An Investigation of the Comparability of Commission-Approved Teaching Performance Assessment Models
Commission Meeting

Presenter: Andrea Sinclair, HumRRO

June ##, 2020

Purpose & Approach of Study

- To "assure that the Commission-approved TPA models are sufficiently comparable that they are equitably assessing candidates working toward a California preliminary multiple or single subject teaching credential" (RFP, p. 5)
  - **Comparable doesn't mean that the models are equal in how they measure the KSAs, but that _all models equitably identify TPE-ready professionals._**
- Using a "Theory of Action" approach (Kane, 2006, 2013) to identify the claims that need to be substantiated to attain this ultimate objective/end goal.
  - **Lack of support for any claim represents a threat to the veracity of this ultimate objective.**
- TAC provided guidance on the design, implementation, and interpretation of the results of the activities.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Claims 1 - 4

Claim 1	The TPA models are sufficiently comparable in their representation of the Commission's Assessment Design Standards 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 are 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.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Claims 5 - 8

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 required to effectively teach the content area(s) authorized by the credential.
Claim 6	Model sponsors for each TPA model conduct 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.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activities

- **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, Score Reports, and Rater Training

- **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

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Final Report

- Year 2 report (Final Report) was delivered in December 2019.
  - The Final Report reflects the status of the models as implemented in 2018-19.
  - Because some updates were made following implementation in 2018-19, findings may not reflect the current status of the models.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Overview of Findings

Activity 1: Overview

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

- Comprehensive review and comparison of documentation and materials to determine adherence to Assessment Design Standards (ADS).

- Cross-check against relevant test design standards from *The Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014)* to ensure that documentation covers industry-wide principles for test design as well.

- Overarching investigation of all eight claims.

Activity 1: Rating Scale

Rating Level	Description
1	No evidence of the Standard/element found in the documentation provided.
2	Little evidence of the Standard/element found in the documentation; less than half of the Standard/element covered in the documentation and/or evidence of key aspects of the Standard/element could not be found.
3	Some evidence of the Standard/element found in the documentation; approximately half of the Standard/element covered in the documentation including some key aspects of the Standard/element.
4	Evidence in the documentation mostly covers the Standard/element; more than half of the Standard/element covered in the documentation, including key aspects of the Standard/element.
5	Evidence in the documentation fully covers all aspects of the Standard/element.

Trained key staff evaluated the strength of evidence for each Standard.

HumRRO Project Director conducted a cross-check on the ratings; discrepancies discussed until consensus was reached.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 1: Findings

Summary Ratings on: Assessment Design Standards

Model	Min.	Max.	Mean	SD
FAST	4	5	**4.83**	**0.38**
edTPA	4	5	**4.76**	**0.44**
CalTPA	4	5	**4.83**	**0.38**

Note. Number of Assessment Design Standard elements = 18.

Summary Ratings on: Joint Standards

Model	Min.	Max.	Mean	SD
FAST	4	5	**4.64**	**0.50**
edTPA	4	5	**4.77**	**0.44**
CalTPA	4	5	**4.71**	**0.47**

Note. Number of relevant test design standards from *Joint Standards* = 14.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 1: Conclusions

- Overall, findings indicate that TPA models are either mostly or fully adhering to the ADS and *Joint Standards* with regard to available technical documentation.

- Supports Claim 1, which states in part that, "*The TPA models are sufficiently comparable in their representation of the Commission's Assessment Design Standards . . .*"

- Technical documentation from Year 1 to Year 2 improved.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 2: Overview

Content Validity Comparability Analysis

- Targeted investigation of **Claim 1** (*TPA models are sufficiently comparable in their representation of the Assessment Design Standards and in their assessment and weighting **of the TPEs***)
- Goes beyond an examination of simple coverage of TPEs
  - Workshop with two, 7-person panels of teacher preparation experts
  - Experts evaluated . . .
    - whether preparation materials, instructions, submission requirements, and actual candidate submissions for each model assess the KSAs for each TPE element
    - how the element is assessed (e.g., video, lesson plan, commentary)
    - the depth and breadth with which the KSAs required by the TPE are assessed by the TPA model
  - Discrepancies between panels discussed and came to consensus

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 2: Screenshot of TPE-to-TPA Linkage Dashboard

TPE 1: Engaging and Supporting all Students in Learning	edTPA				CalTPA			FAST		
Beginning Teachers:	Task 1	Task 2	Task 3	Strength	Cycle 1	Cycle 2	Strength	SVP	TSP	Strength
1. Apply knowledge of students, including their prior experiences, interests, mental health and social-emotional learning needs, as well as their funds of knowledge, cultural, language, and socio-economic backgrounds to engage them in learning.	Cnt Les InM Cmt	Vid Cmt	Cnt	🟡	Cnt Les AsD Cmt	Cnt Cmt	🟢	Cnt Les Cmt	Cnt Cmt Cmp	🟡
2. Maintain ongoing communication with students and parents regarding achievement expectations and support needs.	--[a]	--	Fdk Cmt	🔴	--	AsD Cmt	🔴	--	Asm AsD Fbk Cmt	🟡
3. Connect subject matter to real-life contexts and provide hands-on experiences to engage student interest, support student motivation, and allow students to extend their learning.	Les InM Asm Cmt	Vid Cmt	--	🟡	Les Cmt	Cnt Cmt	🟡	Les Cmt	Les Cmt	🟡
4. Use a variety of developmentally and ability-appropriate instructional strategies, resources, and assistive technology, including principles of Universal Design and a Multi-tiered System of Supports (MTSS), to support access to the curriculum for a wide range of learners within the (general education) classroom (and/or learning environment).	Cnt Les InM Asm Cmt	Vid Cmt	Cnt AsD Cmt	🟡	Les Cmt	Les Asm Cmt	🟡		Les Asm AsD Cmt	🟡
5. Promote students' critical and creative thinking and analysis through activities that provide opportunities for inquiry, collaborative problem solving, responding to and framing meaningful questions, and reflection.	--	--	Cmt	🟡	Les Vid Cmt	Les Vid Asm Cmt	🟢	Les Vid	Les Cmt	🟡
6. Provide a supportive learning environment for students' first and/or second language acquisition by using research-based instructional approaches, including focused English Language Development, Specially Designed Academic Instruction in English (SDAIE), scaffolding across content areas, structured English immersion, and determine communicative intent, particularly with students with low verbal abilities.	Cnt Les InM Asm Cmt	--	AsD Fdk Cmt	🟡	Vid Cnt Les Cmt	Cnt	🟡	Les Vid Cmt	Les Cmt Cmp	🟡

Innovative. Responsive. Impactful.

HumRRO. HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 2: Findings

- ADS 1(a) requires that each TPA task must be "substantively related to <u>two</u> or more major domains of the TPEs" and that "collectively, the tasks and rubrics in the assessment address key aspects of the six major domains of the TPEs."

  - All models exceed this Standard by having all tasks/cycles assess <u>three</u> or more TPE domains.

- Additionally, found that . . .

  - TPE 3 (*Understanding and Organizing Subject Matter for Student Learning*) is the domain assessed most thoroughly by all three TPAs.

  - TPE 6 (*Developing as a Professional Educator*) is the domain assessed least thoroughly by all three TPAs.

- In sum, Activity 2 revealed comparability across models in their assessment of TPE *domains.* There are some differences in emphasis and measurement of TPE *elements* across TPAs, with FAST and CalTPA being more comparable.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 2: Conclusions

- Provides an independent, empirical investigation on the content validity of each model

- There are some differences in the emphasis and measurement of TPEs across the TPA models; however, there is more comparability than dissimilarity across models.

  - Provides partial but not full support for the claim that the "*TPA models are sufficiently comparable in their assessment and weighting of the . . . TPEs*" (Claim 1).

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 3: Overview

Comparison of Stakeholder Input across TPA Models

- Investigates **Claim 2** (*The guidance and supports—e.g., guide/manual/handbook and other resources—provided by model sponsors to candidates and teacher preparation faculty are sufficiently clear and detailed to ensure that the model is implemented as designed and intended.*)

- Also addresses **Claim 1** (*TPA models are comparable in their representation of the Commission's Assessment Design Standards and in their assessment and weighting of the TPEs*)

- Two on-line surveys administered spring 2019
  - Candidate Survey
  - Program Coordinator Survey*

  *FAST refers to these as University Coaches

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION
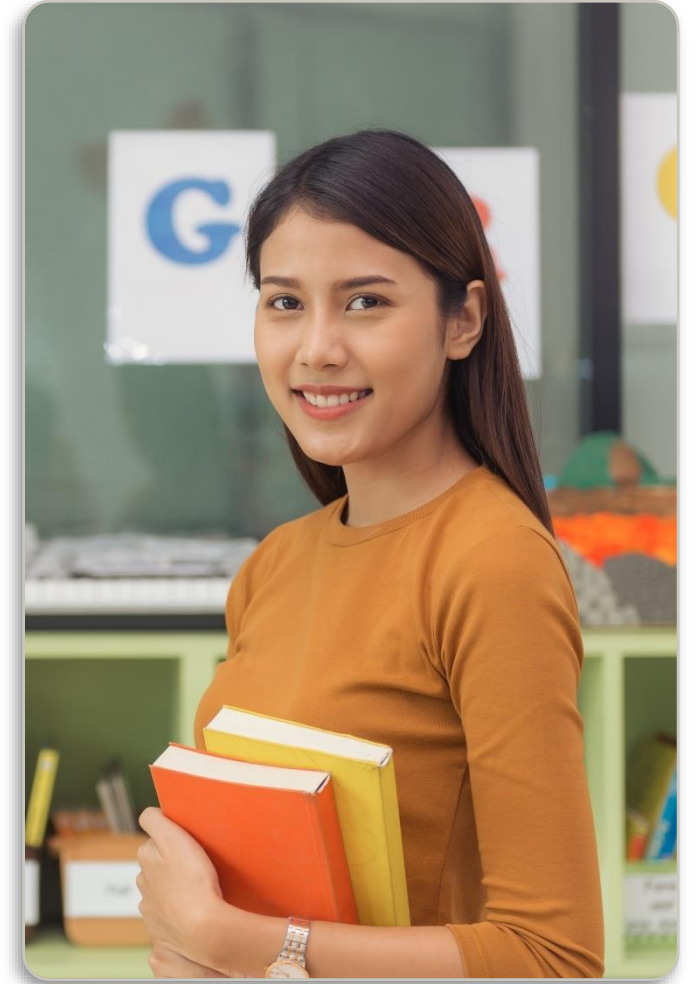
Activity 3: Response Rates

Candidate Response Rate

	FAST		edTPA		CalTPA	
	Year 1	**Year 2**	**Year 1**	**Year 2**	**Year 1**	**Year 2**
Total N	130	236	4,175	4,743	1,736	3,917
N Responded	42	106	103	363	505	407
N Valid Cases	40	94	96	330	474	407
Response Rate (%)	30.77	44.92	2.30	6.96	27.30	10.40

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 3: Response Rates Cont.

Coordinator Response Rate

	FAST		edTPA		CalTPA	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
Total N	39	70	75	58	36	93
N Responded	25	65	31	26	22	54
N Valid Cases	23	54	24	23	20	50
Response Rate (%)	58.97	77.14	32.00	39.66	55.56	53.76

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 3: Claim 2 Findings *(Candidates)*

- Majority of candidates agreed that they understood the requirements for their TPA model
  - FAST: 80% or more agreed
  - edTPA: 59 – 75% agreed
  - CalTPA: 58% agreed

- Majority of candidates agreed that their model's Manual/Handbook/Guide provided sufficient information to assist them throughout the assessment process
  - FAST: 75% agreed
  - edTPA: 66% agreed
  - CalTPA: 75% agreed

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 3: Claim 2 Findings *(Coordinators)*

- Vast majority of responding coordinators across all three models agreed that:
  - they had a clear understanding of their model's purpose (85% or more agreed)
  - they understood their model's requirements (85% or more agreed)
  - they were well informed about their model (90% or more agreed)

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 3: Claim 1 Findings *(Candidates)*

- Majority of FAST, edTPA, and CalTPA candidates agreed that the KSAs assessed by their model are emphasized in their preparation program
  - 85% or more of FAST candidates agreed
  - 72% or more of edTPA candidates agreed
  - 77% of CalTPA candidates agreed

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 3: Claim 1 Findings *(Coordinators)*

- Nearly all coordinator respondents agreed that their TPA model focuses on the appropriate skills and practices necessary for beginning teachers.

- Most coordinator respondents agreed that their TPA model appropriately assesses candidate readiness in the areas measured.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 3: Conclusions

- **Claim 2** *(guidance and support is sufficiently clear and detailed)*
  - Supported by candidates and coordinators

- **Claim 1** *(perceived validity)*
  - TPA models are perceived as valid by both candidates and coordinators

These findings should help to ensure that the models are implemented as intended.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 4: Overview

Scoring Review:
Comparison of Scoring Rubrics, Scorer Training and Score Reports

- Investigated **Claim 3** (*scoring rubrics sufficiently clear and detailed*), **Claim 4** (*comparable, comprehensive process to select, train, calibrate assessors who score submissions*), **Claim 7** (*score reports provide comparable information about candidate outcomes and how score information should be used.*), and **Claim 8** (*rubrics and score reports provide diagnostic information on candidates and programs*).

- Used information obtained via Activities 1 and 2

- Site observation of scorer training and calibration,* including interviews with key scoring personnel

- Evaluation of evidence per ADS and *Joint Standards*

*Participation in on-line training for edTPA

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 4: Findings

Summary Ratings on: Standards Relevant to Claim 3
(Scoring rubrics clear and detailed)

Model	Min.	Max.	Mean	SD
FAST	3	5	**4.14**	**1.07**
edTPA	4	5	**4.57**	**0.53**
CalTPA	4	5	**4.86**	**0.38**

Note. Number of Standards = 7.

Summary Ratings on: Standards Relevant to Claim 4
(Comprehensive process to select, train, and establish calibration of assessors)

Model	Min.	Max.	Mean	SD
FAST	1	5	**3.70**	**1.11**
edTPA	4	5	**4.96**	**0.21**
CalTPA	5	5	**5.00**	**0.00**

Note. Number of Standards = 24.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 4: Findings (Continued)

Summary Ratings on:
Standards Relevant to Claim 7
(Score reports provide similar information about candidate outcomes and clear guidance on how score information should be used)

Model	Min.	Max.	Mean	SD
FAST	3	5	**4.36**	**0.67**
edTPA	4	5	**4.64**	**0.50**
CalTPA	4	5	**4.70**	**0.48**

Note. Number of Standards = 12.

Summary Ratings on:
Standards Relevant to Claim 8
(Rubrics and score reports provide diagnostic information on candidates and on programs)

Model	Min.	Max.	Mean	SD
FAST	3	5	**4.20**	**0.84**
edTPA	4	5	**4.20**	**0.45**
CalTPA	4	5	**4.60**	**0.55**

Note. Number of Standards = 6.

Innovative. Responsive. Impactful.

HumRRO
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 4: Conclusions

- **Claim 3: Scoring rubrics sufficiently clear and detailed to ensure trained scorers can accurately and consistently score submissions**
  - All models have clear and detailed scoring rubrics that help to ensure trained scorers accurately and consistently score submissions, although clearer guidance to FAST scorers on how to weight indicator level ratings on each rubric may help to further strengthen scorer consistency
  - All models use multi-level scoring rubrics that are clearly related to teaching performance expectations (ADS1a), although linkage of scoring rubrics to TPEs are more readily available/transparent for FAST and CalTPA than edTPA

Innovative. Responsive. Impactful.

HumRRO
HUMAN RESOURCES RESEARCH ORGANIZATION

- **Claim 4: There is a comparable, comprehensive process to select, train, and establish calibration of the assessors who score candidate submissions**
  - All models carefully select, train, and calibrate scorers, although in 2018-19 returning FAST scorers attended training, but were not required to re-calibrate.
    - *Recommend requiring returning FAST scorers to recalibrate, particularly when updates are made to the rubric.*
  - edTPA and CalTPA have stronger procedures in place to monitor scoring consistency during the scoring window.
    - *Recommend embedding scorer consistency check during the FAST scoring window.*

28

Innovative. Responsive. Impactful.

HumRRO
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 4: Conclusions (Continued)

- **Claim 7: The score reports (candidate and program) provide similar information about candidate outcomes and include clear guidance on how scores should be used**

  – All models provide rubric level scores to candidates and programs within a timeframe that adheres to ADS requirements.

  – Unlike edTPA and CalTPA, FAST does not include the total score nor an indication of passing status on the score report (edTPA does not include passing status on score report like CalTPA, but does include weblink to where candidates can look it up).

  – All models provide guidance on how candidate score information should be used, although models differ with regard to inclusion of that info on score reports (i.e., appears in other supporting materials).

    - None of the models included guidance *on score reports* that scores should be used in conjunction with other measures for determining candidate readiness.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 4: Conclusions (Continued)

- **Claim 8: Rubrics and score reports provide diagnostic information on candidates and programs such that strengths and weaknesses can be identified**

  - Candidate reports are diagnostic in the sense that they report rubric level scores

  - Only CalTPA score report includes specific guidance on score reports that rubric level scores can be used to identify strengths and areas for improvement

    - Although FAST and edTPA include similar language in other supporting materials

  - At the program level, models make program level results available to programs, but from this study we cannot determine whether programs are using that information diagnostically

Innovative. Responsive. Impactful.

HumRRO
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 4: Overall Conclusion

- More similarities than differences across models in topics related to scoring and score reporting, although differences are notable

- On average, FAST rated slightly lower than edTPA and CalTPA, but lack of comparability may be balanced out, to some extent, by some of the unmeasured benefits that only a local program can achieve.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 5: Overview

Comparison of Standard Setting across TPA Models

- Investigates **Claim 5** *(standard setting procedures for each TPA model are sufficiently comparable and rigorous to ensure that the respective passing standard for each model accurately and consistently identifies candidates possessing the KSAs required to effectively teach the content area(s) authorized by the credential)*

- Observation of standard setting workshop for each TPA model*

- Evaluation of evidence per ADS and *Joint Standards*

*edTPA's standard setting occurred prior to this study

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 5: Findings

Summary Ratings on:
Standards Relevant to Claim 5

(Rigorous standard setting procedures to ensure respective passing standards for each model accurately and consistently identify candidates possessing requisite KSAs)

Model	Min.	Max.	Mean	SD
FAST	2	4	**3.25**	**0.96**
edTPA	5	5	**5.00**	**0.00**
CalTPA	5	5	**5.00**	**0.00**

Note. Number of Standards = 4.

Activity 5: Conclusions

- **edTPA and CalTPA used standard setting procedures that are comparable and rigorous**
  - Briefing Book Method
  - Utilized candidate score profiles
  - Impact data
- **FAST did not use a comparable standard setting procedure**
  - Took the form of verifying, through consensus discussion, that Level 2 descriptor for each rubric accurately describes a just sufficiently qualified candidate
  - No empirical data (e.g., impact data), candidate score profiles, or panelist ratings used to inform judgments related to the passing standard for FAST
  - Different but not necessarily wrong
    - Nonetheless, recommend including performance data and actual student submissions representing the range of performance levels, and enhancing documentation

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 6: Overview

Statistical Analysis and Comparison of Score Data across Models

- Investigates **Claim 6**
  (*Model sponsors conduct statistical analyses to identify differential effects in relation to candidate characteristics . . .*)

- Comparison of pass/fail score patterns across subgroups

  – The subgroups specified in ADS 1(k) are "race, ethnicity, language, gender or disability."

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 6: Overview (Continued)

- Some considerations to note:
    - Multiple subject credential was the only credential area with a large enough sample for all three models to conduct subgroup analyses
    - Only had subgroup information on race/ethnicity and gender (not language or disability)
    - Conducted subgroup analyses on first attempt and final attempt
        - **100% of FAST multiple subject candidates passed on final attempt**
    - Differential impact is a necessary, but insufficient indicator of bias— i.e., differences in pass rates may suggest, but not solely indicate, a construct-irrelevant source that explains the difference

Innovative. Responsive. Impactful.

Activity 6: Findings

- **Before investigating subgroup differences in pass rates, investigated population differences**
  - **Race**
    - Frequency distribution of race categories similar for edTPA and CalTPA
    - Frequency distribution of race categories statistically different for FAST
      - More Hispanic than White candidates
    - Thus, FAST not included in comparison of pass/fail patterns by race
  - **Gender**
    - Gender distribution across models was similar
    - Thus, all models included in comparison of pass/fail patterns by gender

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 6: Findings

- **Pass Rates Overall**
  - **First Attempt**
    - Candidates slightly more likely to pass CalTPA than edTPA and considerably more likely to pass CalTPA than FAST
  - **Final Attempt**
    - FAST not included in pass/fail patterns in Final Attempts because 100% pass final attempt
    - Difference in pass rates still exists between CalTPA and edTPA when considering final attempt

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 6: Findings (Gender)

- Candidates had similar final pass rates by gender across all three models
  - Although some statistically significant results, the effect sizes and actual differences in % passing/failing were small
- Also, investigated differences in mean total scores (Cohen's d)
  - Small differences
    - Mean total scores for females slightly higher than males for edTPA and CalTPA
    - That pattern was reversed for FAST, although only 20 male candidates

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 6: Findings (Race)

- Final pass rates by race were statistically significantly different for edTPA and CalTPA, although actual percentage differences were small.*

- Mean differences in total scores (Cohen's d)
  - White candidates tended to score higher than their non-white counterparts, although differences were near zero or small in all instances, except for the difference between Hispanic and White candidates on FAST.
    - Hispanics earned, on average, about two fewer points

*FAST was not included in this analysis due to a lack of similarity in race distribution with the other models *and* due to the 100% pass rate for final attempts.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 6: Conclusions

- Candidates had similar pass rates by gender across the three models and differences in mean total scores were similarly small across models

- Findings comparing models by race were more complex
  - Distribution of racial categories for FAST differs from edTPA and CalTPA
  - Some small differences in pass rates for edTPA and CalTPA
  - When considering differences in mean total scores, white candidates tend to score slightly higher than non-white counterparts, although there was a large effect size difference between Whites and Hispanics on FAST

- Analyses should be revisited as more data becomes available

- Findings should not be extrapolated beyond the multiple subject credential

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 7: Overview

- Developed a common external referent on which all TPAs can be directly compared
- Purpose was to investigate the ultimate claim/objective:
  - "all models equitably identify TPE-ready professionals"
- Used outcome from Activity 2 as the blueprint to develop a Common Rubric
  - Only measured those TPE elements that were identified as being assessed in substantively the same way across models and for which all models measured the full depth and breadth of the TPE element
- Trained scorers scored a representative sample of multiple subject portfolios from each model using this Common Rubric
- Conducted comparability analyses across TPA models using scores on the Common Rubric as a referent

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 7: Analyses

- Computed correlations between the Common Rubric total scores and TPA Rubric total scores for each model
  - To determine the extent to which models measure a common construct of teaching performance
- Linear regression of the Common Rubric total score on each TPA model total score to predict cut scores on the Common Rubric using each TPA model's cut score
  - To determine whether the candidates' pass/fail rates were similar across models
- Using the model-predicted Common Rubric cut scores, we conducted classification consistency analyses
  - To determine the degree to which the Common Rubric and TPA Rubric classify candidates consistently

Activity 7: Findings (Correlation Analyses)

Correlations between Model Rubric and TPA Rubric Total Scores

edTPA	CalTPA	FAST
.75***	.41*	.46*

- Strong to moderately strong positive correlations for all models
- Supports that each of the TPA models measures a domain that is common across the unique tasks and rubrics for each model

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 7: Findings (Cut Score Comparisons)

TPA Model Predicted Cut Scores on Common Rubric Range

	Predicted Cut Score	Lower 95% Confidence Interval	Upper 95% Confidence Interval
edTPA	21.66	20.77	22.55
CalTPA	22.48	21.07	23.89
FAST	19.49	17.66	21.32

- Overlapping confidence intervals across the TPA models suggest that the models are likely to classify candidates similarly.

- FAST predicted cut score is lower than the predicted cut score for edTPA or CalTPA, but the upper range of the FAST cut score confidence interval (21.32) exceeds the lower range of the edTPA and CalTPA cut scores (20.77 and 21.07, respectively), suggesting that the TPA models would classify candidates comparably based on these common TPE elements.

45

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Classification Consistency Analysis by Common Rubric and TPA Rubric

	edTPA		CalTPA		FAST	
	Pass%	**Fail%**	**Pass%**	**Fail%**	**Pass%**	**Fail%**
Common Rubric Pass	**79.23**	3.44	**63.33**	10	**75.86**	3.44
	(23)	(1)	**(19)**	(3)	**(22)**	(1)
Common Rubric Fail	10.34	**6.90**	20.00	**6.66**	17.24	**3.44**
	(3)	**(2)**	(6)	**(2)**	(5)	**(1)**
Total Classification Consistency %	**86.13%**		**70.00%**		**79.30%**	

- The more classifications in green, the more consistent the classifications across the Common Rubric and the TPA Rubric.

- These strong classification consistency results provide further evidence that the models likely classify candidates in similar ways.

46

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Activity 7: Conclusions

- Despite the unique components and rubrics for each TPA model, **all three models are measuring a highly related construct of teaching performance**.

- 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.

- **Findings support that all three models "equitably identify TPE-ready professionals," BUT important to note that this is based on a small subset of TPEs on which all TPAs could be reliably compared.**

47

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

In Summary . . .

- Overall, the body of evidence provides empirical evidence to support the Commission's decision to approve multiple TPA models as a credentialing requirement for beginning teachers.

- 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.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Future Research Considerations . . .

- 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.

Innovative. Responsive. Impactful.

HumRRO.
HUMAN RESOURCES RESEARCH ORGANIZATION

Questions?

Innovative. Responsive. Impactful.