4D

Action

Educator Preparation Committee

Proposed Adoption of the Passing Score Standard for the Revised California Subject Examinations for Teachers, Multiple Subjects, Subtest III (CSET: MS, Subtest III)

Executive Summary: This agenda item describes the standard setting process used to recommend preliminary passing scores for CSET: Multiple Subjects, Subtest III and presents preliminary passing score standards for potential adoption by the Commission.

Recommended Action: That the Commission adopt one of the proposed preliminary passing score standards for the revised examination subtest.

Presenter: Mike Taylor, Consultant, Professional Services Division

Strategic Plan Goal

I. Educator Quality
   b) Develop, maintain, and promote high quality authentic, consistent educator assessments and examinations that support development and certification of educators who have demonstrated the capacity to be effective practitioners.

Proposed Adoption of the Passing Score Standard for the Revised California Subject Examinations for Teachers, Multiple Subjects, Subtest III (CSET: MS, Subtest III)

Introduction
This agenda item describes the standard setting study for the revised California Subject Examinations for Teachers (CSET), Multiple Subjects, Subtest III (CSET: MS, Subtest III) and provides a recommended initial passing standard for the examination subtest based on the recommendations from the CSET Standard Setting Panel.

Background
Education Code (EC) section 44281 requires the Commission to "administer subject matter examinations....to assure minimum levels of subject matter knowledge by certified personnel." The Commission's CSET examinations serve this statutory purpose. The CSET examinations are required to be aligned with the state-adopted content standards for students. As these content standards change over time, the corresponding CSET examinations are updated to remain in alignment with the most current sets of California TK-12 content standards.

In addition, EC section 44288 specifies the use of subject matter advisory panels to "advise in the selection, administration, and interpretation of examinations." The subject matter advisory panels "shall consist of recognized leaders in the subject matter fields to be examined and shall be composed primarily of full-time public school classroom teachers and full-time college or university classroom teachers." Consistent with statutory requirements, standard Commission practice has always been to use advisory panels of California content experts to advise the Commission in the development of the Commission's subject matter examinations, the CSET.

In January 2019, the State Board of Education adopted new California Arts Standards for Public Schools. In response, Commission staff began working with the examinations contractor, the Evaluation Systems group of Pearson, and panels of California educators to develop Subject Matter Requirements (SMRs) and CSET examinations aligned to the new California Arts Standards for Public Schools. The Commission adopted revised SMRs for Multiple Subjects in August 2021. Because the only revisions to the SMRs occurred within the arts domains, the only CSET: Multiple Subjects subtest that needed revision was Subtest III, as Subtest III contains all the questions measuring the arts domains for CSET: Multiple Subjects.

Updating the CSET examinations requires a two-stage process: first, the revision of the Commission-adopted Subject Matter Requirements (SMRs) that identify the content eligible to be assessed on the examination, and then, following Commission adoption of revised SMRs, the revision, redevelopment, and/or new development of test items that validly and reliably assess candidate levels of knowledge specific to the content area of the credential. In separate work, Commission-approved subject matter preparation programs must also update their coursework

and assessments to align with the revised SMRs and must respond to the Commission documenting the transition to implementing the revised SMRs. The entire examination revision and transition process typically takes a minimum of two years to complete.

The Process for Developing CSET Examinations

Development of a new CSET examination is a multi-stage process that follows testing industry-standard practices and conforms to the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education, 2014), commonly known as the "Joint Standards." Adhering to these national testing standards assures that, consistent with statute, the Commission's examinations are and remain valid and reliable for all examinees. Applicable teacher licensure standards within the full set of Joint Standards align with the foundational and operational portions of the Joint Standards (Part I and Part II) as well as the application standards (Part III) for both credentialing assessments (Chapter 11) and educational assessments (Chapter 12).

The test development process for a CSET examination includes all of the following sequential sets of activities:

1) Recruitment and appointment by the Commission's Executive Director of Subject Matter Advisory Panels of California content experts, in accordance with the provisions of Education Code section 44288.
2) Development and review of draft SMRs.
3) Bias review of the draft SMRs by the Commission's standing Bias Review Committee.
4) Content reviews of the draft SMRs by the Subject Matter Advisory Panels.
5) Review of the draft SMRs by the Commission and direction to proceed to content validation of the draft SMRs.
6) Content validation of the draft SMRs.
7) Review and approval of the draft SMRs by the Commission.
8) Development of new test items for the item bank for each examination.
9) Bias review of new test items.
10) Content review by the Subject Matter Advisory Panels of the revised and new test items.
11) Revisions of test items, as needed.
12) Field testing of new test items.
13) Review of item level field test data to eliminate questions that are not viable.
14) Development of Test Guides available to candidates on the Examinations website.
15) Initial test administration.
16) Scoring of initial test administration.
17) Identification of marker scoring papers, if needed.
18) Standard Setting to establish the passing standard.
19) **Adoption of a passing score standard by the Commission.**
20) Ongoing implementation of the new CSET examination(s).

We are currently at step 19 in this process for CSET: MS, Subtest III.

A California content expert advisory panel was recruited and appointed by the Executive Director to review the new CSET: MS, Subtest III test items. Throughout the fall of 2021, the Commission's standing Bias Review Committee and expert panels of California educators participated in reviews of all the new test questions for appropriateness and for alignment to the 2019 California Arts Standards and to the recently adopted Multiple Subjects SMRs.

Field-testing of the new test items to see how they performed under actual testing conditions was conducted between January and February 2022, after which the results were analyzed, and decisions were made about the final bank of items for operational use with candidates who take the examination. The revised CSET: MS: Subtest III became operational for candidates in August 2022. Following the initial administrations of the new CSET: MS: Subtest III, a standard setting study was conducted. The results of that standard setting are presented later in this item.

CSET: Multiple Subjects Test Structure
Table 1 shows the test structure for CSET: Multiple Subjects, which is divided into three subtests. Each of the subtests measures content knowledge from separate domains. Only Subtest III was revised, and only the test questions measuring candidates' knowledge of visual and performing arts.

Table 1 CSET: Multiple Subjects Test Structure

Subtest	Domains	Number of Multiple-Choice Questions	Number of Constructed-Response Questions
I	Reading, Language, and Literature	26	2
	History and Social Science	26	2
	Subtest I Total	52	4
II	Science	26	2
	Mathematics	26	2
	Subtest II Total	52	4
III	Physical Education	13	1
	Human Development	13	1
	Visual and Performing Arts	**13**	**1**
	Subtest III Total	39	3

The Standard Setting Process
"Standard setting" is the common term used in the large-scale assessment industry to describe the process of determining a minimum passing score, or cut score, for new or revised examinations. The term "standard" as it is used in standard setting refers to a performance standard or minimum level of acceptable performance on an examination.

For criterion-referenced examinations like the CSET, standard setting is a content-focused, structured process in which a panel of content area experts review the content of an examination, and carefully considers the knowledge and skills being measured and relevant data such as question difficulty levels and potential pass rates for various cut scores to make an informed judgment about the minimum level of content knowledge that examinees should demonstrate to "pass" the examination. The standard setting process results in a recommended cut score from the content expert panel to the Commission, which has the authority to establish the minimum passing standard for the CSET examinations.

Standard setting is a common and established process for determining valid and defensible minimum passing scores for standardized examinations. Standard setting allows an authoritative body, in this case the Commission, to make an informed decision when establishing cut scores instead of arbitrarily selecting a minimum passing standard.

There have been many different methods for standard setting published and researched in the field of large-scale assessment over the last 50 years. These standard setting methods are in use today for various types of assessments all over the world. All of the most common standard setting methods for educational assessments involve the informed judgments of "raters," or content area experts. The specific standard setting process used for the CSET is described more fully below.

The CSET: Multiple Subjects, Subtest III Standard Setting Study
The purpose of standard setting studies is to provide the Commission with recommendations, based on the informed judgments of California educators, relevant to the determination of the initial passing standards (in this case, for the CSET: MS, Subtest III). The educators on the Standard Setting Panel represented credentialed TK-12 teachers with experience teaching in elementary classrooms, district-level administrators, and teacher preparation program faculty with experience in elementary education who are responsible for the preparation of Multiple Subjects teachers via the program route.

As with the standard setting study method used for all other Commission examinations, the process employed for the CSET: MS, Subtest III was consistent with recognized psychometric principles and procedures. The standard setting study for CSET: MS, Subtest III was conducted September 8, 2022.

The CSET standard setting meeting began with an orientation and training session. The initial step was to ask the panel members to independently take the examination under simulated test-like conditions. This activity helped the members become familiar with the examination, the knowledge and skills associated with the items, and the perspective of the examinees. The

panel members were then familiarized with the SMRs and the concept of the minimally competent level of content knowledge necessary for a beginning teacher. Panel members were asked to conceptualize the specific content knowledge and skills of a hypothetical beginning teacher who would be competent in the subject area. Panel members used this concept of what a minimally competent beginning teacher would know and be able to do in determining their recommended acceptable score for passing each subtest. Although a number of examinees may exceed the level of acceptable knowledge and skills, none receiving a passing score should fall below this minimally competent level. The panel also reviewed the performance characteristics and score scales used to evaluate the constructed-response items in the CSET: MS, Subtest III. After this extensive training and the simulated test taking, panel members completed the following three rounds of standard setting activities, as described below. These activities focused on arriving at an informed judgment as to what the potential cut score should be that reflects the minimum level of subject matter knowledge necessary for a beginning practitioner just competent to begin professional practice.

Round One: For each multiple-choice item, the panel members were asked to independently rate the percent of minimally competent beginning teachers whom they think would likely answer the item correctly. For each constructed-response item, members were asked to independently indicate the level of response that would likely be achieved by the minimally competent beginning teacher.

Round Two: The Round One ratings, which were displayed anonymously, were distributed, and members discussed the reasoning used in making their determinations. The second round moved the panel from individual item ratings to ratings at the section level (i.e., multiple-choice section and constructed-response section). They were asked the number of multiple-choice items that would be answered correctly and the total score points that would likely be achieved on the constructed-response items by the minimally competent beginning teacher.

Round Three: Panel members were given the results of their Round Two ratings. They were then asked to make final independent recommendations for a passing standard based on the raw score points earned on each section of the test.

Separate ratings for each of the subtests were made during each of the three rounds. The panel's recommendation represents the computed median of the third round results.

Results of the Standard Setting Studies
The Standard Setting Panel followed the procedures outlined above to determine recommendations for the CSET minimum passing score. Following the standard setting study, the Evaluation Systems group of Pearson (Evaluation Systems), the Commission's contractor for examinations, calculated the median panel-recommended passing score based on the individual members' recommendations. Table 2 below provides a summary of the CSET panel recommendations, including the number of scorable items by item type, the total possible score by item type, and the median panel-recommended raw score total.

Table 2: CSET Panel-Recommended Passing Score Standards

Subtest	Item Type	Total Possible Raw Score Points	Panel-Recommended Cut Score
MS, Subtest III	Multiple Choice	36	**22**
	Constructed Response	18	**11**

Table 3 below shows the number of individuals who completed CSET: MS, Subtest III between August 1 and August 28, 2022, and their potential passing rates based on the panel-recommended passing standard.

Table 3: Number of Individuals who completed CSET: MS, Subtest III, with Potential Subtest Passing Rates at the Panel-Recommended Passing Score Standard

MS, Subtest III	Number of Examinees (N)	Percent of N passing at Panel-Recommended Cut Score
All Examinees	317	73%
African American or Black	12	58%
Asian American	15	73%
Southeast Asian American	8	Low N
Pacific Island American	2	Low N
Latin American or Hispanic	129	66%
Native American	4	Low N
White	119	83%
Other	18	67%
Female	246	72%
Male	63	71%

Because this is technically a new examination the potential pass-rate represented in Table 3 is a first-time pass rate. Since the standard setting workshops, CSET: MS, Subtest III has been continued to be administered operationally, and some additional information is available about potential passing rates for examinees who have taken the exam so far. All available impact data is included in Table 4 below to help the Commission to make the most informed decision possible.

A Note on Testing Bias
In large-scale assessment, differential passing rates by subgroups are not considered bias in and of themselves. Commission examinations are designed, in part, to uncover differences in scores according to various subgroups, particularly groups based on race and ethnicity, to help understand gaps in education among the population of candidates coming into teacher preparation. CSET doesn't measure anything that happens in a teacher preparation program. CSET measures content knowledge that candidates bring with them as they begin teacher preparation. Processes to avoid bias are built into the Commission's examination development and administration processes, including a Bias Review Committee which reviews all test content and questions for potential bias, making changes, suggestions, and even eliminating questions if necessary, and differential item functioning (DIF) analysis, which more deeply compares question-level responses of members of various subgroups to flag for potential bias after test administration. The Commission employs these procedures specifically to reduce measurement error that might be caused by bias so that results by gender, race, and ethnicity can be accurately reported.

Standard Error of Measurement
Once the final panel score recommendation is determined, an additional modification may be made to that score before it is recommended to the Commission. This modification is the determination and potential application of an adjustment that takes into consideration the Standard Error of Measurement (SEM). The SEM is a key measurement concept that addresses how accurately the recommended passing score standard reflects the scores likely to be achieved by actual candidates in real-world testing situations. For example, an examinee takes the test one time and receives a score. If that same examinee were to take the same exam several times, with no change in his or her level of knowledge and preparation, it is possible that some of the resulting scores would be slightly higher or slightly lower than the score initially achieved by the examinee the first time he or she took the examination. Given this variation in possible scores on the same test by the same examinee, the examinee's initial score might not reflect the best score that examinee would hypothetically be able to achieve based on his or her actual knowledge and ability in the content area.

The range of scores an examinee would achieve across multiple administrations of the same test, were this activity to take place, includes what is known as the examinee's "true" score (the hypothetical score that would best reflect the examinee's actual ability) and the "observed score" (the actual score received on the first test administration).

A simple way to look at the concept of the SEM is to consider the case of the examinee who takes a CSET examination one time. Many factors affect how the examinee scores on his or her first attempt on the test, including knowledge of the content tested, affective factors such as the examinee's emotional, physical, and/or mental state on that particular day and time, and external factors such as the testing environment. Thus, it is not possible to say with certainty that the score obtained on the initial test taken by the examinee most accurately reflects his or her true level of knowledge, skills, and abilities. The likelihood that the examinee's true score is reflected on his or her first attempt is unknown. Thus, a computed SEM is often applied to

adjust the minimum passing score for an examination in order to account for the difference in the examinee's true score and the examinee's observed score on the assessment.

To account for the difference measurement error may create between examinees' observed scores and true scores the Commission may wish to consider applying a (SEM) adjustment to the panel-recommended minimum passing standard for each of the subtests in both examinations. If the Commission were to consider an SEM adjustment for the revised CSET: MS, Subtest III, the results would be as shown in Table 4 below. The full SEM for each subtest, referred to as +/- 1 SEM, falls between 2 and 4 raw score points, depending on the examination and subtest. When the Commission has previously applied an SEM adjustment to cut scores, it has most often been -1 SEM. More information about Standard Error of Measurement can be found in Appendix B. Table 4 shows potential subtest-level and overall pass rates at -2, -1.5, - 1, -0.5, and 0 SEM.

Table 4: CSET Potential Pass Rates with Application of Different SEM Adjustments

MS, Subtest III	Number of Examinees 8/1/22 – 8/28/22	Passing at -2 SEM	Passing at -1.5 SEM	Passing at -1 SEM	Passing at -0.5 SEM	Passing at Panel-Recommended Cut Score (0 SEM)
All Examinees	317	97%	94%	91%	82%	73%
African American or Black	12	92%	92%	92%	67%	58%
Asian American	15	93%	93%	93%	87%	73%
Southeast Asian American	8	Low N	Low N	Low N	Low N	Low N
Pacific Island American	2	Low N	Low N	Low N	Low N	Low N
Latin American or Hispanic	129	97%	95%	88%	77%	66%
Native American	4	Low N	Low N	Low N	Low N	Low N
White	119	98%	96%	95%	92%	83%
Other	18	94%	89%	83%	67%	67%
Female	246	97%	95%	91%	82%	72%
Male	63	95%	92%	87%	78%	71%

As with any new criterion-referenced assessments it is expected that the pass rates will increase over time as the population of potential examinees becomes familiar with the revised subject matter requirements for this credential.

Staff Recommendation

Whether to apply an SEM adjustment to one or more of the CSET examinations is a policy decision for the Commission's consideration. Staff notes that it is typical when new assessments are introduced, particularly those that measure new content material, for initial passing rates to begin low and rise over time. Staff recommends that the Commission discuss the various options presented in this item for SEM adjustment and set the passing score at the panel-recommended passing score standard with whichever SEM adjustment, if any, the Commission feels is most appropriate.

Next Steps

If the Commission adopts a new passing score standard for the CSET: MS, Subtest III, notification will be posted on the CSET website and distributed to the field. In addition, scores for examinees who recently took the assessment will be calculated based on the adopted passing standard and scaled to a range of 100 to 300, with 220 representing the adopted passing standard for each exam. Individual examinee score reports will then be distributed within three to four weeks of the Commission's decision. The passing standard adopted by the Commission will be applied to all subsequent administrations of the CSET: MS, Subtest III.

Appendix A

Demographics of Standard Setting Panel

CSET MS: Subtest III

Total Number of Participants	**11**
African American or Black	2
Asian American	1
Filipino	1
Southeast Asian American	
Pacific Island American	
Mexican American / Chicano	1
Latin American / Other Hispanic	1
Native American	
White (non-Hispanic)	4
Other Race or Ethnicity	1
Female	10
Male	1
Nonbinary	
Decline to State Gender	
Public School Educator	11
College/University Educator	
Other Profession	

Appendix B

Additional Information About Application of Standard Error of Measurement (SEM)

How Does Applying the SEM Work?

Individual examinee scores on the first attempt of an examination could potentially not represent the examinee's true level of knowledge, skills, and abilities. The score could also potentially represent a "false negative" (i.e., the examinee did have sufficient knowledge, skills, and abilities but the actual score did not closely enough represent the examinee's true abilities) or a "false positive" (i.e., the examinee did not actually have sufficient knowledge, skills, and abilities but was able to earn a higher score than otherwise warranted). In the case of false positives, there is no psychometric approach to mitigating this outcome; in other words, the examinee is fortunate in achieving the passing score. However, there is a standard psychometric technique that does address the case of false negatives. This technique is the application of the SEM to the passing score standard established for a particular examination.

For example, on a CSET examination, an examinee earns a raw score that is then converted to a scaled score. For Commission examinations, raw scores are converted to scaled scores as part of the test equating process. Scaled scores for Commission exams are reported as a whole number between 100 and 300, where 220 represents the minimum scaled score needed to pass. This scaling process allows examination scores to be reported and interpreted as simply as possible across administrations and across examinations. The SEM would be applied to the minimum raw score for a particular exam that equates to the Commission's adopted passing scaled score standard of 220. Thus, if the recommended cut score for an exam were to be a raw score of 30, the SEM would be applied to the raw score of 30. If the calculated SEM was minus 2 raw score points, and was applied to the raw cut score of 30, the raw cut score would now be 28. Raw scores for this exam would then be scaled so that the raw cut score (in this case 28) equated to a scaled score of 220, which is the universal minimum passing scaled score for Commission exams. Examinees would need to achieve at least 28 raw score points to pass the examination. If an examinee whose actual knowledge and ability should have allowed him or her to pass was only able to earn 29 raw score points due to factors other than his/her knowledge of the content such as, for example, emotional upset, application of the SEM to the minimum passing standard would allow him or her to receive a passing score, thereby avoiding a false negative.

The SEM can vary depending on the nature of the particular examination and the range of the candidate population for that examination. Typically, multiple choice examinations that have clear right or wrong responses will have less variability in the range of candidate scores – either the candidate knows or does not know the content being assessed. In the case of constructed response and performance items, where candidates construct their own responses which are scored by trained readers, one might expect a larger range of variability in both responses and the background knowledge and abilities of candidates. It might also differ in the case where an examination is new, or the number of examinees is very low. Thus, a SEM could range from -1 to -5 or even higher. Therefore, each SEM is calculated individually for a particular examination and is then consistently applied to the passing score for that examination.

Consideration of an SEM Adjustment

It is possible for an examinee who does possess the knowledge and skills measured by the exam to receive a failing score, also known as a "false negative," due to measurement error. For example, an examinee takes the test one time and receives a score. If that same examinee were to take the same exam several times, with no change in his or her level of knowledge and preparation, it is possible that some of the resulting scores would be slightly higher or slightly lower than the score initially achieved by the examinee the first time he or she took the examination. Given this variation in possible scores on the same test by the same examinee, the examinee's initial score might not reflect the best score that examinee would hypothetically be able to achieve based on his or her actual knowledge and ability in the content area. Thus, a computed Standard Error of Measurement (SEM) is typically applied to adjust the minimum passing score for an examination in order to account for the difference in the examinee's true score and the examinee's observed score on the assessment.