2D

Action

Educator Preparation Committee

Proposed Adoption of Passing Score Standards for the Redeveloped California Teaching Performance Assessment

Executive Summary: This agenda item provides (a) foundational information about the standard setting process for Commission examinations and (b) recommendations for passing score standards for the redeveloped California Teaching Performance Assessment (CalTPA).

Recommended Action: That the Commission adopt the recommended passing score standards for the redeveloped CalTPA.

Presenters: Amy Reising, Director, and Wayne Bacer, Consultant, Performance Assessment Development, and Heather Klesch, Evaluation Systems group of Pearson

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 Passing Score Standards for the Redeveloped California Teaching Performance Assessment

Introduction
This agenda item provides (a) foundational information about the standard setting process for Commission-developed examinations and (b) recommendations for passing score standards for the redeveloped California Teaching Performance Assessment (CalTPA).

Background
The Commission redeveloped the CalTPA, its approved TPA model, during 2016-18. The membership of the CalTPA Design Team is provided in Appendix A, the description of the groups involved in the redevelopment of the CalTPA is provided in Appendix B, the list of teacher preparation programs using the CalTPA is provided in Appendix C, and the membership of the CalTPA standard setting panel is provided in Appendix D.

The Commission redeveloped the CalTPA for several key reasons, including alignment with the expectations of the state-adopted Common Core State Standards and the Next Generation Science Standards, and the needs of the Commission's Accreditation Data System (ADS) for an outcomes-based set of quality indicators to assure that the data produced on the CalTPA as well as other Commission-developed assessments was consistent and reliable. To increase scoring reliability, the redeveloped CalTPA is now centrally managed and scored and requires using content- specific assessors to ensure that detailed, analytic, feedback based on the individual teaching performance expectations (TPEs) is provided to candidates and programs in a timely manner to guide both candidate development and program improvement.

The CalTPA model has been revised, pilot- and field-tested, and is now ready for consequential statewide use. The Commission now needs to set new passing score standards for the redeveloped CalTPA.

Provisional Passing Standard During the Redeveloped CalTPA Pilot and Field Testing Phases, 2017-2018
The revised CalTPA was piloted by 24 institutions and 250 candidates in early 2017. The CalTPA was further revised during the summer of 2017 based on the pilot test findings, and was subsequently field tested by 27 programs during the 2017-18 academic year. The results of the field test led to another round of revisions of the CalTPA and supporting materials prior to statewide operational administration of the assessment starting fall 2018.

In December 2017, the Commission adopted an expected performance level for the revised CalTPA for candidates participating in the field test. Preparation programs included in the field test were granted a waiver for candidates who took the redeveloped CalTPA, thereby allowing

passage of the redeveloped CalTPA to meet the statutory requirement that candidates pass a Commission-approved TPA as part of their preparation for a teaching credential. The waiver provisional passing standard set by the Commission was a minimum of a score of **2 on each rubric, with only one rubric score of 1 allowable**. This provisional passing standard applied to both Cycle 1 and Cycle 2 submissions for the redeveloped CalTPA.

CalTPA First Operational Year (2018-2019) Data Leading to a 2019 Standard Setting Study
Following the field test, the redeveloped CalTPA became operational in fall 2018. For purposes of standard setting, all candidates who had complete cycle submissions submitted and scored through April 25, 2019 were included in the data analysis.

Summary Overview of 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 assessments. The term "standard" as it is used in "standard setting" refers to a performance standard, or minimum level of acceptable performance on an assessment.

Standard setting is a common and established process for determining valid and defensible minimum passing scores for standardized assessments. 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.

For criterion-referenced assessments such as the CalTPA, standard setting is a content-focused, structured process in which a panel of content area experts reviews the content of an assessment, carefully considers the performance expectations being measured, relevant data and potential pass rates at various cut scores to make an informed judgment about the minimum performance level that candidates would need to demonstrate to "pass" the assessment. The standard setting process conducted by the Evaluation Systems group of Pearson (ES) resulted in a recommended cut score from the expert panel to the Commission for each of the two Instructional cycles of the CalTPA.

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. However, all of the most common standard setting methods for educational assessments involve the informed judgments of qualified "raters," or content-specific pedagogical experts.

As with the standard setting study method used for all other Commission examinations (e.g. CSET, CTEL, RICA, CPACE), the process employed for the CalTPA was consistent with recognized psychometric principles and procedures. The standard setting study for the CalTPA was conducted over a two-day period, May 8-9, 2019, with pre-conference preparatory activities for the content expert panel taking place prior to the meeting.

The specific standard setting process used during the meetings for the CalTPA is described in full detail in Appendix E.

Initial and Final Passing Score Recommendation of the Expert Panel
All of the expert panel's standard setting discussions for the initial and final cut score recommendations, made at the conclusion of the second day's standard setting activities, were framed by the following context statement and guiding question:

- Think about a teacher candidate who is just at the level of knowledge and skills required to perform effectively the job of a new teacher in California public schools.

- What score (the sum of all the rubric scores in the Cycle) represents the level of performance that would be achieved by this individual?

The guiding question addressed candidate performance across all rubrics in each cycle. Cycle 1 has 8 rubrics and Cycle 2 has 9 rubrics. Discussion was also conducted to allow for panel recommendations concerning any "side conditions" such as, for example, placing a limitation on the number of rubric scores of "1" that would ultimately be allowed under the final recommended passing score. The scoring rubrics are provided in in the large appendix that is separately linked.

Initial Passing Score Recommendation
Through a facilitated discussion, panelists were presented with CalTPA descriptive data, the activities described in Appendix E were conducted, and each panelist recommended an initial passing during the early part of Day 2's activities.

To arrive at the final standard setting and passing standard recommendations to the Commission, panelists were provided descriptive and summary data, as shown below, to help guide their recommendations. Descriptive and summary data included the number of submissions scored in each CalTPA Cycle and Content Area (Multiple Subject and Single Subject Areas), a summary of the aggregate rubric, step of the cycle, and total CalTPA performance (mean, standard deviation, median, minimum, maximum) for all candidates.

These performance descriptive statistics were provided both in aggregate and broken out by rubric for each CalTPA Cycle and Content Area. Demographics and total score descriptive performance statistics (number, percent, mean, standard deviation, median, minimum, maximum) were provided by gender, ethnicity, placement setting, and candidate primary language. Finally, a distribution of total scores was provided for the complete data set.

After reviewing the descriptive and summary data, and following discussion with the whole group, panelists were asked to make an initial recommendation for a cut score, as well as any recommendation regarding setting any conditions for that cut score. The initial cut score recommendation was as follows:

Cycle 1 (8 rubrics):

- An initial cut score of 19 points was the initial committee-recommended median.
- All 20 panelists recommended a side condition, with 15 recommending no allowance of rubric scores of "1", and 5 recommending allowing 1 rubric score of "1".

Cycle 2 (9 rubrics):

- An initial cut score of 21 points was the initial committee-recommended median.
- All 20 panelists recommended a side condition, with 15 recommending no allowance for rubric scores of "1", and 5 recommending allowing 1 rubric score of "1".

Panelists were shown the frequencies for individual initial recommendations of a cut score, as well as the mean, median, mode, min and max for the initial cut score recommendations. Panelists were also shown a summary of condition recommendations – both the frequency and associated recommendation.

Final Passing Score Recommendation

Through a facilitated discussion, and after examining the initial recommendations, panelists were presented with CalTPA impact data reflecting the number and percent of candidates who would theoretically pass at each potential recommended level, and a final passing score was ultimately recommended by each panelist. The table below shows the rubric level scores, by content area, for both cycles 1 and 2 during the first year of implementation. Rubric scores are shown for candidates' first attempt at passing each cycle.

CalTPA Rubric Scores for 2018-2019

Cycle 1 Performance across all fields is as follows:

Table 1: Cycle 1 Rubric Scores across Content Areas

Field	N	Mean	SD	Median	Min	Max
Multiple Subject: Literacy	1,607	23.8	4.78	23.0	9	40
Multiple Subject: Math	1,039	23.6	4.44	23.0	12	39
Agriculture	21	20.2	3.32	20.0	15	28
Art	74	22.0	3.25	22.0	15	32
Business	4					
English Language Dev	0					
English	433	25.6	4.88	25.0	11	39
Health Science	22	24.2	6.15	24.0	15	39
Home Economics	2					
Industrial & Tech. Ed	0					
Mathematics	256	24.1	4.94	23.5	10	40

Field	N	Mean	SD	Median	Min	Max
Music	111	23.0	4.55	23.0	12	38
PE	231	25.7	5.12	25.0	10	40
Science	281	22.3	3.59	22.0	12	35
Social Science/History	354	24.1	4.91	23.0	12	39
World Language	103	21.6	3.60	21.0	11	32

Table 2: Cycle 2 Performance across Content Areas

Field	N	Mean	SD	Median	Min	Max
Multiple Subject: Literacy	760	26.0	3.76	26.0	9	44
Multiple Subject: Math	410	24.9	3.61	25.0	15	37
Agriculture	10	24.7	3.37	25.0	17	29
Art	30	27.0	3.46	27.0	18	36
Business						
English Language Dev						
English	158	27.4	3.30	27.0	16	42
Health Science	9	23.0	2.06	23.0	21	27
Home Economics						
Industrial & Tech. Ed						
Mathematics	85	25.1	4.27	25.0	11	32
Music	27	22.6	4.91	22.0	14	32
PE	93	26.1	5.05	26.0	12	38
Science	111	26.1	4.45	25.0	17	37
Social Science/History	121	27.0	3.92	27.0	15	38
World Language	32	25.6	4.40	25.0	17	33

Frequency Distribution of the Cycle 1 and Cycle 2 Scores

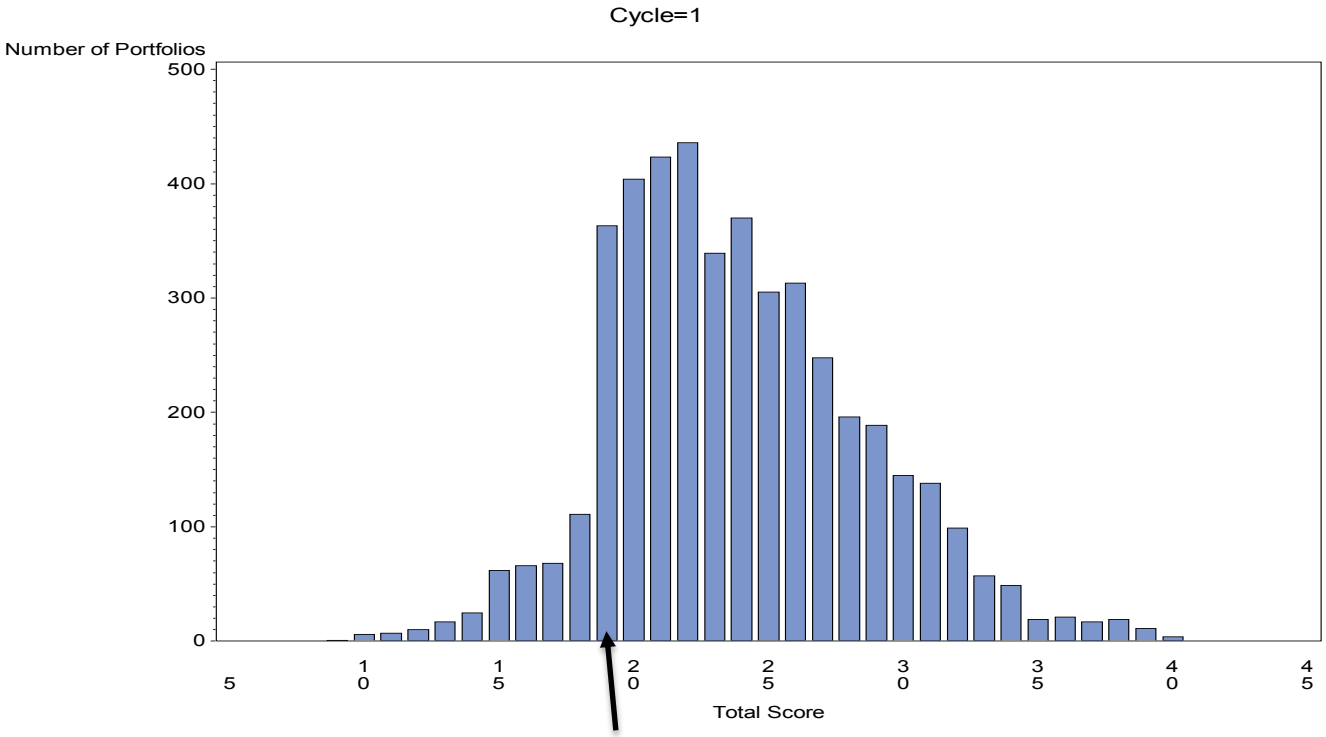
On the next page are two bar charts showing the distribution of the scores for all candidates who had complete cycle submissions submitted and scored through April 25, 2019. An arrow has been added to each chart showing the proposed passing standard (Cycle 1: a score of 19 and Cycle 2: a score of 21).

This chart shows that all Multiple Subject and Single Subject submissions at the passing standard and to the right would pass the cycle if the proposed passing standard were adopted. The Multiple Subject and Single Subject submissions to the left of the arrow would be not pass the cycle. In addition, each of the two charts illustrating the frequency distribution of scores have different total numbers of candidates. Cycle 1 had 4,538 candidates while the chart for

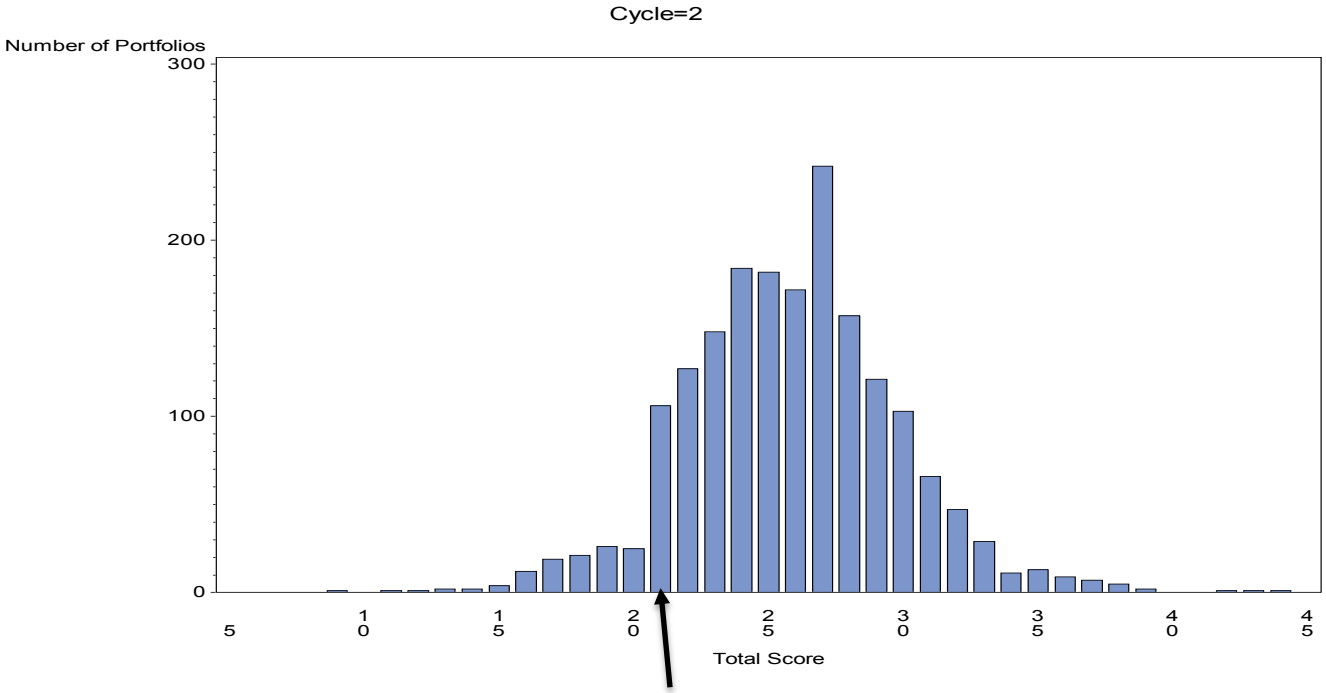
Cycle 2 has 1,848. The reason for this difference in number of candidates in each is that the data informing these charts represents data available as of April 2019. At this April date, less candidates had completed and submitted Cycle 2 for scoring. It is not known at this time how the data might shift as more Cycle 2 submissions are completed, but enough candidate completions are available for a general application of the data to inform a defensible determination of a passing standard for Cycle 2.

It is important to note that candidates must pass both cycles to meet the TPA requirement. These charts illustrate each cycle's performance data separately.

Frequency Distribution of Scores on CalTPA Cycle 1 (n=4,538)



Frequency Distribution of Scores on Cycle 2 (n=1,848)



After reviewing impact data, including the reporting of the modeled passing rate that would have been obtained based on a range of possible cut scores and viewing this information through various demographic variables, the whole group discussed the inferences of the impact data on the initial passing standard recommendation. Following this discussion, panelists were asked to make a final recommendation for a cut score, by individual cycle. The Rubrics for the Multiple Subject cycles can be found beginning on page 16 of the PDF Appendix.

Cycle 1 (8 rubrics):

- A final cut score of **19 points** was the committee-recommended median.
- 19 of the 20 panelists recommended a condition, with 14 recommending not allowing a rubric score of "1", and 5 recommending allowing 1 rubric score of "1".

Cycle 2 (9 rubrics):

- A final cut score of **21 points** was the committee-recommended median.
- 19 of the 20 panelists recommended a condition, with 14 recommending not allowing a rubric score of "1", and 5 recommending allowing 1 rubric score of "1".

To conclude the meeting, panelists were shown the frequencies for individual ratings of a final recommended cut score, as well as the mean, median, mode, minimum and maximum for the final cut score recommendations (See Tables 1 and 2). Panelists were also shown a summary of the condition recommendations – both the frequency and associated recommendation. These tables are presented below, and additional SEM impact tables are provided in Appendix F.

Table 3: Summary of Panel Ratings

Cycle 1 (8 rubrics) – 20 Panelists:

Avg	**18.90**
SD	0.91
Median	**19.00**
Min	18.00
Max	21.00
Mode	18.00

Frequency of Ratings:

15	0
16	0
17	0
18	8
19	7
20	4
21	1
22	0
23	0

Cycle 2 (9 rubrics) – 20 Panelists:

Avg	**20.85**
SD	1.23
Median	**21.00**
Min	18.00
Max	23.00
Mode	21.00

Frequency of Ratings:

17	0
18	1
19	2
20	3
21	8
22	5
23	1
24	0

Consideration of the Standard Error of Measurement (SEM)
Once the final panel score recommendation is determined, an additional modification is sometimes made to that score before it is presented to the Commission for potential adoption. This modification is the determination and potential application of an adjustment known as the Standard Error of Measurement (SEM).

The Standard Error of Measurement takes into consideration the fact that an assessment represents one single point in time when a candidate's knowledge, skills, and abilities are measured. The score obtained on that particular day may or may not be fully reflective of the candidate's true knowledge, skills, and abilities. If, for example, a candidate were to retake the test on multiple occasions, the candidate might well obtain several different scores. Scores are influence by many factors, including, for example, the candidate's health on a particular day, the candidate's frame of mind, the point in the program at which the candidate takes the test, differences in the ratings given by the assessors, and other such factors that may have an influence on the score received on that particular assessment on that particular day. The candidate's "true" score that most accurately reflects the candidate's full set of knowledge, skills, and abilities, may lie somewhere within that total range of scores, and not just in one score obtained on one particular date in time. In addition, a single score could also potentially represent a "false negative" (i.e., the candidate did have sufficient knowledge, skills, and abilities but the actual score did not closely enough represent the candidate's true abilities) or a "false positive" (i.e., the candidate did not actually have sufficient knowledge, skills, and abilities but was able to earn a higher score than otherwise warranted). For these reasons, an adjustment for this "standard error of measurement," or SEM, may sometimes be made to address these factors.

The SEM will vary depending on the total points possible, and the variability of the panelist's ratings as reflected in the panel median. Therefore, the SEM has been calculated for cycle 1 and cycle 2, and the corresponding cut scores at different SEM applications are found in the tables presented below. These tables also provide the impact data on scores and passing rates for both cycles, including the effect of adding an SEM adjustment. Given the panel recommendations, the data was modeled to show not only individual pass rates at the cycle level (as illustrated above) but also modeled pass rates for those 1,814 candidates who completed both cycles 1 and 2, taking into account scenarios of no conditions, and only one "1" (allowing for one rubric score of "1" in both cycles), and allowing no "1s" in either cycle as presented below.

Tables 4-6 provide data on the assessment pass rates with scores on both cycles at the panel recommended cut score. For information purposes, the highlighted boxes in these tables reflect the application of (a) -1 standard error of measure (SEM); (b) the application of -0.5 SEM; and (c) the application of no SEM to the total score. Table 4 depicts what the pass rates on the CalTPA would be if no side conditions (e.g., limitations on the number rubrics scoring a "1") are applied during scoring, Table 5 depicts the pass rates that would occur if the Commission allowed one score of "1" on each cycle, and Table 6 depicts pass rates if no scores of 1 are allowed.

Table 4: % Pass Assessment - No Side Condition for Any Cycle

	Cycle 2 Cut Score										
	17 [a]	**18**	**19 [b]**	**20**	**21 [c]**	**22**	**23**	**24**	**25**	**26**	**27**
Cycle 1 Cut Score											
15 [a]	**98%**	97%	95%	94%	93%	87%	80%	72%	63%	53%	44%
16	97%	96%	94%	93%	92%	86%	80%	72%	62%	53%	44%
17 [b]	95%	94%	**93%**	92%	91%	85%	79%	71%	62%	52%	43%
18	94%	93%	92%	90%	89%	84%	78%	71%	61%	52%	43%
19 [c]	91%	90%	89%	88%	**87%**	82%	76%	69%	60%	51%	42%
20	82%	82%	81%	80%	79%	75%	70%	63%	55%	47%	39%
21	73%	72%	72%	71%	70%	67%	63%	58%	50%	43%	36%
22	63%	63%	62%	62%	61%	59%	55%	51%	44%	38%	32%
23	54%	53%	53%	52%	52%	50%	47%	44%	38%	33%	28%
24	46%	46%	46%	45%	45%	43%	41%	38%	34%	30%	25%

a = -1 SEM b = - 0.5 SEM c = no SEM

Table 5: % Pass Assessment – At Most One 1 in Each Cycle

	Cycle 2 Cut Score										
	17 [a]	18	19 [b]	20	21 [c]	22	23	24	25	26	27
Cycle 1 Cut Score											
15 [a]	**97%**	96%	95%	94%	92%	87%	80%	72%	63%	53%	44%
16	96%	95%	94%	93%	92%	86%	80%	72%	62%	53%	44%
17 [b]	94%	94%	**93%**	92%	90%	85%	79%	71%	61%	52%	43%
18	93%	93%	91%	90%	89%	84%	78%	70%	61%	52%	43%
19 [c]	91%	90%	89%	88%	**87%**	82%	76%	69%	60%	51%	42%
20	82%	81%	81%	80%	79%	75%	70%	63%	55%	47%	39%
21	72%	72%	71%	71%	70%	67%	63%	57%	50%	43%	36%
22	63%	63%	62%	62%	61%	59%	55%	51%	44%	38%	32%
23	53%	53%	53%	52%	52%	50%	47%	44%	38%	33%	28%
24	46%	46%	46%	45%	45%	43%	40%	38%	34%	30%	25%

a = - 1 SEM b = - 0.5 SEM c = no SEM

Table 6: % Pass Assessment - No 1s Allowed in Either Cycle

	Cycle 2 Cut Score										
	17 [a]	18	19 [b]	20	21 [c]	22	23	24	25	26	27
Cycle 1 Cut Score											
15 [a]	**83%**	83%	82%	82%	81%	77%	73%	67%	58%	50%	41%
16	83%	83%	82%	82%	81%	77%	73%	67%	58%	50%	41%
17 [b]	82%	82%	**82%**	81%	80%	77%	72%	66%	58%	49%	41%
18	81%	81%	81%	80%	79%	76%	72%	66%	57%	49%	41%
19 [c]	79%	79%	79%	78%	**78%**	74%	70%	64%	56%	48%	40%
20	72%	72%	72%	72%	71%	68%	65%	59%	52%	45%	37%
21	65%	65%	65%	64%	64%	62%	59%	54%	48%	41%	34%
22	57%	57%	57%	56%	56%	54%	52%	48%	43%	37%	31%
23	48%	48%	48%	48%	48%	46%	44%	41%	37%	32%	27%
24	42%	42%	42%	41%	41%	40%	38%	36%	33%	29%	24%

a = - 1 SEM b = - 0.5 SEM c = no SEM

Tables 7-8 provide the percentage of submissions that would pass at a variety of cut scores. The tables also provide data on how applying side conditions and limiting the number of scores of 1 a candidate may earn, are also shown.

Table 7: Cycle 1-Overall Modeled Passing Rates by Cut Score and Number of Candidate Scores of 1 Allowed

Cut Score	Total N	No Side Condition		At most three 1s		At most two 1s		At most one 1		No 1s allowed	
		N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass
15	4,537	4,471	0.99	4,469	0.99	4,468	0.98	4,448	0.98	4,201	0.93
16	4,537	4,409	0.97	4,409	0.97	4,409	0.97	4,402	0.97	4,201	0.93
17	4,537	4,343	0.96	4,343	0.96	4,343	0.96	4,336	0.96	4,168	0.92
18	4,537	4,275	0.94	4,275	0.94	4,275	0.94	4,271	0.94	4,124	0.91
19*	**4,537**	**4,164**	**0.92**	**4,164**	**0.92**	**4,164**	**0.92**	**4,161**	**0.92**	**4,031**	**0.89**
20	4,537	3,801	0.84	3,801	0.84	3,801	0.84	3,799	0.84	3,705	0.82
21	4,537	3,397	0.75	3,397	0.75	3,397	0.75	3,395	0.75	3,328	0.73
22	4,537	2,974	0.66	2,974	0.66	2,974	0.66	2,973	0.66	2,929	0.65
23	4,537	2,538	0.56	2,538	0.56	2,538	0.56	2,538	0.56	2,508	0.55
24	4,537	2,199	0.48	2,199	0.48	2,199	0.48	2,199	0.48	2,182	0.48

*Panel-recommended median

Table 8: Cycle 2-Overall Modeled Passing Rates by Cut Score and Number of Candidate Scores of 1 Allowed

Cut Score	Total N	No Side Condition		At most three 1s		At most two 1s		At most one 1		No 1s allowed	
		N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass
17	1,848	1,825	0.99	1,825	0.99	1,823	0.99	1,812	0.98	1,628	0.88
18	1,848	1,806	0.98	1,806	0.98	1,806	0.98	1,801	0.97	1,628	0.88
19	1,848	1,785	0.97	1,785	0.97	1,785	0.97	1,781	0.96	1,620	0.88
20	1,848	1,759	0.95	1,759	0.95	1,759	0.95	1,755	0.95	1,604	0.87
21*	**1,848**	**1,734**	**0.94**	**1,734**	**0.94**	**1,734**	**0.94**	**1,730**	**0.94**	**1,585**	**0.86**
22	1,848	1,628	0.88	1,628	0.88	1,628	0.88	1,626	0.88	1,512	0.82
23	1,848	1,501	0.81	1,501	0.81	1,501	0.81	1,499	0.81	1,419	0.77
24	1,848	1,353	0.73	1,353	0.73	1,353	0.73	1,351	0.73	1,298	0.70
25	1,848	1,169	0.63	1,169	0.63	1,169	0.63	1,168	0.63	1,139	0.62
26	1,848	987	0.53	987	0.53	987	0.53	986	0.53	968	0.52
27	1,848	815	0.44	815	0.44	815	0.44	814	0.44	802	0.43

*Panel-recommended median

The Passing Standards Recommendation from the Expert Panel

At the final stage of the Standard Setting process, the panel arrived at the following recommendations for scores for cycle 1 and cycle 2. In addition, the majority of panelists recommended that no scores of 1 on a rubric would be allowed. The expert panel's recommendation to the Commission is as follows:

- Cycle 1 (8 rubrics): A final cut score of **19 points with no scores of 1.**
- Cycle 2 (9 rubrics): A final cut score of **21 points with no scores of 1.**

Staff Recommendation and Rationale

Staff agrees with the standard setting panel that the goal is to have candidates who are eligible for a preliminary teaching credential earn no scores of "1" on the CalTPA. However, only 27 of the 60 CalTPA using institutions participated in the pilot and/or the field test of the new CalTPA. With so many new programs using the CalTPA in 2018-19, it is possible that some of these programs were not as knowledgeable about the evidence required via tasks, the essential questions and the rubric descriptors of performance. Given this consideration of all program's capacity to meet the requirements of the new TPEs and the redeveloped CalTPA, staff through careful discussion and review of data from scored MS and SS cycles, in support of programs, are making an altered recommendation.

Staff recommends that the Commission adopt the following as the passing standard for the CalTPA for the 2019-20 year, effective as of the August 23, 2019 submission window:

- Cycle 1 (8 rubrics): A final cut score of **19 points with one score of 1 allowed.**
- Cycle 2 (9 rubrics): A final cut score of **21 points with one score of 1 allowed.**

Staff suggests that for at least the next two years, a candidate who earns a score of "1" on a rubric but also earns a 19 on Cycle 1 and a 21 on Cycle 2 should pass the assessment. Data will be analyzed during the next two years of administration and staff will, in two years time, report on candidate performance and invite the Commission to reconsider the passing standard at that time.

Finally, because staff is recommending an adjustment to the passing standards through the side condition of allowing a score of "1" on each cycle, staff is not recommending applying an SEM to the passing score standards.

Next Steps

If the Commission adopts the recommended passing score, standards with the side condition of no more than one score of 1 per cycle notification will be posted on the CalTPA website and distributed to the field. The passing standard adopted by the Commission will be applied to all subsequent administrations of the CalTPA upon the effective date, August 23, as determined by the Commission.

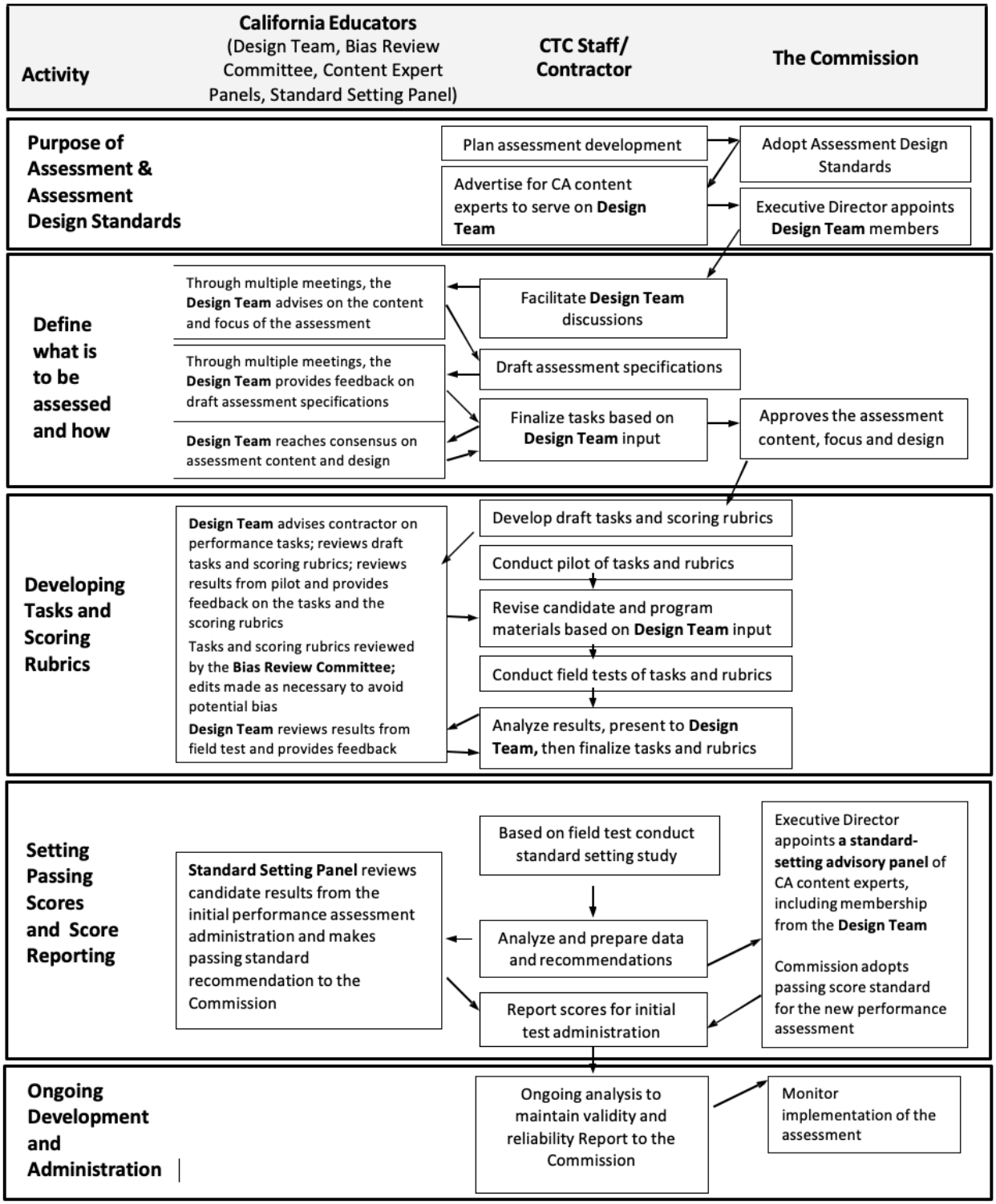
Appendix A

California Teaching Performance Assessment (CalTPA) Design Team

Member	Affiliation
Rebecca Ambrose	University of California, Davis
Paul Boyd-Batstone	California State University, Long Beach
Jorge Colmenero	RFK UCLA K-12 Community School/Los Angeles/LAUSD
Nedra Crow	National University (San Diego)
Brent Duckor	San Jose State University
Karen Escalante	California State University, San Marcos
Meredith Fellows	Cal State TEACH
Annamarie Francois	University of California, Los Angeles
Fred Freking	University of Southern California
Donna Glassman-Sommer	Tulare County Office of Education
Kim Harrison	Washington Unified School District
Jose Lalas	University of Redlands
Edmundo Litton	Loyola Marymount University
Helene Mandell	University of San Diego
Beth Roybal	Salinas Union High School District
Donna Scarlett	Reach Institute for School Leadership
David Sloan	Brandman University
Emily Vazirian	Olive Crest Academy
Mick Verdi	California State University, San Bernardino
Patricia Wick	Brandman University
Tine Sloan	Commission Liaison

Appendix B

Groups Involved in the Redevelopment of the CalTPA



Appendix C

CalTPA Programs 2018-19

California State University
Cal State Poly, Pomona
CSU, Bakersfield
CSU, Chico
CSU, Dominguez Hills
CSU, Fullerton
CSU, Long Beach
CSU, San Bernardino
CSU, San Marcos
CSU, Stanislaus
CalState TEACH
San Jose State University
Sonoma State University

Local Education Agencies (LEAs)
Bay Area School of Enterprise (Reach Institute)
High Tech High
Los Angeles County Office of Education
Los Angeles Unified School District
Mt. Diablo Unified School District
Sacramento County Office of Education
Tulare County Office of Education

University of California (UC)
University of California, Santa Cruz
University of California, Los Angeles

Private Colleges and Universities
Academy of Art University
Antioch University
Azusa Pacific University
Bard College
Biola University
Brandman University
Chapman University
Claremont Graduate University
Concordia University
Dominican University of California
Fresno Pacific University
Hebrew Union College
Holy Names University
Hope International University
Humphreys University
La Sierra University
Mount Saint Mary's University
National University
Pacific Oaks College
Pacific Union College
Point Loma Nazarene University
San Diego Christian College
Santa Clara University
Simpson University
Teachers College of San Joaquin
The Master's University
United States University
University of La Verne
University of Redlands
University of San Diego
University of San Francisco
Vanguard University
Westmont College
Whittier College
William Jessup University

Appendix D

CalTPA Standard-Setting Panel Meeting
Sacramento, CA
May 8-9, 2019

CalTPA Standard-Setting Panelists

Member	Affiliation
Jaime Colly	University of San Francisco
Lara Ervin-Kassab	San Jose State University
Tsehsien (Kelly) Vaughn	Notre Dame de Namur University
Victoria Bisorca	CSU Long Beach
Ernest Black	CalState TEACH, Long Beach
Karen Escalante	CSU Long Beach
Meredith Fellows	Emeritus CalState TEACH, CSU Los Angeles
Robert Frelly	Chapman University
Amy Gimino	Cal Poly Pomona
Melissa Hall	Riverside County Office of Education
Debbie Meadows	CSU Bakersfield
Ricardo Medina	University of San Diego
Laura Pelaez Guzman	CalState TEACH, CSU Los Angeles
Stacy Schmidt	CSU Bakersfield
Nancy Hayes	Rocklin Unified School District
Caryl Hodges	Notre Dame de Namur University
Debra Reeves-Gutierrez	Natomas Unified School District
Beth Roybal	Salinas Union High School District
Daniel Soodjinda	CSU Stanislaus
Patricia Wick	Brandman University

Appendix E

Detailed Description of the Standard Setting Process for the Redeveloped CalTPA

The CalTPA Standard Setting Study Process

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 threshold, or "passing standard." The expert educators on the Standard Setting Panel represented credentialed TK-12 teachers, CalTPA assessors, CalTPA Design Team members, and teacher preparation program faculty who had previously worked with the CalTPA. Demographic information about the educators who served on the standard setting panels is provided in Appendix D.

As with the standard setting study method used for all other Commission examinations (e.g. CBEST, CSET, CTEL, RICA, CPACE), the process employed for the CalTPA was consistent with recognized psychometric principles and procedures. The standard setting study for the CalTPA was conducted on May 8-9, 2019, with pre-conference activities occurring prior to the meeting. An agenda for the meeting is provided in Appendix E.

Prior to the meeting, each invited panelist received CalTPA Guides, rubrics, and six previously scored sample submissions (three from each Cycle) representing different performance levels across various content areas. Panelists were asked to review materials submitted by candidates and the scoring information for the submissions that were assigned to them prior to arriving at the Standard Setting. The purpose of the pre-work was to ensure that participants were able to 1) observe the shared pedagogical content (common trunk of knowledge) across different content areas; 2) gain some exposure to a range of candidate responses; and 3) apply that information in the policy capture activities (activities drawing upon the panelists' experience and discussion) at the meeting.

The CalTPA standard setting meeting began with an orientation and training session. Panelists were informed of the purpose of the assessment and provided with a briefing book to guide their activities.

Throughout the standard setting event, both a context statement and a guiding question were used and revisited to frame all discussions. This statement and question provided a common framework in which all participants could anchor their decisions:

- Think about a teacher candidate who is just at the level of knowledge and skills required to perform effectively the job of a new teacher in California public schools.
- Guiding question: What total score (the sum of all rubric scores in the Cycle) represents the level of performance achieved by this individual?

Panel members were asked to conceptualize the content-specific pedagogy of a hypothetical beginning teacher who would be competent to teach in the subject area. Panel members used

this concept of what a minimally competent beginning teacher would know and be able to demonstrate in determining their recommended acceptable score for passing each of Cycle 1 and Cycle 2. Although a number of candidates may exceed the level of acceptable content-specific pedagogy, none receiving a passing score should fall below this minimally competent level. The panel also reviewed the rubrics used to evaluate the cycle steps in the CalTPA.

After this extensive training and the assessment review, panel members completed the following 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 content-specific pedagogy necessary for a beginning practitioner just competent to begin professional practice as a multiple or single subject teacher.

During the facilitated session, panelists familiarized themselves with the assessment and with the information contained in the briefing book. After a series of policy capture activities, panelists recommended an initial cut score (which may also be referred to as a "passing standard") for each cycle, which was then reviewed and discussed. Following that, panelists individually recommended a final cut score for each cycle.

Policy Capture 1 Activity Overview/Instructions

In this activity, individuals were assigned in table groups with panelists who had reviewed the same submission for the pre-work assignment. To begin, each panelist individually spent some time recalling the specific submission that they reviewed for the pre-work and then provided an individual rating for that cycle submission (see ratings description that follows), completing an individual rating form for the cycle submission reviewed.

Then, the panelists discussed their ratings with other panelists with the goal of arriving at a consensus table rating. Upon reaching consensus, each table completed one consensus rating form for the cycle submission discussed.

After each table completed the table form, panelists moved to the next table assignment and repeated the process two more times for the other submissions they reviewed for pre-work. By the end of the three cycles, individual ratings and table ratings were generated for each of the cycle submissions reviewed by each individual and group.

This process was completed once for Cycle 1, and again for Cycle 2, with six submissions reviewed and discussed by each panelist.

The activities previously described included a rating form with four rating levels from which to select.

Four Rating Levels	
Clearly below the passing standard	CLEARLY NOT performing effectively the job of a new teacher. This teacher has demonstrated one or more major problems in teaching knowledge,

	skills or abilities that require remediation and may need additional time and opportunity for learning and improvement.
Just below the passing standard	APPROACHING but NOT YET effective in performing the job of a new teacher. This teacher has demonstrated some strengths but has one too many issues in teaching knowledge, skills or abilities that will keep him/her from being effective.
Just meets the passing standard	JUST MEETS your definition of performing effectively the job of a new teacher. This teacher has demonstrated some consistent strengths in teaching knowledge and skills and has a foundation on which to build. The teacher may have shown one or more minor flaws in teaching knowledge, skills or abilities that will likely improve with more time and experience.
Clearly above the passing standard	CLEARLY EXCEEDS your definition of performing effectively the job of a new teacher. This teacher has demonstrated clear strengths in teaching knowledge, skills and abilities, and a strong foundation for effective teaching.

All individual and table ratings were tabulated. Data from the individual ratings of the policy capture activity were then presented to the panel. After some discussion of the individual and table ratings, each table discussed a score range (e.g., a lower and upper bound total score) that may include the potential cut score.

The committee's ratings and review determined that score profiles with a range as follows were appropriate for review and discussion.

- Cycle 1: Total scores between 16-20
- Cycle 2: Total scores between 20-23

Given this range, a set of "Candidate Score Profiles" was reviewed by the panelists. Through Standard Setting Policy Capture 1 and the subsequent discussions, panelists began to come to consensus around a common range within which the passing standard would likely be recommended (from widely divergent to less divergent).

Score Profile Review and Discussion Activity

As part of this activity, panelists reviewed a set of "Candidate Score Profiles" within the total score range between:

- Cycle 1: Total scores between 16-20
- Cycle 2: Total scores between 20-23

The Candidate Score Profiles represented a sample of candidate scores (individual rubric scores and total scores), and the rubric descriptors that correspond to each rubric score. See the PDF Appendix for sample Candidate Score Profiles at total scores of 19 (Cycle 1) and 21 (Cycle 2) as well as the Cycle 1 and Cycle 2 Rubrics. Using only the score profiles and rubric descriptors (i.e., not considering the submission itself), panelists evaluated the score profiles against the

common framing of "a teacher candidate who is just at the level of knowledge and skills required to perform effectively the job of a new teacher in California public schools."

All panelists reviewed the same set of Candidate Score Profiles as a group, for each Cycle. The group was asked to review the information to confirm the range of scores within which the passing standard would likely be recommended. Panelists discussed the score profiles and reported out their perception of candidate performance within the upper and lower limits of the score range. Through the Score Profile review and the subsequent discussions, panelists continued to come to consensus around a common range within which the passing standard would likely occur.

Appendix F

Standard Error of Measurement and Panel Recommendations

Panel recommendations (highlighted) and corresponding Standard Error of Measurement (SEM) values at +/- 1.0 SEM

CYCLE 1: 8 Rubrics (Length = 40)

Cut score	SEM	-1	-0.5	+0.5	+1
15.00	3.10	11	13	16	18
16.00	3.14	12	14	17	19
17.00	3.17	13	15	18	20
18.00	3.19	14	16	19	21
19.00	3.20	15	17	20	22
20.00	3.20	16	18	21	23
21.00	3.20	17	19	22	24
22.00	3.19	18	20	23	25
23.00	3.17	19	21	24	26
24.00	3.14	20	22	25	27
25.00	3.10	21	23	26	28

CYCLE 2: 9 Rubrics (Length = 45)

Cut score	SEM	-1	-0.5	+0.5	+1
17.00	3.29	13	15	18	20
18.00	3.32	14	16	19	21
19.00	3.35	15	17	20	22
20.00	3.37	16	18	21	23
21.00	3.38	17	19	22	24
22.00	3.39	18	20	23	25
23.00	3.39	19	21	24	26
24.00	3.38	20	22	25	27
25.00	3.37	21	23	26	28
26.00	3.35	22	24	27	29
27.00	3.32	23	25	28	30