
4B

Action

Professional Services Committee

Report of the 2013-14 Pilot of the edTPA and Review and Potential Approval of the edTPA for use in California

Executive Summary: This agenda item provides information on the 2013-14 pilot of the edTPA and on the standard setting study conducted by the developers of the edTPA during the pilot, and presents a recommendation for the Commission's consideration concerning the potential approval of the edTPA for use in California. Representatives from The Stanford Center for Assessment, Learning, and Equity (SCALE) will be on hand to present and discuss the findings from the standard setting study and recommend a passing standard to the Commission.

Policy Question: Does the edTPA meet the Commission's current adopted Assessment Design Standards? What is an appropriate passing standard for the edTPA?

Recommended Action: That the Commission approve the edTPA for use by California Preliminary Multiple and Single Subject teacher preparation programs, approve a passing standard for the edTPA, and reevaluate the passing standard after two years of implementation.

Presenter: Phyllis Jacobson, Administrator, Professional Services Division

Strategic Plan Goal:

I. Educator Quality

- ◆ 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.

August 2014

Report of the 2013-14 Pilot of the edTPA and Review and Potential Approval of the edTPA for use in California

Introduction

This item presents information about the teaching performance assessment known as the “edTPA,” the 2013-14 pilot of the edTPA, and a recommendation for the Commission’s consideration concerning the potential approval of the edTPA for use in California.

Part 1: Background- The Teaching Performance Assessment Requirement

The Education Code (EC) specifies in section 44320.2(b) that beginning July 1, 2008 all multiple and single subject candidates must pass a teaching performance assessment that is aligned with the California Standards for the Teaching Profession and congruent with state content and performance standards for pupils. Section 44320.2(d) specifies the Commission’s responsibilities with respect to the implementation of the teaching performance assessment requirement, including, under subsection (d)(3) the duty to “establish a review panel to examine each assessment developed by an institution or agency in relation to the standards set by the commission and advise the commission regarding approval of each assessment system.” The Commission also has the statutory duty under subsection (b)(1) to adopt assessment quality standards for the purpose of approval of any locally developed assessment. The Commission’s current assessment quality standards are provided in Appendix A.

Currently there are three teaching performance assessment models approved for use in California teacher preparation programs. The first TPA to be developed was the CalTPA. The CalTPA was developed by the Educational Testing Service under contract with the Commission pursuant to EC §44320.2(b)(2) and (d)(2) which requires the Commission to develop a teaching performance assessment and make a training program for assessors available to any program choosing to use the Commission-developed assessment. Subsequently the Commission reviewed and approved two additional teaching performance assessments under EC § 44320.2(b)(1) and (d)(3): the Performance Assessment for California Teachers (PACT) in 2007 (<http://www.ctc.ca.gov/commission/agendas/2007-10/2007-10-3C.pdf>) and the Fresno Assessment of California Teachers (FAST) in 2008 (<http://www.ctc.ca.gov/commission/agendas/2008-06/2008-06-3G-insert.pdf>). Programs may use any of the three approved TPA models. In 2010-11, approximately 66% of teacher candidates took the CalTPA (55 institutions); approximately 31% of candidates took the PACT (34 institutions); and approximately 3% of candidates took the FAST (1 institution). On July 30, 2013, the Commission received a fourth teaching performance assessment model for review and potential approval, the edTPA.

Overview of the edTPA

The edTPA is owned by Stanford University, through the Stanford Center for Assessment, Learning and Equity (SCALE). The application was submitted jointly by SCALE and Evaluation Systems (group of Pearson), a national testing contractor that administers and manages the scoring of the assessment. The application indicated that the edTPA was developed collaboratively by teachers and teacher educators under the leadership of SCALE and the American Association of Colleges for Teacher Education (AACTE). The edTPA is a national model of teaching performance assessment usable by multiple states. According to the AACTE website (<http://edtpa.aacte.org/faq#17>), the following states are participating in the edTPA: Colorado, Connecticut, Delaware, District of Columbia, Georgia, Hawaii, Idaho, Illinois, Iowa, Indiana, Maryland, Massachusetts, Minnesota, New York, North Carolina, Ohio, Oklahoma, Oregon, South Carolina, Tennessee, Virginia, Washington, Wisconsin, and Wyoming.

The edTPA assessment model provides three key tasks for candidates: Planning, Instruction, and Assessment, all three of which include aspects of academic language and analyzing teaching. Candidates are provided with instructions and prompts that guide their reflections, responses, artifacts and other portfolio evidence, including several video clips of the candidate's instruction, lesson plans, student work samples, analysis of student learning, and reflective commentaries.

Candidate responses are scored according to a total of 15 rubrics (5 per task) except for the Elementary Education task, which is scored according to a total of 18 rubrics (15 rubrics addressing the English Language Arts tasks and 3 addressing the mathematics assessment task) and the World Language/Classical Language tasks, which have 13 rubrics each. All rubrics are on a five point scale, which the developers of the edTPA regard as a developmental scale indicating increasing candidate performance across the scale range.

The edTPA is unique in the history of the implementation of California's teaching performance assessment requirement in that it is the first national assessment to be submitted for review and potential approval. All prior approved TPA models have been developed wholly for use in California and are scored at the campus level by faculty and practitioners recruited by the campus and trained to score. The edTPA was developed in California by the designers of the PACT assessment in collaboration with a variety of national organizations and with educators from multiple states, including California. The edTPA was developed in alignment with InTASC (Interstate New Teacher Assessment and Support Consortium) standards (<http://intascstandards.net/>), which are well aligned with the California Standards for the Teaching Profession, which serve as the basis for the Commission's Teaching Performance Expectations. The edTPA application included a "crosswalk" to demonstrate the alignment between the edTPA tasks and rubrics, the national standards and the California TPEs.

Scorers of the edTPA are drawn from a national "academy" of trained, calibrated assessors. Scorer training and calibration is provided by SCALE; the implementation of candidate registration, portfolio submission, and scoring processes are provided by Evaluation Systems. The fee for the edTPA is \$300 per candidate.

Piloting of the edTPA in California

In September 2012, the Commission approved a small pilot of the edTPA with three California multiple and single subject teacher preparation programs. Three institutions (San Diego State University, UCSB, and USC) that wanted to pilot the edTPA requested that the Commission approve a waiver so that these institutions could pilot the edTPA with some of their candidates. (<http://www.ctc.ca.gov/commission/agendas/2012-09/2012-09-2F.pdf>)

In summer 2013, edTPA applied to the Commission for approval as a TPA model available for use by California preliminary multiple and single subject teacher preparation programs. A technical expert advisory panel reviewed edTPA's submission and reported to the Commission at the September 2013 meeting that the edTPA had not yet fully met the Commission's Assessment Design Standards. (<http://www.ctc.ca.gov/commission/agendas/2013-09/2013-09-4B.pdf>) The key areas where the technical expert advisory panel felt the standards had not yet been fully met were:

- Addressing the teaching of English learners in the assessment
- Addressing the four core content areas for Multiple Subject candidates
- Setting a passing score standard for the assessment's use in California
- Qualifications of national scorers for scoring California candidates

Below is an explanation of the Commission's review process for teaching performance assessment models submitted for potential approval as a model usable by California preliminary teacher preparation programs, as illustrated by the review process for the edTPA.

The edTPA Review Process

Consistent with the Commission's statutory duty to review and approve as appropriate any additional teaching performance assessment model that meets the Commission's standards, an expert psychometric review panel of five members was established to review the materials initially submitted by SCALE and Evaluation Systems. The review panel members included IHE faculty (three members) one K-12 assessment expert, and one out of state teacher licensure assessment expert. The initial edTPA materials were received electronically on July 29, 2013 from Evaluation Systems. The expert review panel met on August 7, 2013 at the Commission office for an orientation to the Education Code related to the TPA requirement, the Commission's adopted Assessment Design Standards, and the other approved TPA models. The panel was then provided with copies of the edTPA materials that had been received from Evaluation Systems and used the remainder of the day for protected review time. The panel was instructed to review the submitted materials against the Commission's Assessment Design Standards and to provide a professional judgment, with supporting rationale and/or evidence, as to whether each standard was met, based on the information provided by Evaluation Systems on behalf of edTPA, or whether the panel needed additional information in order to determine if the standard was met.

After a two-week review period, the panel met for a second time at the Commission office on August 21, 2013 to discuss its findings. The Secondary Mathematics Handbook, which contained all materials related to the tasks, instructions and rubrics, was included in the original

submission. On August 21, the panel was provided by the assessment sponsor with access to the other Handbooks for the edTPA tasks. The panel had a robust discussion concerning its findings, and came to consensus on which standards the panel felt had been met based on the initial submission by edTPA and which standards needed additional information and/or clarification. A consensus report was provided electronically on August 23, 2013 to Evaluation Systems, with a deadline of August 30 for providing the additional information and/or clarifications requested by the panel. Panel members were asked to review each of the Handbooks while awaiting the next edTPA response, since the material in the Handbooks was pertinent to several of the standards that had been deemed initially by the panel to require additional information. A response from Evaluation Systems was received electronically on August 30, 2013, and was immediately forwarded to the members of the expert review panel. The panel then met via conference call on September 5, 2013 to discuss all of the submitted materials and come to a final consensus on their findings.

Expert Review Panel Findings

The edTPA Review Panel indicated in its report to the Commission presented at the September 2013 meeting (<http://www.ctc.ca.gov/commission/agendas/2013-09/2013-09-4B.pdf>) that the edTPA submission had met most of the Commission's Assessment Design Standards, as documented in Appendix E of that agenda item. The panel found the edTPA to be a comprehensive performance assessment potentially approvable for full use in California, but identified for the Commission several issues that needed to be addressed before the panel felt a recommendation for full approval could be made. The panel emphasized, however, that overall the edTPA was a robust assessment consistent with the majority of requirements of the Commission's Assessment Design Standards.

In response to these findings, staff held additional discussions with the edTPA developers to determine whether the identified concerns could be addressed to enable use of the edTPA to move forward in California. Commission staff participated in a meeting organized and facilitated by one of the current edTPA pilot institutions at the University of California Office of the President on September 6, 2013, to discuss issues and questions that had arisen for the piloting institutions during the pilot year. The meeting focused primarily on questions that these institutions had for the potential of edTPA implementation on a statewide basis. At the meeting the edTPA developer provided further details about the edTPA in response to questions from the participants.

Based on these and subsequent discussions with the edTPA developer, staff summarized the unresolved issues identified by the Review Panel and identified those areas where clarification from the edTPA developer had been provided.

Approval of an Expanded Pilot of the edTPA for 2013-2014

At the same Commission meeting in September 2013 where the report from the expert review panel was presented, the Commission approved an expanded pilot of the edTPA to allow additional institutions to try out this TPA model and to allow additional time for edTPA to address the areas where the technical expert advisory committee felt the edTPA had not fully

met the Commission's Assessment Design Standards, as noted above. The Commission established the following conditions for the pilot:

- a) Inclusion throughout the edTPA of a required focus on EL students for all California candidates in all tasks and content areas;
- b) Provision and use of Embedded Signature Assessments or similar assessments in order to address all four core content areas for multiple subject candidates in California;
- c) Looking at the appropriateness of the use of California-only or a mixture of California and scorers from other states to score California edTPA submissions;
- d) Requiring that all scorers, whether national or from California, have demonstrated expertise in the area of English learners and/or preparing teachers to work with English learners;
- e) Establishing through the pilot study the passing score standards for all California multiple and single subject candidates; and
- f) Not allowing California candidates to use the online electronic feedback module provided by Evaluation Systems to receive feedback on the actual materials submitted for review.

In March 2014 Commission staff requested, and edTPA provided, an update on the progress made with respect to conditions a-f above. In June 2014, edTPA held a California standard setting study, which staff and the Executive Director attended as observers. Also in June 2014, staff sent a survey to the eleven institutions participating in the expanded pilot (see below) regarding their experiences with edTPA. In July 2014, staff provided guidance to edTPA on the nature of the responses expected in order to meet the August Commission meeting deadline and to demonstrate that edTPA fully meets the Commission's Assessment Design Standards. This guidance indicated that edTPA should respond to the following issues identified by the expert review panel:

- 1) How the teaching of **English learners** will be addressed within the edTPA for all California candidates.
- 2) How the **four core content areas for Multiple Subject candidates** will be addressed for all California Multiple Subject candidates. Please explain as part of the response which version(s) of the elementary assessment are being proposed for use in California and the rationale for the choice(s), and describe the role of signature embedded assessments/content area assessments.
- 3) The **passing score standard** proposed for the edTPA for all California candidates and the rationale for establishing that standard. (NOTE: The Commission will expect to receive a full technical report when it is available documenting the standard setting study; for purposes of meeting this requirement prior to the August Commission meeting, a summary of the process and rationale will suffice.)
- 4) The **qualifications proposed for the scorers** of edTPA responses submitted by California candidates. (Note: although this is not specifically part of the technical expert panel's recommendations, staff would appreciate it if edTPA could also explain whether it would be willing to work with institutions that (a) might have an interest in local scoring

of candidate responses; (b) would like their own faculty to participate in scoring; or (c) would like only California educators to serve as scorers of their candidates' responses.)

A response was received from edTPA in July 2014 and was reviewed by staff. The information provided in the response informed the staff recommendation presented at the conclusion of this agenda item.

Part 2: Report on the 2013-14 Pilot of the edTPA

A total of eleven institutions participated in the 2013-14 edTPA pilot. Appendix B provides a list of the pilot institutions, how many candidates participated, and which content areas were included in the pilot.

The institutions were asked to respond to survey questions about their experiences with the pilot. Seven of the eleven institutions responded, with multiple responses received from several individuals at the same institution. Some highlights of the responses are described below.

Content areas included in the pilot*

Candidate portfolios from the following content areas were included in the pilot:

- Elementary literacy
- Elementary mathematics
- Elementary education
- Secondary mathematics
- Secondary English
- Secondary science
- Secondary history/social science
- Secondary music
- Secondary art**
- Secondary physical education**
- Secondary world language**

** the pilot data also includes candidate portfolios from the smaller 2012-13 pilot*

***indicates a content area with small numbers of portfolios*

Perceived strengths of the edTPA

Respondents who answered this question commented that the edTPA's online support was responsive to both candidates and institutions; that the handbooks were clear and helpful guides; and that the centralized scoring provided institutions with more time to pay attention to the assessment and assist candidates to prepare.

Challenges during the pilot year

Some respondents indicated that selecting the appropriate assessment and completing the upload of all information were challenges for some candidates. One respondent commented that there were limited resources for integrating the use of feedback into instruction. Another respondent indicated there was not sufficient buy-in from methods

faculty to provide discipline-specific support students in the assessment. The lack of a California-specific passing standard was also a concern to some respondents.

Addressing the teaching of English learners in the pilot

Respondents indicated that candidates selected an English learner student or a student who struggles with academic English as one of the focus students in the assessment section. One respondent commented that this area was not as robust in the edTPA as it had been in the PACT assessment.

Addressing the teaching of students with special needs in the pilot

Respondents noted that one question on the edTPA specifically asked about students with IEP's and/or 504 plans. Several respondents indicated that this area was less strongly addressed, similar to their view of how the English learner area was addressed.

How the institution's faculty was involved in TPA scoring in the pilot and using data for program improvement

Respondents indicated that in general, faculty was not involved in scoring the edTPA during the pilot. Some institutions did their own scoring as an exercise. Some respondents intended, however, to have faculty look at the data from the pilot but this had not yet occurred due to the timing of the assessment and receipt of the results at the end of the program year.

How implementation/use of edTPA compared to the prior experiences with PACT and/or CalTPA

Respondents who are PACT users indicated that using edTPA was a similar process. Several respondents indicated the assessment was still perceived by candidates as overwhelming although slightly more streamlined than PACT. One of the biggest challenges, mentioned by two respondents, was learning to use the different uploading and submission system.

Advice to/further information for the Commission regarding the edTPA pilot

One respondent indicated that it would be helpful if edTPA modified the registration process for California candidates such that only those assessments applicable for use in California were showing. This would eliminate the situation where a candidate registered for an inappropriate version of the assessment.

Several respondents indicated programs should start early in order to address switching to a different scoring system (centralized scoring, different uploading system) as well as to address planning for local evaluation. It was also suggested that programs should attend the annual edTPA conference to learn from other institutions and encourage faculty to go through the assessor training process.

Respondents commented that the scores were received within the promised time frame. Several respondents commented that they preferred the edTPA over the PACT in

terms of its streamlining, organization, and clear information for faculty and students. Finally, there was a concern expressed over the treatment of bilingual candidates in the assessment and the qualifications of scorers for bilingual candidate portfolios, as these differ from the World Language portfolios.

Part 3: Additional Information Provided by edTPA in Response to the Assessment Design Standards

edTPA submitted an initial response to the Assessment Design Standards that the expert review panel had found were not met, as identified above. Staff reviewed the response and had further discussions with the developer/owner of the edTPA on July 22 and July 23, 2014. Subsequently a revised response was received from edTPA on July 24, 2014 and was reviewed by staff. A final revised response, responding to clarifying questions from staff, was received and reviewed on July 31, 2014.

A. Addressing the teaching of English learners in the assessment

edTPA Response: edTPA clarified in its July 24, 2014 response that candidates were reminded in the directions for Tasks 1, 2, and 3 that they needed to consider students who are English learners when formulating their task submissions, and that for California, this instruction would be modified to require California candidates to respond to the Context Information artifact with explicit attention to the English learners they are teaching in the learning segment documented in edTPA. In addition, the Assessment Task in specific requires California candidates to include student work samples from at least one English learner. Further, edTPA will develop California-specific response templates for the Context Information artifact and for the commentary responses used in all three tasks, and will make these available to programs and candidates. These California-specific templates will be used by scorers of California edTPA candidates, who will apply California-specific rubric scores based on evidence provided by candidates. In summary, all of these modifications will ensure that California candidates are explicitly directed to address English learners in their responses.

Staff Analysis: Staff finds that the July 24, 2014 response regarding addressing the teaching of English learners in the assessment and clarifying how edTPA has been modified with respect to the teaching of English learners meets the Commission's current Assessment Design Standards.

B. Addressing the four core content areas for Multiple Subject candidates

edTPA Response: edTPA currently has three approaches to assessing elementary candidates. edTPA is requesting approval for all three approaches, with local programs making the decision as to which approach would be implemented for candidates. The three approaches are:

Handbook Title	Structure	Number of Rubrics
Elementary Literacy	Common Architecture three task model in Literacy – planning, instruction and assessment with academic language and analysis of teaching scoring components	15 rubrics <ul style="list-style-type: none"> – Planning = 3 – Instruction = 4 – Assessment = 3 – Academic Language = 2 – Analyzing Teaching = 3
Elementary Mathematics	Common Architecture Three task model in Mathematics – planning, instruction and assessment with academic language and analysis of teaching scoring components	15 rubrics <ul style="list-style-type: none"> – Planning = 3 – Instruction = 4 – Assessment = 3 – Academic Language = 2 – Analyzing Teaching = 3
Elementary Education	Four Task Model includes the same three tasks described above for Literacy and a fourth task focused on analyzing student learning in Mathematics, including a “re-engagement” lesson designed and taught based on needs identified via student assessment.	18 rubrics <ul style="list-style-type: none"> • 15 from the three task Literacy handbook as above • 3 additional rubrics for the Mathematics Task 4

According to edTPA, programs choosing the Elementary Literacy option would need to use Content Area Tasks (CATs) for the content areas of Mathematics, Science, and History-Social Science; programs choosing the Elementary Mathematics option would need to use CATs for the content areas of English/Language Arts, Science, and History-Social Science; and programs choosing the Elementary combination option would need to use CATs for the content areas of Science and History-Social Science. edTPA describes the Content Area Tasks as a particular type of Embedded Signature Assessment explicitly developed by programs to fulfill the multiple subjects requirement of PACT.

Staff Analysis: Staff finds that the proposal by edTPA to allow the use of all three of the Elementary Education approaches, including the required use of CATs as specified above, meets the Commission’s current Assessment Design Standards.

C. Establishing a California passing score standard

edTPA Response: edTPA conducted a standard setting study for California candidates on July 1, 2014, at Stanford University. Commission staff participated as observers in the standard setting activities and process. The following chart documents the individuals who participated as members of the standard setting panel:

Participant	Affiliation/Representing
Nathan Avani	California State University – San Francisco State University
Lynne Bercaw	California State University – Chico
Nadine Bezuk	California State University – San Diego State University
Jo Birdsell	National University and California Council of Teacher Educators (CCTE)

Marta Fuentes	Teacher and former edTPA Candidate
Vicki Laboskey	Mills College
Antoinette Linton	Washington Preparatory High School and University of Southern California
Helene Mandell	University of San Diego
Noni Reis	California Teachers Association and San Jose State University
Kip Tellez	University of California at Santa Cruz
Keith Walters	California Baptist University
Darby Williams	Sacramento County Office of Education and Beginning Teacher Support and Assessment

Panelists were informed of the purpose of the assessment and provided with a Briefing Book to guide their activity. The Briefing Book Method (Appendix C) is an evidence-based standard setting method intended to develop an appropriate and defensible cut score that can be supported with a validity argument. During the facilitated one-day 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 is also referred to as a “passing standard”), which was then discussed and evaluated. Following that, panelists recommended a final cut score. Discussion was held, and a final “consensus cut score” was obtained from the panelists.

Staff notes that edTPA representatives will be making a presentation to the Commission during this agenda item concerning the methodology of the standard setting study and how the recommended passing score standards were derived and finalized.

Staff Analysis: The method used by edTPA to set a recommended passing score standard for California (Appendix C) is consistent with recognized methodology in the field. Staff notes that panel members were given ample opportunity to make their viewpoints known and to participate in the various rounds of rating activities, as well as a final opportunity to come to consensus on their recommendations. Staff finds that the process for determining the passing score standard proposed for edTPA for California candidates meets the Commission’s current Assessment Design Standards.

D. Qualifications for scorers

edTPA Response: SCALE has committed to working with the Commission to ensure that scoring of edTPA portfolios in California includes participation of educator preparation program faculty and P-12 teachers and that programs have flexibility in choosing a scoring model suited to their needs. Programs can choose to use the current operational scoring model for edTPA with either national or California state scorers, or they can choose a local scoring model that uses local faculty (instructors, supervisors, and school-based educators who participate in the teacher education program) as the primary scorers. For any programs that choose to use national scorers, SCALE has committed to expanding the requirements to include evidence of capacity to evaluate appropriate instruction for English Language Learners.

Staff Analysis: Staff finds that the proposed scoring options meet the Commission’s Assessment Design Standards and address the condition set forth by the Commission in September 2013 that all scorers of the edTPA in California demonstrate expertise in the area of English learners and/or preparing teachers to work with English learners.

Staff finds that the edTPA now meets the Commission’s current Assessment Design Standards and recommends that the Commission approve the use of the edTPA by California Multiple and Single Subject Preliminary Teacher Preparation Programs.

Part 4: Determining a California Passing Score Standard

edTPA recommended passing standard: Based on the Standard Setting activities described in this report, the sample included in the validation study, the median of the panelist recommendations (a score of 43), and the standard error of measurement (SEM) of 4 points, SCALE proposes the following passing standards for the edTPA, as specified below:

- An edTPA California state passing standard of **39 points** for 15-rubric fields
- An edTPA California state passing standard of **34 points** for 13-rubric fields (e.g., World Language, Classical Language)
- An edTPA California state passing standard of **47 points** for 18-rubric fields (e.g., Elementary Education)

Conditions: Conditions are rules around how many score points on the low end of the scoring across the rubric scale candidates are allowed to have and to still pass the assessment. For example, a condition might be that candidates can have no more than a certain number of scores of “1” on a five-point rubric in order to pass the assessment. Based on the Standard Setting activities, feedback received from the panelists, and a preliminary review of the data, SCALE recommends that currently no conditions be applied to the edTPA passing standards indicated above for the first year of implementation. However, SCALE recommends evaluating the potential application of conditions relating to the passing standards the following year when more candidates have taken edTPA and a more robust data set is available for examining patterns related to scores of 1 within a variety of fields and for different rubrics.

Staff Analysis:

Staff notes that initially the raters in the standard setting study recommended a California passing standard between 40 and 44, based on a review of candidate portfolios and both small and whole group discussion. A score of 43 represents the mean of rater recommendations, with individual raters recommending final passing scores ranging from 42 to 44.

SCALE has proposed applying the Standard Error of Measurement (SEM) of -4 points to the mean panel recommended cut score of 43 to reduce potential false negatives, thereby preventing candidates who likely possess the knowledge necessary to pass edTPA from receiving a failing score. Because standard setting is ultimately a policy decision, so too is the decision whether to apply SEM to a panel recommended cut score. As an example, for Commission-owned examinations, SEM is always reviewed but not always applied to panel-

recommended cut scores. Historically the Commission has considered, and at times applied, a SEM when the passing rate would be too low if the SEM were not applied, or if the assessment were a low incidence assessment, or if the assessment were new.

Applying half of the standard error of measurement would result in a passing standard that falls between the current SCALE recommended score of 39 and the standard setting panel mean recommendation of 43. As noted in the discussion of *conditions* attached to the cut score, panelists were concerned about the number of very low scores or “1s” that a given candidate could potentially earn on specific rubrics while still obtaining an overall passing score on edTPA. This was a point of discussion among the panelists as they considered their recommendations. Applying the SEM of -4 to the panel-recommended cut score of 43, thereby reducing it to 39, increases the number of 1s a given candidate could earn on individual rubrics while still achieving an overall passing score.

The table below shows the impact, in terms of pass rates, of three possible passing scores: the panel’s recommendation of 43, SCALE’s SEM adjusted recommendation of 39, and the midpoint of that range. The impact data provided by SCALE includes pass rates for both national and California participants in the edTPA pilot studies at each of these score points.

Overall Pass Rates by Cut Score (15 Rubric Fields)

Cut Score	National Passing Rate	California Passing Rate
43	64%	89%
41	74%	93%
39	78%	96%

The sample for the edTPA California pilot studies was concentrated among a specific segment of institutional type, with more than half of participating institutions being in the University of California (UC) system, and all but two of the institutions being users of the Performance Assessment for California Teachers (PACT). The UC system produces approximately 5% of the new teachers in California and the UC preparation programs are almost universally full time teacher preparation programs. While the sample was large enough to support the standard setting process, it may not be representative of all California candidates and therefore it is not known how these pass rates might differ given a more representative sample of California candidates and institutions. The California candidates who participated in the pilot did outperform their non-California counterparts who participated in the national pilot of edTPA. Though the reasons for California’s scores being higher than the national scores are unknown, it can reasonably be inferred that this was to be expected given California candidates’ status as graduate students (as opposed to mostly undergraduate candidates in the national sample) and the long history of TPA use in California programs.

Approving a passing standard of 39, as recommended by SCALE, would allow for the application of a full standard error of measurement, which could be warranted if the pilot population of candidates is not representative of the population of California candidates and institutions. This passing standard would support institutions seeking to transition to edTPA by allowing time to

ramp up and put in place the supports that candidates will need in order to successfully complete the edTPA. This passing standard falls outside the range recommended by the standard setting panel, however. Approving a passing standard of 41 would fall within the range of scores recommended by the standard setting panel, and allow for the application of one half of the standard error of measurement, based on sample size and uncertainties about the representativeness of the sample. Approving a passing standard of 43 would be consistent with the recommendations of the standard setting panel, but would not take SEM into consideration. Whatever the Commission decides on this matter, staff recommends that the passing standard be revisited after a larger pool of candidates have the opportunity to take the edTPA.

Staff Recommendations

1. Staff finds that the edTPA now meets the Commission's current Assessment Design Standards and recommends that the Commission approve the use of the edTPA by California Multiple and Single Subject Preliminary Teacher Preparation Programs.
2. Staff recommends the Commission approve an appropriate passing score standard for edTPA, as discussed above in this agenda item for a period of two years and that the passing standard be revisited when more California data is available following the first two operational years of edTPA administration in California.
3. Finally, staff recommends that edTPA conduct the further standard setting activities described in its response to the current Assessment Design Standards concerning whether establishing any future conditions for the passing standard are warranted based on candidate outcomes after two years of implementation.

Next Steps

If the Commission takes action to approve edTPA, programs wishing to use the edTPA instead of their current TPA model will need to develop and submit to the Commission a Transition Plan that addresses how the program will inform candidates, provide training to faculty and staff on the edTPA model so that faculty and staff are able to prepare candidates for the assessment, and work with candidates who still need to complete the program's current TPA model. A Program Sponsor Alert (PSA) with information about this process will be prepared by staff and communicated to all approved Multiple and Single Subject Preliminary Teacher Preparation programs.

APPENDIX A

California Teaching Performance Assessment Design Standards, as Adopted by the Commission in December 2006

Assessment Design Standard 1: Assessment Designed for Validity and Fairness

(Assessment Design Standard 1 Applies to Programs that Request Approval of Alternative Assessments)

The sponsor of the professional teacher preparation program requests approval of a Teaching Performance Assessment (TPA) in which complex pedagogical assessment tasks and multi-level scoring scales are linked to the Teaching Performance Expectations (TPEs). The program sponsor clearly states the intended uses of the assessment, anticipates its potential misuses, and ensures that local uses are consistent with the statement of intent. The sponsor maximizes the fairness of assessment design for all groups of candidates in the program, and ensures that the established passing standard on the TPA is equivalent to or more rigorous than the recommended state passing standard.

Required Elements for Assessment Design Standard 1: Assessment Designed for Validity and Fairness

1(a) The Teaching Performance Assessment includes complex pedagogical assessment tasks to prompt aspects of candidate performance that measure the TPEs. Each task is substantively related to two or more major domains of the TPEs. For use in judging candidate-generated responses to each pedagogical task, the assessment also includes multi-level scoring scales that are clearly related to the same TPEs that the task measures. Each task and its associated scales measure two or more TPEs. Collectively, the tasks and scales in the assessment address key aspects of the six major domains of the TPEs. The sponsor of the professional teacher preparation program documents the relationships between TPEs, tasks and scales.

1(b) To preserve the validity and fairness of the assessment over time, the sponsor may need to develop and field-test new pedagogical assessment tasks and multi-level scoring scales to replace or strengthen prior ones. Initially and periodically, the sponsor analyzes the assessment tasks and scoring scales to ensure that they yield important evidence that represents candidate knowledge and skill related to the TPEs, and serves as a basis for determining entry-level pedagogical competence to teach the curriculum and student population of California's K-12 public schools. The sponsor records the basis and results of each analysis, and modifies the tasks and scales as needed.

1(c) Consistent with the language of the TPEs, the sponsor defines scoring scales so different candidates for credentials can earn acceptable scores on the Teaching Performance Assessment with the use of different pedagogical practices that support implementation of the K-12 content standards and curriculum frameworks. The sponsor takes steps to plan and anticipate the appropriate scoring of candidates who use pedagogical practices that are educationally effective but not explicitly anticipated in the scoring scales.

1(d) The sponsor develops scoring scales and assessor training procedures that focus primarily on teaching performance and that minimize the effects of candidate factors that are not clearly related to pedagogical competence, which may include (depending on the circumstances) factors such as personal attire, appearance, demeanor, speech patterns and accents that are not likely to affect student learning.

1(e) The sponsor publishes a clear statement of the intended uses of the assessment. The statement demonstrates the sponsor's clear understanding of the high-stakes implications of the assessment for candidates, the public schools, and K-12 students. The statement includes appropriate cautions about additional or alternative uses for which the assessment is not valid. Before releasing information about the assessment design to another organization, the sponsor informs the organization that the assessment is valid only for determining the pedagogical competence of candidates for initial teaching credentials in California. All elements of assessment design and development are consistent with the intended use of the assessment for determining the pedagogical competence of candidates for Preliminary Teaching Credentials in California.

1(f) The sponsor completes content review and editing procedures to ensure that pedagogical assessment tasks and directions to candidates are culturally and linguistically sensitive, fair and appropriate for candidates from diverse backgrounds. The sponsor ensures that groups of candidates interpret the pedagogical tasks and the assessment directions as intended by the designers, and that assessment results are consistently reliable for each major group of candidates.

1(g) The sponsor completes basic psychometric analyses to identify pedagogical assessment tasks and/or scoring scales that show differential effects in relation to candidates' race, ethnicity, language, gender or disability. When group pass-rate differences are found, the sponsor investigates to determine whether the differences are attributable to (a) inadequate representation of the TPEs in the pedagogical tasks and/or scoring scales, or (b) overrepresentation of irrelevant skills, knowledge or abilities in the tasks/scales. The sponsor acts promptly to maximize the fairness of the assessment for all groups of candidates and documents the analysis process, findings, and action taken.

1(h) In designing assessment administration procedures, the sponsor includes administrative accommodations that preserve assessment validity while addressing issues of access for candidates with disabilities.

1(i) In the course of developing or adopting a passing standard that is demonstrably equivalent to or more rigorous than the State recommended standard, the sponsor secures and reflects on the considered judgments of teachers, the supervisors of teachers, the support providers of new teachers, and other preparers of teachers regarding necessary and acceptable levels of proficiency on the part of entry-level teachers. The sponsor periodically reconsiders the reasonableness of the scoring scales and established passing standard.

Assessment Design Standard 2: Assessment Designed for Reliability and Fairness

(Assessment Design Standard 2 Applies to Programs that Request Approval of Alternative Assessments)

The sponsor of the professional teacher preparation program requests approval of an assessment that will yield, in relation to the key aspects of the major domains of the TPEs, enough collective evidence of each candidate's pedagogical performance to serve as an adequate basis to judge the candidate's general pedagogical competence for a Preliminary Teaching Credential. The sponsor carefully monitors assessment development to ensure consistency with the stated purpose of the assessment. The Teaching Performance Assessment includes a comprehensive program to train and re-train assessors. The sponsor periodically evaluates assessment design to ensure equitable treatment of candidates. The assessment design and its implementation contribute to local and statewide consistency in the assessment of teaching competence.

Required Elements for Assessment Design Standard 2: Assessment Designed for Reliability and Fairness

2(a) In relation to the key aspects of the major domains of the TPEs, the pedagogical assessment tasks and the associated directions to candidates are designed to yield enough evidence for an overall judgment of each candidate's pedagogical qualifications for a Preliminary Teaching Credential. The program sponsor will document sufficiency of candidate performance evidence through thorough field-testing of pedagogical tasks, scoring scales, and directions to candidates.

2(b) Pedagogical assessment tasks and scoring scales are extensively field-tested in practice before being used operationally in the Teaching Performance Assessment. The sponsor of the program evaluates the field-test results thoroughly and documents the field-test design, participation, methods, results and interpretation.

2 (c) The Teaching Performance Assessment system includes a comprehensive program to train assessors who will score candidate responses to the pedagogical assessment tasks. An assessor training pilot program demonstrates convincingly that prospective and continuing assessors gain a deep understanding of the TPEs, the pedagogical assessment tasks and the multi-level scoring scales. The training program includes task-based scoring trials in which an assessment trainer evaluates and certifies each assessor's scoring accuracy in relation to the scoring scales associated with the task. When new pedagogical tasks and scoring scales are incorporated into the assessment, the sponsor provides additional training to the assessors, as needed.

2(d) In conjunction with the provisions of Teacher Preparation Program Standard 19, the sponsor plans and implements periodic evaluations of the assessor training program, which include systematic feedback from assessors and assessment trainers, and which lead to substantive improvements in the training as needed.

2(e) The program sponsor requests approval of a detailed plan for the scoring of selected assessment tasks by two trained assessors for the purpose of evaluating the reliability of scorers during field-testing and operational administration of the assessment. The subsequent assignment of one or two assessors to each assessment task is based on a cautious interpretation of the ongoing evaluation findings.

2(f) The sponsor carefully plans successive administrations of the assessment to ensure consistency in elements that contribute to the reliability of scores and the accurate determination of each candidate's passing status, including consistency in the difficulty of pedagogical assessment tasks, levels of teaching proficiency that are reflected in the multilevel scoring scales, and the overall level of performance required by the Commission's recommended passing standard on the assessment.

2(g) The sponsor ensures equivalent scoring across successive administrations of the assessment and between the Commission's model and local assessments by: using marker performances to facilitate the training of first-time assessors and the further training of continuing assessors; monitoring and recalibrating local scoring through third party reviews of scores that have been assigned to candidate responses; and periodically studying proficiency levels reflected in the adopted passing standard.

2(h) The sponsor investigates and documents the consistency of scores among and across assessors and across successive administrations of the assessment, with particular focus on the reliability of scores at and near the adopted passing standard. To ensure that the overall construct being assessed is cohesive, the sponsor demonstrates that scores on each pedagogical task are sufficiently correlated with overall scores on the remaining tasks in the assessment. The sponsor demonstrates that the assessment procedures, taken as a whole, maximize the accurate determination of each candidate's overall pass-fail status on the assessment.

2(i) The sponsor's assessment design includes an appeal procedure for candidates who do not pass the assessment, including an equitable process for rescoring of evidence already submitted by an appellant candidate in the program.

Appendix B
Profile of edTPA Pilot Participation in California

	Elementary			Secondary								Total
	Literacy	Math	Education (Literacy & Math)	English	Social Science	Math	Science	P.E.	Art	World Languages	Music	
California Polytechnic State University, San Luis Obispo		37		8	10	7	21					83
CalState TEACH			15									15
San Diego State University	19			15	12	17	10	4	2	3	4	86
San Francisco State University							10				12	22
UC Berkeley		19										19
UC Irvine		70										70
UC Los Angeles	48			15	16	11	12					102
UC Riverside				6		16						22
UC San Diego				1		5	8					14
UC Santa Barbara		28		11	4	10	10			7		70
Western Governors University (California)			3			2						5
Grand Total	67	154	18	56	42	68	71	4	2	10	16	508

Appendix C

Overview of the Briefing Book Standard Setting Method

July 27, 2013

Standard setting is a process of determining what score on a test or assessment demonstrates a specified level of performance. Very broadly, the process begins with a statement of the intended *performance standard* – that is, a description of what people meeting the performance standard know and are able to do. The goal is then to determine a *cut score* on an accompanying test or assessment that separates those who meet the performance standard from those who do not. From a technical standpoint it is important that the cut score accurately and reliably distinguish between people who do and do not meet the performance standard. However, because articulation of a performance standard and the accompanying cut score entail value judgments, it is also important to ensure the performance standard and cut score are appropriate for the intended use.

The Briefing Book Method (BBM) is an evidence-based standard setting method intended to develop an appropriate and defensible cut score that can be supported with a validity argument. The BBM provides a framework and approach to standard setting rather than a specific set of steps or procedures that must be followed exactly. The primary aim is to follow a process that allows a body with the appropriate authority and knowledge to reach a defensible and appropriate judgment of a passing cut score.

The BBM proceeds in a number of steps, including an articulation of the purpose for the standard setting, data collection and synthesis, a standard setting session, and continued evaluation.

1. *Define purpose of assessment and standard setting.* Here the purpose of setting a cut score is outlined. This describes how the assessment and cut score will be used. An articulation of the performance standard is formulated. When the performance standard is articulated here, it is essential that the performance standard represent an appropriate level for the intended use and that it be directly aligned to what the assessment measures.
2. *Initial administration and data collection.* The intended use of the assessment will dictate the data that need to be collected during this stage. Minimally, information about the distribution of scores on the target assessment across relevant groups is needed for construction of the briefing book. Additional data might include the results of validity or reliability studies conducted to inform what different scores on the assessment mean and how consistent they are for the intended use.

3. *Briefing book assembly.* The briefing book is the primary source of information for participants who will recommend a cut score. The briefing book describes the nature of the assessment and the goal of the standard setting process. In addition, the briefing book contains evidence to a) characterize the level of performance at different potential cut scores and b) provide contextual information about the likely impact and appropriateness of different potential cut scores (e.g., passing rates). The characterizations of performance at different potential cut scores serve as performance standards corresponding to each cut score. Contextual information informs participants about the likely impact of a potential cut score. Additional information can be included as available and necessary.
4. *Standard setting session.* A group of domain experts and relevant policy makers are convened as panelists for the standard setting session. These panelists are informed of the purpose of the assessment and provided with the briefing book. During a facilitated 1 or 2 day session panelists familiarize themselves with the assessment and with the information contained in the briefing book. Panelists recommend an initial cut score, which is then discussed and evaluated. At least one additional round of recommendations is usually conducted during the session, before the panel recommends a final cut score that best meets the needs of relevant stakeholders and the intended use of the assessment. Ideally this score is reached via consensus.
5. *Follow-up evaluation.* Following adoption of the cut score, subsequent administrations of the assessment are monitored to ensure the cut score is functioning as anticipated and is being used appropriately. This might include determining whether passing rates are at an acceptable level, whether those achieving passing scores demonstrate the intended level of performance in subsequent activities, and whether there is evidence of unequal passing rates or adverse impact across different groups of examinees.

References

- Haertel, E. H. (2002). Standard setting as a participatory process: Implications for validation of standards-based accountability programs. *Educational Measurement: Issues and Practice, 21*, 16–22. doi:10.1111/j.1745-3992.2002.tb00081.x
- Haertel, E. H. (2008). Standard setting. In K. E. Ryan & L. A. Shepard (Eds.), *The future of test-based educational accountability* (pp. 139–154). New York: Taylor & Francis.
- Haertel, E. H., Beimers, J. N., & Miles, J. A. (2012). The briefing book method. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 283–299). New York, NY: Routledge.
- McClarty, K. L., Way, W. D., Porter, A. C., Beimers, J. N., & Miles, J. A. (2013). Evidence-based standard setting: Establishing a validity framework for cut scores. *Educational Researcher, 42*(2), 78–88. doi:10.3102/0013189X12470855