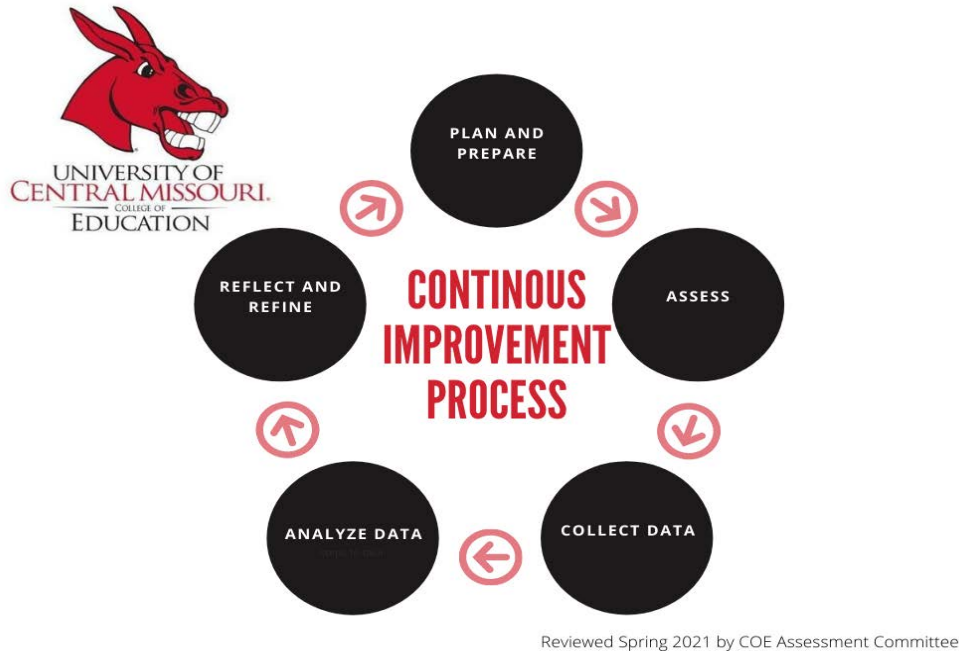


Analyses of Selected Unit Wide Data across Three Cycles



Prepared by Dr. Karrie Snider, Associate Professor
College of Education Assessment Committee, Chair, 2022
Quality Assurance CAEP Standard 5 Workgroup

Acknowledgements for input, review and feedback throughout the evaluation process during initial meeting review, April 14, 2022 & final, May 15, 2022:

Dr. Ann McCoy, Dean

Dr. Yuankun Yao, Professor

*Dr. Matt Thomas, Professor, Chair of School of Teaching and Learning
Quality Assurance CAEP Standard 5 Workgroup members:*

Drs. Rene Burress, Michelle Conrad, Michael Pantleo, Lisa Robbins

Executive Summary

The review of unit wide data from three continuous improvement cycles using varied approaches to analyze the data from initial teacher certification programs rendered positive results related to Educator Preparation Program (EPP) completers.

Two primary EPP summative assessments were targeted for this review: the Missouri Educator Evaluation System (MEES) and the Student Teacher Work Sample (STWS).

The MEES continues to be a reliable and valid measure to assess EPP completers' success towards effective teaching standards established by the State of Missouri. Evaluators (University Supervisors and Cooperating Teachers) demonstrated adequate agreement on teacher candidate performance on the MEES based on correlation analyses results. Further analyses indicated MEES ratings demonstrated statistically significant positive outcomes on means of the nine standards. The MEES standard with the highest mean was standard 1 student engagement in subject matter, the MEES standard with the lowest mean was effective standard 7 use of assessments. There were no significant differences on MEES total mean scores across different demographic characteristics (gender, type of program, and race). There were however differences detected in means between the three cycles (three groups of completers over three semesters). Observations of the up/down/up trend in MEES total means maybe explained due to COVID-19.

The data also revealed promising progress in the development and implementation of the initial then revised Student Teacher Work Sample. The revised STWS yielded adequate reliability statistics, which was an improvement over the initial STWS used to assess teacher candidate performance during the first two cycles.

Ongoing conversations among Professional Education Faculty within various groups will assist in the continued review, reflection, and improved EPP outcomes. Additional analyses of STWS will enable further understandings of reliability and inter-rater agreement.

Educator Preparation Program Program Completion Assessments

Missouri Educator Evaluation System (MEES)

The MEES is a state proprietary measure used at program completion to measure teacher candidates on nine constructs of effective teaching (Table 1). Raters assign scores (0, not evident; 1, Emerging Candidate; 2, Developing Candidate; 3, Skilled Candidate; 4, Exceeding Candidate) which render means across nine standards and one total score for each teacher candidate. University supervisors (USUPs) and cooperating teachers (CTs) separately rate teacher candidates at three time points throughout the student teaching semester. A summative evaluation, given at the last of the time points, is completed by USUPs and CTs, as well as a self-assessment by the teacher candidate.

Reliability

The MEES was used to assess TCs (N=299) on nine constructs of effective teaching across three cycles (Fall 2020, Spring 2021, Fall 2021). University Supervisor, Cooperating Teacher and self-assessed ratings of teacher candidates' teaching were analyzed. The scale had adequate levels of internal consistency (DeVellis, 2003; Kline, 2005) as determined by Cronbach's alpha of 0.763, 0.781 and 0.761 on 10 items respectively. Below are descriptive statistics for the MEES across the three cycles of this self-study (see Table 2 & 3).

Table 1. *Missouri Educator Evaluation System Constructs*

<i>Standard</i>	<i>Description</i>
Student engagement in subject matter	<i>Standard 1: Content knowledge aligned with appropriate instruction. The teacher candidate understands the central concepts, structures, and tools of inquiry of the discipline(s) and creates learning experiences that make these aspects of subject matter meaningful and engaging for students.</i>
Differentiated lesson design	<i>Standard 2: Student Learning, Growth, and Development. The teacher candidate understands how students learn, develop, and differ in their approaches to learning. The teacher candidate provides learning opportunities that are adapted to diverse learners and support the intellectual, social, and personal development of all students.</i>
Implementation of curriculum standards	<i>Standard 3: Curriculum Implementation. The teacher candidate recognizes the importance of long-range planning and curriculum</i>

	<i>development. The teacher candidate implements curriculum based upon student, district and state standards.</i>
Student engagement in critical thinking	<i>Standard 4: Critical Thinking. The teacher candidate uses a variety of instructional strategies and resources to encourage students' critical thinking, problem solving, and performance skills.</i>
Classroom Management	<i>Standard 5: Positive Classroom Environment. The teacher candidate uses an understanding of individual/group motivation and behavior to create a learning environment that encourages active engagement in learning, positive social interaction, and self-motivation.</i>
Verbal and non-verbal communication	<i>Standard 6: Effective Communication. The teacher candidate models effective verbal, nonverbal, and media communication techniques with students, colleagues and families to foster active inquiry, collaboration, and supportive interaction in the classroom.</i>
Effective Use of Assessment	<i>Standard 7: Student Assessment and Data Analysis. The teacher candidate understands and uses formative and summative assessment strategies to assess the learner's progress and uses both classroom and standardized assessment data to plan ongoing instruction.</i>
Self-Assessment	<i>Standard 8: Professionalism. The teacher candidate is a reflective practitioner who continually assesses the effects of choices and actions on others. The teacher candidate actively seeks out opportunities to grow professionally in order to improve learning for all students.</i>
Collaborative Partnerships	<i>Standard 9: Professional Collaboration. The teacher candidate has effective working relationships with students, families, school colleagues, and community members.</i>

Table 2. Descriptive Statistics of the Missouri Educator Evaluation System (MEES) Ratings on across three Self-Study Cycles (N=299)

MEES Standard	USUP		CT		TC self-assessed	
	M	SD	M	SD	M	SD
Student engagement in subject matter	3.39	.51	3.28	.55	3.28	.50
Differentiated lesson design	3.25	.50	3.25	.60	3.29	.53
Implementation of curriculum standards	3.30	.50	3.24	.59	3.24	.54
Student engagement in critical thinking	3.23	.56	3.15	.59	3.20	.54
Classroom Management	3.37	.52	3.38	.62	3.48	.60
Verbal and non-verbal communication	3.35	.50	3.30	.58	3.30	.52
Effective Use of Assessment	3.16	.51	3.22	.54	3.22	.56
Self-Assessment	3.34	.55	3.44	.60	3.44	.51
Cooperative Partnerships	3.33	.54	3.35	.55	3.34	.51
Total MEES Score	29.73	3.25	29.60	4.07	29.80	3.17

Table 3. Frequencies of MEES Ratings by Rater (N=299)

Rater	N	Mean	SD	Mode	Minimum	Maximum
USUP	275	29.77	3.19	27	19	36
CT	275	29.63	4.00	27	15	36
TC*	263	29.75	3.09	27	22	36

*Missing 12

Assessment of MEES Mean Scores

A one-sample t-test compares the mean of a sample to a pre-specified value. The one-sample t-test was applied to examine whether ratings of teacher candidates' performance as assessed by University Supervisors on the MEES Standards differed significantly from the population, using the pre-specified mean of 3 (skilled candidate), the expected level of performance by the end of student teaching semester. Means on all nine MEES standards (N=299) were statistically significantly higher than population mean of 3 (expected level of performance by the end of student teaching).

- Mean MEES standard 1 student engagement score (M=3.39, SD= 0.51) was statistically significantly higher by a mean of 0.39, 95% CI [0.33, 0.45], than the expected level of performance score of 3.0, $t(298) = 13.292, p < .001$. Cohen's $d=.77$.
- The mean MEES standard 2 differentiated lesson design score (M=3.25, SD=0.50) was statistically significantly higher by a mean of 0.25, 95% CI [0.19, 0.30], than the expected level of performance score of 3.0, $t(298) = 8.606, p < .001$. Cohen's $d=.50$.
- The mean MEES standard 3 implementation of curriculum standards score (M=3.30, SD=0.50) was statistically significantly higher by a mean of 0.30, 95% CI [30.245 to 30.36], than the expected level of performance score of 3.0, $t(298) = 10.469, p < .001$. Cohen's $d=.60$.
- The mean MEES standard 4 student engagement in critical thinking score (M=3.23, SD=0.56) was statistically significantly higher by a mean of 0.23, 95% CI [0.17, 0.30], than the expected level of performance score of 3.0, $t(298) = 7.233, p < .001$. Cohen's $d=.42$.
- The mean MEES standard 5 classroom management score (M=3.37, SD= 0.52) was statistically significantly higher by a mean of 0.37, 95% CI [0.17 to 0.30], than the expected level of performance score of 3.0, $t(298) = 12.314, p < .001$. Cohen's $d=.71$.
- The mean MEES standard 6 verbal and non-verbal communication (M=3.35, SD= 0.50) was statistically significantly higher by a mean of 0.35, 95% CI [0.17 to 0.30], than the expected level of performance score of 3.0, $t(298) = 12.175, p < .001$. Cohen's $d=.70$.
- The mean MEES standard 7 effective use of assessment score (M=3.16, SD=.51) was statistically significantly higher by a mean of 0.16, 95% CI [0.11, 0.22], than the expected level of performance score of 3.0, a difference of 0.16, $t(298) = 5.576, p < .001$. Cohen's $d=.32$.

- The mean MEES standard 8 self-assessment score ($M=3.34$, $SD=0.55$) was statistically significantly higher by a mean of 0.34, 95% CI [0.28, 0.41], than the expected level of performance score of 3.0, $t(298)=10.69$, $p<.001$. Cohen's $d=.62$.
- The mean MEES standard 9 cooperative partnerships score ($M=3.33$, $SD=0.54$) was statistically significantly higher by a mean of 0.33, 95% CI [0.17 to 0.30], than the expected level of performance score of 3.0, $t(298)=10.644$, $p<.001$. Cohen's $d=.61$.

Based on interpretation of values of Cohen's d , the effect sizes of each analysis range from small (0.32) to moderate (.77).

Correlation Analysis

To assess the relationship between ratings assigned by both university supervisors and cooperating teachers across nine standards on the MEES, a Pearson correlation analysis was applied. Several positive correlations emerged (see Table 4 below). There was a statistically significant, moderate positive correlation between University Supervisor and Cooperating Teacher total scores ratings, $r(297)=.416$, $p=.01$. These results indicate a moderately strong correlation between raters of teacher candidate performance (See Table 4).

Table 4. *Pearson correlations of University Supervisor, Cooperating Teacher and Teacher Candidate Ratings on the MEES and Teacher Work Sample (N=283)*

	USUP Total MEES Score	CT Total MEES Score	TC Total Score
CT Total MEES Score	.416**		
TC Total MEES Score	.420**	.369**	
TWS Total Score	.247**	.146*	.120

Note: **Correlation is significant at the 0.01 level (two-tailed). *Correlation is significant at the 0.05 level.

Analyses for Differences by Groups

Several tests were performed to assess for differences in dependent variables across various independent variables that produced categorical variables. The following summarizes the examination of differences by gender, race (white/non-white), program type (traditional/alternative), semester of candidates' completion (cycle 1, 2 or 3) and for type of college student (first gen/non-first gen) .

Differences by Gender

An independent samples t-test was conducted to determine if there were differences in MEES ratings based on gender. On average female students had slightly higher ratings across each rater. There was no statistically significant difference in total MEES scores for female or male teacher candidates (see Table 5 for descriptive statistics and Table 6 for test statistics).

Table 5. *Descriptive Statistics of MEES Total Scores by Teacher Candidate Gender*

Rater	TC Gender	N	M	SD
USUP	Female	244	29.78	3.36
	Male	55	29.55	2.72
CT	Female	244	29.69	4.09
	Male	55	29.21	4.00
TC*	Female	235	29.93	3.20
	Male	48	29.12	2.85

Note: *Some teacher candidates did not complete their self-assessment

Table 6. *Independent Samples t test to Assess MEES Total Score Mean Differences by Gender*

Rater	t test statistics			
	<i>t</i>	<i>df</i>	<i>P</i>	<i>F</i>
USUP	.472	297	.082	3.045
CT	.783	297	.539	.378
TC	1.618	281	.161	1.980

Differences by Race/Ethnicity (White TCs/TCs of Color)

A dichotomous variable was created to adjust for uneven groups among teacher candidates' self-reported race/ethnicity (Asian, black, Hispanic/Latinx, white, multi-racial and unreported). There were 272 white and 24 TCs of color across the three data cycles. The uneven nature of these two groups should be taken into account when interpreting the results.

To assess for differences among TCs according to racial and ethnic background, an independent samples t-test was applied. There was no statistically significant difference in MEES mean total

scores between white and non-white candidates by USUP and CT raters. However, there was a statistically significant difference in MEES mean scores on candidates' self-assessments between white and non-white candidates. Non-white candidates' self-assessed MEES total score mean (30.45) was higher than white students' self-assessment MEES total score mean (29.70). MEES total score ratings yielded on self-assessments indicated that the non-white candidates' mean MEES total score was 0.75(95% CI, -2.18 to 0.68) higher than the white candidates' MEES total score mean. (See Tables 7 & 8).

Table 7. *Descriptive statistics of MEES Total Scores of White TCs and TCs by color*

Rate	Group	N	M	SD	Median	Mode	Min	Max
USUP	White	272	29.74	3.16	29.73	27	19	36
	TCs of color	24	29.29	3.96	29.50	27	17	35
CT	White	272	29.62	4.02	29.62	27	15	36
	TCs of color	24	29.20	4.93	31.00	31	17	36
TC	White	261	29.70	3.05	29.70	27	22	36
	TCs of color	24	30.45	3.99	30.45	30.5	23	36

Table 8. *Independent Samples t-test to Assess Differences on MEES Total Score Means of White (n=272) and TCs of Color (n=24)*

t test statistics				
Rater	t	df	P	F
USUP	.650	294	.351	.872
CT	.473	294	.395	.725
TC	-1.031	279	.021	5.377

Differences by Program Type (Traditional vs Alternative Certification)

TCs within initial certification programs may be traditional undergraduate students or students who have earned a four-year degree and seek state certification through the alternative certification pathway. Descriptive statistics indicate that each group of raters' MEES total score mean was higher for alternative certification TCs than traditional. Caution with interpreting scores should be taken, due to uneven groups. Three independent samples t-tests were applied,

by each rater, to test for differences in MEES total score means for traditional and alternative certification TCs. There were no statistically significant differences for either group for either rater. See Table 9.

Table 9. *Descriptive statistics of MEES Total Scores of Traditional vs Alternative Certification TCs by Rater*

Rater	Group	N	M	SD	Median	Mode	Min	Max
USUP	Traditional	266	29.67	3.30	29	27	17	36
	Alt Cert	33	30.37	2.75	30	28	26	36
CT	Traditional	266	29.58	4.07	29	27	15	36
	Alt Cert	33	29.71	4.17	30	31	21	36
TC	Traditional	256	29.75	3.20	29	27	22	36
	Alt Cert	27	30.19	2.76	30	28 ^a	26	35

Note: ^aMultiple modes exist. The smallest value is shown.

Table 10. *Independent Samples t-test to Assess Differences of MEES Total Score Means between Traditional (n=266) and Alternative Certification (n=33) TCs*

<i>t test statistics</i>				
Rater	t	df	P	F
USUP	-1.013	297	.375	.790
CT	-.171	297	.456	.556
TC	-.760	297	.445	.586

Assessment of Differences according to Cycle

Any interpretation of results for the comparisons below, that is of TCs MEES ratings across cycles, must be taken with caution as cycles vary by participant number. Also, one may need to consider the conditions of learning to teach (as a college student) and teaching (as a student teacher) within COVID-19 teaching environments. When looking at the means across cycles, one may see a dip in means at the second cycle across all standards but 7, then an increase across all

standards. These observations may also be explained by the difference in sizes related to each group.

Due to unequal groups across three cycles, an independent samples t-test was applied to assess for differences between two groups: fall 2020 (n=67) and fall 2021 (n=60). There was a statistically significant difference between fall 2020 and fall 2021 total MEES mean scores, as fall 2021 TCs were scored higher than fall 2020 on the total MEES mean, -.98 (CI -2.10 to .14) $t(125)=-1.178, p=.091$. (See Table 11).

Table 11. Descriptive Statistics of the Missouri Educator Evaluation System (MEES) Ratings of TCs performance as assigned by University Supervisors on across three Self-Study Cycles (N=299)

MEES Standard	Fall 2020 (n=67)		Spring 2021 (n=172)		Fall 2021 (n=60)	
	M	SD	M	SD	M	SD
Student engagement in subject matter	3.45	.50	3.34	.50	3.45	.53
Differentiated lesson design	3.30	.49	3.21	.47	3.28	.58
Implementation of curriculum standards	3.33	.53	3.24	.47	3.45	.53
Student engagement in critical thinking	3.16	.59	3.23	.51	3.32	.65
Classroom Management	3.41	.50	3.29	.49	3.53	.57
Verbal and non-verbal communication	3.37	.49	3.31	.49	3.43	.53
Effective Use of Assessment	3.12	.44	3.15	.52	3.27	.54
Self-Assessment	3.31	.56	3.28	.55	3.57	.50
Cooperative Partnerships	3.34	.51	3.28	.54	3.48	.53
Total MEES Score	29.80	2.95	29.34	3.22	30.78	3.45

Student Teacher Work Sample (STWS)

The UCM Student Teacher Work Sample (STWS) is a summative performance evaluation tool to assess candidates immediately prior to program completion. The STWS aligns with national standards (CAEP and InTASC) as well as Missouri Teacher Standards. Table 12 & 13 illustrates the descriptive statistics

Table 12. Descriptive Statistics of STWS “Squishy Pilot”: Tool Development and Field Testing (Davis & Peck, 2021) (N=213)

TWS Indicator/Score	M	SD	Min	Max
Knowledge of community	2.99	.11	2	3
Knowledge of student factors	2.97	.18	2	3
Knowledge of district, school and classroom	2.97	.17	2	3
Analysis interpretation of assessment data	25.03	2.71	7	9
Focus students	14.61	1.37	7	15
Evidence of impact on student learning	9.72	.88	7	10
Instructional strategy based on contextual factors	9.69	.92	7	10
Self-evaluation	9.86	.63	7	10
Implications for future teaching PD and plans	9.79	.77	7	10
Cooperative partnerships	4.85	.60	7	10
Professionalism	4.84	.60	1	5
Total Score	97.33	4.94	77	100

Table 13. *Descriptive Statistics of STWS “Squishy Pilot”: Field Testing of Revised & Adopted STWS (Davis & Peck, 2021) (N=59)*

Item	M	SD	Min	Max
Knowledge of Community	2.00	.00	2	2
Knowledge of Student Factors	1.98	.13	1	2
Knowledge of District/School/and Classroom Factors	1.94	.22	1	2
Measurable Student Objectives Aligned with Standards	2.94	.22	2	3
Lesson Sequence	10.46	1.02	7	11
Resources	1.98	.13	1	2
Differentiation	2.83	.37	2	3
Accommodations and modifications	2.86	.34	2	3
Assessment methods	4.72	.55	3	5
Analysis of Student Learning	14.97	1.66	10	16
Focus students	7.66	.73	5	8
Evidence of Impact on student learning	4.89	.36	3	5
Instructional Strategy based on Contextual Factors	4.79	.48	3	5
Self-evaluation	8.49	1.01	5	9
Implications for Future Teaching/Professional Development	8.33	1.33	3	9
Cooperative partnerships	4.78	.74	1	5
Professionalism	4.67	.72	2	5
Technology to Enhance Student Learning	4.83	.56	2	5
Total Score	95.27	5.96	75	100

Reliability & Validity of STWS

The reliability of the STWS (initial and revised version) was assessed using a reliability analysis in SPSS to render a Cronbach alpha reliability statistic; a generally accepted measure of internal consistency--how close a set of items on a scale are. Reliability coefficient for the 11 item initial STWS scale used in the Fall 2020 and Spring 2021 (n=213) was .545. The revised version of the revised STWS for Fall 2021, which included 7 additional scale items for a total of 18, as well as

utilized an increased variation in scoring across the rubric yielded a reliability coefficient of .756. Improvements are observed from the first scale development to the revised scale, which yielded a higher value of Cronbach's alpha, and one that is in the acceptable range (DeVillis, 2003; Kline, 2005).

Face and Content Validity. Items on the STWS assess and represent knowledge, skills and dispositions that are identified by the InTASC, MTS and MEES standards as relevant to highly effective teaching. It is a performance-based tool that assesses student teachers on how to select, plan, implement, differentiate and engage students during instruction and within positive learning environments, as well as use communication, professionalism technology and collaboration in their professional role. The STWS developed first from a standardized framework used in teacher education programs across the United States. Historically, the STWS had been adopted into the EPP prior to 2010 when University of Central Missouri were charter members of a consortium that created the Renaissance Student Teacher Work Sample model (<https://www.wku.edu/rtwsc/>). After changes were made in teacher education requirements in 2018 at the state level, members of an interdisciplinary team of clinical educators and faculty members convened and developed the current tool to serve as a unit wide performance based summative assessment of student teachers. As the tool moved from a performance-based support for making student teacher learning visible towards an assessment or measure of teacher candidate summative performance, the instrument was reviewed multiple times by a core workgroup with additional stakeholder feedback and input. Final revisions were adopted formally during the three cycles of this accreditation cycle. (See Figure 1 for summary of the development).

In Spring 2015, the STWS was replaced by a requirement of the State of Missouri, the MoPTA. The MoPTA was used from Fall 2015 through Spring 2018. In Fall 2018, the MoPTA was removed as a unit wide assessment at the EPP. At that time, the EPP's Teacher Education Council decided to investigate bringing back the STWS as a performance-based indicator. While this was not being mandated by the State's Department of Elementary and Secondary Education, it was being strongly recommended. In Fall 2018, a preliminary draft of the new TWS was presented to the TEC. This version was intended to focus on the unit plan and was piloted in the Art, Middle School, and Early Childhood programs. It would be required for all student teachers the following Spring. In December 2018, the TEC voted to approve the use of the TWS with the MEES. In Spring 2019, it was noted that the TWS was not a "one size fits all" format but overall the sections and requirements should be uniform across programs. Specifically, the requirements for lesson plans were such that the plan needed to be detailed enough that any educator could teach from it, with or without knowledge of prior lessons. Spring 2019 was the second pilot of the TWS, with the initial version rolled out in the Fall 2020 being approved by the TEC and implemented by the end of that semester with the decision that the TWS would be a pass/fail assignment and failure to submit would result in a "U" grade and responsibility for completion of the student teaching semester would lie between the student and their assigned university supervisor. In Spring 2021, final revisions were completed after input from various education stakeholders, including clinical educators, advisory groups--including partnership district leaders and practitioners, and university faculty and supervisors.

Figure 1. Development of the Student Teacher Work Sample