held in Sacramento to determine the initial passing standard recommendations of California educators. These recommendations were subsequently adopted by the Commission at the January-February 2006 meeting.

Description of the CSET: American Sign Language Examination

The CSET: ASL Examination is comprised of subtests differentiated by content area. The subtests consist of both multiple-choice and constructed-response items. Constructed-response items are of two types: *extended* constructed-response items that are scored using a four-point scale, and *focused* constructed-response items that are scored using a three-point scale. For subtest I, candidates may respond in ASL or written English. For subtest II, candidates must respond in written English. For subtest III, candidates must respond in ASL. Constructed-response performance characteristics and scoring scales are provided in Appendix A. The test structure for the CSET: ASL is shown in Table 1 below.

Table 1: Test Structure of the CSET: American Sign Language Examination

Subtest	Domains	Number of Multiple-Choice Items (MC)	Number of Constructed-Response Items (CR)
I	Literary and Cultural Texts and Traditions	10	1 focused
	Cultural Analysis and Comparisons	20	1 extended
	<i>Subtest total</i>	30	1 focused 1 extended
II	General Linguistics	10	none
	Linguistics of the Target Language – American Sign Language (Language Structures; Contrastive Analysis; Sociolinguistics and Pragmatics)	20	2 focused
	<i>Subtest total</i>	30	2 focused

III	Linguistics of the Target Language – American Sign Language (Error Analysis)	none	1 focused
	Language and Communication: Receptive Comprehension	18	1 focused
	Language and Communication: Expressive Production	none	2 focused
	<i>Subtest total</i>	<i>18</i>	<i>4 focused</i>
	<i>Total Items</i>	<i>78</i>	<i>7 focused 1 extended</i>

Each CSET testing session is five hours in length. Examinees can choose to take any one or all subtests within a single testing session. Individual subtests are not timed. The CSET: ASL is administered two times each year. For the first administration of the CSET: ASL in November 2005, fifteen examinees completed all three subtests.

Procedures for the Original Standard Setting Study (2005)

The original standard setting study for the CSET: ASL was conducted on December 6-7, 2005, with a Commission-appointed standard setting panel of ASL specialists. The purpose of the standard setting procedure is to provide the Commission with recommendations, based on the informed judgments of California educators and subject specialists, relevant to the determination of the passing standard for this examination.

The standard setting study began with an orientation to the program and the examination and a training session on the review task. Panel members were provided with the subject matter requirements and the subtest forms used for the most recent test administration. To help the panel members become familiar with the examination, the knowledge and skills associated with the items, and the perspective of the examinee, panel members were asked to take the test under simulated test-like conditions. They were asked to read and answer each item independently, without reference to the answer key, and then to score their own performance on the multiple-choice items.

Panel members were then asked to consider the “just acceptable” candidate. Although many of the examinees will exceed the level of knowledge and skills of the acceptably qualified candidate, none should fall below that level. For this reason, panel members were trained to make judgments based on candidates just at the level of knowledge and skills required of an entry-level teacher candidate for ASL to successfully satisfy the subject matter requirement.

After extensive training and the simulated test taking, panel members were asked to complete three rounds of standard setting tasks based on the test structure. This process is briefly described below. A detailed description of the process is found in Appendix B.

In Round One, panel members were asked to individually rate each item on each subtest. They were asked to rate the percent of correct responses that would be expected from a group of

“just acceptable” candidates for each multiple-choice item and the level of response that would be achieved by the “just acceptable” candidate for each constructed-response item.

Using the item statistics produced from Round One to inform judgments, Round Two moved the panel from individual item ratings to ratings at the component level (i.e., multiple-choice component and constructed-response component). They were asked the number of multiple-choice items that would be answered correctly and the total score points that would be achieved on the constructed-response items. Panel members were also asked to consider the “component score combination rule,” or the percentage of points that should be allocated to each component (e.g., 80% multiple-choice and 20% constructed-response, 70% multiple-choice and 30% constructed-response).

In the final round of ratings, the panel members were asked to make independent recommendations for a passing standard for each component and “component score combination rule.”

Following the standard setting studies, Pearson staff calculated for each subtest the median and the distribution of individual Round Three panel recommendations for the multiple-choice and constructed-response test components. Panel recommendations on component score combination rules were also tabulated.

Commission-Adopted Initial Passing Standards, 2005

As shown in Table 2 below, the Commission adopted the following initial passing standards for the subtests of the CSET forms administered on November 5, 2005.

The staff-recommended raw score points for multiple-choice and constructed-response components reflected adjustments made for standard errors of measurement as appropriate. A candidate’s passing status is determined on the basis of total subtest performance. Test results are reported as scaled scores. A scaled score is based on the number of raw score points earned on each component (i.e., multiple-choice and/or constructed-response) and the weighting of each component. For the CSET, raw scores are converted to a scale from 100 to 300, with a score of 220 representing the passing score as set by the Commission. Scaled scores are used to help ensure that the level of competence required to pass a given test is independent of the particular form of the test taken.

Table 2: Commission-Adopted Initial Passing Standards for CSET: American Sign Language (2005)

Subtest	Multiple-Choice Raw Score Points	Constructed-Response Raw Score Points	Component Score Combination Rule MC/CR	Passing Rate Nov. 2005 Administration by Subtest	Overall Passing Rate Nov. 2005 Administration ³
I ¹	17	10	70/30	47%	20%
II ¹	16	9	70/30	63%	
III ²	10	19	40/60	38%	

¹ Subtest I and II adjusted by -1 Standard Error of Measurement (S. E. M.) on the Multiple-Choice Raw Score Points and by -1 score-point on the Constructed-Response Raw Score Points

² Subtest III adjusted by -1 S. E. M. on the Multiple-Choice Raw Score Points and by -2 score-point on the Constructed-Response Raw Score Points

³ Percent of examinees taking and passing all subtests for November 5, 2005 test administration.

Recommendations from the 2009 Passing Standard Review Panel

A new CSET: ASL standard setting review advisory panel was appointed by the Executive Director from the most qualified applicants. For historical and continuity purposes, the panel included two of the members of the original 2005 ASL standard setting advisory panel.

The panel followed all of the same standard setting procedures previously described above with respect to the work of the 2005 standard setting advisory panel. Because the Passing Standard Review Panel meeting was scheduled for January 15-16, 2009, which is after the agenda publication deadline date, the panel's report and recommendations will be provided as an in-folder attachment to this agenda item.

APPENDIX A

**PERFORMANCE CHARACTERISTICS
AND
SCORING SCALES**

Appendix A
Performance Characteristics and Scoring Scales

Subtest I Focused Constructed Response Items

PERFORMANCE CHARACTERISTICS

PURPOSE	The extent to which the response addresses the constructed-response assignment's charge in relation to relevant CSET subject matter requirements.
SUBJECT MATTER KNOWLEDGE	The application of accurate subject matter knowledge as described in the relevant CSET subject matter requirements.
SUPPORT	The appropriateness and quality of the supporting evidence in relation to relevant CSET subject matter requirements.

SCORE SCALE

SCORE POINT	SCORE POINT DESCRIPTION
3	<p>The "3" response reflects a command of the relevant knowledge and skills as defined in the CSET subject matter requirements.</p> <ul style="list-style-type: none"> • The purpose of the assignment is fully achieved. • There is an accurate application of relevant subject matter knowledge. • There is appropriate and specific relevant supporting evidence.
2	<p>The "2" response reflects a general command of the relevant knowledge and skills as defined in the CSET subject matter requirements.</p> <ul style="list-style-type: none"> • The purpose of the assignment is largely achieved. • There is a largely accurate application of relevant subject matter knowledge. • There is acceptable relevant supporting evidence.
1	<p>The "1" response reflects a limited or no command of the relevant knowledge and skills as defined in the CSET subject matter requirements.</p> <ul style="list-style-type: none"> • The purpose of the assignment is only partially or not achieved. • There is limited or no application of relevant subject matter knowledge. • There is little or no relevant supporting evidence.
U	The "U" (Unscorable) is assigned to a response that is unrelated to the assignment, unreadable, not in American Sign Language or written English, or does not contain a sufficient amount of original work to score.
B	The "B" (Blank) is assigned to a response that is blank.

Subtest I Extended Constructed Response Items

PERFORMANCE CHARACTERISTICS

PURPOSE	The extent to which the response addresses the constructed-response assignment's charge in relation to relevant CSET subject matter requirements.
SUBJECT MATTER KNOWLEDGE	The application of accurate subject matter knowledge as described in the relevant CSET subject matter requirements.
SUPPORT	The appropriateness and quality of the supporting evidence in relation to relevant CSET subject matter requirements.
DEPTH AND BREADTH OF UNDERSTANDING	The degree to which the response demonstrates understanding of the relevant CSET subject matter requirements.

SCORE SCALE

SCORE POINT	SCORE POINT DESCRIPTION
4	<p>The "4" response reflects a thorough command of the relevant knowledge and skills as defined in the CSET subject matter requirements.</p> <ul style="list-style-type: none"> • The purpose of the assignment is fully achieved. • There is a substantial and accurate application of relevant subject matter knowledge. • The supporting evidence is sound; there are high-quality, relevant examples. • The response reflects a comprehensive understanding of the assignment.
3	<p>The "3" response reflects a general command of the relevant knowledge and skills as defined in the CSET subject matter requirements.</p> <ul style="list-style-type: none"> • The purpose of the assignment is largely achieved. • There is a largely accurate application of relevant subject matter knowledge. • The supporting evidence is adequate; there are some acceptable, relevant examples. • The response reflects an adequate understanding of the assignment.
2	<p>The "2" response reflects a limited command of the relevant knowledge and skills as defined in the CSET subject matter requirements.</p> <ul style="list-style-type: none"> • The purpose of the assignment is partially achieved. • There is limited accurate application of relevant subject matter knowledge. • The supporting evidence is limited; there are few relevant examples. • The response reflects a limited understanding of the assignment.
1	<p>The "1" response reflects little or no command of the relevant knowledge</p>

	<p>and skills as defined in the CSET subject matter requirements.</p> <ul style="list-style-type: none"> • The purpose of the assignment is not achieved. • There is little or no accurate application of relevant subject matter knowledge. • The supporting evidence is weak; there are no or few relevant examples. • The response reflects little or no understanding of the assignment.
U	The "U" (Unscorable) is assigned to a response that is unrelated to the assignment, unreadable, not in American Sign Language or written English, or does not contain a sufficient amount of original work to score.
B	The "B" (Blank) is assigned to a response that is blank.

Subtest II Focused Constructed Response Items

PERFORMANCE CHARACTERISTICS

PURPOSE	The extent to which the response addresses the constructed-response assignment's charge in relation to relevant CSET subject matter requirements.
SUBJECT MATTER KNOWLEDGE	The application of accurate subject matter knowledge as described in the relevant CSET subject matter requirements.
SUPPORT	The appropriateness and quality of the supporting evidence in relation to relevant CSET subject matter requirements.

SCORE SCALE

SCORE POINT	SCORE POINT DESCRIPTION
3	The "3" response reflects a command of the relevant knowledge and skills as defined in the CSET subject matter requirements. <ul style="list-style-type: none">• The purpose of the assignment is fully achieved.• There is an accurate application of relevant subject matter knowledge.• There is appropriate and specific relevant supporting evidence.
2	The "2" response reflects a general command of the relevant knowledge and skills as defined in the CSET subject matter requirements. <ul style="list-style-type: none">• The purpose of the assignment is largely achieved.• There is a largely accurate application of relevant subject matter knowledge.• There is acceptable relevant supporting evidence.
1	The "1" response reflects a limited or no command of the relevant knowledge and skills as defined in the CSET subject matter requirements. <ul style="list-style-type: none">• The purpose of the assignment is only partially or not achieved.• There is limited or no application of relevant subject matter knowledge.• There is little or no relevant supporting evidence.
U	The "U" (Unscorable) is assigned to a response that is unrelated to the assignment, unreadable, not in written English, or does not contain a sufficient amount of original work to score.
B	The "B" (Blank) is assigned to a response that is blank.

Subtest III Focused Constructed Response Items

Assessment Structure

Candidates are presented with an exercise in which they must correct five sentences in American Sign Language. There are ten scorable tasks that are presented in sentence context in which corrections to language structure and/or articulation must be made.

The tasks included in the assignment are as follows:

Error Analysis	Five sentences, each containing two errors that require each sentence to be resigned, correcting the contained errors— 10 tasks
-----------------------	--

The score point assigned and reported for the assignment is a composite score based on the number of errors identified in the candidate's responses.

SCORE POINT	ERROR TABULATION
3	0–2 errors
2	3–6 errors
1	7–10 errors
U	10 errors— <u>all</u> responses unscorable (i.e., unrelated to the assignment, unreadable, not in American Sign Language or not containing a sufficient amount of original work to score)
B	10 errors— <u>all</u> responses blank

Subtest III Focused Constructed Response Items

PERFORMANCE CHARACTERISTICS

PURPOSE	The extent to which the response addresses the constructed-response assignment's charge in relation to relevant CSET subject matter requirements.
SUBJECT MATTER KNOWLEDGE	The application of accurate subject matter knowledge as described in the relevant CSET subject matter requirements.
SUPPORT	The appropriateness and quality of the supporting evidence in relation to relevant CSET subject matter requirements.

SCORE SCALE

SCORE POINT	SCORE POINT DESCRIPTION
3	<p>The "3" response reflects a command of the relevant knowledge and skills as defined in the CSET subject matter requirements.</p> <ul style="list-style-type: none"> • The purpose of the assignment is fully achieved. • There is an accurate application of relevant subject matter knowledge. • There is appropriate and specific relevant supporting evidence.
2	<p>The "2" response reflects a general command of the relevant knowledge and skills as defined in the CSET subject matter requirements.</p> <ul style="list-style-type: none"> • The purpose of the assignment is largely achieved. • There is a largely accurate application of relevant subject matter knowledge. • There is acceptable relevant supporting evidence.
1	<p>The "1" response reflects a limited or no command of the relevant knowledge and skills as defined in the CSET subject matter requirements.</p> <ul style="list-style-type: none"> • The purpose of the assignment is only partially or not achieved. • There is limited or no application of relevant subject matter knowledge. • There is little or no relevant supporting evidence.
U	The "U" (Unscorable) is assigned to a response that is unrelated to the assignment, unreadable, not in American Sign Language, or does not contain a sufficient amount of original work to score.
B	The "B" (Blank) is assigned to a response that is blank.

APPENDIX B

DETAILED STANDARD SETTING PROCESS

Appendix B Detailed Standard Setting Process

Round One Standard Setting Ratings

In Round One, panel members independently provided item-by-item ratings, first for the multiple-choice items and then for the constructed-response items.

Multiple-Choice Items

For Round One, panel members were provided the following materials:

- subject matter requirements
- the subtest forms used for the current test administration
- the accompanying subtest form answer keys
- the Round One Rating Form for multiple-choice items
- the Round One Rating Form for multiple-choice items

Round One began with a set of practice, multiple-choice items for each panel member to rate. This set of items represented a range of item difficulties. Panel members were asked to rate each item by responding to the following question.

Imagine a hypothetical group of candidates for the Single Subject Teaching Credential in American Sign Language, each of whom is just at the level of knowledge and skills important for effective job performance as a beginning teacher in a departmentalized classroom in California public schools.

What percent of this group would answer the item correctly?

0% – 10% = 1	51% – 60% = 6
11% – 20% = 2	61% – 70% = 7
21% – 30% = 3	71% – 80% = 8
31% – 40% = 4	81% – 90% = 9
41% – 50% = 5	91% – 100% = 10

Panel members were polled as to how they rated each item and as a panel discussed, when necessary, expected performance of the “just acceptable” candidate and the standard setting procedure.

Following the practice set, panel members began the same rating process with the multiple-choice items used on the current operational test forms. Pearson staff analyzed the individual and group results from these item judgments (percentage of “just acceptable” candidates who would answer the item correctly) for use in Round Two of the standard setting process.

Constructed-Response Items

For Round One of the constructed-response item ratings, panel members were provided the following materials:

- the subtest description
- the subtest form used for the November 2005 test administration
- the appropriate set of performance characteristics and scoring scale
- the Subject Matter Advisory Panel-approved marker responses for each score point on the scoring scale, where available
- the Round One Rating Form for constructed-response items

To begin the Round One constructed-response ratings, panel members rated a practice set of sample items. They were asked to rate each item by responding to the following question.

Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning teacher of American Sign Language in California public schools.

For this constructed-response item, which of the points on the scoring scale represents the level of response that would be achieved by this individual?

After panel members completed the practice set of constructed-response items, Pearson staff polled them regarding their item ratings; facilitated a discussion to review the concept of the “just-acceptable candidate”; discussed how to make the standard setting judgment; discussed how to review and consider the marker responses; and answered questions about the rating process.

Following the practice set, panel members began the same rating process with the actual constructed-response items used on the current operational test forms. In responding to the standard setting question, panel members were asked to refer to the score point descriptions that are appropriate for the type of constructed-response item under consideration (i.e., the descriptions associated with a three-point scale or those associated with a four-point scale). They were also asked to refer to the marker responses for each score point for each assignment. Pearson staff analyzed the individual results from these item judgments for use in Round Two of the standard setting process.

Round Two Standard Setting Ratings

Round Two of the standard setting process moved the panel from providing ratings at the item level to ratings made at the component level (i.e., the multiple-choice component and the constructed-response component) of each subtest. Panel members were asked to provide, for each subtest, (1) separate preliminary passing score recommendations for the set of multiple-choice items and the set of constructed-response items on each subtest and (2) the percent of points to be allocated for each component in the subtest.

For Round Two, panel members were provided the following materials:

- subject matter requirements
- the subtest descriptions

- the Round One multiple-choice item rating summary sheet, which provided the sum of the median rating for all items across all panel members and, for each panel member, the sum of their Round One ratings listed in descending order by score value
- the Round One constructed-response item rating summary sheet, which provided the sum of the median rating for all items across all panel members, doubled to reflect the actual combined scores examinees will receive from two scorers. The sheet also provided the sum of each panel member's Round One constructed-response item ratings doubled to reflect the actual combined scores examinees will receive from two scorers. These individual ratings were listed in descending order by score value
- the Round Two subtest component standard setting recommendation form for multiple-choice items
- the Round Two subtest component standard setting recommendation form for constructed-response items

(NOTE: Results of individual panel members were provided by identification number only to maintain the confidentiality of each person's ratings.)

Multiple-Choice Items

Panel members were given an opportunity to discuss the results of the Round One ratings and to provide their thoughts on the merits of various multiple-choice component "cut scores" at the subtest level (understanding that candidates will not "pass" the multiple-choice component alone; candidates' pass/fail status will be determined at the subtest level, which involves the combination of multiple-choice component and constructed-response component performance). The concept of the multiple-choice component "cut score" was used as a temporary convenience to discuss the aggregated panel member ratings.

Working independently, and considering their own aggregated rating from Round One and the group median, each panel member provided a Round Two multiple-choice component "cut score" recommendation for each subtest by responding to the following question.

Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning teacher of American Sign Language in California public schools.

What is the number of multiple-choice items on the subtest (out of the total number of scorable items) that would be answered correctly by this individual?

Constructed-Response Items

Panel members were given an opportunity to discuss the ratings and to provide their thoughts on the merits of various constructed-response component "cut scores" at the subtest level (understanding that candidates will not "pass" the constructed-response component alone; candidates' pass/fail status will be determined at the subtest level, which involves the combination of multiple-choice component and constructed-response component performance). The concept of the constructed-response component "cut score" was used as a temporary convenience to discuss the aggregated panel member ratings.

Working independently, and considering their own ratings from Round One and the results of the group's ratings, each panel member provided a Round Two constructed-response component "cut score" recommendation for each subtest by responding to the following question.

Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning teacher of American Sign Language in California public schools.

What is the total score for the constructed-response items on the subtest (out of the total number of score points) that would be obtained by this individual?

Combined Component Scores

The panel was provided with the concept of combining subtest component scores in terms of determining the percent of the total points available to be allocated to each component of a subtest. Key issues that are relevant to this determination were discussed, such as the concept of reliability, the length of each component, and the nature of the information about a candidate's knowledge and skills that is to be provided by each component. The following options that were provided to panel members are intended to yield reliable results and are psychometrically defensible.

For subtests I and II, panel members were given two alternatives for allocating points consistent with psychometric standards and the structure of each examination: (a) multiple-choice component 80% and the constructed-response component 20% or (b) the multiple-choice component 70% and the constructed-response component 30%. Panel members were given the opportunity to discuss the options, with advice from Commission staff and Pearson staff.

Following the discussion, each panel member was asked to independently make a recommendation by responding to the following question.

80%-20% or 70%-30% example:

In combining scores on the multiple-choice component and the constructed-response component to yield a total subtest score, what percent of points should be allocated to each component?

Check one of the following:

_____ *80% multiple-choice component and
20% constructed-response component*

_____ *70% multiple-choice component and
30% constructed-response component*

A similar process was used for subtest III, which assesses the ability to comprehend signed ASL and the ability to sign ASL correctly. The panel members were given two alternatives for allocating points: (a) multiple-choice component 40% and the constructed-response component 60% or (b) the multiple-choice component 50% and the constructed-response component 50%. These choices are consistent with psychometric standards and the structure of each examination.

Following this combined component score rating activity, Pearson collected and analyzed the panel members' recommendations and informed the panelists of the results.

Round Three Standard Setting Ratings

The goal of Round Three of the standard setting process was to produce a passing standard recommendation for each component of each subtest and a set of panel-recommended rules for combining scores from the multiple-choice and constructed-response components.

For Round Three, panel members were provided the following materials:

- subject matter requirements
- the subtest descriptions
- the Round Two multiple choice results summary sheet, which included the panel's computed median, and each panel member's Round Two multiple choice rating listed in descending order by score value
- the Round Two constructed-response results summary sheet, which included the panel's computed median, and each panel member's Round Two constructed-response item rating
- the Round Two tabulated panel recommendations on component score combinations
- the Round Three subtest standard setting recommendation form

These materials helped to facilitate a discussion among each panel about their ratings, the nature of the examinee sample, the options for combining component scores, the goal of Round Three, the purpose of the CSET program, and the concept of the just-acceptable candidate.

Panel members were cautioned about making judgments based on small numbers of examinees, and were advised that the examinees at the specific test administration may or may not reflect the same proportions of all the types and capabilities of examinees in the population that will take the test in the future.

After a discussion, panel members were asked to independently recommend a passing standard and score combination rule for each subtest in their field by responding to the following questions.

Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning teacher of American Sign Language in California public schools.

What is the number of multiple-choice items on the subtest (out of the total number of scorable items) that would be answered correctly by this individual?

What is the total score for the constructed-response items on the subtest (out of the total number of score points) that would be obtained by this individual?

In combining scores on the multiple-choice component and the constructed-response component to yield a total subtest score, what percent of points should be allocated to each component?

80% multiple-choice component and 20% constructed-response component

OR

70% multiple-choice component and 30% constructed-response component

(The same questions were asked for subtest III using the 50/50 and 40/60 options)

As the final step to the standard setting study, each panel member was asked to complete independently a meeting evaluation form regarding the training provided and the task in general. Pearson staff compiled the results of the standard setting panel work for use in the determination of the staff-recommended passing standards that were adopted by the Commission.

APPENDIX C

CSET STANDARD SETTING CONSIDERATIONS

CSET Standard Setting Considerations

As described in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education, 1999), the standard setting process is a key piece of validity evidence supporting a testing program.

Defining the minimum level of knowledge and skill required for licensure or certification is one of the most important and difficult tasks facing those responsible for credentialing. Verifying the appropriateness of the cut score or scores on the tests is a critical element in validity. The validity of the inference drawn from the test depends on whether the standard for passing makes a valid distinction between adequate and inadequate performance. Often, panels of experts are used to specify the level of performance that should be required. Standards must be high enough to protect the public, as well as the practitioner, but not so high as to be unreasonably limiting. Verifying the appropriateness of the cut score or scores on a test used for licensure or certification is a critical element of the validity of test results (p.157).

In making initial and subsequent recommendations to the Commission on passing standards for the CSET: ASL, staff considered the following factors and options that affect the standard setting process in determining the staff-recommended passing standards.

Subtest Scoring Model

The subtest scoring model used with CSET is a non-compensatory subtest model in which all subtests in a subject area must be passed independently. The original Subject Matter Advisory Panel considered this model when determining the subtest structure of the examination.

Professional Judgments

The recommended passing standards for the CSET are based upon the professional judgments provided by the members of the Standard Setting Panels. Since these panel recommendations are criterion-referenced—based on expert judgment of the minimum required subject matter knowledge for beginning teachers—examinee performance data provides supplemental, though not necessary, information. Performance data is provided to inform those judgments when there are at least 20 examinees.

Standard Error of Measurement

Standard error of measurement is one way to express test reliability and addresses the imprecision of test data. Measurements are not perfectly reliable. In testing, for example, only one score from a single test administration is available for each examinee. An individual examinee's score may, or may not, be the same as the examinee's hypothetical "true score". However, the standard error allows us to determine a range within which the examinee's true score is likely to lie. Within reasonable limits, the standard error of measurement provides a safeguard against placing undue emphasis on a single numerical score. This is just one index of reliability, and should be applied to the standard setting process in combination with other test-specific characteristics.