2A

Action

Educator Preparation Committee

Proposed Adoption of the Passing Score Standard for the California Administrator Performance Assessment

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

Recommended Action: That the Commission adopt the recommended passing score standard for the CalAPA.

Presenters: Amy Reising, Director, Performance Assessment Development, and Gay Roby, Consultant, Professional Services Division, 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 the Passing Score Standard for the California Administrator Performance Assessment

Introduction
This agenda item provides (a) foundational information about the standard setting process for Commission-developed examinations and assessments and (b) recommendations for the passing score standard for the California Administrator Performance Assessment (CalAPA) for Preliminary Administrative Services Credentialing programs (PASC).

Background
PASC programs, in partnership with the Commission, an appointed Design Team and Evaluation Systems group of Pearson (ES) have been engaged in the development and implementation of the CalAPA for the past four years, 2015-present. In June, a standard setting panel was convened to conduct a review of the 2018-19 candidate score data and determine a passing standard to recommend to the Commission for the 2019-20 administration. The membership of the CalAPA Design Team is provided in Appendix A, the description of the process and the groups involved in the development of the CalAPA is provided in Appendix B, the list of administrator preparation programs using the CalAPA is provided in Appendix C, and the membership of the CalAPA standard setting panel is provided in Appendix D.

The Commission began discussing the development of a performance assessment for administrative services credential candidates in September 2012, passing an action item approving the idea in September 2013. In 2015 the project was funded through the state budget act and development work began. To ensure scoring reliability, the CalAPA is centrally managed and scored by calibrated assessors to ensure that detailed, analytic, feedback based on the California Administrators' Performance Expectations (CAPE) is provided to candidates and programs in a timely manner to guide both candidate development and program improvement. As a result, CalAPA candidate data is consistent and reliable and aligns with the needs of the Commission's Accreditation Data System (ADS) providing an outcomes-based set of quality indicators to guide review of administrator preparation programs.

The CalAPA model has been developed, pilot- and field-tested, and implemented in a non-consequential year (candidates did not pay an assessment fee and were not required to meet a passing standard but were required to register, complete, and submit each of the three cycles for scoring), and is now ready for consequential statewide use (candidates register, pay an assessment fee for each of the three cycles, and must meet a Commission-adopted passing standard). After four years of development and testing, it is now the appropriate time for the Commission to set passing score standard for the CalAPA for the 2019-20 consequential administration.

During the Redeveloped CalAPA Pilot (2016-17) and Field-Testing Phases (2017-18)
The CalAPA was piloted by 23 institutions and 304 candidates in the 2016-17 academic year, then revised during summer 2017 based on the pilot test findings, and subsequently field tested by 23 programs and 438 candidates during the 2017-18 academic year. The results of the field test led to another round of revisions of the CalAPA cycles and rubrics prior to statewide, non-consequential operational administration of the assessment in the 2018-19 year. While the majority of the 23 PASC programs joined both the pilot and the field tests, not all the same programs participated in both tests.

CalAPA First Operational/Non-Consequential Year (2018-19) Data Leading to a 2019 Standard Setting Study
Following the field test conducted in 2017-18, the CalAPA became operational for all PASC programs in fall 2018. During the non-consequential year, implementation was required of all programs, with candidates required to fully complete and submit all three cycles of the CalAPA. Candidates were not required to meet a passing standard in order to be recommended for the preliminary Administrative Services Credential or Certificate of Eligibility. For purposes of the 2019 June standard setting study, all candidates who completed cycle submissions submitted and scored through June 6, 2019 are included in the data analysis.

Description of Three Cycles

Leadership Cycle 1
Leadership Cycle 1 focuses on analyzing multiple sources of school data for the purpose of identifying equity gaps to inform an initial draft plan for equitable improvement in line with the school's vision, mission, and goals. Within the cycle of investigate, plan, act, and reflect, candidates collect and analyze multiple sources of longitudinal quantitative and qualitative data. They then conduct an equity gap analysis to identify potential factors, institutional and/or structural, all culminating in a problem statement defining a specific area of educational need related to equity and seek input from a stakeholder(s) at the school site. To close, the candidate reflects on equitable leadership.

Leadership Cycle 2
Leadership Cycle 2 focuses on facilitating collaborative professional learning within a community of practice for the purpose of improving teaching and student learning or well-being. Within the cycle of investigate, plan, act, and reflect, candidates begin by identifying and working with a small group of educators to identify a problem of practice. That group identifies an evidence-based instructional strategy to address the problem of practice that will strengthen and increase equitable learning and/or well-being for all students. They then facilitate meetings with the group and collaboratively lead the professional learning of the community of practice. In addition, candidates reflect on how their facilitation supports the group to address the problem of practice, understands early implementation findings, and how they responded to the group's feedback on their facilitation.

Leadership Cycle 3

Leadership Cycle 3 focuses on coaching an individual teacher to strengthen teaching practices and improve student learning and/or well-being. Within the cycle of investigate, plan, act, and reflect, candidates familiarize themselves with coaching and observation practices at the school; identify a volunteer teacher with whom they coach; and conduct a full coaching cycle, including a pre-observation meeting, a focused classroom observation to collect CSTP-related evidence of practice, and conduct a post-observation meeting. Throughout this leadership cycle, candidates reflect on their strengths and areas for professional growth as a coach and an equity minded leader.

Comparison between CalAPA and CalTPA Requirements

While there are many similarities in the structure of the CalAPA and CalTPA (inquiry-based cycles made up of multiple tasks, analytic rubrics, centralized scoring) it should be noted that there are important differences.

- The Teaching Performance Assessment (CalTPA) centers around a candidate's ability to teach and assess one group of students, while the CalAPA deals with multiple levels of stakeholder groups (e.g. district and schools administrators, staff members, teachers, and PreK-12th grade students, family units and community members). For the majority of California's administrator candidates, this is a growth role for them, requiring preparation, and knowledge, before candidates can accomplish the tasks that the CalAPA asks of them.
- PASC programs need to work with district and school partners to inform them about the new CalAPA requirement and support districts to understand their increased role to support candidates. Revised program standards establish the heightened responsibilities of both partners.
- Two of the three CalAPA cycles require a video-recording with annotation, a new skill set.
- The participation of other educators is included in two of the CalAPA cycles. Cycle 2 asks for a group of three to five teachers, who are willing to devote time to improving their own practice and their students' achievements while providing honest feedback to the candidate. Cycle 3's volunteer is a colleague who is willing to reflect upon their own classroom practice and explore ways to improve and align with the standards for the teaching profession.

Due to these differences and expectations for implementation, PASC programs need time to embed the complex cycles in their coursework and clinical practice. Programs are working to retool the expectations for districts and schools and prepare faculty to understand the expectations of the cycles and rubrics so that in turn, they can support candidates as they engage with the CalAPA.

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 establishing 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[1] such as the CalAPA, standard setting is a content-focused, structured process in which a panel of education experts reviews the content of the assessment, carefully considers the performance expectations being measured as well as 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 ES resulted in a recommended cut score from the expert panel to the Commission for each of the three instructional cycles of the CalAPA.

There have been many different methods for standard setting developed, researched and published 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 CalAPA was consistent with recognized psychometric principles and procedures. The standard setting study for the CalAPA was conducted over a two-day period, June 25-26, 2019, with pre-conference preparatory activities taking place for the expert panel prior to the meeting. The specific standard setting process used during the meeting for the CalAPA 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 an administrator candidate who is just at the level of knowledge and skills required to perform effectively the job of a new administrator 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?

[1] Criterion-referenced assessments are designed to measure a candidate's knowledge, skills, and abilities in a specific domain of content. Candidates are assessed on the basis of their performance relative to a specific, predetermined criterion or standard.

The guiding questions addressed candidate performance across all rubrics in each cycle. Cycle 1 has eight rubrics while Cycles 2 and 3 have seven 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 for a Cycle under the final recommended passing score. The 22 scoring rubrics are provided in the large appendix.

Initial Passing Score Recommendation

Through a facilitated discussion, panelists were presented with CalAPA descriptive data, the activities described in Appendix E were conducted, and each panelist recommended an initial passing score during the early part of Day Two's activities.

To arrive at the final standard setting and passing standard recommendations to be presented 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 CalAPA cycle, a summary of the aggregate rubric, step of the cycle, and total CalAPA 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 CalAPA cycle. 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, including any recommendation regarding setting any side conditions for that cut score. Please note one panel member left the meeting prior to any initial or final rating discussions so the total number of panelists providing initial and final ratings is 18, not 19.

The **initial** cut score recommendation was as follows:

Cycle 1 (8 rubrics):

- A cut score of 19 points was the initial panel-recommended median.
- 13 panelists indicated a condition should be included, as follows:
    - 1 panelist indicated no scores of "1" allowed,
    - 11 panelists indicated no more than one score of "1" allowed, and
    - 1 panelist indicated no more than two scores of "1" allowed.
- 4 panelists indicated no conditions should be included
- 1 panelist abstained from a condition decision

Cycle 2 (7 rubrics):

- A cut score of 17 points was the initial panel-recommended median.
- 12 panelists indicated a condition should be included, as follows:
    - 2 panelists indicated no scores of "1" allowed,

    - 9 panelists indicated no more than one score of "1" allowed, and
    - 1 panelist indicated no more than two scores of "1" allowed.
- 5 panelists indicated no conditions should be included
- 1 panelist abstained from a condition decision

Cycle 3 (7 rubrics):

- A cut score of 17 points was the initial panel-recommended median.
- 13 panelists indicated a condition should be included, as follows:
    - 4 panelists indicated no scores of "1" allowed, and
    - 9 panelists indicated no more than one score of "1" allowed.
- 4 panelists indicated no conditions should be included
- 1 panelist abstained from a condition decision

Panelists were then 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 CalAPA 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, for all cycles during the first year of implementation. Rubric scores are shown for candidates' first attempt at passing each cycle.

CalAPA Rubric Scores for 2018-2019

Table 1: Cycle 1 Rubric Scores

Field	N	Mean	SD	Median	Min	Max
Cycle 1	1,385	23.1	5.50	23.0	8	40
Cycle 2	975	19.2	4.39	19.0	7	35
Cycle 3	906	20.1	3.53	20.0	8	33

Total N in Table 1 for each cycle is different because data was identified for use for the standard setting study as of June 6, 2019. N's in Table 1 represent the total number of scored cycles available at the June date. In early June, less candidates had submitted Cycles 2 and 3 but would go on to submit in late June and July. Programs do not need to complete cycles in any particular order, but these numbers do reflect that order of cycles started with completion of Cycle 1, and then 2 and 3.

Frequency Distribution of the Cycle 1, Cycle 2, Cycle 3 Scores

The following three bar charts show the distribution of the scores for all candidates who had complete cycle submissions submitted and scored through June 6, 2019. An arrow has been added to each chart showing the proposed passing standard (Cycle 1--a score of 19; Cycle 2—a score of 17; and Cycle 3—a score of 17).

These three charts show that all submissions at the passing standard and to the right would pass the cycle if the proposed passing standard were adopted. The submissions to the left of the arrow illustrate candidates who would not pass the cycle. In addition, each of the charts illustrating the frequency distribution of scores have different total numbers of candidates. As of the June date, fewer candidates had completed and submitted Cycles 2 and 3 for scoring. It is not known at this time how the data might shift, as more Cycle 2 and Cycle 3 submissions are completed, but sufficient candidate completions are available for a general application of the data to inform a defensible determination of a passing standard for Cycles 2 and 3.

It is important to note that each candidate must pass all three cycles to meet the APA requirement for earning the PASC. The following Figures 1.1-1.3 illustrate each cycle's performance data separately.

Figure 1.1 - Frequency Distribution of Scores on CalAPA Cycle 1 (n=1,385)

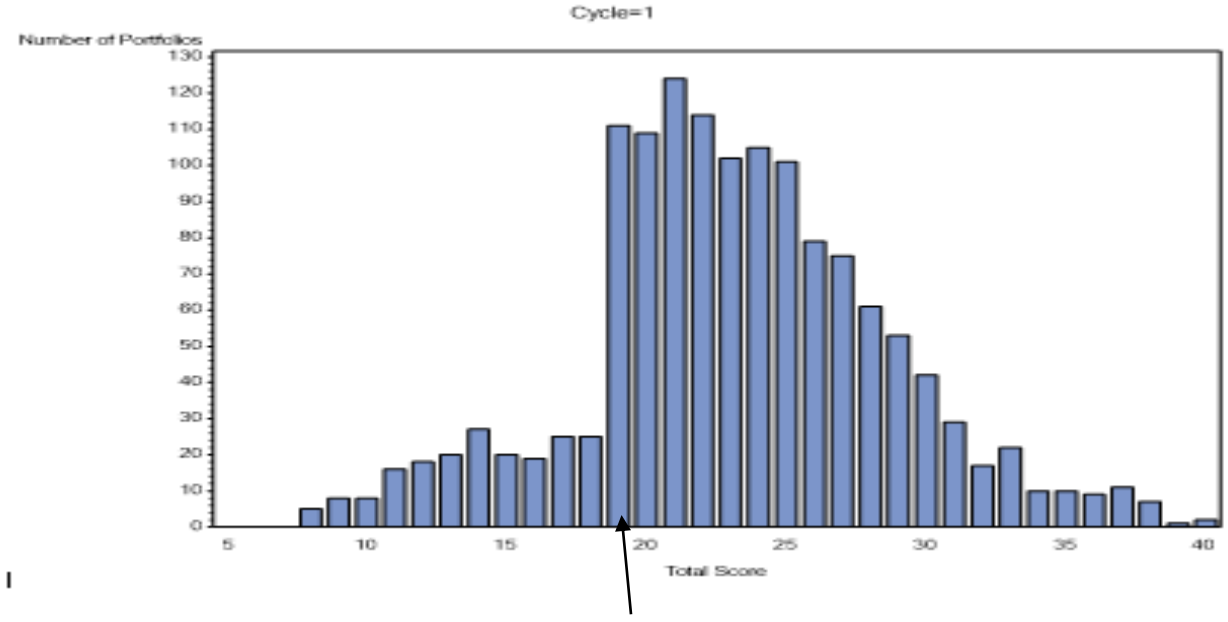


Figure 1.2 - Frequency Distribution of Scores on CalAPA Cycle 2 (n=975)

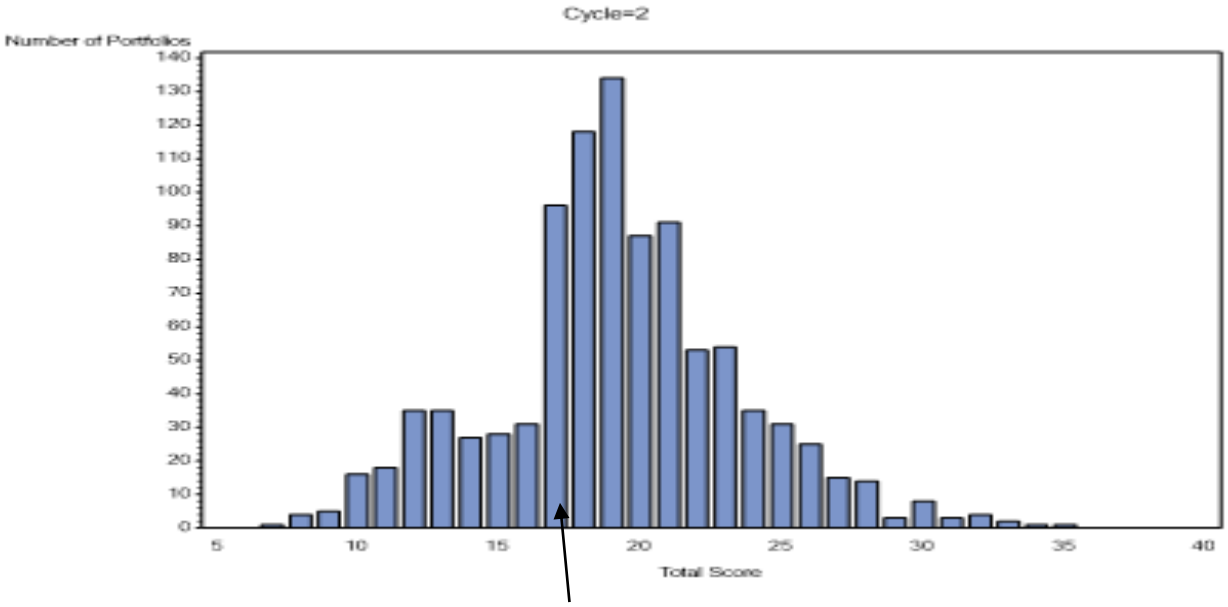
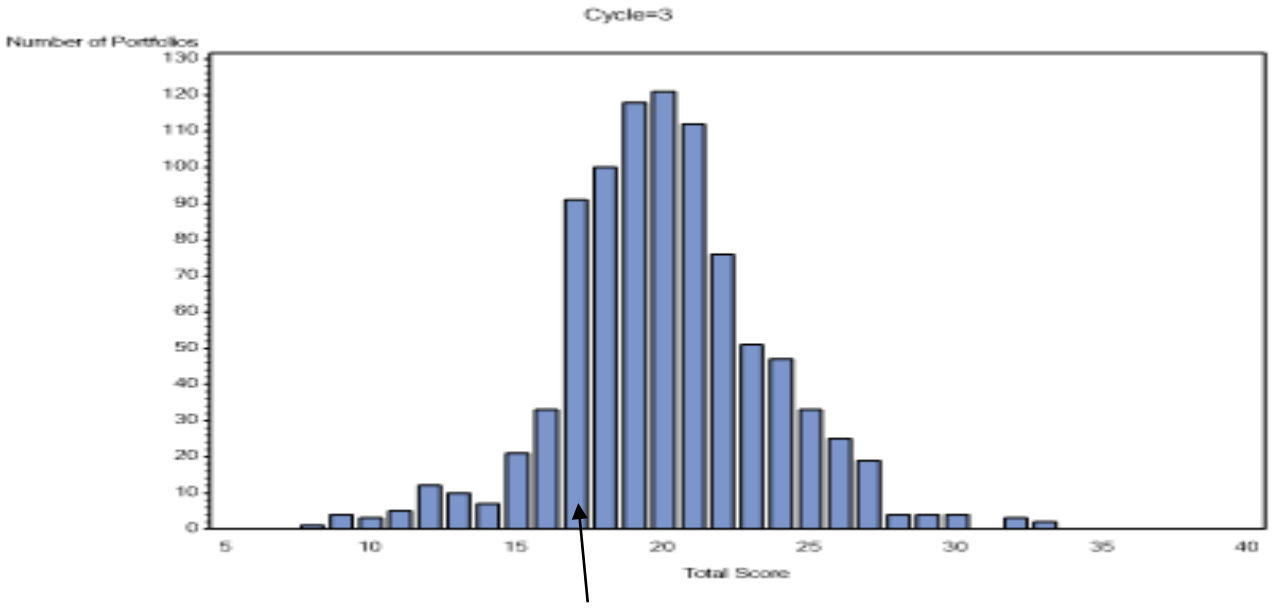


Figure 1.3 - Frequency Distribution of Scores on CalAPA Cycle 3 (n=906)



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 cycles can be found beginning on page 33 of the Supplemental Appendix.

Cycle 1 (8 rubrics):

- A final cut score of **19 points** was the panel-recommended median.
- 2 of the 18 panelists recommended a side-condition, with 1 recommending the allowance of a single rubric score of "1," and 1 panelist recommending allowing 2 rubric scores of "1." The majority of panelists did not recommend a side-condition.

Cycle 2 (7 rubrics):

- A final cut score of **17 points** was the panel-recommended median.
- 1 of the 18 panelists recommended a condition, with 1 recommending the allowance of a single rubric score of "1." The majority of panelists did not recommend a side-condition.

Cycle 3 (7 rubrics):

- A final cut score of **17 points** was the panel-recommended median.
- 1 of the 18 panelists recommended a condition, with 1 recommending the allowance of a single rubric score of "1." The majority of panelists did not recommend a side-condition.

Next, 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. 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 2: Summary of Panel Final Recommendations

Cycle 1 (8 rubrics) – 18 Panelists:

Avg	**19.00**
SD	0.00
Median	**19.00**
Min	19.00
Max	19.00
Mode	19.00

Frequency of Ratings:

Score	# Panelists
15	0
16	0
17	0
18	0
19	**18**
20	0
21	0
22	0
23	0

Cycle 2 (7 rubrics) – 18 Panelists:

Avg	**16.94**
SD	0.24
Median	**17.00**
Min	16.00
Max	17.00
Mode	17.00

Frequency of Ratings:

Score	# Panelists
13	0
14	0
15	0
16	**1**
17	**17**
18	0
19	0
20	0
21	0

Cycle 3 (7 rubrics) – 18 Panelists:

Avg	**17.00**
SD	0.00
Median	**17.00**
Min	17.00
Max	17.00
Mode	17.00

Frequency of Ratings:

Score	# Panelists
13	0
14	0
15	0
16	0
17	**18**
18	0
19	0
20	0
21	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 as it is recommended 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 SEM represents the uncertainty of our reliability in the measure of candidate performance and quantifies some error in measurement by taking into consideration the fact that an assessment represents a 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 reflective of the candidate's true knowledge, skills, and abilities. If, for example, a candidate was to retake the assessment on multiple occasions, the candidate might well obtain several different scores. A candidate's "true" score, the 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. We model the SEM to demonstrate the impact of making this statistical adjustment to the recommended passing score from the panel.

CalAPA cycles, 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 the three 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 635 candidates who completed all three cycles, considering scenarios of no side conditions, only one "1" (allowing for one rubric score of "1" in both cycles), and allowing no "1s" in any cycle as presented below.

Tables 3-6 provide data on the assessment pass rates with scores on all cycles at the panel recommended cut score, and adjustments in applying the SEM.

Table 3: % Pass Assessment – No Side-Condition for Any Cycle, Modeled at Cycle 1 Cut Score = 19 (panel recommended median)

Cycle 3 Cut Score

Cycle Two Cut Score	**12** (1.5 SEM)	**13**	**14** (1.0 SEM)	**15** (0.5 SEM)	**16**	**17**	**18**	**19**	**20**
10	84%	83%	82%	82%	81%	78%	70%	62%	50%
11	84%	83%	82%	82%	80%	77%	70%	62%	50%
12 (1.5 SEM)	83%	82%	81%	81%	79%	77%	69%	62%	50%
13	81%	80%	79%	79%	77%	75%	67%	60%	48%
14 (1.0 SEM)	78%	78%	77%	77%	76%	73%	66%	59%	48%
15 (0.5 SEM)	76%	76%	75%	75%	74%	71%	65%	58%	46%
16	74%	73%	73%	72%	71%	69%	62%	56%	46%
17	71%	70%	70%	69%	69%	66%	60%	55%	45%
18	62%	62%	61%	61%	60%	58%	53%	48%	39%
19	53%	53%	53%	52%	52%	50%	46%	42%	34%

A score of 17 is the panel's recommendation for both cycle 2 and 3 with no SEM

Table 4: % Pass Assessment – No Side-Condition for Any Cycle, Modeled at Cycle 1 Cut Score = 17 (-.5 SEM)

Cycle 3 Cut Score

Cycle Two Cut Score	**12** (1.5 SEM)	**13**	**14** (1.0 SEM)	**15** (0.5 SEM)	**16**	**17**	**18**	**19**	**20**
10	88%	87%	86%	85%	84%	81%	72%	64%	51%
11	84%	83%	82%	82%	80%	77%	70%	62%	50%
12 (1.5 SEM)	87%	86%	86%	85%	84%	80%	72%	64%	51%
13	86%	85%	84%	84%	82%	79%	71%	63%	50%
14 (1.0 SEM)	83%	83%	82%	81%	80%	77%	69%	61%	49%
15 (0.5 SEM)	81%	80%	79%	79%	78%	75%	67%	60%	48%
16	79%	78%	77%	77%	76%	73%	66%	59%	47%
17	76%	75%	75%	75%	73%	71%	64%	57%	47%
18	73%	72%	72%	72%	71%	68%	62%	56%	45%
19	64%	64%	63%	63%	62%	60%	54%	49%	40%

A score of 17 is the panel's recommendation for both cycle 2 and 3 with no SEM

Table 5: % Pass Assessment—No Side-Condition for Any Cycle, Modeled at Cycle 1 Cut Score = 15 (-1 SEM)

Cycle 3 Cut Score

Cycle Two Cut Score	**12** (1.5 SEM)	**13**	**14** (1.0 SEM)	**15** (0.5 SEM)	**16**	**17**	**18**	**19**	**20**
10	91%	90%	89%	88%	87%	83%	74%	65%	52%
11	84%	83%	82%	82%	80%	77%	70%	62%	50%
12 (1.5 SEM)	91%	90%	89%	88%	87%	83%	74%	65%	52%
13	89%	88%	87%	87%	85%	82%	73%	64%	51%
14 (1.0 SEM)	86%	86%	85%	84%	83%	79%	71%	62%	50%
15 (0.5 SEM)	83%	83%	82%	82%	80%	77%	69%	61%	49%
16	81%	80%	80%	79%	78%	75%	68%	60%	48%

17	78%	77%	77%	76%	75%	72%	65%	58%	47%
18	75%	74%	74%	73%	73%	70%	63%	56%	46%
19	66%	65%	65%	65%	64%	61%	55%	50%	40%

A score of 17 is the panel's recommendation for both cycle 2 and 3 with no SEM

Table 6: % Pass Assessment – No Side-Condition for Any Cycle, Modeled at Cycle 1 Cut Score = 14 (-1.5 SEM)

Cycle 3 Cut Score

Cycle Two Cut Score	**12** (1.5 SEM)	**13**	**14** (1.0 SEM)	**15** (0.5 SEM)	**16**	**17**	**18**	**19**	**20**
10	93%	92%	91%	90%	88%	85%	75%	66%	53%
11	92%	91%	91%	90%	88%	84%	75%	66%	52%
12 (1.5 SEM)	91%	90%	89%	88%	86%	83%	74%	65%	52%
13	88%	87%	86%	86%	84%	80%	71%	63%	50%
14 (1.0 SEM)	85%	84%	83%	83%	81%	78%	69%	61%	49%
15 (0.5 SEM)	82%	82%	81%	81%	79%	76%	68%	60%	48%
16	79%	79%	78%	77%	76%	73%	65%	58%	47%
17	76%	75%	75%	74%	73%	70%	63%	57%	46%
18	67%	67%	66%	66%	65%	62%	55%	50%	40%
19	57%	57%	56%	56%	55%	53%	48%	44%	36%

A score of 17 is the panel's recommendation for both Cycle 2 & 3 with no SEM

To further explore the data and candidate performance, panelists were also shown modeled pass rates for cut score for each demographic group. The next three tables (Tables 7, 8, and 9) provide the percentage of submissions that would pass at a variety of cut scores, for each individual Cycle, by demographic group with 15 or more submissions.

Table 7: Cycle 1-Overall Modeled Passing Rates by Cut Score

	N/A	N/A	Black	Black	Asian	Asian	SE Asian	SE Asian	Hispanic	Hispanic	White	White	Other	Other
Cut Score	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass
12	58	0.98	102	0.91	84	0.99	62	0.97	321	0.97	642	0.98	60	0.98
13	56	0.95	102	0.91	84	0.99	61	0.95	317	0.96	631	0.97	60	0.98
14	56	0.95	98	0.88	83	0.98	60	0.94	313	0.95	621	0.95	60	0.98
15	55	0.93	94	0.84	83	0.98	59	0.92	306	0.92	608	0.93	60	0.98
16	53	0.90	94	0.84	83	0.98	59	0.92	300	0.91	598	0.92	59	0.97
17	52	0.88	93	0.83	82	0.96	59	0.92	295	0.89	589	0.90	58	0.95
18	51	0.86	92	0.82	82	0.96	57	0.89	285	0.86	580	0.89	57	0.93
19 *	51	0.86	90	0.80	79	0.93	55	0.86	275	0.83	572	0.88	57	0.93
20	48	0.81	75	0.67	69	0.81	54	0.84	250	0.76	524	0.80	49	0.80
21	42	0.71	70	0.63	62	0.73	48	0.75	225	0.68	478	0.73	39	0.64

*19 is the Panel-recommended median

Table 8: Cycle 2-Overall Modeled Passing Rates by Cut Score

	N/A	N/A	Black	Black	Asian	Asian	SE Asian	SE Asian	Hispanic	Hispanic	White	White	Other	Other
Cut Score	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass
12	47	0.96	75	0.89	62	0.98	49	0.94	211	0.95	434	0.96	41	0.98
13	47	0.96	68	0.81	60	0.95	47	0.90	206	0.93	416	0.92	40	0.95
14	43	0.88	62	0.74	58	0.92	45	0.87	201	0.91	403	0.90	38	0.90
15	41	0.84	58	0.69	57	0.90	41	0.79	194	0.87	395	0.88	37	0.88
16	38	0.78	57	0.68	55	0.87	41	0.79	183	0.82	386	0.86	36	0.86
17 *	36	0.73	52	0.62	54	0.86	39	0.75	175	0.79	373	0.83	36	0.86
18	30	0.61	46	0.55	45	0.71	37	0.71	148	0.67	330	0.73	33	0.79
19	21	0.43	36	0.43	36	0.57	31	0.60	125	0.56	274	0.61	29	0.69

*17 is the Panel-recommended median

Table 9: Cycle 3-Overall Modeled Passing Rates by Cut Score

		N/A	Black	Black	Asian	Asian	SE Asian	SE Asian	Hispanic	Hispanic	White	White	Other	Other
Cut Score	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass	N Pass	% Pass
12	41	1.00	74	0.96	54	0.98	46	1.00	202	0.97	427	0.99	37	1.00
13	41	1.00	74	0.96	54	0.98	45	0.98	198	0.95	423	0.98	35	0.95
14	41	1.00	71	0.92	54	0.98	43	0.93	195	0.94	421	0.98	35	0.95
15	39	0.95	71	0.92	54	0.98	43	0.93	192	0.92	419	0.97	35	0.95
16	38	0.93	69	0.90	52	0.95	43	0.93	185	0.89	414	0.96	31	0.84
17 *	38	0.93	67	0.87	51	0.93	41	0.89	172	0.83	400	0.93	30	0.81
18	31	0.76	60	0.78	45	0.82	36	0.78	149	0.72	364	0.85	24	0.65
19	27	0.66	49	0.64	42	0.76	32	0.70	118	0.57	318	0.74	24	0.65

*17 is the Panel-recommended median

The Passing Standard Recommendation from the Expert Panel

At the final stage of the Standard Setting process, the panel arrived at the following recommendations for scores for Cycles 1 (19), 2, and 3 (17). As most panelists did not recommend any side-conditions, no side-condition scenarios are presented. Panelists discussed pass-rates in relation to demographic group performance, gender, and placement. The discussion focused on the significant difference findings. A significant difference finding means that the performance difference between two groups (in this case ethnicity groupings) cannot be explained by chance. The following significant performance differences were found ($p<.05$):

- Cycle 1: White/Black
- Cycle 2: Asian/Black; White/Black
- Cycle 3: White/Hispanic; White/Black

On the second day of the study, the panel did discuss the opportunity to use an SEM in the final score decisions with panelists making the following recommendations on their Final Passing Standard Recommendation forms:

- Revisit the passing score annually for the first three years. Apply a -1/2 SEM for policy
- Consider applying an SEM for Cycles 2 and 3
- Include an SEM adjustment
- Apply a -1/2 SEM adjustment to Cycle 2
- Apply a Cycle 2 SEM, open to application of SEM in other Cycles
- SEM adjustment for Cycles 2 and 3, for at least 1 to 2 years
- Apply an SEM adjustment to Cycles 2 and 3
- Apply a -1/2 SEM adjustment to Cycles 2 and 3

- Consider SEM adjustment to Cycle 2 (-1/2 or -1)
- Apply a -1/2 SEM adjustment
- Consider an SEM adjustment applied to all cycles for the initial years of consequential participation
- Apply a -1/2 SEM adjustment to Cycle 2
- Apply a -1/2 SEM adjustment to Cycles 2 and 3
- Apply a -1 SEM to Cycle 2, and -1/2 SEM adjustment to Cycle 3, recommend a review in Years 1 and 2
- Apply a -1/2 SEM adjustment to Cycle 2
- Consider applying an SEM for Cycles 2 and 3
- Recommend -1 SEM for Cycle 2 as a policy consideration for the unknown factors for significantly lower scores for Black candidates

In addition, panelists discussed various concerns related to a non-consequential year administration including:

- PASC programs are early in their implementation and still learning how to embed the assessment cycles in their programs and clinical practice (of the 65 PASC programs, only 23 participated in the pilot and field test years leaving 42 programs to have implemented the CalAPA for the first time in the non-consequential year of 2018-19).
- Candidates and faculty may not have fully engaged due to the non-consequential year expectation that completion for scoring, not quality of the completion, was the requirement
- Districts and schools were learning about the embedded assessment and how to support candidates.

The expert panel's recommendation to the Commission is as follows:

- Cycle 1 (8 rubrics): A final cut score of **19 points**
- Cycle 2 (7 rubrics): A final cut score of **17 points**
- Cycle 3 (7 rubrics): A final cut score of **17 points**

Staff Recommendation and Rationale

As a result of panel discussion around candidate performance by demographic group and the significant differences found in Cycles 1, 2, and 3 for Black and Hispanic candidates, Commission staff worked with ES and requested modeled data be provided at the assessment level for candidates submitting all three cycles, by demographic group with 15 or more submissions. Staff examined performance by cycle, by individual rubric (including examining rubrics associated with video evidence, and rubrics associated with narrative writing), by campus affiliation (including an examination of campuses involved in operational administration as compared to the field test and pilot test). Tables 10.1 – 10.6 represent the modeled assessment pass rates at various combinations of Cycles 1, 2, and 3 cut scores.

Tables 10.1-10.6: Modeled Assessment Pass-rates By Demographic Group

10.1 – Total N candidates by Demographic Group attempting Cycles 1, 2, and 3

All	N/A	Black	Asian	SE Asian	Pacific Islander	Hispanic	Native Amer	White	Other
635	32	58	48	43	3	136	5	286	24

Race/Ethnicity

10.2 - % Pass Assessment - No Side-Condition and Cut Scores of Cycle 1=19, Cycle 2=17, Cycle 3=17

All	N/A	Black	Asian	SE Asian	Hispanic	White	Other
66%	69%	50%	81%	58%	57%	71%	75%

Race/Ethnicity

10.3 - % Pass Assessment - No Side-Condition and Cut Scores of Cycle 1=19, Cycle 2=14, Cycle 3=14

All	N/A	Black	Asian	SE Asian	Hispanic	White	Other
77%	84%	60%	90%	70%	74%	79%	79%

Race/Ethnicity

10.4 - % Pass Assessment - No Side-Condition and Cut Scores of Cycle 1=17, Cycle 2=14, Cycle 3=14

All	N/A	Black	Asian	SE Asian	Hispanic	White	Other
79%	84%	64%	94%	71%	78%	81%	79%

Race/Ethnicity

10.5 - % Pass Assessment - No Side-Condition and Cut Scores of Cycle 1=15, Cycle 2=14, Cycle 3=14

All	N/A	Black	Asian	SE Asian	Hispanic	White	Other
82%	88%	65%	94%	71%	84%	84%	83%

Race/Ethnicity

10.6 - % Pass Assessment - No Side-Condition and Cut Scores of Cycle 1=14, Cycle 2=12, Cycle 3=12

All	N/A	Black	Asian	SE Asian	Hispanic	White	Other
91%	94%	81%	98%	88%	92%	91%	92%

Staff suggests that for at least the next year, the Commission consider applying an SEM of -1.5 to the panel's recommended cut score for each of the three cycles. With this passing standard, PASC candidates would need to meet an actual score of 14 for Cycle 1, and a score of 12 for each of Cycles 2 and 3. In addition, staff recommends that data be analyzed during the next year in preparation for a second standard setting panel in spring 2020. The standard setting panel would discuss the candidate score data collected during the 2019-20 administration, the first consequential year in which candidates pay for the assessment and meet the passing standard set for the 2019-20 year. The panel would, based on this new data potentially recommend a revised passing standard for the 2020-21 administration and beyond.

Staff agrees with the standard setting panel that the goal is to have candidates who are eligible for a preliminary administrator credential earn scores of 19 on Cycle 1 (Cycle 1 has 8 rubrics), and a score of 17 on Cycles 2 and 3 (these two cycles each have 7 rubrics). However, only 23 of the 65 CalAPA using institutions participated in the pilot and/or the field test. With 42 programs engaging in using the CalAPA in 2018-19 for the first time, it is possible that some of these programs were not as knowledgeable about the evidence required via tasks, the essential questions and the rubric description of performance. In addition, three new PASC programs are beginning to accept candidates, and will implement the CalAPA in 2019-20 for the first time. Given the importance of developing all programs' capacity to meet the requirements of the new CAPE and the CalAPA, and the concern that this is not yet the case, staff, after careful discussion and review of data from scored cycles, in support of programs and candidates, are making a different recommendation.

Staff recommends that the Commission adopt the following as the passing standard for the CalAPA for the 2019-20 year, effective as of June 1, 2019:

- Cycle 1 (8 rubrics): A final cut score of **14 points**
- Cycle 2 (7 rubrics): A final cut score of **12 points**
- Cycle 3 (7 rubrics): A final cut score of **12 points**

This passing standard reflects the application of a -1.5 SEM to the panel recommended passing standard of 19 for Cycle 1, 17 for Cycle 2, and 17 for Cycle 3. Staff agrees with the panel that no side conditions should be applied at this time; all candidates will be required to attain the passing score across the rubrics of a cycle and may do so in a number of ways based on total score points attained.

Next Steps

If the Commission adopts the recommended passing scores with applied -1.5 SEM for each cycle, standards will be posted on the CalAPA websites and distributed to the field. The passing standard adopted by the Commission will be applied to all candidates as of June 1, 2019 for the next year of consequential administration of the CalAPA as determined by the Commission.

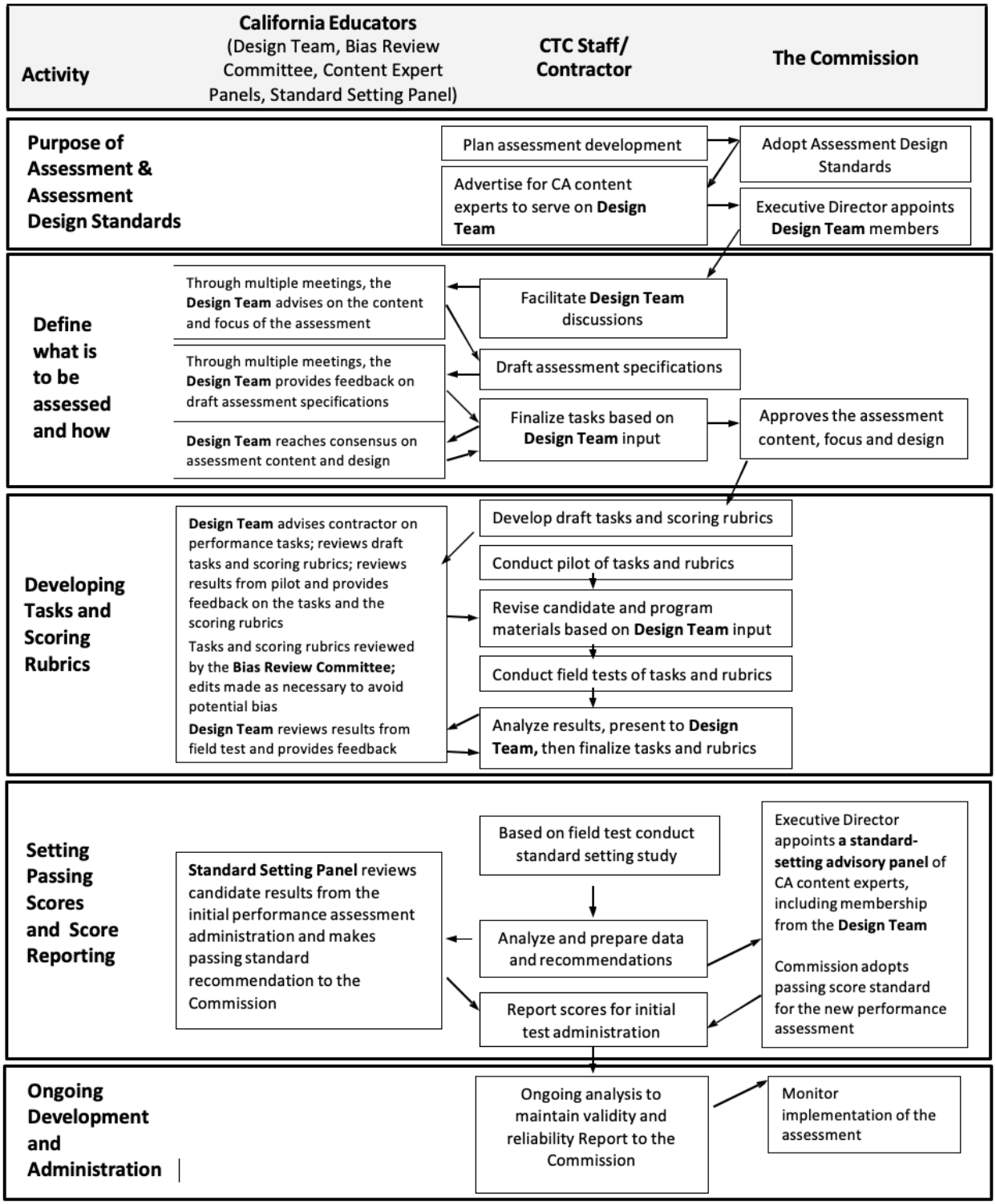
Appendix A

California Administrator Performance Assessment (CalAPA) Design Team

Design Team Member	Affiliation
Susan Belenardo	Superintendent, La Habra School District; University of California, Irvine
Rebecca Cheung	University of California, Berkeley
Kathy Condren	Madera County Superintendent of Schools
Janice Cook	University of San Diego; San Diego State University
Katrine Czajkowski	Sweetwater Union High School District
Ardella Dailey	California State University, East Bay
Alan Enomoto	Brandman University
Deborah Erickson	Pt. Loma Nazarene University
Ursula Estrada-Reveles	Azusa Pacific University; Riverside County Office of Education
Doug Fisher	San Diego State University
Lanelle Gordin	Riverside County Office of Education
Keith Myatt	California State University, Dominguez Hills
Kelli Seydewitz	California Teachers Association
James Webb	Wm. S. Hart Union High School District
Charles Weis	California State University, Channel Islands

Appendix B

Groups Involved in the Redevelopment of the CalAPA



Appendix C

Approved Preliminary Administrative Services Programs 2018-19

California State University
Cal State Poly, Pomona
Cal State Poly, San Luis Obispo
CSU, Bakersfield
CSU, Channel Islands
CSU, Chico
CSU, Dominguez Hills
CSU, East Bay
CSU, Fresno
CSU, Fullerton
CSU, Long Beach
CSU, Los Angeles
CSU, Monterey Bay
CSU, Northridge
CSU, Sacramento
CSU, San Bernardino
CSU, San Marcos
CSU, Stanislaus
Humboldt State University
San Diego State University
San Francisco State University
San Jose State University*
Sonoma State University

Local Education Agencies (LEAs)
Animo Leadership Charter High Schools (Green Dot Public Schools)
Bay Area School of Enterprise (REACH Institute)
Madera County Office of Education
Mount Diablo Unified School District
Orange County Department of Education
Placer County Office of Education
Riverside County Office of Education
Sacramento County Office of Education
San Diego County Office of Education
San Mateo County Office of Education
Santa Barbara County Education Office
Santa Clara County Office of Education
Shasta County Office of Education
Sonoma County Office of Education
Tulare County Office of Education

University of California (UC)
University of California, Berkeley
University of California, Irvine
University of California, Los Angeles

Private Colleges and Universities
Azusa Pacific University
Brandman University
California Baptist University*
California Lutheran University*
Concordia University
Fresno Pacific University
Hope International University
La Sierra University
Loyola Marymount University
Mills College
National University
Notre Dame de Namur University
Pepperdine University*
Point Loma Nazarene University
Santa Clara University
Simpson University
Saint Mary's College of California
Teachers College of San Joaquin
Touro University
University of La Verne
University of Redlands
University of San Francisco
University of Southern California
University of the Pacific*

* indicate institutions who did not participate in the implementation of CalAPA in 2018-19 as they had no eligible candidates or have an inactive PASC program.

Appendix D

CalAPA Standard-Setting Panel Meeting
Sacramento, CA
June 25-26, 2019

CalAPA Standard-Setting Panelists

Member	Affiliation
Cheryl Argawal	San Mateo County Office of Education
Susan Belenardo	University of California, Irvine
Leticia Bradley	Santa Barbara County Education Office
Dana Coleman	Loyola Marymount University
Kathy Condren	Madera County Superintendent of Schools
Ardella Dailey	California State University, East Bay
Ellen Edeburn	California State University, Northridge
Delia Estrada	Los Angeles Unified School District
Ursula Estrada-Reveles	Riverside County Office of Education
Toni Faddis	Chula Vista Elementary School District
Charles Flores	California State University, Los Angeles
Joe Frescatore	San Diego County Office of Education
Lanelle Gordin	Riverside County Office of Education
Jason Lea	Sonoma County Office of Education
Maria Montgomery	San Diego Unified School District
Tonikiaa Orange	University of California, Los Angeles
Noni Reis*	San Jose State University, retired California Association of Professors of Education Administration
Glenn Sewell	National University
Nichole Walsh	Fresno State University

* did not participate in the scoring as she left the meeting early

Appendix E

Detailed Description of the Standard Setting Process for the Redeveloped CalAPA

The CalAPA 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 CalAPA assessors, CalAPA Design Team members, County Office of Education administrators, principal mentors/coaches, and administration preparation program faculty who had previously worked with the CalAPA. The names and affiliations of 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 CalAPA was consistent with recognized psychometric principles and procedures. The standard setting study for the CalAPA was conducted on June 25-26, 2019, with pre-conference activities occurring prior to the meeting. An agenda for the meeting is provided in the Supplemental Appendix.

Prior to the meeting, each invited panelist received CalAPA Guides, rubrics, and nine previously scored sample submissions (three from each Cycle) representing different performance levels. 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) gain some exposure to a range of candidate responses and 2) apply that information in the policy capture activities (activities drawing upon the panelists' experience and discussion) at the meeting.

The CalAPA 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 an administrator credential candidate who is just at the level of knowledge, skills, and abilities required to perform effectively the job of a new administrator 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 used this concept of what a minimally competent beginning administrator would know and be able to demonstrate in determining their recommended acceptable score for Cycle 1, Cycle 2, and Cycle 3. Although a number of candidates may exceed the level of

acceptable knowledge, skills, and abilities, 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 CalAPA.

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 knowledge, skills, and abilities necessary for a beginning practitioner just competent to begin professional practice as a public-school administrator.

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, and again for Cycle 3, with nine submissions reviewed and discussed by each panelist.
The activities previously described included a rating form with four rating levels from which to select.

Rating Levels	Definitions of Each Rating Level
Clearly below the passing standard	CLEARLY NOT performing effectively the job of a new administrator. This candidate has demonstrated one or more *major* problems in administrative 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 administrator. This candidate has demonstrated some strengths but has one too many issues in administrative 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 administrator. This candidate has demonstrated some consistent strengths in administrative knowledge, skills, or abilities and has a foundation on which to build. The candidate may have shown one or more minor flaws in administrative knowledge, skills, or abilities that will likely improve with more time and experience.
Clearly above the passing standard	CLEARLY MEETS your definition of performing effectively the job of a new administrator. This candidate has demonstrated clear strengths in administrative knowledge, skills, and abilities, and a strong foundation for effective administration.

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

- Cycle 1: Total scores between 20-24
- Cycle 2: Total scores between 17-21
- Cycle 3: Total scores between 17-21

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 20-24
- Cycle 2: Total scores between 17-21
- Cycle 3: Total scores between 17-21

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 Supplemental Appendix for sample Candidate Score Profiles at total scores of 15, 16, and 19 (Cycle 1), 14 and 17 (Cycle 2), and 14 and 17 (Cycle 3) as well as the rubrics for each cycle. 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 "an administrative services credential candidate who is just at the level of knowledge and skills required to perform effectively the job of a new administrator 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 and corresponding Standard Error of Measurement (SEM) values at +1 to- 2.0 SEM

CYCLE 1: 8 Rubrics (Total Possible = 40) Panel recommendation for Cycle 1 is 19.

Cut score	SEM	-2	-1.5	-1	-0.5	+0.5	+1
5.00	3.10	8	10	11	13	16	18
16.00	3.14	9	11	12	14	17	19
17.00	3.17	10	12	13	15	18	20
18.00	3.19	11	13	14	16	19	21
19.00	**3.20**	**12**	**14**	**15**	**17**	**20**	**22**
20.00	3.20	13	15	16	18	21	23
21.00	3.20	14	16	17	19	22	24
22.00	3.19	15	17	18	20	23	25
23.00	3.17	16	18	19	21	24	26
24.00	3.14	17	19	20	22	25	27
25.00	3.10	18	20	21	23	26	28

CYCLE 2: 7 Rubrics (Total Possible = 35) Panel recommendation for Cycle 2 is 17.

Cut score	SEM	-2	-1.5	-1	-0.5	+0.5	+1
13.00	2.90	7	8	10	11	14	15
14.00	2.94	8	9	11	12	15	16
15.00	2.97	9	10	12	13	16	17
16.00	2.99	10	11	13	14	17	18
17.00	**3.00**	**11**	**12**	**14**	**15**	**18**	**20**
18.00	3.00	12	13	15	16	19	21
19.00	2.99	13	14	16	17	20	21
20.00	2.97	14	15	17	18	21	22
21.00	2.94	15	16	18	19	22	23
22.00	2.90	16	17	19	20	23	24
23.00	2.85	17	18	20	21	24	25

CYCLE 3: 7 Rubrics (Total Possible = 35) Panel recommendation for Cycle 3 is 17.

Cut score	SEM	-2	-1.5	-1	-0.5	+0.5	+1
13.00	2.90	7	8	10	11	14	15
14.00	2.94	8	9	11	12	15	16
15.00	2.97	9	10	12	13	16	17
16.00	2.99	10	11	13	14	17	18
17.00	**3.00**	**11**	**12**	**14**	**15**	**18**	**20**
18.00	3.00	12	13	15	16	19	21
19.00	2.99	13	14	16	17	20	21
20.00	2.97	14	15	17	18	21	22
21.00	2.94	15	16	18	19	22	23
22.00	2.90	16	17	19	20	23	24
23.00	2.85	17	18	20	21	24	25