
7A

Action

Professional Services Committee

Recommended Initial Passing Standards for the California Subject Examinations for Teachers (CSET) in Languages Other Than English in German, Japanese, Korean, Mandarin, Punjabi, Russian, and Vietnamese

Executive Summary: The CSET: Languages Other Than English in German, Japanese, Korean, Mandarin, Punjabi, Russian, and Vietnamese were first administered on November 6, 2004. This report describes the standard setting studies conducted on December 2-3, 2004, the results of those studies, and provides staff-recommended passing standards.

Recommended Action: Staff recommends the Commission adopt the proposed passing standards found on page 11 of this report for CSET: Languages Other Than English in German, Japanese, Korean, Mandarin, Punjabi, Russian, and Vietnamese.

Presenters: Diane Tanaka, Assistant Consultant, Amy Jackson, Administrator, Professional Services Division

Strategic Plan Goal(s):

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 Examinations for Teachers (CSET): Languages Other Than English in German, Japanese, Korean, Mandarin, Punjabi, Russian, and Vietnamese

Introduction

This report describes the standard setting studies for the remaining Phase II California Subject Examinations for Teachers (CSET): Languages Other Than English in German, Japanese, Korean, Mandarin, Punjabi, Russian, and Vietnamese conducted on December 2-3, 2004, and provides recommended initial passing standards for each examination.

Background

In spring 2003, the Commission's Executive Director appointed subject matter advisory panels for Phase II of the California Subject Examinations for Teachers (CSET) for the single subject areas of art, languages other than English, music, physical education, and for preliminary educational technology to advise Commission staff on the development of new subject matter program standards and examinations in these subject areas. National Evaluation Systems, Inc. (NES), the Commission's CSET testing contractor, and Commission staff have worked with these panels since then to facilitate this work. These subject matter advisory panels consisted of diverse groups of classroom teachers; subject area specialists; college and university faculty; teacher educators; and members of relevant professional organizations and committees, each with a specialty in the subject area of the panel.

From spring through fall of 2003, each panel developed the subject matter requirements (SMRs) for their specific subject area that were aligned with available state and national student content standards and frameworks, and standards of national professional organizations. The Commission approved those SMRs in its January 2004 meeting. The SMRs specify the content that is to be taught in Commission-approved subject matter preparation programs and the content that is eligible for assessment on the Commission's subject matter examinations.

The panels also developed new program standards for each subject area based upon the content in the SMRs that will be utilized by California accredited colleges and universities to develop single subject matter preparation programs in these academic areas. Those program standards were approved by the Commission in its May 2004 meeting.

The development of thirteen new examinations began in spring 2004. The new Phase II examinations are the CSET: Art; Language Other Than English in French, German, Japanese, Korean, Mandarin, Punjabi, Russian, Spanish, and Vietnamese; Music; Physical Education; and Preliminary Educational Technology. Test structures were approved by the advisory panels and multiple-choice and constructed-response items were drafted, reviewed and revised as needed by

both the Bias Review Committee and the appropriate subject matter advisory panel. Once these items were field-tested, the panels selected marker responses and scored the constructed-responses from the field test. Additionally, test guides including the subject matter requirements, test structures, and sample questions were developed to assist candidates in preparing to take the new examinations.

The last step in the test development process is to conduct standard setting studies for use in the establishment of the initial passing standards for the examinations. At its December 2004 meeting, the Commission adopted the initial passing standards for the CSET: Art, French, Music, Physical Education, Spanish, and Preliminary Educational Technology which were based upon the September 18, 2004 test administration. For more information regarding the establishment of those passing standards, please refer to the Commission's website at <http://www.ctc.ca.gov/aboutctc/agendas/december-2004/december-2004-8D.pdf>.

On November 6, 2004, the first test administrations of the remaining Phase II examinations for German, Japanese, Korean, Mandarin, Punjabi, Russian and Vietnamese were held. On December 2-3, 2004, the standard setting studies for these examinations were held in Sacramento.

The CSET: German, Japanese, Korean, Mandarin, Punjabi, Russian and Vietnamese

Each of the seven new language tests in the CSET program is comprised of three subtests differentiated by content area. Due to the nature of the languages, the subject matter advisory panels established two test structure models dependent upon the origin of the language. Although the content to be assessed for each examination was the same, there were slight variations in the number and type of multiple-choice and constructed response items, as well as scoring rubrics. The subtest structures are shown in Tables 1 and 2, and the performance scales and scoring scales are found in Appendix A.

Currently, CSET language subtests are administered in two testing sessions: the performance-based Subtest III is administered in the morning session, and paper and pencil formatted Subtests I and II are administered in the afternoon. Each CSET testing session is five hours in length. Examinees may choose to take one or both of the afternoon subtests within a single testing session. Individual subtests are not timed. The CSET in German, Japanese, Korean, Mandarin, Punjabi, Russian, and Vietnamese will be administered two times each year. The numbers of examinees who completed⁶ subtests at the first test administration of the tests in November are provided in Appendix B.

⁶ Completion is defined as having attempted at least five multiple-choice items AND provided a scorable response to each constructed-response item.

Table 1: Subtest Structure of the CSET: German, Russian

Subtest	Number of Multiple-Choice Items	Number of Constructed-Response Items
I: General Linguistics	15	1 focused
Linguistics of the Target Language	25	2 focused
<i>Subtest total</i>	40	3 focused
II: Literary & Cultural Texts & Traditions	20	1 extended
Cultural Analysis & Comparisons	20	
<i>Subtest total</i>	40	1 extended
III: Language & Communication:		
Listening	15	1 focused
Reading	15	1 focused
Written	none	1 focused 1 extended
Oral Expression	none	2 focused
<i>Subtest total</i>	30	5 focused 1 extended
Total Items	110	8 focused 2 extended

Table 2: Subtest Structure of the CSET: Japanese, Korean, Mandarin, Punjabi, and Vietnamese

Subtest	Number of Multiple-Choice Items	Number of Constructed-Response Items
I: General Linguistics	5	1 focused
Linguistics of the Target Language	15	3 focused
<i>Subtest total</i>	20	4 focused
II: Literary & Cultural Texts & Traditions	10	2 focused
Cultural Analysis & Comparisons	10	1 focused 1 extended
<i>Subtest total</i>	20	3 focused 1 extended
III: Language & Communication:		
Listening	10	1 focused
Reading	10	1 focused
Written	none	1 focused 1 extended
Oral Expression	none	2 focused
<i>Subtest total</i>	20	5 focused 1 extended
Total Items	60	12 focused 2 extended

The Standard Setting Studies

Standard setting studies for the remaining examinations of the CSET program were conducted December 2-3, 2004 with language-specific panels (see Appendix C). The purpose of the standard setting procedure is to provide the Commission with recommendations, based on the informed judgments of California educators, relevant to the determination of the initial passing standards for the CSET: German, Japanese, Korean, Mandarin, Punjabi, Russian, and Vietnamese. Panel members selected from across the state, including curriculum specialists, public school teachers, teacher educators, school administrators, mentor teachers, and superintendents, participated in the studies.

The standard-setting procedures used for these examinations are the established methods used for each of the other examinations in the CSET program. Each standard setting study began with an orientation and training session. Panel members were provided the subject matter requirements, the subtest forms used for the November 2004 test administration and the accompanying subtest form answer keys, and for subtests with two or more examinees, item statistics displaying the percent of the November 2004 examinees who answered each test item correctly. To help the panel members become familiar with the examinations, the required knowledge and skills

assessed, 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, and then to score their own performance on the multiple-choice items.

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

In Round One, panel members individually rated 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 or response that would be achieved by the “just acceptable” candidate for the constructed-response items.

Using the item statistics produced from Round One to inform judgements, Round Two moved the panel from item ratings to independent 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 by the “just acceptable” teacher candidate. 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., 50% multiple-choice and 50% constructed-response, 60% multiple-choice and 40% constructed-response).

In the final round of ratings, the panel members were asked to make independent recommendations for a passing standard and “component score combination rule”. To inform their discussions, they were provided the results of the component-level statistics generated from Round Two; and for subtests with more than 15 examinees, any applicable examinee demographic information and data analyses on the percent of examinees that would pass at particular raw score combinations that were available from the first administration of the test. Due to the low number of examinees who took the first administration of examinations in November, only the Japanese panel was presented with the subtest data analyses.

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 Tables 3-9.

Table 3: Panel-Recommended Initial Passing Standards for CSET: German

German	Item Type ^{1/}	Scorable Items	Possible Score Points	Computed Median based on Panel Recommendations	Component Score Combination Rule ^{2/}	
					80/20	70/30
I. General Linguistics; Linguistics of the Target Language	MC	32	32	20.0		✓
	CR	3	18	10.0		
II. Literary & Cultural Texts & Traditions; Cultural Analysis & Comparisons	MC	32	32	18.0	✓	
	CR	1	8	4.0		
					60/40	50/50
III. Language & Communication: Listening; Reading; Written; Oral Expression	MC	24	24	17.1		✓
	CR	6	38	24.0		

Table 4: Panel-Recommended Initial Passing Standards for CSET: Japanese

Japanese	Item Type ^{1/}	Scorable Items	Possible Score Points	Computed Median based on Panel Recommendations	Component Score Combination Rule ^{2/}	
					50/50	40/60
I. General Linguistics; Linguistics of the Target Language	MC	16	16	12.0		✓
	CR	4	18	14.1		
II. Literary & Cultural Texts & Traditions; Cultural Analysis & Comparisons	MC	16	16	11.3		✓
	CR	4	26	17.5		
III. Language & Communication: Listening; Reading; Written; Oral Expression	MC	16	16	13.1		✓
	CR	6	38	31.0		

^{1/} MC = multiple-choice, CR = constructed-response

^{2/} The component score combination rule is formatted as multiple-choice percent/constructed-response percent (e.g., 60/40 is 60% multiple choice / 40% constructed response).

Table 5: Panel-Recommended Initial Passing Standards for CSET: Korean

Korean	Item Type ^{1/}	Scorable Items	Possible Score Points	Computed Median based on Panel Recommendations	Component Score Combination Rule ^{2/}	
					50/50	40/60
I. General Linguistics; Linguistics of the Target Language	MC	16	16	10.0	✓	
	CR	4	18	12.5		
II. Literary & Cultural Texts & Traditions; Cultural Analysis & Comparisons	MC	16	16	9.0	✓	
	CR	4	26	18.2		
III. Language & Communication: Listening; Reading; Written; Oral Expression	MC	16	16	12.0	✓	
	CR	6	38	30.0		

Table 6: Panel-Recommended Initial Passing Standards for CSET: Mandarin

Mandarin	Item Type ^{1/}	Scorable Items	Possible Score Points	Computed Median based on Panel Recommendations	Component Score Combination Rule ^{2/}	
					50/50	40/60
I. General Linguistics; Linguistics of the Target Language	MC	16	16	12.1	✓	
	CR	4	18	12.3		
II. Literary & Cultural Texts & Traditions; Cultural Analysis & Comparisons	MC	16	16	11.1	✓	
	CR	4	26	15.9		
III. Language & Communication: Listening; Reading; Written; Oral Expression	MC	16	16	12.0	✓	
	CR	6	38	25.8		

^{1/} MC = multiple-choice, CR = constructed-response

^{2/} The component score combination rule is formatted as multiple-choice percent/constructed-response percent (e.g., 40/60 is 40% multiple choice / 60% constructed response).

Table 7: Panel-Recommended Initial Passing Standards for CSET: Punjabi

Punjabi	Item Type ^{1/}	Scorable Items	Possible Score Points	Computed Median based on Panel Recommendations	Component Score Combination Rule ^{2/}	
					50/50	40/60
I. General Linguistics; Linguistics of the Target Language	MC	16	16	12.0	✓	
	CR	4	18	15.3		
II. Literary & Cultural Texts & Traditions; Cultural Analysis & Comparisons	MC	16	16	12.0	✓	
	CR	4	26	21.0		
III. Language & Communication: Listening; Reading; Written; Oral Expression	MC	16	16	13.8	✓	
	CR	6	38	32.0		

Table 8: Panel-Recommended Initial Passing Standards for CSET: Russian

Russian	Item Type ^{1/}	Scorable Items	Possible Score Points	Computed Median based on Panel Recommendations	Component Score Combination Rule ^{2/}	
					80/20	70/30
I. General Linguistics; Linguistics of the Target Language	MC	32	32	21.3		✓
	CR	3	18	14.9		
II. Literary & Cultural Texts & Traditions; Cultural Analysis & Comparisons	MC	32	32	20.1		✓
	CR	1	8	4.8		
					60/40	50/50
III. Language & Communication: Listening; Reading; Written; Oral Expression	MC	24	24	17.9	Split decision	
	CR	6	38	27.0		

^{1/} MC = multiple-choice, CR = constructed-response

^{2/} The component score combination rule is formatted as multiple-choice percent/constructed-response percent (e.g., 40/60 is 40% multiple choice / 60% constructed response).

Table 9: Panel-Recommended Initial Passing Standards for CSET: Vietnamese

Vietnamese	Item Type ^{1/}	Scorable Items	Possible Score Points	Computed Median based on Panel Recommendations	Component Score Combination Rule ^{2/}	
					50/50	40/60
I. General Linguistics; Linguistics of the Target Language	MC	16	16	11.0		✓
	CR	4	18	10.3		
II. Literary & Cultural Texts & Traditions; Cultural Analysis & Comparisons	MC	16	16	9.8		✓
	CR	4	26	16.0		
III. Language & Communication: Listening; Reading; Written; Oral Expression	MC	16	16	11.0		✓
	CR	6	38	22.0		

^{1/} MC = multiple-choice, CR = constructed-response

^{2/} The component score combination rule is formatted as multiple-choice percent/constructed-response percent (e.g., 40/60 is 40% multiple choice / 60% constructed response).

Staff Recommended Initial Passing Standards

Based on the previously approved guidelines for the establishment of CSET standards (see Appendix E), staff recommends that the Commission adopt the initial passing standards for the subtests of the CSET forms administered on November 6, 2004 that:

- are equivalent to the raw score points on the multiple-choice component and on the constructed-response component as shown in Table 10;
- are based on the component score combination rules as shown in Table 10; and
- reflect passing standards that are as equivalent as possible 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 demonstrated that they have met the subject matter requirement. Since none of the first administration of these new subtests yielded 150 examinees, a subsequent passing standard activity will be held to review the initial passing standards. 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 for each CSET subtest 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. A passing scores on each of the required subtest of the CSET in the subject area is required to meet the subject matter requirement for the credential.

If the Commission adopts the staff-recommended initial passing standards for these new language examinations as indicated in Table 10, it will mark the completion of the CSET Phase II test development, and NES will mail the official score reports for the November 2004 test administration to examinees by February 28, 2005. The CSET is offered in these languages two times per year. The next scheduled test administration of the CSET in these languages is on May 21, 2005.

Table 10: Staff-Recommended Initial Passing Standards for CSET

Subtest	Multiple-Choice Raw Score Points	Constructed-Response Raw Score Points	Component Score Combination Rule MC/CR
CSET: German			
Subtest I	18	10	70/30
Subtest II	16	4	80/20
Subtest III	15	22	50/50
CSET: Japanese			
Subtest I	10	12	40/60
Subtest II	10	16	40/60
Subtest III	11	29	40/60
CSET: Korean			
Subtest I	8	11	50/50
Subtest II	7	16	50/50
Subtest III	10	28	50/50
CSET: Mandarin			
Subtest I	10	10	50/50
Subtest II	9	14	50/50
Subtest III	10	24	50/50
CSET: Punjabi			
Subtest I	10	13	50/50
Subtest II	10	19	50/50
Subtest III	12	30	50/50
CSET: Russian			
Subtest I	19	13	70/30
Subtest II	18	5	70/30
Subtest III	16	25	50/50
CSET: Vietnamese			
Subtest I	9	10	40/60
Subtest II	8	14	40/60
Subtest III	9	20	40/60

APPENDIX A

PERFORMANCE CHARACTERISTICS AND SCORING SCALES

CSET SCORING RUBRIC: ALL LANGUAGES
THREE-POINT SCORE SCALE
 (10-15 Minute Responses)

PERFORMANCE CHARACTERISTICS

PURPOSE	The extent to which the examinee responds to 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, illegible, not in the target language or 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.

CSET SCORING RUBRIC: ALL LANGUAGES
FOUR-POINT SCORE SCALE
 (30-45 Minute Response)

PERFORMANCE CHARACTERISTICS

PURPOSE	The extent to which the examinee responds to 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, illegible, not in the target language or 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.

**CSET SCORING RUBRIC: JAPANESE, KOREAN, MANDARIN, PUNJABI, AND
VIETNAMESE ONLY
SUBTEST I CRI ASSIGNMENTS**

Domain 2 – Linguistics of the Target Language: Language Structures and Error Analysis

Assessment Structure

Candidates are presented with two exercises in which they must transform and correct sentences in the target language. First, there are 4 scorable tasks in which the candidates must make transformations. Second, there are 8 scorable tasks that are presented in sentence context in which grammatical corrections must be made and, in the case of Japanese, a rationale for the correction must be given. Additionally, candidates are required to answer the charge made in a constructed-response item that will be holistically scored.

The tasks included in each language structures assignment are as follows:

Transformation	4 sentences that require making a specified transformation — 4 tasks
Error Analysis	For Korean, Mandarin, Punjabi, Vietnamese: 4 sentences, each containing two errors, that require each sentence to be rewritten correcting the contained errors — 8 tasks
	For Japanese only: 4 sentences, each containing one error, that require each sentence to be rewritten correcting the contained error and a written rationale for each corrected error – 8 tasks

Scoring the Assignments

A "Correct" response is described by one of the following:

TASKS	NOTATION DESCRIPTION
Transformation: Tasks 1-4	Correct. The candidate's response correctly transforms the sentence as instructed, i.e., the response follows all applicable grammatical rules of the target language and accurately supplies a transformed word, phrase, or clause construction as directed.
Error Analysis: Tasks 5-12	Correct. For Korean, Mandarin, Punjabi, Vietnamese : The candidate's response acceptably corrects the syntactic or linguistic error(s) found in the sentence, i.e., the response has located and corrected the error(s) of language structure found within the sentence and has retained the original meaning.
	For Japanese only: The candidate's response, i.e., the provided rationale, correctly states the syntactic or linguistic reason(s) for the correction of the identified error contained in the sentence.

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

Assigned Score Point	Error Tabulation
3	0-2 errors
2	3-6 errors
1	7-12 errors

**CSET SCORING RUBRIC: ALL LANGUAGES
SUBTEST III FOCUSED CRI
Domain 7 – Oral Expression**

PERFORMANCE CHARACTERISTICS

- Construct connected oral discourse that communicates a message effectively.
- Construct connected oral discourse that demonstrates a wide range of vocabulary.
- Construct connected oral discourse that demonstrates idiomatic expressions.
- Construct connected oral discourse that demonstrates linguistic structures.

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	<p>The "U" (Unscorable) is assigned to a response that is unrelated to the assignment, inaudible/incomprehensible, not in the target language, or does not contain a sufficient amount of original work to score.</p>
B	<p>The "B" (Blank) is assigned to a response that is blank.</p>

**CSET SCORING RUBRIC: ALL LANGUAGES
SUBTEST III FOCUSED CRI
Domain 8 – Written Expression**

PERFORMANCE CHARACTERISTICS

- Compose a well-organized passage in the target language employing styles and levels of diction appropriate for a given audience.
- Compose a well-organized passage in the target language employing styles and levels of diction appropriate for a given purpose.
- Compose a well-organized passage in the target language employing styles and levels of diction appropriate for a given occasion.
- Compose a well-organized passage in the target language demonstrating command of a wide range of vocabulary.
- Compose a well-organized passage in the target language demonstrating command of idiomatic expressions.
- Compose a well-organized passage in the target language demonstrating command of linguistic structures.

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	<p>The "U" (Unscorable) is assigned to a response that is unrelated to the assignment, illegible, not in the target language, or does not contain a sufficient amount of original work to score.</p>
B	<p>The "B" (Blank) is assigned to a response that is blank.</p>

CSET SCORING RUBRIC: ALL LANGUAGES
SUBTEST III Extended CRI
Domain 8 – Written Expression

PERFORMANCE CHARACTERISTICS

- Compose a well-organized passage in the target language employing styles and levels of diction appropriate for a given audience.
- Compose a well-organized passage in the target language employing styles and levels of diction appropriate for a given purpose.
- Compose a well-organized passage in the target language employing styles and levels of diction appropriate for a given occasion.
- Compose a well-organized passage in the target language demonstrating command of a wide range of vocabulary.
- Compose a well-organized passage in the target language demonstrating command of idiomatic expressions.
- Compose a well-organized passage in the target language demonstrating command of linguistic structures.

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	<p>The "U" (Unscorable) is assigned to a response that is unrelated to the assignment, illegible, not in the target language, or does not contain a sufficient amount of original work to score.</p>
B	<p>The "B" (Blank) is assigned to a response that is blank.</p>

APPENDIX B

**CSET:
NOVEMBER 6, 2004 TEST ADMINISTRATION**

Appendix B

CSET: NOVEMBER 6, 2004 TEST ADMINISTRATION NUMBERS OF EXAMINEES BY SUBTEST

SUBTEST	EXAMINEES PER SUBTEST	EXAMINEES TAKING ALL SUBTESTS
CSET: Languages Other Than English: German		
I: General Linguistics; Linguistics of the Target Language	6	6
II: Literary and Cultural Texts and Traditions; Cultural Analysis and Comparisons	6	
III: Language and Communication: Listening Comprehension; Reading Comprehension; Written Expression; Oral Expression	7	
CSET: Languages Other Than English: Japanese		
I: General Linguistics; Linguistics of the Target Language	17	12
II: Literary and Cultural Texts and Traditions; Cultural Analysis and Comparisons	13	
III: Language and Communication: Listening Comprehension; Reading Comprehension; Written Expression; Oral Expression	15	
CSET: Languages Other Than English: Korean		
I: General Linguistics; Linguistics of the Target Language	5	4
II: Literary and Cultural Texts and Traditions; Cultural Analysis and Comparisons	4	
III: Language and Communication: Listening Comprehension; Reading Comprehension; Written Expression; Oral Expression	4	
CSET: Languages Other Than English: Mandarin		
I: General Linguistics; Linguistics of the Target Language	9	5
II: Literary and Cultural Texts and Traditions; Cultural Analysis and Comparisons	8	
III: Language and Communication: Listening Comprehension; Reading Comprehension; Written Expression; Oral Expression	10	
CSET: Languages Other Than English: Punjabi		
I: General Linguistics; Linguistics of the Target Language	3	2
II: Literary and Cultural Texts and Traditions; Cultural Analysis and Comparisons	3	
III: Language and Communication: Listening Comprehension; Reading Comprehension; Written Expression; Oral Expression	2	

**CSET: NOVEMBER 6, 2004 TEST ADMINISTRATION
 NUMBERS OF EXAMINEES BY SUBTEST
 (CONTINUED)**

SUBTEST	EXAMINEES PER SUBTEST	EXAMINEES TAKING ALL SUBTESTS
CSET: Languages Other Than English: Russian		
I: General Linguistics; Linguistics of the Target Language	1	1
II: Literary and Cultural Texts and Traditions; Cultural Analysis and Comparisons	1	
III: Language and Communication: Listening Comprehension; Reading Comprehension; Written Expression; Oral Expression	1	
CSET: Languages Other Than English: Vietnamese		
I: General Linguistics; Linguistics of the Target Language	1	1
II: Literary and Cultural Texts and Traditions; Cultural Analysis and Comparisons	1	
III: Language and Communication: Listening Comprehension; Reading Comprehension; Written Expression; Oral Expression	1	

APPENDIX C

STANDARD SETTING PANEL DEMOGRAPHICS

Appendix C
CSET STANDARD SETTING PANELS

	German	Japanese	Korean	Mandarin	Punjabi	Russian	Vietnamese	Total
Total Number								
Appointed	10	10	8	9	5	7	3	52
Participated	5	10	4	9	4	6	3	41
Ethnicity								
African American	0	0	0	0	0	0	0	0
Asian	0	8	4	9	0	0	2	23
Hispanic	0	0	0	0	0	0	0	0
White	4	1	0	0	0	6	0	11
Other/Not Provided	1	1	0	0	4	0	1	7
Sex								
Female	4	8	4	5	4	4	2	31
Male	1	2	0	4	0	2	1	10
Region								
North	3	5	0	5	4	2	1	20
South	2	5	4	4	0	4	2	21
Profession								
Public School Educators	4	7	2	5	4	2	1	25
College/University Educators	1	3	2	4	0	4	2	16
Years of Experience								
0-6	1	1	1	0	3	0	1	7
7-10	1	4	0	1	0	1	1	8
11+	3	4	3	8	1	5	1	25
Not Provided	0	1	0	0	0	0	0	1

APPENDIX D

DETAILED STANDARD SETTING PROCESS

Appendix D

Standard Setting Rating Tasks

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 2004 test administration;
- the accompanying subtest form answer keys;
- the Round One Rating Form for multiple-choice items; and
- if appropriate, the item statistics displaying the percent of examinees who answered each test item correctly (i.e., for tests in which 2 or more examinees took all subtests).

Round One began with a set of approximately ten 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 the following question.

Imagine a hypothetical group of candidates for the Single Subject Teaching Credential in (INSERT FIELD NAME), 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. The group also reviewed item statistics (p-values) on each practice test item, where applicable, which provided an indicator of the difficulty level of the item.

Following the practice set, panel members began the same rating process with the multiple-choice items used on the November 6, 2004 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 2004 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; 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 two 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 (INSERT FIELD NAME) 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 6, 2004 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 panels 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. Additionally, panel members were asked to provide 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 description;

- Round One Multiple-Choice Item Rating Summary Sheet, which provided the sum of the median rating for each item 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 each item 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 typically 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 single, holistic, Round Two multiple-choice component “cut score” recommendation for each subtest, representing the total number of scorable items at the subtest level that would, in their judgment, be answered correctly by the just-acceptable candidate. To make this recommendation, panel members responded 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 (INSERT FIELD NAME) 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 typically 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 single, holistic, Round Two constructed-response component “cut score” recommendation for each subtest, representing the total number of points at the subtest level that would, in their judgment, be earned by the just-acceptable candidate. To make this recommendation, panel members responded 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 (INSERT FIELD NAME) 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

Panels were provided 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 panels members are intended to yield reliable results and are psychometrically defensible.

For German and Russian 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%. For Subtest III, panel members were given the following two alternatives for allocating points: (a) multiple-choice component 60% and the constructed-response component 40% or (b) the multiple-choice component 50% and the constructed-response component 50%.

For all Japanese, Korean, Mandarin, Punjabi, and Vietnamese subtests, panel members were given the following two alternatives for allocating points: (a) multiple-choice component 60% and the constructed-response component 40% or (b) the multiple-choice component 50% and the constructed-response component 50%.

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:

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*

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;
- the Round Three Subtest Standard Setting Recommendation Form;
- Summary Statistics Report for subtests taken by 15 or more examinees, which included the following for each subtest:
 - Descriptive (demographic) information characterizing the sample of examinees who took the first test administration; and
 - A set of analyses showing in tabular form the percent of examinees from the first test administration who would pass each subtest, given possible multiple-choice component and constructed-response component raw score combinations and each selected component combination rule (e.g., 80–20 and 70–30).

(Note: these reports were only provided to the Japanese panel, as their subtests were the only subtests to meet this threshold)

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 and measurement error.

Panels 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 much 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 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 (INSERT FIELD NAME) 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%-20% or 70%-30% used as an example):

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

OR

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

As the final step to the standard setting studies, 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 standards setting panels for use in the determination of the staff-recommended passing standards presented in this report.

APPENDIX E

CSET STANDARDS SETTING CONSIDERATIONS

CSET Standards 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: German, Japanese, Korean, Mandarin, Punjabi, Russian and Vietnamese, staff considered the following factors and options that affect the standard setting process in determining the staff-recommended passing standards. These are the same considerations used for all examinations in the CSET testing program.

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 Panels considered this model when determining the subtest structures of each examination.

Professional Judgments

The recommended passing standards the CSET are based upon the professional judgments provided by the members of the Subject Matter Advisory Panels. Since these panel recommendations are criterion-referenced—based on expert judgment of the minimum required subject matter knowledge for beginning teachers—any examinee performance data provides supplemental, though not necessary, information.

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 accurate. However, the standard error allows us to determine a range within which the examinee's 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.