
2C

Information

Educator Preparation Committee

Report on the Teaching Performance Assessment Comparability Study

Executive Summary: The agenda item presents the findings of the Comparability Study of the three Commission-approved Teaching Performance Assessment models for the Commission's consideration.

Recommended Action: For information only

Presenter: Mike Taylor, Consultant, Professional Services Division

Strategic Plan Goal

I. Educator Quality

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

Report on the Teaching Performance Assessment Comparability Study

Background

The 2015-16 state budget included an allocation of \$1 million to support an equating/comparability study of all Commission-approved Teaching Performance Assessment (TPA) models. A Request for Proposals (RFP) to secure a contractor to conduct an equating/comparability study was released on February 8, 2017.

There are three Commission-approved TPA models in use within California's Preliminary Multiple and Single Subject teacher preparation programs. These models are:

- The FAST (Fresno Assessment of Student Teachers), owned and operated by California State University, Fresno (Fresno State);
- The edTPA, owned by Stanford University, with the Evaluation Systems group of Pearson serving as the operational contractor; and
- The CalTPA (California Teaching Performance Assessment), owned by the Commission, with the Evaluation Systems group of Pearson serving as the operational contractor.

The Commission has an interest in determining the level of comparability across these approved TPA models and assuring that they each equitably assess candidates working toward a California Preliminary Multiple or Single Subject teaching credential. For this comparability study project the Commission was specifically interested in determining the following:

- The degree to which scores on the TPA models reliably demonstrate comparable levels of candidate competence with respect to the Teaching Performance Expectations (TPEs); and
- The degree to which the TPA models have been found to be comparable in terms of constructs measured, the relative weighting given to these constructs, if any, within the scoring process, the passing score standards, and candidate scores across models.

This report also provides guidance to the Commission regarding the potential implications of these findings, along with any other factors that should be considered by the Commission based on the findings. It is important to note that the statutory framework for performance assessments expressly allows variations in models as long as each model can be found to meet the Commission's Assessment Design Standards. There is no expectation, based on statute as well as Commission standards, that models be revised to achieve a greater degree of comparability.

The RFP was posted publicly on the Commission's website and following the public bidding process The Human Resources Research Organization (HumRRO) was selected as the contractor for the TPA Comparability Study.

Comparability Study Design

About HumRRO

Founded in 1951, HumRRO is an independent, U.S.-based, non-profit organization with experience collaborating with government, private, and nonprofit agencies to design, implement, and evaluate education, training, and personnel systems. Their professional staff is composed primarily of education researchers, statisticians, and measurement experts. Several staff members are nationally recognized experts in their fields.

HumRRO has recognized expertise in large scale assessment, including performance assessments and comparability studies. HumRRO staff has experience working in teacher certification and had worked on projects similar to the scope requested by the Commission for this study.

For this project, HumRRO subcontracted with POTHOS, Inc. to provide meeting and travel logistics support for onsite and Technical Advisory Panel meetings. POTHOS, Inc. is a veteran-owned, certified Disabled Veteran Business Enterprise (DVBE). POTHOS, Inc. is a globally recognized, full service, strategic meeting management, incentive and corporate travel company in the United States. The company has been in business for over 15 years and is headquartered in San Diego, California.

Summary of the Scope of Work

HumRRO used a “Theory of Action approach” (Kane, 2006; 2013) to identify eight claims that needed to be substantiated for this comparability study. The claims designed to be substantiated by the study are listed below.

- Claim 1: The TPA models are sufficiently comparable in their representation of the Commission’s Assessment Design Standards (ADS) and in their assessment and weighting of the Commission-adopted Teaching Performance Expectations (TPEs).
- Claim 2: The guidance and supports (e.g., guide/manual/handbook and other resources) provided by model sponsors to candidates and teacher preparation faculty is sufficiently clear and detailed to ensure that the model is implemented as designed and intended.
- Claim 3: The scoring rubrics for each TPA model are sufficiently clear and detailed to ensure that trained scorers can accurately and consistently score candidate submissions.
- Claim 4: For each TPA model, there is a comparable, comprehensive process to select, train, and establish calibration of the assessors who score candidate submissions.
- Claim 5: The standard-setting procedures used for each TPA model are sufficiently comparable and rigorous to ensure that the respective passing standards for each model accurately and consistently identify candidates possessing the requisite knowledge, skills, and abilities (KSAs) required to effectively teach the content area(s) authorized by the credential, regardless of the TPA model on which the candidate was assessed.

- Claim 6: The model sponsor for each TPA model conducts statistical analyses to identify differential effects in relation to candidates' race, ethnicity, language, gender or disability. Any differences are documented, and processes implemented to eliminate sources of construct-irrelevant variance.
- Claim 7: For each TPA model, the score reports (candidate-level and program-level) provide similar information about candidate outcomes and include clear guidance on how candidate score information should be used.
- Claim 8: The rubrics and score reports provide diagnostic information on candidates and on programs such that the strengths and weaknesses of each can be identified.

Seven distinct activities (studies) were designed to investigate these claims. The seven activities conducted for this comparability study are listed below. Details of each of these activities can be found in the final Comparability Study Report.

Activity 1: Evaluation and Comparison of Evidence across TPA Models for Adherence to Assessment Design Standards

Activity 2: Content Validity Comparability Analysis

Activity 3: Comparison of Stakeholder Input across TPA Models

Activity 4: Scoring Review: Comparison of Scoring Rubrics, Scorer Training, and Score Reports across TPA Models

Activity 5: Comparison of Standard Setting across TPA Models

Activity 6: Statistical Analysis and Comparison of Score Data across TPA Models

Activity 7: Comparison of TPA Models to a Common Criterion Measure

Technical Advisory Committee

As part of the contract, the Contractor (HumRRO) established a Technical Advisory Committee (TAC) to provide guidance on the design, implementation, and interpretation of results for the seven proposed activities. The TAC consisted of two representatives from each of the three TPA model sponsors, as well as two other recognized experts in the field, and Commission staff. The members of the TAC are listed below. The Executive Director and a number of staff were also present for each meeting of the TAC.

- Tom Guskey: Professor Emeritus in the College of Education, University of Kentucky
- Laurie Wise: Principal Scientist, HumRRO
- Ray Pecheone: Executive Director of the Stanford Center for Assessment, Learning, and Equity (SCALE)
- Nicole Merino: Director of Early Learning and Teacher Performance Assessment, SCALE
- Jeanie Behrend: Professor, Department of Curriculum and Instruction at Fresno State
- Amy Reising: Director of Performance Assessment Development, Commission on Teacher Credentialing

- Helene Mandell: Instructor (retired) for the Department of Learning and Teaching, University of San Diego

Dates and details of the TAC meetings held throughout the duration of the contract are contained in the final report, linked in this agenda item.

HumRRO staff also met with Commission staff every other week throughout the duration of the project to provide updates on the implementation and findings of the study.

Results

[The final Comparability Study Report](#) can be found on HumRRO's website. The final report contains an executive summary, detailed descriptions of all claims and activities, and detailed list of findings and implications for the Commission to consider. Selected findings by activity are extracted from the report and included below.

Activity 1: Evaluation and Comparison of Evidence across TPA Models for Adherence to Assessment Design Standards

Findings: The technical documentation indicates that all three TPA models mostly or fully adhere to the ADS. (Claim 1)

Activity 2: Content Validity Comparability Analysis

Findings: Overall, the findings for Activity 2 indicate that there are some differences in the emphasis and measurement of TPEs across the TPA models; however, there is more comparability than dissimilarity. (Claim 1)

Although they are different assessments with different designs, all three models comprehensively assess a number of specific TPEs and indicators in common, thus supporting the content validity of these assessments. CalTPA, edTPA, and FAST all assess TPEs 1.8, 3.1, 3.2, 3.5, 5.1, and 6.1 with a high level of content validity. In addition, both CalTPA and FAST also comprehensively measure TPEs 2.5, 4.1, and 5.2 with a similarly high level of content validity. These TPEs are provided in [Appendix A](#).

Activity 3: Comparison of Stakeholder Input across TPA Models

Findings: Survey results indicate that the majority of candidates across all three models agree that they understand the requirements (e.g., directions, rubrics, evidence requirements) for their TPA model. (Claim 2)

There was also consistently strong agreement across models that program coordinators have a clear understanding of their model's purpose and requirements, and that they felt well informed during the assessment process. (Claim 2)

Activity 4: Scoring Review: Comparison of Scoring Rubrics, Scorer Training, and Score Reports across TPA Models

Findings: A review of each model's scoring rubrics reveals that they are sufficiently clear and detailed to ensure that trained scorers can accurately and consistently score candidate submissions. (Claim 3)

Scorer trainings for all TPA models address key aspects of the ADS and Joint Standards related to scorer training. (Claim 4)

Score reports for all models are diagnostic.

Activity 5: Comparison of Standard Setting across TPA Models

Findings: edTPA and CalTPA procedures are sufficiently comparable and rigorous to ensure that their passing standards accurately and consistently identify candidates possessing the requisite KSAs required to effectively teach the content area(s) authorized by the credential. (Claim 5)

The FAST model used a nontraditional standard setting method whereby teacher preparation staff at Fresno State reviewed the Level 1 (“Does Not Meet Expectations”) and Level 2 (“Meets Expectations”) rubric descriptors to ensure that the Level 2 descriptors adequately described the KSAs of a just-sufficiently-qualified beginning teacher. (Claim 5)

Activity 6: Statistical Analysis and Comparison of Score Data across TPA Models

Findings: There is no evidence to suggest substantive differences in pass rates for males and females within TPA models. Moreover, the pattern of pass rates for males and females was comparable across models. In addition, when differences in mean total scores were examined, the magnitude of the differences between males and females were similarly small for all three models.

The findings indicate that the pass rates for the various race categories were similar both within and across models for edTPA and CalTPA. (Claim 6)

Activity 7: Comparison of TPA Models to a Common Criterion Measure

Findings: Despite the unique components and rubrics for each TPA model, all three models are measuring a highly related construct of teaching performance (based on the subset of TPEs that could be reliably compared).

Regardless of which teaching performance assessment a candidate completes, his/her performance is likely to be consistently classified as passing or failing by all three models. (Claim 5)

The complete set of findings by claim can be found in the [final report](#). The final report contains some recommendations for each model owner about ways in which each model might be improved, as well as some recommendations for the Commission regarding potential future research about the compatibility of multiple TPA models. The primary practical implication of this investigation “...is that it provides empirical evidence to support the Commission’s decision to approve multiple TPA models as a credentialing requirement for beginning teachers. Again, this is not to say that the models are equal, but rather that all models are likely to equitably identify teacher candidates who are “ready”—that is, possess the KSAs required for beginning teaching.” (Final Report, p. 226)

The recommendations below are also contained in the executive summary, along with some longer-range ideas for additional study. Based on this study, HumRRO suggests the Commission may want to consider the following additional activities:

- Conduct another content validity investigation (Activity 2) but expand upon it by having teacher preparation experts identify which aspects of each TPE element are assessed by each model.
- The above bullet point could be further extended by creating a new common rubric that more fully addresses the construct space, and then updating Activity 7 using this more robust common rubric.
- Activity 6 (investigation of score patterns for subgroups) could be conducted for other credential areas beyond the Multiple Subject credential. Also, Activity 6 could be expanded upon by investigating subgroup differences in score patterns for language and disability, assuming the models capture this demographic information in their score data. If multiple years of data were combined, then this would help to circumvent concerns regarding small samples.
- When/if notable changes are made to a model(s), any number of the seven activities could be repeated to evaluate the improved model.

Next Steps

Staff will continue to work with model sponsors to identify the data that needs to be submitted on an annual basis from each model sponsor to support the development of annual reports on each TPA model for the Commission. Staff will continue to collect data from model sponsors for ongoing monitoring of the findings of this comparability study, and ongoing monitoring of the implementation of multiple TPA models in California.

Appendix A

Teaching Performance Expectations (TPEs) Measured by TPA Models

While each of the TPA models approved for use by the Commission measures multiple TPEs, there are a number of TPEs that the comparability study determined were measured well by all of the TPA models. The definition of “strong evidence” used in the study is shown below.

- Strong evidence = The model measures this TPE and the evidence requires evaluation of multiple sources of information or application of significant conceptual understanding and higher-order thinking, *and* it covers the full breadth and depth of the TPE.

Each of the three TPA models requires strong evidence of knowledge of the following TPEs from candidates to achieve a passing score.

- 1.8 Monitor student learning and adjust instruction while teaching so that students continue to be actively engaged in learning.
- 3.1 Demonstrate knowledge of subject matter, including the adopted California State Standards and curriculum frameworks.
- 3.2 Use knowledge about students and learning goals to organize the curriculum to facilitate student understanding of subject matter, and make accommodations and/or modifications as needed to promote student access to the curriculum.
- 3.5 Adapt subject matter curriculum, organization, and planning to support the acquisition and use of academic language within learning activities to promote the subject matter knowledge of all students, including the full range of English learners, Standard English learners, students with disabilities, and students with other learning needs in the least restrictive environment.
- 5.1 Apply knowledge of the purposes, characteristics, and appropriate uses of different types of assessments (e.g., diagnostic, informal, formal, progress-monitoring, formative, summative, and performance) to design and administer classroom assessments, including use of scoring rubrics.
- 6.1 Reflect on their own teaching practice and level of subject matter and pedagogical knowledge to plan and implement instruction that can improve student learning.

In addition, it was determined that both CalTPA and FAST require strong evidence of knowledge of these additional TPEs:

- 2.5 Maintain high expectations for learning with appropriate support for the full range of students in the classroom.
- 4.1 Locate and apply information about students' current academic status, content- and standards-related learning needs and goals, assessment data, language proficiency status, and cultural background for both short-term and long-term instructional planning purposes.
- 5.2 Collect and analyze assessment data from multiple measures and sources to plan and modify instruction and document students' learning over time.