

Measure 3 Candidate Competency at program Completion

Student Teaching Final Evaluation Interrater Reliability Results 2020-2021

Student Teaching Final Evaluation: Clinical Instructors and College Supervisors evaluate candidates' content knowledge, pedagogical knowledge, and performance on the student teaching assessment rubric. The student teaching assessment measures candidates' progression on competencies aligned to the Virginia Uniform Performance Standards and the InTASC standards. The expectation is that candidates score a 6.0, the target rating of "Proficient," on the final set of evaluations. Candidates participate in a final conference with their coach and mentor to review feedback and establish professional development goals for their first year of teaching.

Data from this cycle demonstrates Randolph College raters agree our students meet program expectations and are prepared to enter the classroom.

Table 1

2020-2021 Intra Class Correlation (ICC) Results reflect Absolute Agreement Values and F values (95%) obtained from Two-way Mixed Models

	Student Teaching Final Evaluation		Sig	Mean	
	Intraclass Correlation (ICC)	95% Confidence Interval (CI)Upper Bound		CI	CS
Overall rating scores	.866*	.93	.000	7.5	7.2

*high reliability

Results in Table 1 show a high degree of reliability was found between raters (College Supervisors and Clinical Instructors) scoring measurements. The average measure ICC was .866 with a 95% confidence interval from .69 lower bound to .93 upper bound $p < .001$.

Note the means between the clinical instructor (classroom teacher) and the college supervisor (retired veteran teachers and principals) are very close. The 2020 fall orientation session included a focused training with the College Supervisors and the Clinical Instructors on the Student Teaching Evaluation rubric. In addition, the College Supervisors to meet regularly with the Clinical Instructors to review the weekly evaluations. The Randolph College student teaching model requires weekly observations by the College Supervisor and the Clinical Instructor. With the additional training we have seen a higher rate of inter rater agreement. Both the College Supervisors (M=7.2) and the Clinical Instructors (M=7.5) scored on average above the target rating of 6.

Subjectivity in scoring was reduced by using standardized scoring criteria via a student teaching rubric, which is based on the InTASC standards. An increase in objective scoring was achieved by training the clinical instructors and college supervisors to correctly apply scoring rubrics.

Scoring rubrics ranges (see Table 2) outline the possible scores an individual could receive on an assessment, and the levels of performance that must be demonstrated for each score to be given.

Table 2

Scoring Protocol for Student Teaching Final Evaluation Rubric

	8	7	6	5	4	3	2	1
Sample performance indicators: Examples of teacher work conducted in the performance of the standard may include, but are not limited to:	Proficient		Satisfactory		Developing		Unsatisfactory	
	Effective performance independently		Performs well with assistance		Requires additional support		Unsuccessful performance	

Table 3 summarizes disaggregated data on the seven InTASC standards, and show high levels of agreement in all areas. Rater agreement on the student teacher observation rubric, a performance assessment, established how closely the clinical instructor and the college supervisor agree about the student teacher’s instructional performance in a classroom setting. The clinical instructor and the college supervisor independently code an observation. If observer’s codes agree this is evidence that the coding scheme is objective (i.e. similar coding for both raters). Generally, we want our data to be objective, so it is important to establish that inter-rater reliability is high. The student teaching observation rubric includes an eight-point scale with 1 indicating and 8 indicating. When raters scored a 7 (proficient) or an 8 (proficient) these were coded as agreement because they indicate proficient performance.

Table 3

Table 2. Rater agreement and means by rater role for student teaching final evaluations combined items and subcategories

Categories	Means		
	Percent Agreement	CI	CS
All items	94.8%	7.7	7.6
Instruction & Subject Knowledge	100%	7.5	7.3
Instructional Planning	100%	7.5	7.2
Active Learning & Differentiation	100%	7.5	7.1
Expectations and Assessment	98%	7.6	7.55
Cultural Competence & Environment	92.5%	7.5	7.2
Professionalism	87%	7.8	7.4
Sets & Measures Learning Goals	100%	7.5	7.0

The lowest rater agreement occurred in Professionalism where agreement, though high, at 87% was low compared to the other areas where agreement ranged from 98% to 100%. Raters agreed 100% in four of the seven categories. Similar results are reported in the analysis of the dispositions rubric where the College Supervisors rated candidates lower on participation in professional development. Moving forward, we will meet with College Supervisors and Clinical Instructors to develop clear definitions and examples of the expectations for student teachers related to professional behaviors and professional development. This will allow us to determine if the current expectation is unclear to the raters or if we need to increase clarification of professional behaviors to the student teachers.