Teacher Candidate Observation Evaluation Rubric Report:

Assessing Inter-Rater Reliability

Prepared by: Melissa Goldsmith and Brooke Barrigar

The Urban Institute for Teacher Education (UITE) is working to ensure greater inter-rater reliability among the Faculty Cohort Leaders, Site Teacher Educators, and University Supervisors who evaluate Teacher Candidates' student teaching performances at mid-term and upon completion of student teaching. To this end, the UITE Faculty Cohort Leaders, Site Teacher Educators, and University Supervisors engaged in a pilot study that allowed for measurement and open discussion of the ratings given by these groups on the Teacher Candidate Observation Form.

The specific aims of this pilot study were to develop clear understandings of what defines a successful Teacher Candidate in each of the Utah Effective Teaching Standards. Raters of Teacher Candidates' student teaching performances intended to come to a shared understanding to clearly distinguish between the rating of a Teacher Candidate who is passing with a rating of a "3," "4," or "5" on the evaluation form. Raters also intended to ensure that all individuals using the evaluation form are interpreting the question item, which is a statement of a Utah Effective Teaching Standard, in the same way. The pilot study of this project have been conducted with the Elementary and Secondary program options, however, future evaluation form testing discussions will include the Special Education Department. This report includes a summary of the data collected from fall 2012 through spring 2014.

Method

Instrument Development

The instrument developed for the testing of inter-rater reliability consists of a subset of questions from the Teacher Candidate Observation Form used by Elementary, Secondary, and

Special Education Cohort Leaders, Site Teacher Educators/Cooperating Teachers, and University Supervisors in the Teacher Education Program. The Teacher Candidate Observation Form and the study evaluation instrument are directly aligned with the Utah Effective Teaching Standards. To develop the instrument used for this study, core competencies from the study evaluation instrument were selected and adapted. The items selected reflect core elements of Teacher Candidates' skills and competencies. The study evaluation instrument was programmed using the on-line survey tool, Qualtrics, so that quantitative data could be collected electronically.

The study evaluation instrument directions asked each rater to rate the teacher in the lesson using a 5-point scale. After each numerical rating, raters were given the opportunity to offer comments on the reasoning behind their rating. The rating scale, which is also used on the Teacher Candidate Observation Form used in the Teacher Education program, is as follows:

- 1-2: Unsatisfactory. The Teacher Candidate lacks basic knowledge and an ability to meet this performance standard
- 3: Practicing. The Teacher Candidate possesses developing competencies in his/her knowledge and ability to meet performance standard.
- 4: Effective. The Teacher Candidate displays a general understanding of linkages between knowledge and content and executes sound lessons on a consistent basis.
- 5: Highly Effective. The Teacher Candidate exhibits exemplary performance, beyond that of a novice teacher.
- N/A: Not Applicable

The instrument asked each rater to rate the teacher in the lesson using the following question items:

- **Q1** The Teacher Candidate creates developmentally appropriate and challenging learning experiences based on individual students' strengths, interests and needs.
- **Q2** The Teacher Candidate encourages students to use speaking, listening, reading, writing, analysis, synthesis and decision-making skills in various real-world contexts.
- **Q3** The Teacher Candidate designs, adapts, and delivers instruction to address students' diverse learning strengths and needs.

- **Q4** The Teacher Candidate allows students different ways to demonstrate learning sensitive to their multiple experiences and diversity.
- **Q5** The Teacher Candidate incorporates tools of language development into planning and instruction for English Language Learners, and supports development of English proficiency.
- **Q6** The Teacher Candidate provides multiple opportunities for students to develop higher order and meta-cognitive skills.
- **Q7** The Teacher Candidate uses a variety of classroom management strategies to effectively maintain a positive learning environment.
- **Q8** The Teacher Candidate equitably engages students in learning by organizing, allocating, and managing the resources of time, space, and attention.
- **Q9** The Teacher Candidate knows the content of discipline and conveys accurate information and concepts.
- **Q10** The Teacher Candidate uses multiple representations and explanations of concepts that capture key ideas.
- **Q11** The Teacher Candidate designs, and/or selects pre-assessments, formative, and summative assessments in variety of formats that match learning objectives and engage learner in demonstrating knowledge and skills.
- **Q12** The Teacher Candidate understands and practices a range of developmentally, culturally, and linguistically appropriate instructional strategies.
- **Q13** The Teacher Candidate uses data to assess the effectiveness of instruction and to make adjustments in planning and instruction.
- **Q14** The Teacher Candidate plans instruction based on the approved state curriculum.
- **Q15** The Teacher Candidate individually and collaboratively selects and creates learning experiences that are appropriate for reaching content standards, relevant to learners, based on principles of effective instruction.
- **Q16** The Teacher Candidate differentiates instruction for individuals and groups of students by choosing appropriate strategies and accommodations, resources, materials, sequencing, technical tools, and demonstrations of learning.
- **Q17** The Teacher Candidate uses appropriate strategies and resources to adapt instruction and vary his or her role to meet the needs of individual and group learners.
- **Q18** The Teacher Candidate uses a variety of instructional strategies to support and expand learners' communication skills.
- **Q19** The Teacher Candidate supports content and skill development by using multiple media and technology resources and knows how to evaluate these resources for quality, accuracy, and effectiveness.
- **Q20** The Teacher Candidate uses a variety of questioning strategies to promote engagement and learning.

There were other data collected on the instrument including the role of the rater (University Faculty, Site Teacher Educator (STE), or a University Supervisor) and whether or not they were an educator from the Elementary Education or Secondary Education program option.

The instrument developed for this pilot has a very high level of internal reliability (Chronbach's alpha=.980). The instrument is internally consistent meaning that the data results across items within the instrument are consistent with one another.

Data Collection

The data collection method consisted of two components. Raters (Faculty Cohort Leaders, Site Teacher Educators, and University Supervisors) completed the study evaluation instrument by rating the 20 question items after they had watched a video lesson. There were 3 possible video lessons of practicing teachers available to be rated for this study. The video lessons were gathered from publicly accessible teaching websites and represent a variety of teaching styles, lesson contents, and grade levels. The 3 videos were of a 5th grade elementary social studies lesson, an 8th grade secondary science lesson, and a 5th grade elementary language arts lesson.

Table 1 indicates the program option and role of the raters of each video lesson. There were 22 unique raters, but to achieve more of a cross-level design, many raters watched multiple videos regardless of the rater's program option (i.e., a rater from the elementary education program rating the secondary lesson or a rater from the secondary education program rating the elementary lesson).

Table 1: Number of Raters by Program Option and Role for Each Video Lesson

Video Lesson	Elementary	Secondary	Faculty	STEs	Supervisor
Elementary Social Studies	10	2	5	5	2
Secondary Science	7	4	4	4	3
Elementary Language Arts	5	10	5	0	10

The quantitative data collection informed the subsequent discussions among raters. After the quantitative data were collected, the raters participated in one or more group discussions about their interpretation of question items and the reasoning behind their ratings. These discussion groups provided an opportunity to discuss whether or not the question items were being interpreted by the raters in the same way and, if not, to identify the heart of the discrepancy, as the group moderators chose to follow up on those questions that had the most variation in responses among raters, as evidenced by the quantitative data.

There were five discussions over the course of the data collection period. The discussions provide the basis for the qualitative data presented in this report. Importantly, the instrument has been tested and discussed by Cohort Leaders, Site Teacher Educators, and Supervisors. These groups are integral to the learning and skill development of Teacher Candidates and represent both the Elementary and Secondary program options. Also, the instrument was tested in a variety of settings because there were variations among the video lessons in terms of lesson content, grade level, and teaching style.

Cohort Leaders

The instrument was tested multiple times with Cohort Leaders from the Teacher Education Program's Elementary Education and Secondary Education program options. Cohort leaders watched the chosen video lessons and used the instrument to rate the teaching presented in the lessons. The cohort leaders watched and rated an elementary 5th grade social studies lesson, a secondary 8th grade science lesson, and an elementary 5th grade English lesson. They met twice to discuss their ratings of the three video lessons.

Site Teacher Educators

The instrument was tested multiple times with Site Teacher Educators from the Teacher Education Program's Elementary Education program option. Site Teacher Educators watched the selected video lessons and used the instrument to rate the teaching presented in the lessons. The Elementary Site Teacher Educators watched the elementary 5th grade social studies lesson and the secondary 8th grade science lesson. In an effort to collect a robust amount of data without burdening the Site Teacher Educators, we asked only Elementary Education Site Teacher Educators, and not the Secondary Education Site Teacher Educators, to watch 2, and not all 3, of the video lessons.

Supervisors

The instrument was tested multiple times with Supervisors from the Teacher Education Program's Elementary Education and Secondary Education program options. Supervisors watched the chosen video lessons and used the instrument to rate the teaching presented in the lessons. The Elementary and Secondary Supervisors watched the elementary 5th grade social studies lesson, the secondary 8th grade science lesson, and the elementary 5th grade English lesson. Members of this group met four times to discuss their ratings for the videos.

Table 2 summarizes the qualitative group discussions by identifying the role and program options of those who watched the video lessons, the lesson(s) observed, and the number of people who participated in each discussion about Teacher Candidate observations and lesson ratings.

Table 2: Characteristics of Each Discussion Group

Group Discussion	Role of Raters	Program Option of Raters	Lesson(s) Observed	Number of Discussants
1	Cohort Leaders & Supervisors	Elementary & Secondary	5th grade Social Studies & 8th grade Science	11
2	Cohort Leaders & Supervisors	Elementary & Secondary	5th grade English	6

3	Supervisors	Secondary	5th grade English	9
4	Supervisors	Elementary	5th grade English	3
5	STEs	Elementary	5th grade Social Studies & 8th grade Science	3

Findings

Throughout the ratings process, raters discussed the particular video lessons that they watched as part of this inter-rater reliability study. Raters, however, also discussed the process of observing and rating the Program's Teacher Candidates. Therefore, the findings of this study are not restricted to the video lessons that observers watched to participate in this study, but that this inter-rater reliability pilot study also has implications for the Teacher Education Program more broadly.

Inter-Rater Reliability among Raters

Although the Teacher Education Program strives to make the observation process as objective as possible, the raters in the program have an overall awareness of where biases in their ratings may occur. When rating Teacher Candidates, raters have an awareness that they are comparing one Teacher Candidate to another when selecting their observation score. For example, in watching two video lessons one after another, raters discussed that their observations and ratings could have been influenced by the comparisons they were making between the two lessons presented in the videos. Noting that raters do have an awareness of their biases is important to note when attempting to quantify inter-rater reliability, an area to which this report now turns.

Based on the ordinal level of the scale used to rate the question items, the number of raters, and that the raters who participated in the study do not constitute the entire population of raters in the Teacher Education Program, the Intraclass Correlation Coefficient was chosen as the proper statistic to assess inter-rater reliability.

An Intraclass Correlation Coefficient is appropriate to determine the level of agreement among all raters (regardless of program option or role). This statistic is used to describe the consistency among multiple measurements of the same data (i.e. multiple raters watching the same video). It yields a coefficient based on the average scores between the raters and is interpreted as a percentage of agreement after excluding that the raters could have agreed by chance.

Since the Intraclass Correlation Coefficient is a test for consistency of measurement, these data must be analyzed taking into account the particular video lesson. Raters were asked to watch and rate up to three lessons.   As Table 3, indicates the Elementary Social Studies Lesson has the greatest amount of agreement among raters, 78%. The secondary science lesson has 61% agreement among its raters. For the Elementary Language Arts lesson, any agreement between the raters might have occurred only by chance. As Table 1 showed, this video was not rated by STEs and the majority of the raters for this video are University Supervisors, who, as this report will explain, tend to rate differently than the other groups.

Table 3: Inter-Rater Reliability for Each Video Lesson

Video Lesson	Intraclass Correlation Coefficient (0-1.0)	Statistically Significant Agreement Among Raters
Elementary Social Studies	.78	Yes (p=.000)
Secondary Science	.61	Yes (p=.000)
Elementary Language Arts	.22	No (p=.131)

The Intraclass Correlation Coefficients shown above from the pilot study are based on absolute agreement between raters, which is a strict measurement. Although this is a strict measurement, it is appropriate to these data in the pilot stage of measuring inter-rater reliability since one of the goals of this research is to more firmly identify the reasons for why a rater selects the particular rating that he/she does.

In the group discussions, raters had a high level of agreement that one lesson in particular was deemed to be unsuccessful by the majority of raters (i.e., the elementary social studies lesson). As Table 3 indicates, this lesson has the highest level of inter-rater reliability. Other reasons for the level of agreement among the video lessons are found in the quantitative and qualitative data collected as part of the pilot study, as explained in the following sections of this report. The data collected offers sources of agreement and disagreement on the quality of the rating scale, the interpretability of the question items, and the utility of the comments section on Teacher Candidate Observation Form.

Teacher Candidate Observation Form

The discussions focused on the utility of the form's rating scale, "n/a" option, the accompanying USOE rubric, and the comments section. The goal in discussing these aspects of the form was to understand whether the form's components were assisting observers in capturing the Teacher Candidate's performance and offering feedback to the Teacher Candidate.

Rating Scale

Raters thought that the 5-point scale with the option of an "n/a" is a good scale for raters because it has enough variability and range to accommodate rating a myriad of lessons and teaching styles. Raters discussed the appropriate time to use an "n/a" as there were discrepancies among raters in this area. Cohort Leaders said they use an "n/a" when there is no opportunity to observe that item. Over time, the "n/a" will likely turn into a numerical score in future student teaching observations as the Teacher Candidate's continue their learning in the Program. Supervisors and Site Teacher Educators expressed that they were not always certain when the use of an "n/a" was appropriate.

Although ratings of a "1" or a "2" are both considered unsatisfactory, there appears to be a difference between a rating of a "1" and a "2," with a "1" being a "detrimental" rating. For example, some Cohort Leaders will never rate a question item as a "1" because they feel it is demoralizing to a Teacher Candidate. Across all groups, raters would only give a "1" if they observed the Teacher Candidate to be doing something damaging to a classroom student or if the Teacher Candidate was behaving in such a way as that was making student learning more difficult. Some raters, particularly Supervisors, said that they almost never give a Teacher Candidate a rating of a "1" or a "2" because a rating of a "1" or "2" on an item leads to remediation for a Teacher Candidate. Supervisors indicated that they would rather rate a question item as an "n/a" rather than a "1" or "2".

In general, raters noted that the difference between a lower score and a higher score on the rating scale might be that the teacher missed the opportunity to perform better in a certain area, particularly in the areas of meeting the diverse needs of student. As one rater noted, needs, "if there is an ELL in the room, it is always possible to incorporate language development in the classroom." Raters indicated that it is hard to give a rating of a "4" or "5" if the instructor does not use visual aids, tries to do too much in a lesson, and/or if the lesson is not engaging. Raters also noted that the difference between a rating of a "4" and a "5" is whether or not higher order thinking is required ("5"). Supervisors said that they do not give a "5" for any question item for the mid-term evaluations.

Raters indicated that the rating scale is appropriate for measuring change in a Teacher Candidate's performance. The vast majority of raters indicated that they aim to capture any improvements in the Teacher Candidate's performance as they progress through the Program. Raters also noted that the criteria for what is a good lesson at mid-term observation and what is

good enough at final observation is different. This criteria changes as student teaching progresses, and expectations increase.

Variations on the level of agreement between raters may be, in part, related to a difference discovered among raters during the discussions. The discussions indicated that some raters adhere very strictly to the rubric created by the Utah State Office of Education that accompanies the Utah Effective Teaching Standards when selecting a rating on the scale and others less so. Specifically, some raters, in particular University Supervisors, consistently refer to the USOE's definitions of "practicing," "effective," "highly effective," and "distinguished" for each standard when selecting a rating for how well the Teacher Candidate represented that teaching standard. University Supervisors use the rubric. Other raters, in particular the Site Teacher Educators, indicated that they rarely rely on this rubric when rating lessons.

Question Items

Appendix A features the frequency distribution and descriptive statistics for each question on the study instrument. The question items are parsed out by the program option and role of the raters.

Across question items, the question that has the highest mean score across all raters is Question 9, "The Teacher Candidate knows the content of discipline and conveys accurate information and concepts" (mean=4.21). The question that has the lowest mean score is Question 19, "The Teacher Candidate supports content and skill development by using multiple media and technology resources and knows how to evaluate these resources for quality, accuracy, and effectiveness" (mean=3.00). As Table 4 illustrates, the range of mean scores broken down by the rater's program option and their role. Judging by the mean scores, it appears that Supervisors generally give higher scores than Faculty or Site Teacher Educators.

Table 4: Summary of Mean Scores by Rater Program Option and Rater Role

Rater	Question with the Lowest Mean	Question with the Highest Mean	Range of Mean Scores
Elementary	Q19: The Teacher Candidate supports content and skill development by using multiple media and technology resources and knows how to evaluate these resources for quality, accuracy, and effectiveness.	Q9: The Teacher Candidate knows the content of discipline and conveys accurate information and concepts.	2.81-4.27
Secondary	Q5: The Teacher Candidate incorporates tools of language development into planning and instruction for English Language Learners, and supports development of English proficiency.	Q7: The Teacher Candidate uses a variety of classroom management strategies to effectively maintain a positive learning environment.	3.00-4.25
Faculty	Q13: The Teacher Candidate uses data to assess the effectiveness of instruction and to make adjustments in planning and instruction.	Q9: The Teacher Candidate knows the content of discipline and conveys accurate information and concepts.	2.50-4.07
Site Teacher Educator	Q19: The Teacher Candidate supports content and skill development by using multiple media and technology resources and knows how to evaluate these resources for quality, accuracy, and effectiveness.	Q14: The Teacher Candidate plans instruction based on the approved state curriculum.	2.11-4.13
Supervisor	Q5: The Teacher Candidate incorporates tools of language development into planning and instruction for English Language Learners, and supports development of English proficiency.	Q15: The Teacher Candidate individually and collaboratively selects and creates learning experiences that are appropriate for reaching content standards, relevant to learners, based on principles of effective instruction.	3.40-4.29

The quantitative data indicate that although there is a certain level of agreement among raters, there are some systematic differences between groups across all video lessons. There is a statistically significant difference between the ratings of Elementary and Secondary Educators on one question item, "Q7: The Teacher Candidate uses a variety of classroom management strategies to effectively maintain a positive learning environment" ($t=-1.84$, $p<.10$). Secondary Educators gave slightly higher ratings to the teacher performance in the videos for this question.

For all other question items, there were no statistically significant difference between the ratings from Elementary Educators and Secondary Educators.

There are statistically significant differences between the ratings of Faculty Cohort Leaders, Site Teacher Educators, and University Supervisors in three areas. The question with the greatest statistical significance among raters is for Question 19, "The Teacher Candidate supports content and skill development by using multiple media and technology resources and knows how to evaluate these resources for quality, accuracy, and effectiveness." Supervisors gave the highest ratings in this area (mean=3.67), followed by Faculty Cohort Leaders (mean=3.00), and then Site Teacher Educators, who gave the lowest rating to this question (mean=2.11) (F=5.56, p<.01).

A second area of difference in ratings is for Question 11, "The Teacher Candidate designs, and/or selects pre-assessments, formative, and summative assessments in variety of formats that match learning objectives and engage learner in demonstrating knowledge and skills." For this question item, Supervisors gave significantly higher ratings (mean=4.08) than Faculty Cohort Leaders (mean=3.50), and Site Teacher Educators, who gave the lowest rating to this question (mean=2.86) (F=4.09, p<.05).

A third area where raters in different roles differed is on Question 10, "The Teacher Candidate uses multiple representations and explanations of concepts that capture key ideas." Supervisors gave the highest rating to teachers in the video lessons (mean=4.33), followed by Faculty Cohort Leaders (mean=3.71), and then Site Teacher Educators, who gave the lowest rating to this question (mean=3.22) (F=2.92, p<.10).

The qualitative discussions also revealed some sources of variability among raters. For example, some Utah Effective Teaching Standards address more than one teaching skill within a standard. Therefore some of the questions on the Teacher Candidate observation form are double

or triple barreled and ask raters about more than one concept at a time. For example, Question 12 reads, "The Teacher Candidate understands and practices a range of developmentally, culturally, and linguistically appropriate instructional strategies." This makes rating a challenge because some raters may focus more on the "developmentally appropriate instructional strategies," while other raters may offer their rating on aspect of the question asking about "culturally appropriate instructional strategies," etc. This raises the question of whether there is some inconsistency in the way a standard is being understood and therefore inconsistency between ratings for the observed lesson.

Rating Diverse Learning Question Items

Several of the question items on the Teacher Candidate Evaluation Form center around the needs of diverse learners. Based upon the quantitative data collected, raters were asked to explain what they were looking for in a lesson to see if the instructor was meeting the needs of diverse learners. Supervisors, in particular, were asking for a consensus as to how to assess a diverse learner, whether these types of questions applied solely to English Language Learners or whether diverse learning defined as student learning disabilities or student difference in learning aptitudes is also being rated with the evaluation form questions. Elementary Supervisors noted that every Elementary Education student lesson has language built into it, so it was harder for them to make differentiations between a language-rich instruction versus an English Language Learners instruction (for example, see Question 5, "The Teacher Candidate incorporates tools of language development into planning and instruction for English Language Learners, and supports development of English proficiency"). Elementary Site Teacher Educators also felt that Question 5 was not just specific to English Language Learners and also noted difficulty in differentiating oral language capacity versus English Language Learning for this question item.

Discussions indicate that raters were interpreting Question 20 in multiple ways. The statement reads, "The Teacher Candidate uses a variety of questioning strategies to promote engagement and learning." Some raters were observing the lesson with an eye to whether or not the instructor asked different types of questions (i.e. asking for a definition, asking for an application of a concept, asking for a synthesis between multiple concepts, etc.) versus whether or not the instructor utilized a questioning strategy.

Raters noted that the overlap between questions on the Teacher Candidate Observation Forms, where a specific skill is rated more than once on the evaluation form. Even though there appears to be some overlap between dimensions on the Teacher Candidate Observation Form, raters said that they are able to differentiate the question items through multiple observations of a Teacher Candidate, where they may focus more heavily on one part of the question when observing one lesson and then focus on another part of the same question more heavily when observing a subsequent lesson. Some raters use the comments section to tease out the differences between the similar question items by sharing exactly what aspect of the question they were focusing on when determining a rating.

Also, in the cases where question items are similar in multiple sections of the form, raters make an effort to be consistent in their ratings between them. Raters agreed that noting the question item's category can sometimes help in sorting out the specific aspect of the skill to be rated, but having a similar question item in two different categories does complicate the ratings process.

Comments Section

When asked about the comments section on the Teacher Candidate Observation Form, all the cohort leaders generally used the comments form in the same way. For the mid-term

evaluations, faculty members used the comment sections to note the Teacher Candidate's attributes as well as the noting areas of improvement. Then, at the time for final evaluation, cohort leaders would use the comments section to be more summative. The comments offered for the final evaluation focus almost entirely on the Teacher Candidate's strengths. Negative feedback is offered to the Teacher Candidate, but it is written up in a separate, less formal document. Supervisors and Site Teacher Educators indicated that they tended to avoid negative comments on the Final Student Teaching Evaluation because the Final Student Teaching Observation Form data might affect next steps in the Teacher Candidate's career path.

Areas for Future Consideration

- According to this pilot study, there are some areas where Cohort Leaders, Site Teacher Educators, and University Supervisors rate Teacher Candidates' lessons differently, with University Supervisors consistently offering higher ratings and Site Teacher Educators consistently offering lower ratings.
- Determining the aspects of "diverse learning" that are being evaluated during the Teacher Candidate Observation Form.
- How raters should handle overlapping questions, where questions appear multiple times on the Teacher Candidate Observation Form, but are tapping into different Teaching Standards.
- How raters should handle double-barreled questions, where a question item asks about more than one skill at a time.
- How raters should use an "n/a".
- When it is appropriate for raters to give a Teacher Candidate a rating of a "1".
- How raters are to present feedback on the Final Student Teaching Evaluation (versus the mid-term evaluations).
- Determining whether all raters should be using the rubric that accompanies the Utah Effective Teaching Standards to rate each question item.

Appendix A: Descriptive Statistics for Questions By Total, Rater's Program Option, and Rater's Role

Q1 The Teacher Candidate creates developmentally appropriate and challenging learning experiences based on individual students' strengths, interests and needs.

	Total	Rater's Program Option		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	3%	5%	-	-	11%	-
2 Unsatisfactory	8%	5%	13%	21%	-	-
3 Practicing	18%	19%	19%	14%	22%	20%
4 Effective	39%	45%	31%	43%	44%	33%
5 Highly Effective	29%	23%	38%	21%	22%	40%
Not Applicable	3%	5%	-	-	-	7%
n	37	21	16	14	9	14
Mean	3.86	3.81	3.94	3.64	3.67	4.21
SD	1.03	1.03	1.06	1.08	1.22	.80

Q2 The Teacher Candidate encourages students to use speaking, listening, reading, writing, analysis, synthesis and decision-making skills in various real-world contexts.

	Total	Rater's Program Option		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	-	-	-	-	-	-
2 Unsatisfactory	13%	14%	13%	21%	22%	-
3 Practicing	21%	23%	19%	14%	33%	20%
4 Effective	37%	41%	31%	50%	11%	40%
5 Highly Effective	26%	23%	31%	14%	33%	33%
Not Applicable	3%	-	6%	-	-	7%
n	37	22	15	14	9	14
Mean	3.78	3.72	3.87	3.57	3.56	4.41
SD	1.00	0.98	1.06	1.02	1.24	.77

Q3 The Teacher Candidate designs, adapts, and delivers instruction to address students' diverse learning strengths and needs.

	Total	Rater's Program Option		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	5%	5%	7%	7%	11%	-
2 Unsatisfactory	16%	23%	7%	21%	33%	-
3 Practicing	27%	18%	40%	21%	11%	43%
4 Effective	32%	36%	27%	36%	33%	29%
5 Highly Effective	19%	18%	20%	14%	11%	29%

	Total	Rater's Program Option		Rater's Role		
		Elem	Sec	Fac	STE	Sup
Not Applicable	-	-	-	-	-	-
n	37	22	15	14	9	15
Mean	3.43	3.41	3.47	3.29	3.00	3.86
SD	1.14	1.18	1.12	1.20	1.32	.86

Q4 The Teacher Candidate allows students different ways to demonstrate learning sensitive to their multiple experiences and diversity.

	Total	Rater's Program Option		Rater's Role		
		Elem	Sec	Fac	STE	Sup
1 Unsatisfactory	3%	-	6%	7%	-	-
2 Unsatisfactory	29%	36%	19%	21%	56%	20%
3 Practicing	11%	5%	19%	14%	-	13%
4 Effective	26%	32%	19%	36%	11%	27%
5 Highly Effective	32%	27%	38%	21%	33%	40%
Not Applicable	-	-	-	-	-	-
n	38	22	16	14	9	15
Mean	3.55	3.50	3.62	3.43	3.22	3.87
SD	1.29	1.26	1.36	1.28	1.48	1.19

Q5 The Teacher Candidate incorporates tools of language development into planning and instruction for English Language Learners, and supports development of English proficiency.

	Total	Rater's Program Option		Rater's Role		
		Elem	Sec	Fac	STE	Sup
1 Unsatisfactory	3%	-	6%	7%	-	-
2 Unsatisfactory	13%	18%	6%	14%	22%	7%
3 Practicing	29%	14%	50%	36%	22%	27%
4 Effective	26%	32%	19%	7%	44%	33%
5 Highly Effective	3%	5%	-	-	11%	-
Not Applicable	26%	32%	19%	36%	-	33%
n	28	15	13	9	9	10
Mean	3.18	3.33	3.00	2.67	3.44	3.40
SD	.90	0.98	0.82	.87	1.01	.70

Q6 The Teacher Candidate provides multiple opportunities for students to develop higher order and meta-cognitive skills.

	Total	Rater's Program Option		Rater's Role		
		Elem	Sec	Fac	STE	Sup
1 Unsatisfactory	3%	5%	-	-	11%	-
2 Unsatisfactory	3%	5%	-	7%	-	-

	Total	Elem	Sec	Fac	STE	Sup
3 Practicing	32%	32%	31%	36%	44%	20%
4 Effective	34%	32%	38%	29%	22%	47%
5 Highly Effective	29%	27%	31%	29%	22%	33%
Not Applicable	-	-	-	-	-	-
n	38	22	16	14	9	15
Mean	3.84	3.72	4.00	3.79	3.44	4.13
SD	.97	1.08	0.82	.97	1.24	.74

Q7 The Teacher Candidate uses a variety of classroom management strategies to effectively maintain a positive learning environment.

	Total	Rater's Program Option[*]		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	-	-	-	-	-	-
2 Unsatisfactory	13%	18%	6%	14%	33%	-
3 Practicing	11%	9%	13%	7%	11%	13%
4 Effective	47%	59%	31%	50%	33%	53%
5 Highly Effective	29%	14%	50%	29%	22%	33%
Not Applicable	-	-	-	-	-	-
n	38	22	16	14	9	15
Mean	3.92	3.68	4.25	3.93	3.44	4.20
SD	.97	0.96	0.93	1.00	1.24	.68

Q8 The Teacher Candidate equitably engages students in learning by organizing, allocating, and managing the resources of time, space, and attention.

	Total	Rater's Program Option		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	-	-	-	-	-	-
2 Unsatisfactory	13%	18%	6%	14%	33%	-
3 Practicing	18%	14%	25%	21%	11%	20%
4 Effective	37%	45%	25%	29%	44%	40%
5 Highly Effective	29%	23%	38%	36%	11%	33%
Not Applicable	3%	-	6%	-	-	7%
n	37	22	15	14	9	14
Mean	3.84	3.72	4.00	3.86	3.33	4.14
SD	1.01	1.03	1.00	1.10	1.12	.77

Q9 The Teacher Candidate knows the content of discipline and conveys accurate information and concepts.

	Total	Rater's Program Option		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**

	Total	Elem	Sec	Fac	STE	Sup
1 Unsatisfactory	-	-	-	-	-	-
2 Unsatisfactory	3%	-	6%	7%	-	-
3 Practicing	13%	14%	13%	14%	22%	7%
4 Effective	45%	45%	44%	43%	22%	60%
5 Highly Effective	39%	41%	38%	36%	56%	33%
Not Applicable	-	-	-	-	-	-
n	38	22	16	14	9	15
Mean	4.21	4.28	4.13	4.07	4.33	4.27
SD	.78	0.70	0.89	.92	.87	.59

Q10 The Teacher Candidate uses multiple representations and explanations of concepts that capture key ideas.

	Total	Rater's Program Option		Rater's Role[*]		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	5%	5%	6%	7%	11%	-
2 Unsatisfactory	13%	18%	6%	14%	33%	-
3 Practicing	5%	9%	-	14%	-	-
4 Effective	45%	36%	56%	29%	33%	67%
5 Highly Effective	32%	32%	31%	36%	22%	33%
Not Applicable	-	-	-	-	-	-
n	38	22	16	14	9	15
Mean	3.84	3.72	4.00	3.71	3.22	4.33
SD	1.17	1.24	1.10	1.33	1.48	.49

Q11 The Teacher Candidate designs, and/or selects pre-assessments, formative, and summative assessments in variety of formats that match learning objectives and engage learner in demonstrating knowledge and skills.

	Total	Rater's Program Option		Rater's Role[**]		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	3%	5%	-	-	11%	-
2 Unsatisfactory	5%	9%	-	7%	11%	-
3 Practicing	24%	18%	31%	21%	33%	20%
4 Effective	26%	32%	19%	21%	22%	33%
5 Highly Effective	13%	5%	25%	7%	-	27%
Not Applicable	29%	32%	25%	43%	22%	20%
n	27	15	12	8	7	12
Mean	3.59	3.33	3.92	3.50	2.86	4.08
SD	1.01	1.05	0.90	.93	1.07	.79

Q12 The Teacher Candidate understands and practices a range of developmentally, culturally, and linguistically appropriate instructional strategies.

	Total	Rater's Program Option		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	11%	14%	6%	7%	22%	7%
2 Unsatisfactory	18%	18%	19%	36%	11%	7%
3 Practicing	26%	14%	44%	29%	22%	27%
4 Effective	26%	32%	19%	14%	44%	27%
5 Highly Effective	8%	9%	6%	-	-	20%
Not Applicable	11%	14%	6%	14%	-	13%
n	34	19	15	12	9	13
Mean	3.03	3.05	3.00	2.58	2.89	3.54
SD	1.17	1.31	1.00	.90	1.27	1.20

Q13 The Teacher Candidate uses data to assess the effectiveness of instruction and to make adjustments in planning and instruction. *[Note: The sample size is lower for this question due to a programming error in the electronic survey].*

	Total	Rater's Program Option		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	3%	5%	-	-	11%	-
2 Unsatisfactory	13%	10%	20%	33%	11%	-
3 Practicing	7%	5%	10%	-	11%	8%
4 Effective	17%	15%	20%	11%	22%	17%
5 Highly Effective	3%	-	10%	-	-	8%
Not Applicable	57%	65%	40%	56%	44%	67%
n	13	7	6	4	5	4
Mean	3.08	2.86	3.33	2.50	2.80	4.00
SD	1.19	1.21	1.21	1.00	1.30	.82

Q14 The Teacher Candidate plans instruction based on the approved state curriculum.

	Total	Rater's Program Option		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	-	-	-	-	-	-
2 Unsatisfactory	-	-	-	-	-	-
3 Practicing	16%	9%	27%	29%	11%	7%
4 Effective	38%	45%	27%	50%	56%	14%
5 Highly Effective	14%	18%	7%	7%	22%	14%
Not Applicable	32%	27%	40%	14%	11%	64%
n	25	16	9	12	8	5
Mean	3.96	4.13	3.67	3.75	4.13	4.20

SD	.68	0.62	0.71	.62	.64	.84

Q15 The Teacher Candidate individually and collaboratively selects and creates learning experiences that are appropriate for reaching content standards, relevant to learners, based on principles of effective instruction.

	Total	Rater's Program Option		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	-	-	-	-	-	-
2 Unsatisfactory	16%	18%	13%	29%	22%	-
3 Practicing	8%	5%	13%	-	11%	13%
4 Effective	42%	55%	25%	50%	33%	40%
5 Highly Effective	29%	18%	44%	21%	22%	40%
Not Applicable	5%	5%	6%	-	11%	7%
n	36	21	15	14	8	14
Mean	3.89	3.76	4.07	3.64	3.63	4.29
SD	1.04	0.99	1.10	1.15	1.19	.73

Q16 The Teacher Candidate differentiates instruction for individuals and groups of students by choosing appropriate strategies and accommodations, resources, materials, sequencing, technical tools, and demonstrations of learning.

	Total	Rater's Program Option		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	5%	9%	-	-	22%	-
2 Unsatisfactory	13%	14%	13%	14%	33%	-
3 Practicing	18%	14%	25%	14%	-	33%
4 Effective	24%	23%	25%	14%	22%	33%
5 Highly Effective	18%	14%	25%	29%	22%	7%
Not Applicable	21%	27%	13%	29%	-	27%
n	30	16	14	10	9	11
Mean	3.47	3.25	3.72	3.80	2.89	3.64
SD	1.22	1.34	1.07	1.23	1.62	.67

Q17 The Teacher Candidate uses appropriate strategies and resources to adapt instruction and vary his or her role to meet the needs of individual and group learners.

	Total	Rater's Program Option		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	6%	9%	-	-	22%	-
2 Unsatisfactory	22%	23%	21%	36%	22%	8%
3 Practicing	14%	9%	21%	14%	11%	15%
4 Effective	39%	50%	21%	21%	44%	54%
5 Highly Effective	17%	9%	29%	29%	-	15%
Not Applicable	3%	-	7%	-	-	8%

	Total			
n	35	22	13	14	9	12
Mean	3.40	3.28	3.61	3.43	2.78	3.83
SD	1.19	1.20	1.19	1.28	1.30	.83

Q18 The Teacher Candidate uses a variety of instructional strategies to support and expand learners' communication skills.

	Total	Rater's Program Option		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	5%	9%	-	-	22%	-
2 Unsatisfactory	18%	14%	25%	29%	22%	7%
3 Practicing	24%	23%	25%	14%	11%	40%
4 Effective	39%	45%	31%	50%	44%	27%
5 Highly Effective	13%	9%	19%	7%	-	27%
Not Applicable	-	-	-	-	-	-
n	38	22	16	14	9	15
Mean	3.37	3.32	3.44	3.36	2.78	3.73
SD	1.10	1.13	1.09	1.01	1.30	.96

Q19 The Teacher Candidate supports content and skill development by using multiple media and technology resources and knows how to evaluate these resources for quality, accuracy, and effectiveness.

	Total	Rater's Program Option		Rater's Role[***]		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	13%	18%	6%	7%	33%	7%
2 Unsatisfactory	18%	23%	13%	29%	33%	-
3 Practicing	13%	14%	13%	7%	22%	13%
4 Effective	39%	41%	38%	43%	11%	53%
5 Highly Effective	3%	-	6%	-	-	7%
Not Applicable	13%	5%	25%	14%	-	20%
n	33	21	12	12	9	12
Mean	3.00	2.81	3.33	3.00	2.11	3.67
SD	1.20	1.21	1.15	1.13	1.05	.98

Q20 The Teacher Candidate uses a variety of questioning strategies to promote engagement and learning.

	Total	Rater's Program Option		Rater's Role		
		Elem	**Sec**	**Fac**	**STE**	**Sup**
1 Unsatisfactory	-	-	-	-	-	-
2 Unsatisfactory	5%	9%	-	-	22%	-
3 Practicing	34%	36%	31%	43%	33%	27%

4 Effective	16%	9%	25%	21%	-	20%
5 Highly Effective	45%	45%	44%	36%	44%	53%
Not Applicable	-	-	-	-	-	-
n	38	22	16	14	9	15
Mean	4.00	3.90	4.13	3.93	3.67	4.27
SD	1.01	1.11	0.89	.92	1.32	.88

[***] statistically significant difference between groups at the 99% confidence level ($p<.01$); [**] statistically significant difference between groups at the 95% confidence level $p<.05$; [*] statistically significant difference between groups at the 90% confidence level $p<.10$