
7B

Information/Action

Professional Services Committee

Recommended Initial Passing Standards for the California Subject Examinations for Teachers (CSET): Agriculture, Business, Health Science, Home Economics, and Industrial and Technology Education

Executive Summary: This report provides the Commission with recommendations, based on the informed judgments of California educators, for passing standards for the CSET: Agriculture, Business, Health Science, Home Economics, and Industrial and Technology Education.

Recommended Action: That the Commission act on the proposed passing standards for the CSET examinations in Agriculture, Business, Health Science, Home Economics, and Industrial and Technology Education.

Presenters: Amy Jackson, Administrator, Professional Services Division and Dr. John Mattar, 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.

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 Examinations for Teachers (CSET): Agriculture, Business, Health Science, Home Economics, and Industrial and Technology Education

Introduction

This report describes the standard setting studies for the California Subject Examinations for Teachers (CSET): Agriculture, Business, Health Science, Home Economics, and Industrial and Technology Education and provides recommendations for the adoption of initial passing standards for each examination.

Background

In spring 2004, the Commission's Executive Director appointed subject matter advisory panels for the California Subject Examinations for Teachers (CSET) for the single subject areas of agriculture, business, health science, home economics, and industrial and technology education. The purpose of the panel was 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, all with a specialty in the subject area of the panel.

From spring through fall of 2004, each panel developed subject matter requirements (SMRs) for their specific subject area that were aligned with available state student content standards and curriculum frameworks, and standards of national professional organizations. Staff from the California Department of Education (CDE) participated in all of the panel development activities. The SMRs specify the content that is to be taught in Commission-approved subject matter preparation programs and constitute the test specifications for the subject matter examinations. The Commission approved those SMRs in its January-February 2005 meeting.

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 areas. Those program standards will be presented to the Commission for consideration at a future meeting.

Following the adoption of the SMRs, the advisory panels worked with NES to develop the examinations. 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.

On September 10, 2005, the first test administrations of these new examinations in Agriculture, Business, Health Science, Home Economics, and Industrial and Technology Education were conducted. On October 17-19, 2005, the standard setting studies for these examinations were held in Sacramento to determine the initial passing standard recommendations of California educators.

The CSET for Languages Other Than English: American Sign Language will be administered for the first time in November 2005. A similar standard setting study will be conducted for this examination in December, and staff recommended initial passing standards will be brought to the Commission at its January-February 2006 meeting.

The CSET: Agriculture, Business, Health Science, Home Economics, and Industrial and Technology Education

Each of the five new tests in the CSET program is comprised of subtests differentiated by content area. Each of these examinations are paper-and-pencil tests that 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 scoring scale. Constructed-response performance characteristics and scoring scales are provided in Appendix A. Test structures for the CSET in Agriculture, Business, Health Science, Home Economics, and Industrial and Technology Education are shown in Tables 1-5 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 examinations for Agriculture, Business, Health Science, Home Economics, and Industrial and Technology Education will be administered four times each year. The numbers of examinees who completed¹ subtests at the first test administration of the tests in September are provided in Appendix B. On October 17-19, 2005, Commission staff and NES conducted standard setting studies for the new examinations. The standard-setting procedures used and the results of these studies are described in Part II of this report.

¹ 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: Agriculture

Subtest	Domains	Number of Multiple-Choice Items	Number of Constructed-Response Items
I	Plant and Soil Science	25	2 short (focused)
	Ornamental Horticulture	15	1 short (focused)
	<i>Subtest total</i>	40	3 short (focused)
II	Animal Science	25	2 short (focused)
	Environmental Science and Natural Resource Management	15	1 short (focused)
	<i>Subtest total</i>	40	3 short (focused)
III	Agricultural Business and Economics	20	2 short (focused)
	Agricultural Systems Technology	20	1 short (focused)
	<i>Subtest total</i>	40	3 short (focused)
	Total Items	120	9 short (focused)

Table 2: Subtest Structure of the CSET: Business

Subtest	Domains	Number of Multiple-Choice Items	Number of Constructed-Response Items
I	Business Management	20	1 extended
	Marketing	20	1 short (focused)
	<i>Subtest total</i>	40	1 extended 1 short (focused)
II	Accounting and Finance	25	1 short (focused)
	Economics	15	1 short (focused)
	<i>Subtest total</i>	40	2 short (focused)
III	Information Technology	25	1 short (focused)
	Business Environment and Communication	15	1 extended
	<i>Subtest total</i>	40	1 short (focused) 1 extended
	Total Items	120	4 short (focused) 2 extended

Table 3: Subtest Structure of the CSET: Health Science

Subtest	Domains	Number of Multiple-Choice Items	Number of Constructed-Response Items
I	Foundations of Health Education	10	1 extended
	Human Growth and Development	10	none
	Chronic and Communicable Diseases	20	1 short (focused)
	<i>Subtest total</i>	40	<i>1 extended</i> <i>1 short (focused)</i>
II	Nutrition and Fitness	15	1 short (focused)
	Mental and Emotional Health	10	none
	Alcohol, Tobacco, and Other Drugs	15	1 extended
	<i>Subtest total</i>	40	<i>1 short (focused)</i> <i>1 extended</i>
III	Family Life and Interpersonal Relationships	15	1 short (focused)
	Consumer and Community Health	15	none
	Environmental Health	10	none
	<i>Subtest total</i>	40	<i>1 short (focused)</i>
	Total Items	120	3 short (focused) 2 extended

Table 4: Subtest Structure of the CSET: Home Economics

Subtest	Domains	Number of Multiple-Choice Items	Number of Constructed-Response Items
I	Personal, Family, and Child Development	40	1 extended
	<i>Subtest total</i>	40	<i>1 extended</i>
II	Nutrition, Foods, and Hospitality	40	2 short (focused)
	<i>Subtest total</i>	40	<i>2 short (focused)</i>
III	Fashion and Textiles	12	1 short (focused)
	Housing and Interior Design	12	1 short (focused)
	Consumer Education	16	2 short (focused)
	<i>Subtest total</i>	40	<i>4 short (focused)</i>
	Total Items	120	1 extended 6 short (focused)

Table 5: Subtest Structure of the CSET: Industrial and Technology Education

Subtest	Domains	Number of Multiple-Choice Items	Number of Constructed-Response Items
I	Nature of Technology	45	2 short (focused) 1 extended
	<i>Subtest total</i>	45	2 short (focused) 1 extended
II	Power and Energy	25	1 short (focused)
	Information and Communication	25	1 short (focused)
	Project and Product Development	25	1 short (focused)
	<i>Subtest total</i>	75	3 short (focused)
	Total Items	120	5 short (focused) 1 extended

The Standard Setting Studies

Standard setting studies for the new examinations of the CSET program were conducted October 17-19, 2005 with independent panels for each subject area (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: Agriculture, Business, Health Science, Home Economics, and Industrial and Technology Education. A total of 61 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.

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 September 2005 test administration, and item statistics displaying the percent of examinees who answered each test item correctly (for tests with a minimum of 20 examinees). 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 the subject area 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 structures. This process is briefly described below. A detailed description of the process is found in Appendix D.

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 or 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”. To aid in their discussions, panelists were provided with additional data for examination sections that were taken by a minimum of 20 candidates. This information included the results of the component-level statistics generated from Round Two, applicable examinee demographic information for fields of 20 or more candidates, 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.

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 6-10 below.

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

Agriculture	Item Type ^{1/}	Scorable Items	Possible Score Points	Computed Median Based on Panel Recommendations	Component Score Combination Rule ^{2/}	
					80/20	70/30
I. Plant and Soil Science; Ornamental Horticulture	MC	40	32	24.0		✓
	CR	3	18	14.3		
II. Animal Science; Environmental Science and Natural Resource Management	MC	40	32	24.2		✓
	CR	3	18	14.3		
III. Agricultural Business and Economics; Agricultural Systems Technology	MC	40	32	24.2		✓
	CR	3	18	14.3		

^{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).

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

Business	Item Type ^{1/}	Scorable Items	Possible Score Points	Computed Median Based on Panel Recommendations	Component Score Combination Rule ^{2/}	
					80/20	70/30
I. Business Management; Marketing	MC	40	32	22.9		✓
	CR	2	14	10.7		
II. Accounting and Finance; Economics	MC	40	32	22.9		✓
	CR	2	12	9.9		
III. Information Technology; Business Environment and Communication	MC	40	32	23.1		✓
	CR	2	14	10.9		

^{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).

Table 8: Panel-Recommended Initial Passing Standards for CSET: Health Science

Health Science	Item Type ^{1/}	Scorable Items	Possible Score Points	Computed Median based on Panel Recommendations	Component Score Combination Rule ^{2/}	
					80/20	70/30
I. Foundations of Health Education; Human Growth and Development; Chronic and Communicable Diseases	MC	40	32	22.2		✓
	CR	2	14	10.6		
II. Nutrition and Fitness; Mental and Emotional Health; Alcohol, Tobacco, and Other Drugs	MC	40	32	24.1	✓	
	CR	2	14	11.1		
III. Family Life and Interpersonal Relationships; Consumer and Community Health; Environmental Health	MC	40	32	23.4	✓	
	CR	1	6	4.3		

^{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).

Table 9: Panel-Recommended Initial Passing Standards for CSET: Home Economics

Home Economics	Item Type ^{1/}	Scorable Items	Possible Score Points	Computed Median based on Panel Recommendations	Component Score Combination Rule ^{2/}	
					80/20	70/30
I. Personal, Family, and Child Development	MC	40	32	24.3		✓
	CR	1	8	6.1		
II. Nutrition, Foods, and Hospitality	MC	40	32	24.9		✓
	CR	2	12	9.1		
III. Fashion and Textiles; Housing and Interior Design; Consumer Education	MC	40	32	23.4		✓
	CR	4	24	18.1		

Table 10: Panel-Recommended Initial Passing Standards for CSET: Industrial and Technology Education

Industrial and Technology Education	Item Type ^{1/}	Scorable Items	Possible Score Points	Computed Median based on Panel Recommendations	Component Score Combination Rule ^{2/}	
					80/20	70/30
I. Nature of Technology	MC	45	36	23.8		✓
	CR	3	20	13.5		
II. Power and Energy; Information and Communication; Project and Product Development	MC	75	60	37.9		✓
	CR	3	18	12.0		

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

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

Initial Passing Standards Recommendations

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 September 10, 2005 that:

- are equivalent to the raw score points on the multiple-choice component and on the constructed-response component as shown in Table 11;
- are based on the component score combination rules as shown in Table 11; 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 tests for Agriculture, Business, Home Economics, and Industrial and Technology Education yielded fewer than 150 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 recommended initial passing standards, as indicated in Table 11, NES will release score reports for the September 2005 test administration by December 31, 2005. The next test administration of the CSET in these subject areas is scheduled for January 21, 2006.

Table 11: Staff-Recommended Initial Passing Standards for CSET*

*Adjusted by -1 Standard Error of Measurement on the Multiple-Choice Raw Score Points

Subtest	Multiple-Choice Raw Score Points	Constructed-Response Raw Score Points	Component Score Combination Rule MC/CR	Passing Rate for September 10, 2005 Test Administration by Subtest	Overall Passing Rate for September 10, 2005 Test Administration
CSET: Agriculture					29%
Subtest I	22	14	70/30	43%	
Subtest II	22	14	70/30	53%	
Subtest III	22	14	70/30	50%	
CSET: Business					33%
Subtest I	20	11	70/30	66%	
Subtest II	20	10	70/30	52%	
Subtest III	21	11	70/30	69%	
CSET: Health Science					32%
Subtest I	20	11	70/30	32%	
Subtest II	22	11	80/20	72%	
Subtest III	21	4	80/20	83%	
CSET: Home Economics					63%
Subtest I	22	6	70/30	83%	
Subtest II	22	9	70/30	91%	
Subtest III	21	18	70/30	82%	
CSET: Industrial and Technology Education					73%
Subtest I	21	14	70/30	79%	
Subtest II	35	12	70/30	73%	

APPENDIX A

PERFORMANCE CHARACTERISTICS AND SCORING SCALES

STANDARD CSET SCORING RUBRIC
THREE-POINT SCORE SCALE
 (10-15 Minute Responses)

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, illegible, primarily in a language other than 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.

STANDARD CSET SCORING RUBRIC
FOUR-POINT SCORE SCALE
(30-45 Minute Response)

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, illegible, primarily in a language other than 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.

APPENDIX B

CSET:

SEPTEMBER 10, 2005 TEST ADMINISTRATION

Appendix B

CSET: SEPTEMBER 10, 2005 TEST ADMINISTRATION NUMBERS OF EXAMINEES BY SUBTEST

SUBTEST	EXAMINEES PER SUBTEST	EXAMINEES TAKING ALL SUBTESTS
CSET: Agriculture		
I: Plant and Soil Science; Ornamental Horticulture	14	14
II: Animal Science; Environmental Science and Natural Resource Management	17	
III: Agricultural Business and Economics; Agricultural Systems Technology	14	
CSET: Business		
I: Business Management; Marketing	53	40
II: Accounting and Finance; Economics	48	
III: Information Technology; Business Environment and Communication	48	
CSET: Health Science		
I: Foundations of Health Education; Human Growth and Development; Chronic and Communicable Diseases	148	125
II: Nutrition and Fitness; Mental and Emotional Health; Alcohol, Tobacco, and Other Drugs	147	
III: Family Life and Interpersonal Relationships; Consumer and Community Health; Environmental Health	136	
CSET: Home Economics		
I: Personal, Family, and Child Development	12	8
II: Nutrition, Foods, and Hospitality	11	
III: Fashion and Textiles; Housing and Interior Design; Consumer Education	11	
CSET: Industrial and Technology Education		
I: Nature of Technology	14	11
II: Power and Energy; Information and Communication; Project and Product Development	11	

APPENDIX C

STANDARD SETTING PANEL DEMOGRAPHICS

Appendix C
CSET STANDARD SETTING PANELS

	Agriculture	Business	Health Science	Home Economics	Industrial and Technology Education	Total
Total Number						
Appointed	14	18	16	19	19	86
Participated	11	13	10	13	14	61
Ethnicity						
African American	0	1	0	0	0	1
Asian	0	1	1	1	0	3
Hispanic	1	1	1	0	2	5
White	8	9	6	8	9	40
Other/Not Provided	2	1	2	4	3	12
Sex						
Female	1	10	7	13	1	32
Male	10	3	3	0	13	29
Region						
North	5	1	4	5	3	18
South	6	12	6	8	11	43
Profession						
Public School Educators	9	12	6	11	11	49
College/University Educators	2	1	4	2	3	12
Years of Experience						
0-6	1	2	2	1	0	6
7-10	2	2	2	1	4	11
11+	5	8	6	10	9	38
Not Provided	3	1	0	1	1	6

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 September 2005 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 20 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 to one of the following questions, depending on the type of panel.

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 related to the percentage of candidates who answered each question correctly (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 September 10, 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 September 2005 test administration;
- the appropriate set of performance characteristics and scoring scale;
- the Subject Matter Advisory Panel-approved marker responses¹ 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 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.

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 September 10, 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 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.

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 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 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 (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 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 (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.

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:

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;
- Round Three Subtest Standard Setting Recommendation Form;
- Summary Statistics Report for subtests taken by 20 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).

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.

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 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 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% 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: Agriculture, Business, Health Science, Home Economics, and Industrial and Technology Education, 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 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 Standard Setting Panels. Since these panel recommendations are criterion-referenced—based on expert judgment of the minimum required subject matter knowledge for beginning teachers—examinee performance data provides supplemental, though not necessary, information. Performance data is provided to inform those judgments when there are at least 20 examinees.

Standard Error of Measurement

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