Table 5.2.1 List of Assessments and Validation/Reliability Outcomes

List of EPP Assessments	Proprietary	Used as Evidence for Standards	Validation	Reliability
1.1.1 Modified Danielson Framework for Teaching Candidate Data	NO	1.1, 1.2, 1.4	Mean Content Validity ratio = 0.656 **see Attachment 1**	Kappa coefficients for all 16 items aligned to INTASC **see Attachment 1**
1.1.2 EdTPA Candidate Data	YES	1.1, 1.2, 1.4, 3.6,	Lalley, J. P. (2017). Reliability and validity of edTPA. In Teacher performance assessment and accountability reforms (pp. 47-78). Palgrave Macmillan, New York.	
1.1.3 Educating All Students (EAS) Candidate Data	YES	1.1, 1.4, 3.6	http://www.nystce.nesinc.com/content/docs/NYSTCE_Validation_Reliability.pdf	
1.1.4 Content Specialty Test (CST) Candidate Data	YES	1.1	http://www.nystce.nesinc.com/content/docs/NYSTCE_Validation_Reliability.pdf	
1.1.5 Teacher Work Sample Candidate (TWS)	NO	1.1, 1.2, 1.4	Mean Content Validity ratio = 0.078 **see Attachment 2**	In process **see Attachment 2**
1.1.6 Mid-Point Assessment (Unit Plans and Course Grades)	NO	1.1, 1.2, 1.5	Q1 Validity = 79%– see attachment 3 Unit Plan **See Attachment 3**	Q2 Inter-rater Reliability = 86% **See Attachment 3**
1.1.7 Candidate Exit Survey Data	NO	1.1, 1.2, 1.4, 1.5	NA	NA
1.5.2 Modified Effective Teacher Technology Use	NO	1.5	Mean Content Validity ratio = 0.033 **See Attachment 5**	In process **See Attachment 5**

2.1.2 Teacher Mentor Survey Data	NO	2.1	NA	NA
3.3.1 Classroom Disposition Instrument	NO	3.3	3.3.1 Report: 1st Validation Process and Outcomes for Dispositions Summer 2014 to Spring 2017	3.3.2 Report: 2nd Validation Process and Outcomes for Dispositions Fall 2018
3.3.2 Field Disposition Instrument	NO	3.3	3.3.1 Report: 1st Validation Process and Outcomes for Dispositions Summer 2014 to Spring 2017	3.3.2 Report: 2nd Validation Process and Outcomes for Dispositions Fall 2018
4.1.2 NYS APPR	YES	4.1		
4.2.1 Pilot Administrator Observation (NYSUT, Danielson, APPR)	YES	4.2		

Attachment 1: Process for Validating the Modified Danielson Framework for Teaching Rubric Items

On Feb. 1, 2019, EPP faculty gathered by program to conduct an initial validation of the Danielson Framework for Teaching indicators. Program faculty discussed each Danielson indicator aligned with INTASC standards and commented individually on each item. Using Lawshe's process on content validity, Danielson indicators were determined to be either essential to teacher candidate's knowledge and skills, useful, but not essential, or not necessary. Outcomes of the faculty input were analyzed using Lawshe's Content Validity Ratio revealing a Mean Content Validity Ratio (CVR) of .656. Since the total number of faculty exceeded 15 (N=24), the acceptable ratio for determining valid items was .49. Our mean CVR of 0.656 showed that the Danielson Framework was a valid tool to measure teacher candidate competencies in the field. Since this was the first administration of the Lawshe's validation process, faculty found that several Danielson indicators showed outcomes that would suggest faculty revisit the individual items and either re-word or eliminate specific indicators. The next step in the process was to have the Ruth S. Ammon School of Education Assessment Committee review the Danielson Framework to re-tool the items and bring forward to the next EPP faculty retreat for further validation.

Reliability of the Danielson Framework was conducted using Kappa agreement. Danielson Summative Evaluations from Teacher Mentors and University Supervisors were collected by the Ruth S. Ammon School of Education Office of Professional Experiences and Community Engagement (PECE) and analyzed by the of College of Education and Health Sciences Office Assessment and Accreditation. Performance ratings of teacher candidates on the 16 Danielson items aligned to INTASC standards from four domains were collected across all programs. Teacher Mentors and University Supervisors rated teacher candidates across four performance levels, "Unsatisfactory," "Basic," "Proficient," and "Distinguished" for each of the 16 items. Items (N=254) rated by University Supervisors and Teacher Mentors were analyzed using Cohen's Kappa analysis. Data revealed a range of agreement from slight to fair. At an alpha level ≤ .01, all values were significant. Within all four levels of Danielson's domains, there were no significant variations in agreement between the Teacher Mentor and University Supervisor on the Summative Danielson Rubric that exceeded more than one performance level rating. For example, there were none to few cases of items where a Teacher Mentor (or University Supervisor) would rate a teacher candidate's performance as "Distinguished" while the same candidate's performance on the same measure would be rated as "Basic" by the University Supervisor (or Teacher Mentor). The same were true for "Proficient" and "Unsatisfactory."

Table 1: Summary table of k values and agreement level between University Supervisors and Teacher Mentors on the Summative Danielson Rubric (N=254).

Item	Kappa Coefficients (k)	Agreement Level
Q1	.236	Fair
Q2	.258	Fair
Q3	.170	Slight
Q4	.349	Fair
Q5	.329	Fair
Q6	.130	Slight
Q7	.250	Fair
Q8	.298	Fair
Q9	.281	Fair
Q10	.309	Fair
Q11	.230	Fair
Q12	.234	Fair
Q13	.127	Slight
Q14	.201	Fair
Q15	.200	Fair
Q16	.249	Fair

Attachment 2: Validation of Teacher Work Sample Rubric Items

The Teacher Work Sample (TWS) developed initially by the Renaissance Partnership for Improving Teacher Quality and adapted by the Ruth S. Ammon School of Education contains competencies identified by research and best practices as fundamental to improving student learning. Competencies assessed during the teacher candidate's clinical practice are: Instructional Decision Making, Analysis of Student Learning, and Reflection and Self-Evaluation. In February 2019, a survey was distributed on-line to EPP program faculty in order to validate the indicator and tasks aligned to INTASC using the Lawshe's process of content validation. Individual faculty indicated if the TWS competencies were either essential to a teacher candidate's knowledge and skills, useful, but not essential, or not necessary. Twenty-eight percent (N=9) of the EPP faculty participated in the validation study. Outcomes of the faculty input were analyzed using Lawshe's Content Validity Ratio revealing a Mean Content Validity Ratio (CVR) of 0.78. Our mean CVR of 0.78 showed that the TWS was a valid tool to measure teacher candidate competencies in the field. The next step in the process is to have EPP faculty participate in a reliability study similar to the study conducted for our mid-point assessment of unit plans (KA3 of SPA's – see attachment 3). In the reliability study, faculty will review samples of the TWS for key items Q1, Q2, and Q3 (see below) to address if the teacher candidate sample score was accurate based on the rubric. Rater agreement will then be independently assessed.

Q1: TWS Standard: The candidate designs instruction for specific learning goals, student characteristics and needs, and learning contexts. Task: Provide two examples of instructional decision-making based on students' learning or responses.

Q2: TWS Standard: The candidate uses assessment data to profile student learning and communicate information about student progress and achievement. Task: Analyze your assessment data, including pre/post assessments and formative assessments to determine students' progress related to the unit learning goals. Use visual representations and narrative to communicate the performance of the whole class, subgroups, and two individual students.

Q3: TWS Standard: The candidate analyzes the relationship between his or her instruction and student learning in order to improve teaching practice. Task: Reflect on your performance as a candidate and link your performance to student learning results. Evaluate your performance and identify future actions for improved practice and professional growth.

Attachment 3: Validation and Reliability Study of Unit Plans (Key Assessment 3 of SPA) for Mid-Point Assessment Aligned to INTASC Standards

A validation and reliability study was conducted by the College of Education and Health Sciences by EPP faculty on May 22 and May 23, 2018.

Goals

The first goal of this study was to validate the accuracy of content and ensure that samples were accurately assessed with the corresponding rubric. The second goal was to collect inter-rater reliability on the Unit Plans. The third goal was to provide unit level data for the mid-point assessment of INTASC standards using the Key Assessment that measures the ability to plan instruction.

Each SPA requires a Key Assessment that measures the ability to plan instruction. This key assessment is administered at the mid-point of a Teacher Candidates' program which is at the Exploration/Synthesis phase of the Ruth S. Ammon School of Education. The Teacher Education programs submitted 46 samples of student work related to the Key Assessment, the mid-point assessment aligned with INTASC. All of the rubrics and data were part of the SPA reports.

Methods

Question 1. Does this rubric measure the topic that it is supposed to measure?

1	2	3	4
Rubric & scoring guide did not adequately assess the topic of interest.	Rubric & scoring guide minimally assessed the topic of interest.	Rubric & scoring guide sufficiently assessed the topic of interest.	Rubric & scoring guide exceptionally assessed the topic of interest.

Question 2. Is the student sample score accurate based on the rubric?

1	2	3	4

Scores are not accurate at all.	Scores are marginally accurate (less than or equal to 25%)	Scores are somewhat accurate (26-75%)	Scores are completely accurate (76 – 100%).

Results:

The construct validity (Question 1: Does this rubric measure the topic that it is supposed to measure?) of the rubrics from each program measuring what it purports to measure was high (79%). The interrater reliability (Question 2: Is the student sample score accurate based on the rubric?) based on rater agreement who independently assessed the scoring of samples was also high (86%).

	Question 1: Does this rubric measure the topic that it is supposed to measure?		Question 2: Is the student sample score accurate based on the rubric?	
	Count	**Percent Agreement**	**Count**	**Percent Agreement**
score 4	84	79%	95	86%
score 3	22	21%	13	12%
score 2	0	0%	1	1%
score 1	0	0%	1	1%

Attachment 4: Please refer to the following evidences;

- 3.3.1 Report: 1st Validation Process and Outcomes for Dispositions Summer 2014 to Spring 2017
- 3.3.2 Report: 2nd Validation Process and Outcomes for Dispositions Fall 2018

Attachment 5: Validation and Reliability Study of Technology Competencies Aligned to INTASC Standards

The competencies for effective use of technology by teachers in the classroom were developed by Johnson and Mielke in (2013)[1]. Johnson and Mielke aligned their technology competencies to the four domains of the Danielson Framework. The Ruth S. Ammon School of Education is in the process of adapting the technology competencies by validating the indicators. In February 2019, a survey detailing 34 technology competencies and aligned with Danielson's Framework was distributed on-line to EPP program faculty in the College of Education and Health Sciences using the Lawshe process of content validation. Individual faculty indicated if the technology competencies were either essential to a teacher candidate's knowledge and skills, useful, but not essential, or not necessary. Thirty-two percent of the faculty (N=14) participated in the technology competency validation process. Outcomes of the faculty input were analyzed using Lawshe's Content Validity Ratio revealing a Mean Content Validity Ratio (CVR) of 0.033. The range of validation across competencies was -.42 to .85. The data reveal that as is, the technology competencies in its current form is not a valid tool. Our Ruth S. Ammon School of Education (RSASOE) expected as much, particularly in view of the overwhelming number of competencies being assessed. The RSASOE undertook this initial validation study to have the College of Education and Health Sciences Technology Committee review the technology indicators and eliminate items that measured poorly. Once a more concise set of technology competencies are created (from the set being validated), the technology assessment will undergo a second validation study by the RSASOE Assessment Committee. Items that are found to be relevant and valid will be incorporated into the four domains of the Danielson Framework for Teaching Rubric which will be administered to teacher candidates during their student teaching/practicum field experience. After the RSASOE has assessed the technology competencies in the field, the RSASOE Assessment Committee will conduct a reliability study to determine the level of interrater agreement between University Supervisors and Teacher Mentors who will be observing teacher candidates on the technology competencies.

[1] Johnson, D., & Mielke, N (2013). Power Up! Technology Skills Every Teacher Needs. *Education Leadership. 70*(6), pp. 84-85.

College of Education and Health Sciences
Teacher Education Retreat
Friday, February 1, 2019
Alumni House

8:30 AM-9:00 AM	Coffee and Breakfast
9:00 AM-9:15 AM	Dean's Welcome and Agenda
9:15 AM-10:15 AM	Provider Quality Assurance and Continuous Improvement (CAEP Standard 5)
	Quality Assurance System Program Discussion
10:15 AM-11:45 AM	Content and Pedagogical Knowledge (CAEP Standard 1)
	Initial Programs – Program Actions on InTASC Standards Advanced Programs – Review Phase-ins
11:15 - 11:45 AM	Technology (CAEP Cross Cutting Theme)
	Capturing our Students' Use of Technology
11:45 AM-12:15 PM	Lunch
12:15 PM-1:00 PM	Candidate Quality, Recruitment, and Selectivity (CAEP Standard 3)
	Presentation of Recruitment, Marketing and Promotion Strategies

Communication Plan for Employment Opportunities
Faculty Review and Discussion

NEXT Teacher Education Retreat – March 29, 2019 (9 AM – 1 PM)