
7A

Information/Action

Professional Services Committee

Recommended Initial Passing Standards for the California Subject Matter Examinations for Teachers (CSET): American Sign Language

Executive Summary: This report provides the Commission with recommendations relevant to the determination of passing standards for the CSET: American Sign Language (ASL).

Staff Direction: Commission staff is seeking direction from the Commission regarding passing standards for the CSET ASL examination.

Presenters: Amy Jackson, Administrator, Professional Services Division and Dr. Marty Karlin, Senior Area Director, National Evaluation Systems, Inc.

Strategic Plan Goal: 1

Promote educational excellence through the preparation and certification of professional educators.

- ◆ Sustain high quality standards for the preparation of professional educators.
- ◆ Sustain high quality standards for the performance of credential candidates.

Recommended Initial Passing Standards for the California Subject Matter Examinations for Teachers (CSET): American Sign Language

Introduction

This report describes the standard setting studies for the California Subject Examinations for Teachers (CSET): American Sign Language (ASL) and provides staff-recommended initial passing standards for each subtest of the examination.

Background

In spring 2004, the Commission's Executive Director appointed a subject matter advisory panel for the California Subject Examinations for Teachers (CSET) for the single subject area of American Sign Language (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. (NES), the Commission's CSET testing contractor, and Commission staff have worked with this panel since then to facilitate this work. The 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 have expertise in ASL.

From spring through fall of 2004, the panel developed the subject matter requirements (SMRs) for ASL that were aligned with available state and national student content standards and frameworks, and standards of national professional organizations. Staff from the California Department of Education (CDE) participated in all of the panel development activities. The Commission approved those SMRs in its January-February 2005 meeting. The SMRs specify the content that is to be taught in Commission-approved subject matter preparation programs and that is eligible for assessment on the Commission's subject matter examination.

The panel also developed standards based upon the content in the SMRs for institutions that wish to offer undergraduate subject matter programs in ASL, that will be utilized by California accredited colleges and universities to develop single subject matter preparation programs in ASL. Those program standards will be presented to the Commission for consideration at a future meeting.

The development of the new examination began in spring 2004. Test structures were developed by the advisory panel and 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 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 new CSET: ASL.

On November 5, 2005, the first test administration of this new examination was conducted. On December 6-7, 2005, the standard setting study for this examination was held in Sacramento to determine the initial passing standard recommendations of California educators.

The CSET: American Sign Language

The new CSET: ASL 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 of this agenda item.

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 will be administered two times each year. For the first administration of the CSET: ASL in November, fifteen examinees completed subtest I, II and III. Fifteen examinees completed subtest I, and sixteen examinees completed each of subtests II and III. On December 6-7, 2005, Commission staff and NES conducted a standard setting study for the new examination. The standard-setting procedures used and the results of these studies are described later in this report.

Table 1: Subtest Structure of the CSET: American Sign Language

Subtest	Domains	Number of Multiple-Choice Items	Number of Constructed-Response Items
I	Literary and Cultural Texts and Traditions	10	1 short (focused)
	Cultural Analysis and Comparisons	20	1 extended
	<i>Subtest total</i>	30	1 short (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 short (focused)
	<i>Subtest total</i>	30	2 short (focused)
III	Linguistics of the Target Language – American Sign Language (Error Analysis)	none	1 short (focused)
	Language and Communication: Receptive Comprehension	18	1 short (focused)
	Language and Communication: Expressive Production	none	2 short (focused)
	<i>Subtest total</i>	18	4 short (focused)
	<i>Total Items</i>	78	7 short (focused) 1 extended

The Standard Setting Study

The standard setting study for the CSET: ASL was conducted December 6-7, 2005, with the standard setting panel of ASL specialists. Information about the panel is shown in Appendix B. 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 initial passing standards. A total of 8 panel members selected from across the state who have varied backgrounds related to ASL, including college and university faculty; classroom teachers; and ASL interpreters participated in the study.

Each standard setting study began with an orientation and training session. Panel members were provided with the subject matter requirements and the subtest forms used for the November 2005 test administration. To help the panel members become familiar with the examinations, 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 C.

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.”

Results

Following the standard setting studies, NES 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.

A summary of the panel-based passing score recommendations, including the number of scorable items and the weighting of each component in the total subtest score, is provided in Table 2 below.

Table 2: Panel Recommended Initial Passing Standards for CSET: American Sign Language

Subtest	Item Type ¹	Scorable Items	Possible Score Points	Computed Median Based on Panel Recommendations	Component Score Combination Rule ²	
					80/20	70/30
I. Literary and Cultural Texts and Traditions; Cultural Analysis and Comparisons	MC	24	24	19.50		√
	CR	2	14	11.25		
II. General Linguistics; Linguistics of the Target Language – American Sign Language (Language Structures; Contrastive Analysis; Sociolinguistics and Pragmatics)	MC	24	24	18.50		√
	CR	2	12	10.07		
Subtest	Item Type ¹	Scorable Items	Possible Score Points	Computed Median Based on Panel Recommendations	Component Score Combination Rule ²	
					50/50	40/60
III. Linguistics of the Target Language – American Sign Language (Error Analysis); Language and Communication: Receptive Comprehension; Language and Communication: Expressive Production	MC	15	15	11.30		√
	CR	4	24	20.50		

¹ MC = multiple-choice, CR = constructed-response

² The component score combination rule is formatted as multiple-choice percent/constructed-response percent (e.g., 80/20 is 80% multiple choice / 20% constructed response).

Staff-Recommended Initial Passing Standards

Based on the previously approved guidelines for the establishment of CSET standards (see Appendix D), staff recommends that the Commission adopt the initial passing standards for the subtests of the CSET forms administered on November 5, 2005 that:

- are equivalent to the raw score points on the multiple-choice component and on the constructed-response component as shown in Table 3;
- are based on the component score combination rules as shown in Table 3; and
- reflect passing standards that are equivalent for future forms of the test.

The staff-recommended raw score points for multiple-choice and constructed-response components reflect adjustments made for standard errors of measurement as appropriate.

For the CSET, it is appropriate to review passing standards periodically to verify that the standards are fulfilling the responsibility of the Commission to award teaching credentials only to those candidates who have fulfilled the subject matter requirement. The first administration of the new CSET: ASL yielded fewer than 20 examinees. For this reason, a subsequent passing standard activity will be held to review the passing standards in light of the increased number of examinees once there are at least 150 examinees. Following further review, recommendations for any change in the standards will be presented to the Commission for consideration and adoption.

Passing status will be determined on the basis of total subtest performance. Test results will be 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.

If the Commission adopts the staff-recommended initial passing standards, as indicated in Table 3, NES will release score reports for the November 2005 test administration by February 27, 2006. The next test administration of the CSET in these subject areas is scheduled for May 20, 2006.

Table 3: Staff-Recommended Initial Passing Standards for CSET: American Sign Language*

- * 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
- * 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

Subtest	Multiple-Choice Raw Score Points	Constructed-Response Raw Score Points	Component Score Combination Rule MC/CR	Passing Rate for November 5, 2005 Test Administration by Subtest	Overall Passing Rate for November 5, 2005 Test Administration
Subtest I	17	10	70/30	47%	20%
Subtest II	16	9	70/30	63%	
Subtest III	10	19	40/60	38%	

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

APPENDIX A

PERFORMANCE CHARACTERISTICS AND SCORING SCALES

Appendix A: Performance Characteristics and Scoring Scales

Subtest I Focused CRI

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 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 CRI

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 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 CRI

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 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 CRI

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
-----------------------	--

Scoring the Assignment

To denote a "Correct" response, make no mark in the "ITEMS INCORRECT" box on the scoring form for each item if, and only if, the candidate's response is described by the following:

NOTATION	NOTATION DESCRIPTION
	Correct. The candidate's response acceptably corrects the syntactic or linguistic errors found in the sentence, i.e., the response has located and corrected the errors of language structure and/or articulation found within the sentence and has retained the original meaning.

To denote an "Incorrect" response, mark the "ITEMS INCORRECT" box on the scoring form for each item if, and only if, the candidate's response is described by the following:

NOTATION	NOTATION DESCRIPTION
X	Incorrect. The candidate's response does not accurately follow American Sign Language's rules of structure and/or articulation or the response is incomplete or is blank.

When all 10 tasks have been scored, count the number of items marked as incorrect, and fill in the numbered circle, which indicates the "TOTAL INCORRECT."

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 CRI

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

STANDARD SETTING PANEL DEMOGRAPHICS

Appendix B

CSET: ASL Standard Setting Panel Demographics

Total Number	
Appointed	10
Participated	8
Ethnicity	
African American	
Asian	
Hispanic	
White	8
Other/Not Provided	
Sex	
Female	5
Male	3
Region	
North	2
South	6
Profession	
Public School Educators	5
College/University Educators	3
Years of Teaching Experience	
0-6	1
7-10	4
11+	3
Not Provided	

APPENDIX C

DETAILED STANDARD SETTING PROCESS

Appendix C 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 November 2005 test administration;
- the accompanying subtest form answer keys;
- the Round One Rating Form for multiple-choice items; and

Round One began with a set of 8 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 November 5, 2005 operational test forms. NES 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; and
- the Round One Rating Form for constructed-response items.

To begin the Round One constructed-response ratings, panel members rated a practice set of 2 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, NES 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 November 2005 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. NES 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;

- 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.
- 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.
- Round Two Subtest component Standard Setting Recommendation Form for multiple-choice items; and,
- 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 XX—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 XX—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 NES 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:

<i>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?</i>	
<i>Check <u>one</u> of the following:</i>	
<input type="checkbox"/>	<i>80% multiple-choice component and 20% constructed-response component</i>
<input type="checkbox"/>	<i>70% multiple-choice component and 30% constructed-response component</i>

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, NES 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;
- 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;
- 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;
- Round Two tabulated panel recommendations on component score combinations; and
- 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 first 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 XX—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 XX—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.

NES compiled the results of the standard setting panel work for use in the determination of the staff-recommended passing standards presented in this report.

APPENDIX D

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, & 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 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 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 Panel. 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.