

*California
Commission on Teacher Credentialing*

*Meeting of
April 2-3, 2003*

AGENDA ITEM NUMBER: PERF - 1

COMMITTEE: Performance Standards Committee

TITLE: Recommended Passing Standards for the California Subject Examinations for Teachers (CSET): Multiple Subjects, English, Mathematics, Science, and Social Science

 X Action

 Information

 Report

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.

Presented By: Dr. Nicole A. Amador

Prepared By: _____

Nicole A. Amador, Ph.D.
Consultant, Professional Services Division

Date: _____

Approved By: _____

Amy Jackson
Administrator, Professional Services Division

Date: _____

Approved By: _____

Mary Vixie Sandy
Director, Professional Services Division

Date: _____

Authorized By: _____

Dr. Sam W. Swofford
Executive Director

Date: _____

**Recommended Passing Standards for
California Subject Examinations for Teachers (CSET): Multiple Subjects,
English, Mathematics, Science, and Social Science**

Professional Services Division

April 2-3, 2003

Executive Summary

The purpose of the standard setting studies is to provide the Commission with recommendations, based on the informed judgments of California educators, relevant to the determination of passing standards for the CSET: Multiple Subjects, English, Mathematics, Science, and Social Science.

The CSET was first administered on January 25, 2003, and in March 2003, Commission staff, and National Evaluation Systems, Inc. (NES), conducted standard setting studies for the examinations in multiple subjects (for the Multiple Subject Teaching Credential), English, mathematics, science, and social science. This report describes the standard setting studies and the results of the studies, and provides staff-recommended passing standards.

Policy Issue(s) to be Considered

What passing standards should be established for each of the examinations of the CSET: Multiple Subjects, English, Mathematics, Science, and Social Science?

Fiscal Impact Statement

National Evaluation Systems, Inc. (NES) is developing the CSET at no cost to the Commission; the contractor will be compensated directly from examinee fees. This test development work includes the standard setting studies described in this report.

Recommendation(s)

Staff recommends that the Commission adopt the proposed passing standards for each of the examinations of the CSET: Multiple Subjects, English, Mathematics, Science, and Social Science. These recommended passing standards are on pages 22 and 23 of this report.

Recommended Passing Standards for California Subject Examinations for Teachers (CSET): Multiple Subjects, English, Mathematics, Science, and Social Science

Professional Services Division

April 2-3, 2003

Overview of the Report

This report describes the standard setting studies for the California Subject Examinations for Teachers (CSET): Multiple Subjects, English, Mathematics, Science, and Social Science, and provides staff-recommended passing standards for each examination. Part I provides background information on the development of the CSET. Part II describes the standard setting procedures used and the results of the studies. Part III includes staff recommendations for action.

Part I: Background Information

The Subject Matter Competence Requirement for a Teaching Credential

Teacher candidates in California are required to demonstrate competence in the subject matter they will be authorized to teach. Candidates have two options available for satisfying this requirement. They can either complete a Commission-approved subject matter preparation program or they can pass the appropriate Commission-adopted subject matter examination(s). Because they satisfy the same requirement, these two options are as aligned and congruent as possible.

In the early 1990s, the Commission developed and adopted (a) standards for subject matter preparation programs and, at the same time, (b) specifications for the subject matter examinations. This work was based on the advice of subject matter advisory panels and data from validity studies, and resulted in program standards and examination specifications (defining the subject matter competence requirement).

The validity of the subject matter competence requirement (i.e., program standards and examination specifications) is not permanent, however. The need for periodic validity studies of the subject matter requirement is directly related to one of the Commission's most fundamental missions: to provide a strong assurance that teaching credentials are awarded to individuals who have learned the most important knowledge, skills, and abilities that are actually needed in order to succeed in California public school teaching positions. The validity of the examination specifications and program standards used by the Commission has been established in conjunction with their initial development. Professional practice and legal defensibility require, however, that the validity of these policies be periodically re-established, as job requirements and expectations may change over time.

In the late 1990s, the State Board of Education adopted K-12 student content standards in English, mathematics, science, and social science. These new standards have obvious and direct implications for the subject matter competence requirement of prospective teachers. This was recognized in SB 2042 (Alpert, 1998), which requires the Commission to ensure that subject matter program

standards and examinations are aligned with the K-12 student content standards adopted by the State Board.

Development of Content Specifications for the Multiple Subject Teaching Credential

In 1998, the Commission authorized the Executive Director to establish a panel of elementary school teachers, principals, curriculum specialists, teacher educators, and college faculty members to advise the Commission on content specifications for the examination and related program standards for the Multiple Subject Teaching Credential. The Elementary Subject Matter Advisory Panel was formed, consisting of twenty-six members with expertise in the seven subject areas required by Education Code Section 44282: history and social science; human development; mathematics; physical education; reading, language, and literature; science; and the visual and performing arts.

The development of the content specifications was a three-phase process. Phase I, conducted by WestEd and MPR Associates, Inc., consisted of a job analysis to determine the knowledge, understanding, and skills important for competent teaching in K-8 classrooms in California. In Phase II, the Elementary Subject Matter Advisory Panel drafted content specifications in the seven subject matter areas. These draft specifications were a series of statements that reflect the level of knowledge and application expected of candidates for the Multiple Subject Teaching Credential. Phase III, conducted by American Institutes for Research (AIR), was a study designed to provide content-oriented evidence of the validity of the draft content specifications by demonstrating the relationship between the specifications and the jobs of effective, entry-level teachers in grades K-8 in California's public schools.

The recommended draft specifications were reviewed by an independent panel of educators for alignment to the state-adopted student academic content standards, by the Commission's Bias Review Committee for issues of potential bias, and by representatives of the Commission and State Board of Education to ensure consistency with other significant policy reforms. The Elementary Subject Matter Advisory Panel finalized the draft content specifications in light of the results of the validity study and these reviews. In September 2001, the Commission formally adopted the content specifications as part of the *Standards of Program Quality and Effectiveness for the Subject Matter Requirement for the Multiple Subject Teaching Credential*. This document is available at: http://www.ctc.ca.gov/aboutctc/agendas/september_2001/PREP_3_SEP2001.pdf.

Development of Subject Matter Requirements for Single Subject Teaching Credentials in English, Mathematics, Science, and Social Science

In January 2001, the Executive Director appointed subject matter advisory panels in English, mathematics, science, and social science to advise Commission staff on the development of new subject matter program standards and examinations in these subject areas. Each panel consisted of:

- classroom teachers of the subject area,
- subject area specialists in school districts, county offices of education, and postsecondary institutions,
- professors in the subject area teaching in subject matter preparation programs,
- teacher educators,
- members of relevant professional organizations,
- members of other relevant committees and advisory panels, and
- a liaison from the California Department of Education.

With leadership from Commission staff and assistance from AIR staff, the subject matter advisory panels developed draft subject matter requirements (SMRs), the subject-specific knowledge, skills, and abilities needed by beginning teachers of English, mathematics, science, and social science. These preliminary sets of SMRs included the following eight content areas:

- English
- Mathematics
- Social Science
- Science, including
 - General Science
 - Biology
 - Chemistry
 - Earth and Planetary Science (Geoscience)
 - Physics

In January 2002, AIR launched a statewide survey-based validity study of the preliminary SMRs developed by the panels. The purpose of this study was to provide content-oriented evidence of the validity of the draft SMRs by demonstrating the relationship between them and the jobs of effective, entry-level teachers of these subject areas in California's public schools.

The preliminary SMRs were independently reviewed by two separate groups. An Alignment and Congruence Committee reviewed the SMRs for alignment with the state-adopted student content standards. The Commission's Bias Review Committee reviewed the SMRs for issues of potential bias. Changes suggested by these two groups were presented to and acted on by the appropriate subject matter advisory panel.

The results of the validation study and the reviews for alignment and congruence and potential bias informed the subject matter advisory panels and staff in their recommendations to the Commission. In May of 2001, the Commission adopted Subject Matter Requirements for English, Mathematics, Science, and Social Science (http://www.ctc.ca.gov/aboutctc/agendas/june_2002/June_2002_PERF-1.pdf).

The California Subject Examinations for Teachers (CSET): Multiple Subjects, English, Mathematics, Science, and Social Science

Given the new subject matter requirements for the Multiple Subject Teaching Credential and Single Subject Teaching Credentials in English, Mathematics, Science, and Social Science, the Commission implemented a new examination program called the California Subject Examinations for Teachers (CSET). The development of the CSET began in the spring of 2002. Multiple-choice and constructed-response items were drafted based on the subject matter requirements, and 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 subject matter advisory 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. These guides as well as Internet registration and other CSET information is available at www.cset.nesinc.com. The CSET program will, over the next four years, take the place of the current SSAT and Praxis II subject matter examinations. The CSET is developed and administered by National Evaluation Systems, Inc. (NES®).

Each of the five new tests in the CSET program is comprised of subtests differentiated by content area. The paper-and-pencil tests consist of both multiple-choice and constructed-response items. Constructed-response items are of two types: *extended* constructed-responses items that are scored using a four-point scale, and *focused* constructed-response items that are scored using a three-point scoring scale. Refer to the Appendix A for the constructed-response performance characteristics and scoring scales. Test structures for the CSET in Multiple Subjects, English, Mathematics, Science, and Social Science, are shown in Tables 1-5 below.

Table 1: Subtest Structure of the CSET: Multiple Subjects

Subtest	Number of Multiple-Choice Items per Test Form	Number of Constructed-Response Items per Test Form
I: Reading, Language, and Literature; History and Social Science	52 (26 in each area)	4 (2 in each area; focused)
II: Science; Mathematics	52 (26 in each area)	4 (2 in each area; focused)
III: Physical Education; Human Development; Visual and Performing Arts	39 (13 in each area)	3 (1 in each area; focused)
Total Items	143	11

Each CSET testing session is five hours in length. Examinees can choose to take any one or all subtests within a single testing session. Individual subtests are not timed. The CSET in Multiple Subjects, English, Mathematics, Science, and Social Science will be administered six times each year, the initial administration of which occurred on January 25, 2003. The numbers of examinees who completed¹ subtests at the first test administration, are provided in Appendix B. The week of March 3, 2003, Commission staff and NES conducted standard setting studies for each new examination in the CSET program. 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 2: Subtest Structure of the CSET: English

Subtest	Number of Multiple-Choice Items per Test Form	Number of Constructed-Response Items per Test Form
I: Literature and Textual Analysis; Composition and Rhetoric	50	none
II: Language, Linguistics, and Literacy	50	none
III: Composition and Rhetoric; Literature and Textual Analysis	none	2 (1 based on literary text, 1 based on non-literary text: extended)
IV: Communications: Speech, Media, and Creative Performance	none	4 (focused)
Total Items	100	6

Note. Subtest I tests the two domains of (a) Literature and Textual Analysis and (b) Composition and Rhetoric separately, using multiple-choice items. Subtest III tests the two domains in an integrated fashion, using constructed-response items.

Table 3: Subtest Structure of the CSET: Mathematics

Subtest	Number of Multiple-Choice Items per Test Form	Number of Constructed-Response Items per Test Form
I: Algebra; Number Theory	30 (inc. 10 “enhanced” items)	4 (3 in Algebra, 1 in Number Theory; extended)
II: Geometry; Probability and Statistics	30 (inc. 10 “enhanced” items)	4 (3 in Geometry, 1 in Probability and Statistics; extended)
III: Calculus; History of Mathematics	30 (inc. 10 “enhanced” items)	4 (3 in Calculus, 1 in History of Mathematics; extended)
Total Items	90	12

Note. “Enhanced” multiple-choice items are more complex items requiring 2.5 minutes each on average. Examinees will be able to use graphing calculators while taking Subtest II. No calculators will be allowed for Subtests I and III.

Table 4: Subtest Structure of the CSET: Science

Subtest	Number of Multiple-Choice Items per Test Form	Number of Constructed-Response Items per Test Form
I: General Science: Astronomy; Dynamic Processes of the Earth; Earth Resources; Waves; Forces and Motion; Electricity and Magnetism	58	2 (focused)
II: General Science: Ecology; Genetics and Evolution; Molecular Biology and Biochemistry; Cell and Organismal Biology; Heat Transfer and Thermodynamics; Structure and Properties of Matter	58	2 (focused)
III: Concentration: Biology, Chemistry, Earth and Planetary Science, or Physics	50	3 (focused)
Total Items	166	7

Note. A candidates must complete all parts of Subtests I and II, but need only complete Subtest III in his/her area of concentration.

Table 5: Subtest Structure of the CSET: Social Science

Subtest	Number of Multiple-Choice Items per Test Form	Number of Constructed-Response Items per Test Form
I: World History; Geography	39	1 (World History; extended) 2 (1 in World History, 1 in Geography; focused)
II: U.S. History; Geography	39	1 (U.S. History; extended) 2 (1 in U.S. History, 1 in Geography; focused)
III: Civics; Economics; California History	40	3 (1 in each domain; focused)
Total Items	118	9

Part II: Standard Setting Studies

Standard setting studies for the new examinations of the CSET program were conducted March 4-7, 2003 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 passing standards for the CSET: Multiple Subjects, English, Mathematics, Science, and Social Science. A total of 115 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.

Standard setting studies began with an orientation and training session, during which panel members were 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 to successfully satisfy the subject matter requirement.

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 provided with a copy of the test for their field that was administered in January 2003. Under test-like conditions, panel members were asked to read and answer each item independently, and then to score their own performance on the multiple-choice items.

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. The subtests of the CSET program typically combine results from both multiple-choice and constructed-response components. The exception is the English examination; its multiple-choice and constructed-response items appear on separate subtests. The three rounds of standard setting procedures are described in further detail below.

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:

- content specifications / subject matter requirements;
- the subtest forms used for the January 2003 test administration;
- the accompanying subtest form answer keys;
- round One Rating Form for multiple-choice items; and
- item statistics displaying the percent of examinees who answered each test item correctly, where appropriate (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 the question that follows.

For Multiple Subjects: *Imagine a hypothetical group of candidates for the Multiple Subject Teaching Credential, each of whom is just at the level of knowledge and skills important for effective job performance as a beginning teacher in any of grades K-8 in California public schools.*

OR

For Single Subjects: *Imagine a hypothetical group of candidates for the Single Subject Teaching Credential in (SUBJECT AREA), 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, facilitating a discussion about the performance of the just acceptable candidate and the standard setting procedure. The group also reviewed item statistics (p -values) on each practice test item as an indicator of the difficulty level for all of the examinees who took the test, not only the just acceptable candidates.

Following the practice set, panel members began the same rating process with the multiple-choice items used on the January 25, 2003 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:

- content specifications / subject matter requirements;
- the subtest forms used for the January 2003 test administration;
- the appropriate set(s) of performance characteristics and scoring scale(s);
- Subject Matter Advisory Panel-approved marker responses² for each score point on the scoring scale; and
- 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.

² Each subject matter advisory panel selected responses as marker responses. Marker responses are score-point exemplars used in the training and calibration of scorers.

For Multiple Subjects: *Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning elementary school teacher in California public schools.*

OR

For Single Subjects: *Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning teacher of (SUBJECT AREA) 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 the panel had completed the practice set, they were polled for their item ratings, facilitating a discussion of the concept of the just acceptable candidate, the use of the marker responses and scoring scales, and the standard setting procedure.

Following the practice set, panel members began the same rating process with each of the constructed-response items used on the January 25, 2003 operational test forms. NES analyzed the individual and group 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. For the CSET: English, panel members were asked to provide only subtest recommendations because the subtests in this field contain only one component. Additionally, panel members were asked to provide for each subtest the percent of points to be allocated for each component in the subtest.

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

- Content specifications / subject matter requirements
- Round One Multiple-Choice Item Rating Summary Sheet
 - Sum of the median rating for each item across all panel members
 - 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
 - Sum of the median rating for each item across all panel members and doubled to reflect the actual combined scores examinees will receive from two scorers
 - 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.

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

- Round Two Subtest component Standard Setting Recommendation Form

Multiple-Choice Items

At this stage in the study, panel members were provided the Round One Item Rating Summary Sheet for discussion and to use that item-level data to inform their component-level recommendations. They were advised that candidates will not “pass” the multiple-choice component alone; a candidate’s passing status will be determined at the subtest level, which, with the exception of the English examination, involves the combination of multiple-choice component and constructed-response component performance.

Panel members worked independently, considering their own aggregated Round One rating and the group median. Each member recommended a single, holistic, Round Two multiple-choice component “cut score” for each subtest, representing the total number of scorable items at the subtest level that would, in his or her judgment, be answered correctly by the just acceptable candidate. To make this recommendation, panel members responded to the following question:

For Multiple Subjects: *Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning elementary school teacher in California public schools.*

OR

For Single Subjects: *Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning teacher of (SUBJECT AREA) 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

For the Round Two constructed-response ratings, panel members were again provided the opportunity to discuss the results of the Round One ratings and the merits of various constructed-response component cut scores at the subtest level. Panel members were advised that candidates will not “pass” the constructed-response component alone; a candidate’s passing status is determined at the subtest level, which, with the exception of the English examination, is based on the combination of both the multiple-choice and constructed-response components.

Panel members worked independently, considering their own aggregated Round One rating and the group median. Each member recommended a single, holistic, Round Two constructed-response component cut score for each subtest, representing the total number of points at the subtest level that would, in his or her judgment, be answered correctly by the just acceptable candidate. To make this recommendation, panel members responded to the following question:

For Multiple Subjects: *Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning elementary school teacher in California public schools.*

OR

For Single Subjects: *Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning teacher of (SUBJECT AREA) 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

Panel members were provided two alternative rules for allocating points consistent with psychometric standards and the structure of each examination: a) multiple-choice component weighting of 80% and constructed-response component weighting of 20%; or b) multiple-choice component weighting of 70% and constructed-response component weighting of 30%. These two options are intended to yield reliable results and are psychometrically defensible. In orienting the panels to this task, they were asked to consider issues of reliability, the length of each component, and the nature of the information provided by each component. Panel members discussed the score combination rules and the rationales for each. Following this discussion, panel members independently made recommendations by responding to the following question:

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:

- content specifications / subject matter requirements;
- Round Two Results Summary Sheet;
- Round Three Subtest Standard Setting Recommendation Form; and
- Summary Statistics Report for subtests, where appropriate (i.e., for subtests taken by 20 more examinees).

- Demographic (descriptive) information characterizing the sample of examinees that took the subtest at the January 25, 2003 test administration.
- A set of analyses showing 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 for both component combination rules (i.e., 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 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 questions:

For Multiple Subjects: *Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning elementary school teacher in California public schools.*

OR

For Single Subjects: *Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning teacher of (SUBJECT AREA) 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.

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 Passing Standards for CSET: Multiple Subjects

Subtest	Item Type	Scorable Items	Possible Score Points	Median of Task Force Recommendations
I: Reading, Language, and Literature; History and Social Science	MC	46	46	29.6
	CR	4	24	15.8
II: Science; Mathematics	MC	46	46	28.5
	CR	4	24	15.6
III: Physical Education; Human Development; Visual and Performing Arts	MC	36	36	24.4
	CR	3	18	11.9

Component score combination rule for all subtests: 70% multiple-choice, 30% constructed-response.

Table 7: Panel-Recommended Passing Standards for CSET: English

Subtest	Item Type	Scorable Items	Possible Score Points	Median of Task Force Recommendations
I: Literature and Textual Analysis; Composition and Rhetoric	MC	46	46	32.3
II: Language, Linguistics, and Literacy	MC	46	46	32.6
III: Composition and Rhetoric; Literature and Textual Analysis	CR	2	16	12.7
IV: Communications: Speech, Media and Creative Performance	CR	4	24	19.8

Component score combination rule for all subtests: 70% multiple-choice, 30% constructed-response.

Table 8: Panel-Recommended Passing Standards for CSET: Mathematics

Subtest	Item Type	Scorable Items	Possible Score Points	Median of Task Force Recommendations
I: Algebra; Number Theory	MC	27	27	17.1
	CR	4	32	22.4
II: Geometry; Probability and Statistics	MC	27	27	18.4
	CR	4	32	23.7
III: Calculus; History of Mathematics	MC	27	27	19.1
	CR	4	32	23.9

Component score combination rule for all subtests: 70% multiple-choice, 30% constructed-response.

Table 9: Panel-Recommended Passing Standards for CSET: Science

Subtest	Item Type	Scorable Items	Possible Score Points	Median of Task Force Recommendations
I: General Science: Astronomy; Dynamic Processes of the Earth; Earth Resources; Waves; Forces and Motion; Electricity and Magnetism	MC	52	52	31.7
	CR	2	12	7.9
II: General Science: Ecology; Genetics and Evolution; Cell and Organismal Biology; Molecular Biology and Biochemistry; Heat Transfer and Thermodynamics; Structure and Properties of Matter	MC	52	52	35.8
	CR	2	12	8.0
III: One of the following concentration areas:				
• Biology/Life Science	MC	46	46	28.8
	CR	3	18	11.8
• Chemistry	MC	40	40	29.3
	CR	3	18	11.5
• Earth and Planetary Science	MC	40	40	24.8
	CR	3	18	10.0
• Physics	MC	40	40	30.5
	CR	3	18	12.9

Component score combination rule for Subtests I & II: 80% multiple-choice, 20% constructed-response.
Component score combination rule for Subtest III: 70% multiple-choice, 30% constructed-response.

Table 10: Panel-Recommended Passing Standards for CSET: Social Science

Subtest	Item Type	Scorable Items	Possible Score Points	Median of Task Force Recommendations
I: World History; World Geography	MC	35	35	19.2
	CR	3	20	13.7
II: U.S. History; U.S. Geography	MC	35	35	20.2
	CR	3	20	13.8
III: Civics, Economics; California History	MC	36	36	19.5
	CR	3	18	13.0

Component score combination rule for all subtests: 70% multiple-choice, 30% constructed-response.

Part III: Staff-Recommended Passing Standards

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: Multiple Subjects, English, Mathematics, Science, and Social Science, staff considered several factors and options that affect the standard setting process. Each consideration is discussed below, followed by staff-recommended passing standards.

Compensatory Scoring

One approach to setting passing standards involves a policy and process by which the subtest scores on an examination are compensatory. Three basic models of compensatory scoring follow.

Fully Compensatory Scoring.

In a fully compensatory scoring model, a passing score is established for each subtest as well as a total passing score across all subtests. High performance on one subtest may compensate for low performance on any other subtest, without having to meet subtest minimum scores.

Advantages. This model allows candidates to compensate for performance on one subtest with performance on any other subtest.

Disadvantages. This model largely negates the original intention of the subtest structure by NOT assuring minimum content knowledge in each domain. A very high score on one subtest can compensate for a very low score on another subtest.

Fully Noncompensatory Scoring.

In a fully noncompensatory scoring model, a passing standard is established for each subtest. This subtest score stands alone, unaffected by the score on any other subtest. Examinees must meet or exceed the passing standard on each subtest. Higher performance on one subtest can NOT compensate for lower performance on another subtest.

Advantages. This model preserves, to the fullest extent possible, the assurance of minimum content knowledge in each domain and the integrity of the examination as a criterion-referenced test. It is the most straightforward model in the use and interpretation of scores. This model allows for adjustment of panel-recommended passing standards for measurement error specifically by subtest and in a way that does not impact the interpretation and use of scores. The fully noncompensatory scoring model also allows for a single subtest passing score to be adjusted in the future without adjustments to other subtests in the set.

Disadvantages. This model does not allow candidates to use higher performance on one subtest to compensate for lower performance on another subtest.

Partially Compensatory Scoring.

One example of a partially compensatory scoring model is one in which three types of scores are established: a passing score for each subtest; a total passing score across all subtests; and minimum scores on each subtest, below which a candidate cannot pass even if the total score across all subtests is achieved. In this way, a high score on one subtest can compensate for a low score on another subtest, as long as the lower score meets that subtest minimum score.

Advantages. This model allows for score compensation across subtests to some extent.

Disadvantages. In allowing for partial compensation across subtests, the assurance of minimum content knowledge in each domain is diminished. This model results in different minimum scores by subtest, which complicates the use and interpretation of scores, particularly for Title II reporting.

Additionally, some examinations (i.e., Mathematics and Science) have subtest structures that support differentiated credentials. These variations on the use of subtests complicates subtest minimum scores and compensation. For example, Subtests I and II of the mathematics examination can be used toward the Single Subject Teaching Credential in Foundational-Level Mathematics. It is possible to have contradictory results, whereby a candidate could pass all three subtests of the

mathematics examination, qualifying for a Single Subject Teaching Credential in Mathematics, but not for a credential in foundational-level mathematics.

Review of Passing Standards

Determining the passing standard(s) for an examination is a careful, conscientious process. 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. Additionally, it is expected that the passing standards of those examinations that did not at the first administration achieve at least 150 examinees taking all subtests will be reviewed. Any recommendations for change in the standards will be presented to the Commission for consideration and adoption.

Phased-in Passing Standards

At times, it is appropriate to phase-in passing standards over time. This conservative approach is sometimes used with new testing programs, when substantial changes have been made in the content of a testing program, or for tests with low examinee volumes. The information available on a test from its first administration is not necessarily representative of the population and performance of the test over time. Phasing-in standards allows for lower standards to be raised over time.

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.

Staff-Recommended Passing Standards

Based on these considerations, staff recommends that the Commission adopt the passing standards for the subtests of the CSET forms administered on January 25, 2003 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 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.

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

If the Commission adopts the staff-recommended passing standards, as indicated in Table 11, at their April 2003 meeting, NES will release score reports for the January 2003 test administration on April 18th and for the March 2003 test administration on April 21st. The next test administration of the CSET is scheduled for May 17, 2003.

Table 11: Staff-Recommended Passing Standards for CSET

Subtest	Multiple-Choice Raw Score Points	Constructed-Response Raw Score Points	Component Score Combination Rule MC/CR	Passing Rate for January 25, 2003 Test Administration by Subtest	Overall Passing Rate for January 25, 2003 Test Administration
CSET: Multiple Subjects					50%
Subtest I	27	14	70/30	74%	
Subtest II	26	14	70/30	59%	
Subtest III	22	10	70/30	83%	
CSET: English					39%
Subtest I	29	--	--	77%	
Subtest II	30	--	--	71%	
Subtest III	--	11	--	70%	
Subtest IV	--	18	--	60%	
CSET: Mathematics					31%
Subtest I	15	19	70/30	45%	
Subtest II	16	21	70/30	61%	
Subtest III	17	21	70/30	40%	
CSET: Science					67%
Subtest I	29	6	80/20	75%	
Subtest II	33	6	80/20	79%	
Subtest III (one of the following concentration areas)					
Biology/Life Science	26	10	70/30	66%	
Chemistry	27	10	70/30	77%	
Earth & Planetary Science	22	8	70/30	79%	
Physics	28	11	70/30	65%	
CSET: Social Science					42%
Subtest I	19	14	70/30	62%	
Subtest II	20	14	70/30	58%	

Subtest III	20	13	70/30	59%	
-------------	----	----	-------	-----	--

Appendix A

CSET: MULTIPLE SUBJECTS THREE-POINT PERFORMANCE CHARACTERISTICS AND SCORING SCALE

PERFORMANCE CHARACTERISTICS

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

SCORING 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 Content Specifications.</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 Content Specifications.</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 Content Specifications.</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.

CSET: SINGLE SUBJECT
THREE-POINT PERFORMANCE CHARACTERISTICS AND SCORING SCALE

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.

SCORING 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.

CSET FOUR-POINT PERFORMANCE CHARACTERISTICS AND SCORING SCALE

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.

SCORING 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: JANUARY 25, 2003 TEST ADMINISTRATION NUMBERS OF EXAMINEES BY SUBTEST

Subtest	Examinees Per Subtest	Examinees Taking All Subtests
CSET: Multiple Subjects		
I: Reading, Language, and Literature; History and Social Science	2015	1,276
II: Science; Mathematics	1571	
III: Physical Education; Human Development; Visual and Performing Arts	1828	
CSET: English		
I: Literature and Textual Analysis; Composition and Rhetoric	635	405
II: Language, Linguistics, and Literacy	576	
III: Composition and Rhetoric; Literature and Textual Analysis	502	
IV: Communications: Speech, Media, and Creative Performance	478	
CSET: Mathematics		
I: Algebra; Number Theory	274	39
II: Geometry; Probability and Statistics	303	
III: Calculus; History of Mathematics	72	
CSET: Science		
I: General Science: Astronomy; Dynamic Processes of the Earth; Earth Resources; Waves; Forces and Motion; Electricity and Magnetism	309	
II: General Science: Ecology; Genetics and Evolution; Molecular Biology and Biochemistry; Cell and Organismal Biology; Heat Transfer and Thermodynamics; Structure and Properties of Matter	300	
III: Biology	176	110
Chemistry	66	39
Earth and Planetary Science	29	22
Physics	20	13
CSET: Social Science		
I: World History; Geography	490	391
II: U.S. History; Geography	537	
III: Civics; Economics; California History	467	

Appendix C

CSET STANDARD SETTING PANELS

	Multiple Subjects	English	Mathematics	Science	Social Science	Total
Total Number						
Appointed	48	20	23	36	20	147
Participated	35	18	17	31	14	115
Ethnicity						
African American	2	0	0	0	1	3
Asian	3	0	2	0	0	5
Hispanic	1	1	3	1	1	7
White	27	11	9	29	11	87
Other/Not Provided	2	1	3	1	1	8
Sex						
Female	23	11	7	14	6	61
Male	12	2	10	17	8	49
Region						
North	13	2	4	11	4	34
South	22	11	13	20	10	76
Profession						
Public School Educators	29	10	11	26	9	85
College/University Educators	6	3	6	5	5	25
Years of Experience						
0-6	3	1	0	5	2	11
7-10	2	1	2	6	1	12
11+	29	11	14	19	11	84
Not Provided	1	0	1	1	0	3

