

# Final Report on North Carolina DPI's Observation Calibration Training 2015-2016

Empirical Education Inc.

*July 20, 2016*

## Table of Contents

Background .....	3
OVERVIEW OF THE OBSERVATION CALIBRATION TRAINING .....	3
OCT PROGRAM 2015-16 .....	3
Results.....	5
PARTICIPATION.....	5
Overview of LEA Participation.....	5
Scoring Study Participation.....	8
Lesson Participation .....	9
BloomBoard Marketplace Usage .....	11
OBSERVER PERFORMANCE.....	14
Comparison of SS1 and SS2 Based on Reports Generated by Observation Engine.....	14
Statistical Analysis of Scoring Accuracy Improvement.....	17
Statistical Analysis: Did Lesson Completion Affect Performance on Scoring Studies? .....	18
FEEDBACK FROM PARTICIPANT SURVEY .....	20
Discussion .....	29
Appendix A. North Carolina Educator Evaluation System Rubric .....	30

## Background

### OVERVIEW OF THE OBSERVATION CALIBRATION TRAINING

The Observation Calibration Training (OCT) provides North Carolina school districts with access to a suite of calibration and training activities for school administrators across the state to improve the accuracy and reliability of teacher evaluations. The online platform combines BloomBoard's professional development resources with Empirical Education's training and calibration tool, Observation Engine™. The BloomBoard Marketplace contains thousands of streaming videos, eBooks, articles, presentations, and self-paced courses. For the OCT, a selection of full-length classroom videos and short video clips in Observation Engine were master scored by a team of experts using the North Carolina Educator Evaluation System (NCEES; see appendix for full rubric). These videos were then made available to observers as rater calibration events called *Scoring Studies* and element-specific learning exercises called *Lessons*. Observation Engine Scoring Studies and Lessons are described below:

- *Scoring Studies*: Scoring Studies help build consensus and inter-rater reliability among a group of evaluators. A study assigns a video (or set of videos) to observers who must watch and rate the video using the NCEES rubric. A Scoring Study report provides helpful information about observer agreement with both target and modal scores, as well as the general distribution of scores across a group of observers.
- *Lessons*: Lessons provide targeted, self-paced online learning activities for evaluators and/or teaching staff. Designed for professional development activities associated with the NCEES rubric, Lessons provide immediate on-screen feedback for observers that appears as soon as they have submitted their scores.

The OCT aims to improve observation skills, increase rater agreement, and to provide a common experience for local education agencies (LEAs) to host collaborative conversations to improve instructional leadership skills.

### OCT PROGRAM 2015-16

Following a successful pilot implementation of the OCT during the 2014/15 school year, the North Carolina Department of Public Instruction (NCDPI), BloomBoard, and Empirical Education initiated a statewide implementation of the OCT for the 2015/16 school year. The program followed the structure and flow of the pilot. By providing Scoring Studies at the beginning and at the end of the program, it was possible to measure improvement as a result of the activities during the program.

The table below shows the activities included in the OCT. On December 10, the initial cohort of participants was provided with an introduction to the project and available resources via a live webinar. They were also given access to written instructional materials and a short demonstration video. Later registrants were provided access to the recorded webinar and the same support materials. Participants were instructed to first complete Scoring Study 1 (which served as a "pretest") and to then complete the 19 observable Lessons over the course of approximately five months at their own pace. Scoring Study 2 was then administered as a "posttest". Several webinars were offered throughout the

program period to provide feedback on participation and performance, as well as to offer tips and strategies for increasing calibration and scoring accuracy.

**TABLE 1. OCT SCHEDULE OF ACTIVITIES AND TASKS**

Date/completion window	OCT task or activity	Description
<b>12/10/15</b>	OCT kick-off	Live webinar introduction to program and OCT platform
<b>12/10/15 – 1/31/16</b>	Scoring Study 1 (SS1)	One-video observation where participants rated all observable NCEES elements. No immediate feedback was provided.
<b>2/1/16 – 4/1/16 (Continuous)</b>	Element-specific Lessons	Short video clips focused on one specific NCEES element (17 Lessons available with 2 clips for each observable element). Observers watched and rated the clip, then received immediate feedback on their scores.
<b>2/1/16 – 4/1/16 (Continuous)</b>	Full-observation Lessons	Longer videos (2) focused on all 17 observable elements. Observers watched and rated the clip, then received immediate feedback on their scores.
<b>2/8/16</b>	OCT webinar: Results of SS1 and Facilitator Support	Reviewed results of Scoring Study 1 and provided tips for increasing calibration. Immediately following the webinar, facilitators stayed on to review available support resources.
<b>4/1/16 – 4/29/16*</b>	Scoring Study 2 (SS2)	One-video observation where participants rated all observable NCEES elements
<b>5/16/16 – 7/8/16</b>	Feedback survey	Online survey eliciting feedback from participants on their experience during the program.
<b>6/17/16</b>	OCT Recorded Webinar: Results of SS2	Reviewed results of Scoring Study 2. Also provided a comparison of Scoring Study 1 and Scoring Study 2, providing feedback on particularly challenging elements.

\* Two LEAs, Catawba County Schools and Carteret County Schools, were given an extension until June 28<sup>th</sup> due to scheduling difficulties.

This report presents participation results, performance outcomes for Scoring Study 2 in comparison to Scoring Study 1, and feedback from the participant survey. Usage information from BloomBoard’s Marketplace is also provided.

## Results

### PARTICIPATION

#### Overview of LEA Participation

Following the success of the 2014/15 pilot, NCDPI reached out to LEA personnel across the state to elicit participation. After the initial registration window in November, two more registration opportunities were opened for participants who continued to express interest in the OCT but were unable to sign up by the previous deadline. The final registration window ended January 15.

Overall, 72 LEAs (37 of which included just one participant each) agreed to participate (see Table 2). In total, 1003 principals<sup>1</sup> and other evaluators (observers) were added to the platform. Of the initial 72 LEAs, 39 did not participate (i.e. no observers completed tasks in the platform). By the end of the program, 354 observers across 33 LEAs completed at least one task in the system. It should be noted that there are many reasons why users initially added to the platform may not have participated. Over-inclusion of staff members in the original list submitted to NCDPI, personnel changes, and shifts in LEA priorities and resources likely all contributed to participant attrition.

**TABLE 2. PARTICIPATING LOCAL EDUCATION AGENCIES**

LEA name	No. of observers initially added	No. of observers completing one or more tasks
ACE Academy Charter	1	0
Anson County Schools	1	1
Arapahoe Charter School	2	0
Beaufort County Schools	4	2
Bertie County Schools	1	0
Bladen County Schools	1	0
Buncombe County Schools	1	0
Burke County Public Schools	5	3
Cabarrus County Schools	6	1
Caldwell County Schools	3	3
Carolina International School	1	0
Carteret County Public Schools	43	40
Catawba County Schools	75	68

<sup>1</sup> This number included 455 participants registered with Charlotte-Mecklenburg Schools, which did not end up participating in this year's program.

**TABLE 2. PARTICIPATING LOCAL EDUCATION AGENCIES**

<b>LEA name</b>	<b>No. of observers initially added</b>	<b>No. of observers completing one or more tasks</b>
Central Park School for Children	1	0
Chapel Hill-Carrboro City Schools	7	1
Charlotte-Mecklenburg Schools	455	5
Chatham County Schools	1	0
Cleveland County Schools	2	1
Columbus County Schools	1	0
Craven County Schools	27	26
Cumberland County Schools	8	1
Currituck County Schools	1	1
Dare County Schools	2	0
Davie County Schools	1	1
Durham Public Schools	1	0
Edenton-Chowan Schools	16	16
Elizabeth City-Pasquotank County Schools	1	0
Elkin City Schools	1	0
Excelsior Classical Academy	1	0
Franklin County Schools	4	0
Gaston County Schools	21	18
Gates County Public Schools	1	0
Granville County Public Schools	6	1
Guilford County Schools	3	0
Halifax County Schools	1	0
Haliwa-Saponi Tribal School	1	0
Healthy Start Academy	1	0
Henderson County Schools	1	0
Hyde County Schools	1	0
Jackson County Public Schools	1	0
Johnston County Schools	4	0
Kannapolis City Schools	1	0
Martin County Schools	2	0
Mitchell County Schools	1	1

**TABLE 2. PARTICIPATING LOCAL EDUCATION AGENCIES**

<b>LEA name</b>	<b>No. of observers initially added</b>	<b>No. of observers completing one or more tasks</b>
Montgomery County Schools	1	0
Moore County Schools	3	0
Nash-Rocky Mount Public Schools	39	31
North Carolina Department of Corrections	1	0
Northeast Regional School of Biotech and Agriscience	1	1
Person County Schools	1	0
Pitt County Schools	6	1
Randolph County School System	1	0
Research Triangle High School	2	0
Roanoke Rapids Graded School District	14	0
Rockingham County Schools	62	49
Sampson County Schools	2	1
Stanly County Schools	1	1
Stokes County Schools	34	32
The Capital Encore Academy	1	1
Transylvania County Schools	17	7
Triangle Math and Science Academy	1	1
Two Rivers Community School	1	0
Tyrell County Schools	1	0
Vance County Schools	9	0
Wake County Public School System	19	15
Washington County Schools	5	0
Watauga County Schools	1	0
Wayne County Public Schools	25	20
Weldon City Schools	8	1
Wilson County Schools	1	0
Winston-Salem/Forsyth County Schools	26	2
Yadkin County Schools	1	1
<b>TOTAL</b>	<b>1003</b>	<b>354</b>

### Scoring Study Participation

There were two Scoring Studies administered during the program. Scoring Studies are online calibration events where a group of observers first watch and rate a video independently, and then a report is run to compare scores from observers to target scores. The report displays scoring distributions, agreement trends, and performance metrics to facilitate conversations around the scores and observation rubric. For the purposes of measuring improvement during the program, Scoring Study 1 (SS1) was considered a “pretest” at the beginning of the program, and Scoring Study 2 (SS2) was considered a “posttest” at the end of the program.

Table 3 shows Scoring Study participation by LEA. Overall, 331 participants across 31 LEAs participated in Scoring Study 1. 172 participants across 21 LEAs participated in Scoring Study 2.

**TABLE 3. SCORING STUDY PARTICIPATION BY LEA**

LEA name	No. of observers completing SS1	No. of observers completing SS2
Anson County Schools	1	1
Beaufort County Schools	1	1
Burke County Public Schools	3	2
Cabarrus County Schools	1	0
Caldwell County Schools	3	1
Carteret County Public Schools	35	30
Catawba County Schools	65	9
Chapel Hill-Carrboro City Schools	1	0
Charlotte-Mecklenburg Schools	5	1
Cleveland County Schools	1	0
Craven County Schools	26	25
Cumberland County Schools	1	0
Currituck County Schools	1	1
Davie County Schools	1	0
Edenton-Chowan Schools	16	16
Gaston County Schools	18	4
Granville County Public Schools	1	0
Mitchell County Schools	1	0
Nash-Rocky Mount Public Schools	30	9



**TABLE 3. SCORING STUDY PARTICIPATION BY LEA**

LEA name	No. of observers completing SS1	No. of observers completing SS2
Northeast Regional School of Biotech and Agriscience	1	1
Pitt County Schools	1	0
Rockingham County Schools	44	25
Stokes County Schools	31	26
The Capital Encore Academy	1	1
Transylvania County Schools	7	3
Triangle Math and Science Academy	1	0
Wake County Public School System	13	6
Wayne County Public Schools	18	8
Weldon City Schools	1	0
Winston-Salem/Forsyth County Schools	1	1
Yadkin County Schools	1	1
<b>TOTAL</b>	<b>331</b>	<b>172</b>

Note. LEAs with no OCT participation are excluded from this table.

### Lesson Participation

There were 19 total Lessons available in the OCT: 17 element-specific Lessons and 2 full-observation Lessons. Participants were asked to complete all available Lessons. The table below shows Lesson participation rates by LEA. Overall, 263 participants across 29 LEAs completed at least one Lesson. Of all observers that participated in the OCT, 32% (115 of 354) completed all available Lessons.

**TABLE 4. LESSON PARTICIPATION BY LEA**

LEA name	No. of observers completing at least 1 element-specific Lesson	No. of observers completing at least 1 full-observation Lesson	No. of observers completing all 19 available Lessons
Anson County Schools Count	1	1	1
Beaufort County Schools Count	1	1	1
Burke County Public Schools Count	3	1	1
Cabarrus County Schools Count	1	1	0
Caldwell County Schools Count	3	3	3
Carteret County Public Schools Count	40	3	2

**TABLE 4. LESSON PARTICIPATION BY LEA**

<b>LEA name</b>	<b>No. of observers completing at least 1 element-specific Lesson</b>	<b>No. of observers completing at least 1 full-observation Lesson</b>	<b>No. of observers completing all 19 available Lessons</b>
Catawba County Schools Count	44	25	12
Chapel Hill-Carrboro City Schools Count	1	0	0
Charlotte-Mecklenburg Schools Count	2	1	1
Cleveland County Schools Count	1	0	0
Craven County Schools Count	2	14	0
Currituck County Schools Count	1	1	1
Davie County Schools Count	1	0	0
Edenton-Chowan Schools Count	5	2	1
Gaston County Schools Count	4	0	0
Granville County Public Schools Count	1	0	0
Mitchell County Schools Count	1	1	1
Nash-Rocky Mount Public Schools Count	21	9	6
Northeast Regional School of Biotech and Agriscience Count	1	1	0
Rockingham County Schools Count	49	45	38
Sampson County Schools Count	1	1	1
Stanly County Count	1	0	0
Stokes County Schools Count	31	27	23
The Capital Encore Academy Count	1	1	1
Transylvania County Schools Count	7	1	0
Wake County Public School System Count	4	3	3
Wayne County Public Schools Count	19	19	18
Winston-Salem/Forsyth County Schools Count	2	1	0
Yadkin County Schools Count	1	1	1
<b>TOTAL</b>	<b>250</b>	<b>163</b>	<b>115</b>

Note. LEAs with no OCT participation are excluded from this table.

Participation rates for SS2 were lower than for SS1. This may be due to SS2 being administered at the end of the year, a busy time for most administrators. However, for a completely voluntary project, this level of participation is quite encouraging. Participation in Lessons was greater than participation in

SS2 possibly because of the availability of immediate on-screen feedback, which may make Lessons inherently more attractive to observers.

### BloomBoard Marketplace Usage

In addition to the Observation Engine resources, the OCT platform did include access to BloomBoard's Marketplace. In December 2015, BloomBoard finalized the integration with NCDPI's IAM/Home Base, which gave all NC educators (in addition to OCT participants) access to BloomBoard via Home Base.

The table below shows usage metrics for the Marketplace by LEA. The first column lists the number of sessions. Sessions were counted as any period where a user logged in and was actively navigating the system. The second column breaks down the number of resources that were accessed by users. Overall, there was a significant increase in engagement with the Marketplace this year compared to the pilot year.

**TABLE 5. BLOOMBOARD MARKETPLACE USAGE**

LEA name	Marketplace Session Count (2015-16)	Previewed & Consumed Resources (2015-16)
Anson County Schools	62	2
Arapahoe Charter School	6	13
Beaufort County Schools	78	29
Bertie County Schools	35	12
Bladen County Schools	39	7
Buncombe County Schools	56	34
Burke County Schools	96	12
Cabarrus County Schools	124	4
Caldwell County Schools	59	5
Carolina International School	1	0
Carteret County Public Schools	456	56
Catawba County Schools	530	44
Chapel Hill-Carrboro City Schools	67	12
Charlotte-Mecklenburg Schools	373	216
Chatham County Schools	5	0
Cleveland County Schools	24	24
Columbus County Schools	8	0
Craven County Schools	269	53
Cumberland County Schools	512	124
Currituck County Schools	6	0
Dare County Schools	23	28
Davie County Schools	63	41

**TABLE 5. BLOOMBOARD MARKETPLACE USAGE**

<b>LEA name</b>	<b>Marketplace Session Count (2015-16)</b>	<b>Previewed &amp; Consumed Resources (2015-16)</b>
Durham Public Schools	47	30
Edenton-Chowan Schools	206	20
Elizabeth City-Pasquotank Public Schools	21	13
Elkin City	3	0
Excelsior	3	0
Franklin County Schools	79	16
Gaston County Schools	227	20
Gates County Schools	1	0
Granville County Schools	62	10
Guilford County Schools	151	233
Halifax County Schools	12	6
Haliwa-Saponi Tribal School	1	0
Healthy Start Academy	4	0
Henderson County Schools	11	3
Hyde County Schools	10	9
Jackson County Public Schools	22	5
Johnston County Schools	170	31
Kannapolis City Schools	33	8
Martin County Schools	3	0
Mitchell County Schools	32	4
Montgomery County Schools	56	1
Moore County Schools	31	5
Nash-Rocky Mount Schools	398	62
North Carolina Department of Corrections	8	0
Pitt County Schools	154	32
Randolph County Schools	38	64
Research Triangle Charter	1	4
Roanoke Rapids Grade School District	6	0
Rockingham County Schools	617	54
Sampson County Schools	81	23
Stanly County Schools	63	16
Stokes County Schools	162	18
The Capitol Encore Academy	25	9

**TABLE 5. BLOOMBOARD MARKETPLACE USAGE**

<b>LEA name</b>	<b>Marketplace Session Count (2015-16)</b>	<b>Reviewed &amp; Consumed Resources (2015-16)</b>
<b>Transylvania County Schools</b>	28	71
<b>Triangle Math and Science Academy</b>	19	0
<b>Two Rivers Community School 95A</b>	1	0
<b>Tyrell County Schools</b>	4	0
<b>Vance County Schools</b>	42	8
<b>Wake County Schools</b>	486	113
<b>Washington County Schools</b>	21	0
<b>Watauga County Schools</b>	6	3
<b>Wayne County Public Schools</b>	428	56
<b>Weldon City Schools</b>	11	0
<b>Wilson County Schools</b>	115	4
<b>Winston Salem/Forsyth County Schools</b>	324	62
<b>Yadkin County Schools</b>	24	15
<b>TOTAL (2015-16)</b>	<b>7252</b>	<b>1917</b>
<b>Previous Total (2014-15)</b>	<b>1023</b>	<b>352</b>

## OBSERVER PERFORMANCE

### Comparison of SS1 and SS2 Based on Reports Generated by Observation Engine

Examining the automatically-generated Observation Engine scoring reports for both Scoring Studies, there appeared to be improvement in scoring accuracy between SS1 and SS2. To interpret these graphs and subsequent analyses, three performance metrics are defined:

- *Percent target agreement*: the percent of an observer's scores that agrees exactly with the target scores
- *Percent target discrepant*: the percent of an observer's scores that disagrees with the target scores by 2 or more performance levels (e.g. when the target score is 2 and the score provided is a 4)
- *Scoring bias*: when an observer has a statistically significant tendency to rate higher or lower than the target score

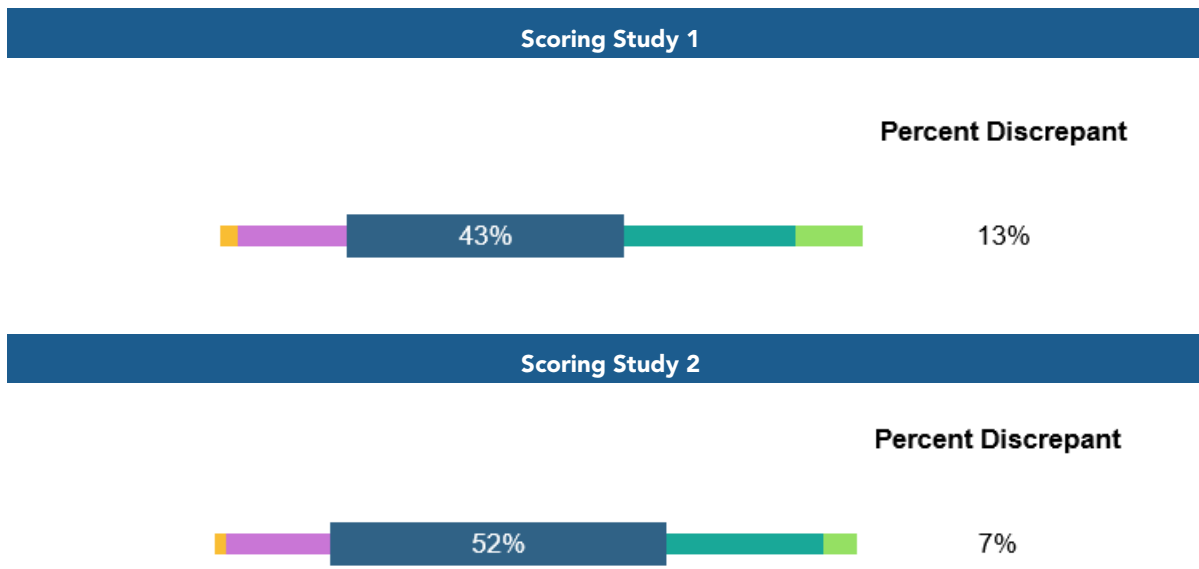
The figures below that report the results use the graphic convention shown here:



This multicolored bar represents all scores submitted by observers. The length of each colored section represents the portion of scores that fall in that particular category: the longer the section, the higher the percent of total scores. The value reported in the dark blue section of the bar is the percent target agreement. Scores that were one score adjacent to the target score are represented by the purple ("1 below target") and teal ("1 above target") colored sections. The percent target discrepant is reported to the right of each bar and are also represented by the orange ("2+ below target") and lime green ("2+ above target") colored sections. Any scoring bias would be reported as an upwards-facing or downwards-facing arrow next to the bar.

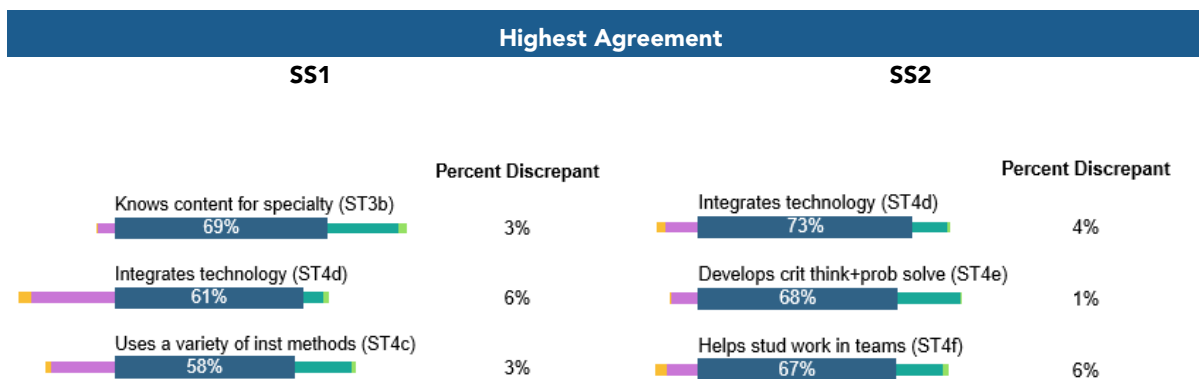
Figure 1 summarizes the agreement to target scores for all observers who completed each Scoring Study. You can see that in SS2, exact agreement to target scores was higher by 9%. In addition, 93% of all scores in SS2 were either on target or directly adjacent to the target scores (as opposed to 87% for SS1). There was no significant bias towards rating higher or lower than target scores in either Scoring Study.

**FIGURE 1. SCORING STUDY COMPARISON: OVERALL AGREEMENT**

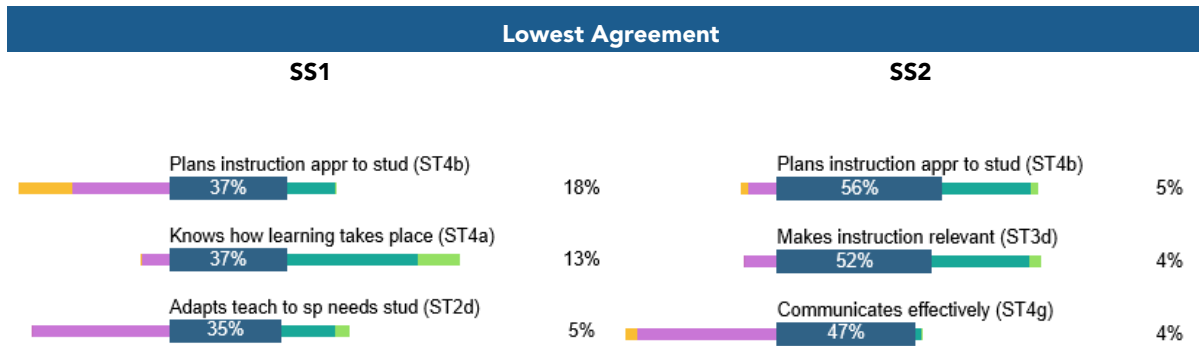


The Scoring Study reports also show agreement by individual NCEES element. Figure 2 shows the agreement graphs for the three elements with the highest and lowest levels of agreement for each Scoring Study. Examining these graphs show that there is some overlap between SS1 and SS2. Observers had high agreement with target scores on Element 4d in both Scoring Studies; Element 4b was challenging for observers in both Studies. This could mean that this element was particularly difficult to rate in a video observation context. This could also mean that observers should revisit the language of this element in the NCEES rubric to clarify any confusion or misinterpretation of the language.

**FIGURE 2. SCORING STUDY COMPARISON: AGREEMENT BY ELEMENT**

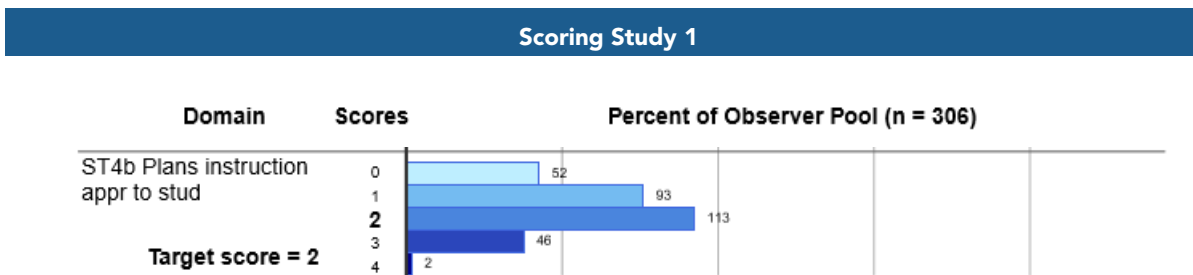


**FIGURE 2. SCORING STUDY COMPARISON: AGREEMENT BY ELEMENT**



Examining the distribution graphs in the Scoring Study reports provides a useful snapshot of scoring that helps elucidate the nature of disagreement with target scores. Figure 3 below shows the distribution of scores in Scoring Study 2 for element 4b – *Teachers plan instruction appropriate for their students*.

**FIGURE 3. ANALYZING DISAGREEMENT: SAMPLE SCORING DISTRIBUTION GRAPH**



The Observation Engine report of the score distribution for element 4b shows that 30% of the observers (93 of 306) thought that the score should have been a 1 rather than a 2, and 32% thought the score should be either a 0 or 3. The Observation Engine gives observers the opportunity to examine the justifications for the target score to see why the expert scoring committee believed the score to be a 2, and then to examine their own evidence for the score. The report can also be of value to trainers and other evaluation program personnel in highlighting the need for clarification/more training in particular elements, investigating the possibility that observers misinterpreted the video, or uncovering that observers had developed a bias or general impression of a teacher that affected their scores on particular elements. In cases where the majority of observers disagreed with the target score, it could also mean that the target score should be re-evaluated.

The use of these reports in some of the districts shows that these kinds of explorations of the scoring data can inspire useful collaborative conversations—around the NCEES rubric and evaluation practices—that can contribute to calibration and rater agreement.



### Statistical Analysis of Scoring Accuracy Improvement

Although the reports for both Scoring Studies generated by Observation Engine showed higher ratings in SS2, statistical analysis allows us to determine whether that improvement was likely due to chance fluctuation. Table 6 shows the results using data from the 160 observers that completed both Scoring Study 1 and Scoring Study 2. The statistical analysis showed that agreement to target scores was significantly higher in SS2 and discrepancy was lower. There was no significant change in scoring bias as there was very little bias to start with.

**TABLE 6. SCORING STUDY 1 & SCORING STUDY 2 GROUP PERFORMANCE**

Metric	Scoring Study 1	Scoring Study 2
<b>Mean Percent Target Agreement</b>	43.9%	53.0%*
<b>Mean Target Discrepant</b>	12.2%	7.1%*
<b>Average Scoring Bias</b>	0.18	0.19

\*Difference from SS1 results statistically significant at  $p < .0001$

*Percent Target Agreement:* The percentage of scores that exactly match the target score.

*Percent Target Discrepant:* The percentage of scores that disagree with the target score by two or more performance levels.

*Scoring Bias:* Scorer has a statistically significant bias towards rating either higher or lower than the target score.

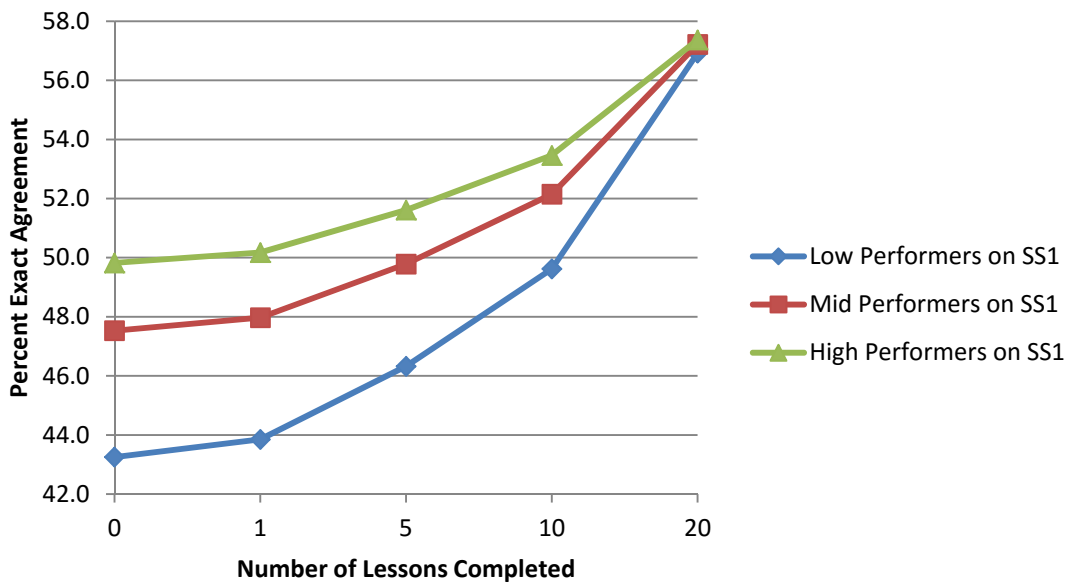
This means that the improvement seen in the reports generated by Observation Engine was not due to chance, and rater accuracy did, in fact, improve from the beginning of the program to the end.

**Statistical Analysis: Did Lesson Completion Affect Performance on Scoring Studies?**

As previously reported, 19 total Lessons were available to participants in the OCT platform. An important question is whether or not the number of Lessons completed by the observers is associated with improvement between SS1 and SS2. To answer this question, regression analysis was used to measure the strength of the association between the number of Lessons completed between SS1 and SS2 (based on timestamps in Observation Engine) and the two measurements of performance: percent target agreement and percent target discrepant.

Regression results are shown graphically below in Figures 4 and 5. Figure 4 shows that for low- and mid-performing observers on SS1, the more Lessons completed, the higher the percent target agreement. The effect is greatest for the lowest performing observers, but all observers benefit from completing Lessons.

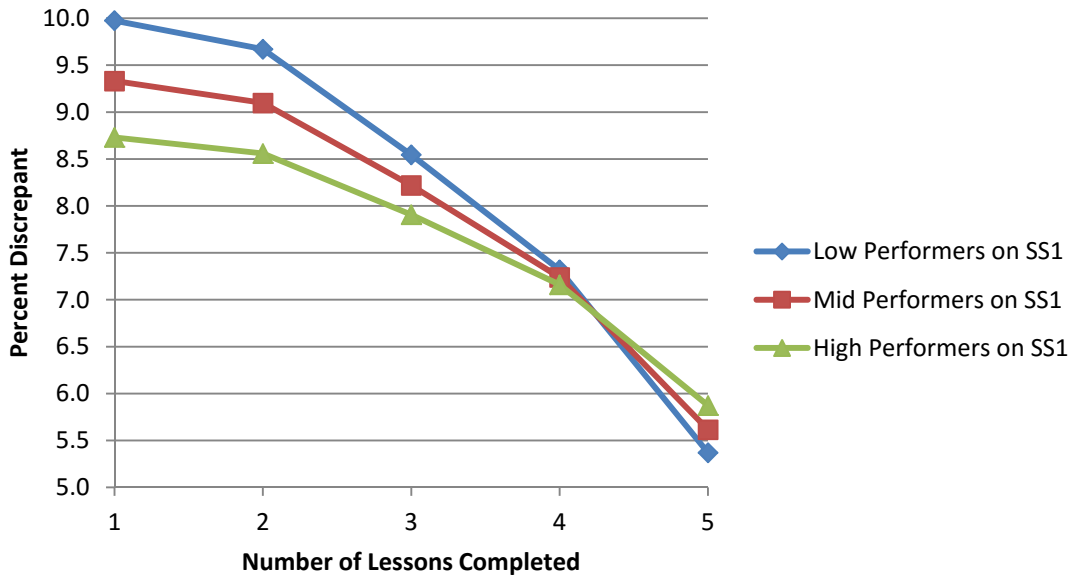
**FIGURE 4. REGRESSION MODEL: EFFECT OF LESSONS ON TARGET AGREEMENT**



Note. These exact data points are theoretical and not representative of any individual observer. Performance trend lines indicate performance rankings on SS1 (i.e. "pre-test" performance). Low = 25<sup>th</sup> percentile, Mid = 50<sup>th</sup> percentile, High = 75<sup>th</sup> percentile.

The effect of completing Lessons was even stronger on the percent target discrepant metric. Figure 5 shows that for all SS1 performance levels, the more Lessons completed, the lower the percent of discrepant scores. This means that all observers, regardless of their initial performance on SS1, benefited from completing Lessons. Similar to the effect on percent target agreement, completing Lessons benefited the lowest scoring observers the most (the trend line for the low performers is steepest).

**FIGURE 5. REGRESSION MODEL: EFFECT OF LESSONS ON TARGET DISCREPANT**



Note. These exact data points are theoretical and not representative of any individual observer. Performance trend lines indicate performance rankings on SS1 (i.e. "pre-test" performance). Low = 25<sup>th</sup> percentile, Mid = 50<sup>th</sup> percentile, High = 75<sup>th</sup> percentile.

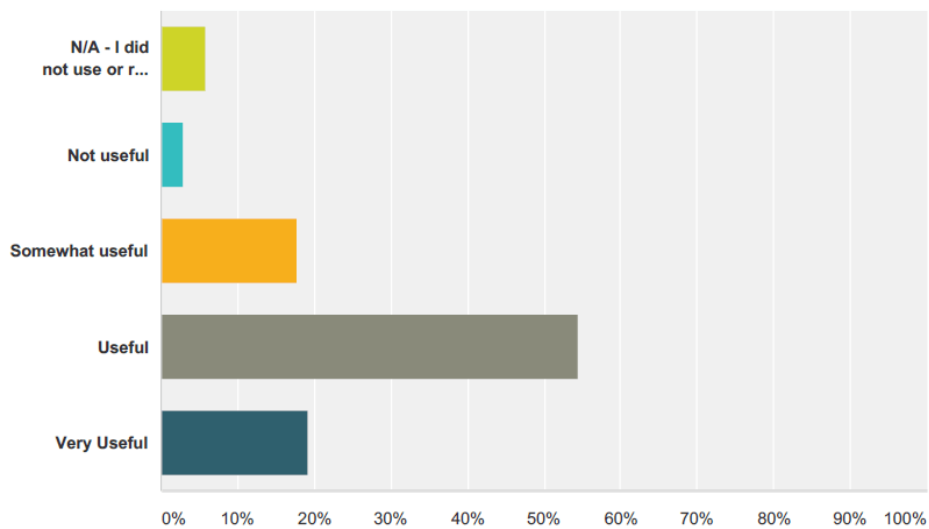
### FEEDBACK FROM PARTICIPANT SURVEY

Following the completion of the program, an online survey was sent out to all participants who completed both Scoring Studies. As of July 8, 68 participants completed the survey. The following section displays responses for selected portions of the feedback survey.

Table 7 shows response distributions for questions related to the introductory material and training. Most respondents (74%) thought that the introductory materials and the kick-off webinar were either useful or very useful. Two respondents thought the introductory email was not useful; however, all respondents thought the introductory webinar was useful in various degrees. In comments, two participants requested that the details about the OCT tasks be made more explicit: “Explain more that there are two videos per standard. I finished the whole thing and thought I was done and then later discovered something called Study 2 and didn't realize I had to do it too”; “[Include] a description of the standards as far as what should be looked at.” The majority of the comments noted that the introduction materials were easy to understand.

**TABLE 7. SURVEY RESULTS: OCT TRAINING AND INTRODUCTORY MATERIAL**

Question	Response	Frequency	Percent
<b>You were sent an introductory email explaining the Observation Calibration Training (OCT) platform and how to login. How useful was it?</b>	Did not use	4	5.9%
	Not useful	2	2.9%
	Somewhat useful	12	17.7%
	Useful	37	54.4%
	Very useful	13	19.1%



**TABLE 7. SURVEY RESULTS: OCT TRAINING AND INTRODUCTORY MATERIAL**

Question	Response	Frequency	Percent
<b>How useful was the webinar introducing the OCT?</b>	Did not see webinar	7	10.3%
	Not useful	0	0%
	Somewhat useful	22	32.3%
	Useful	32	47.1%
	Very useful	7	10.3%

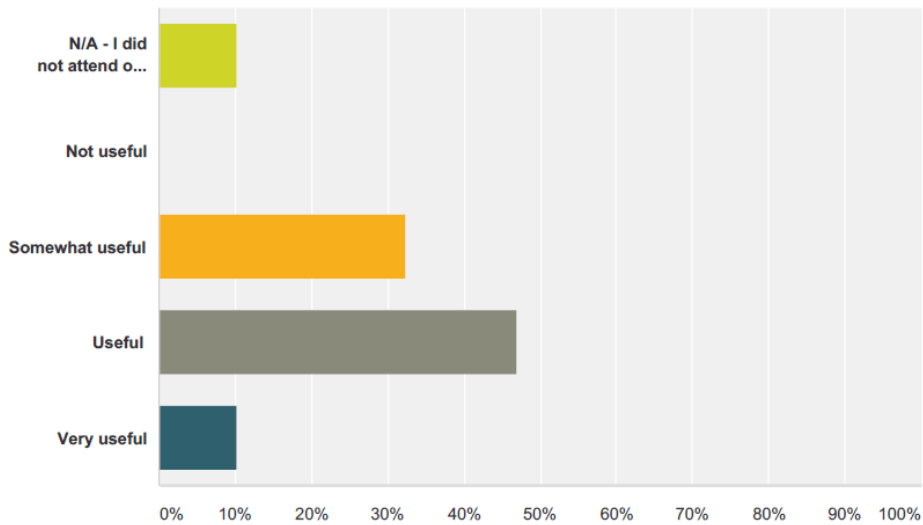
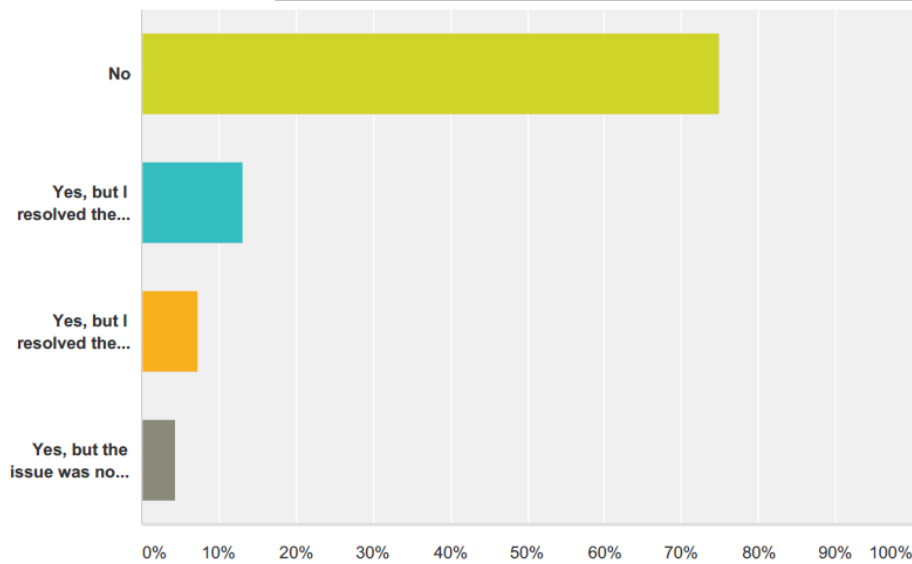


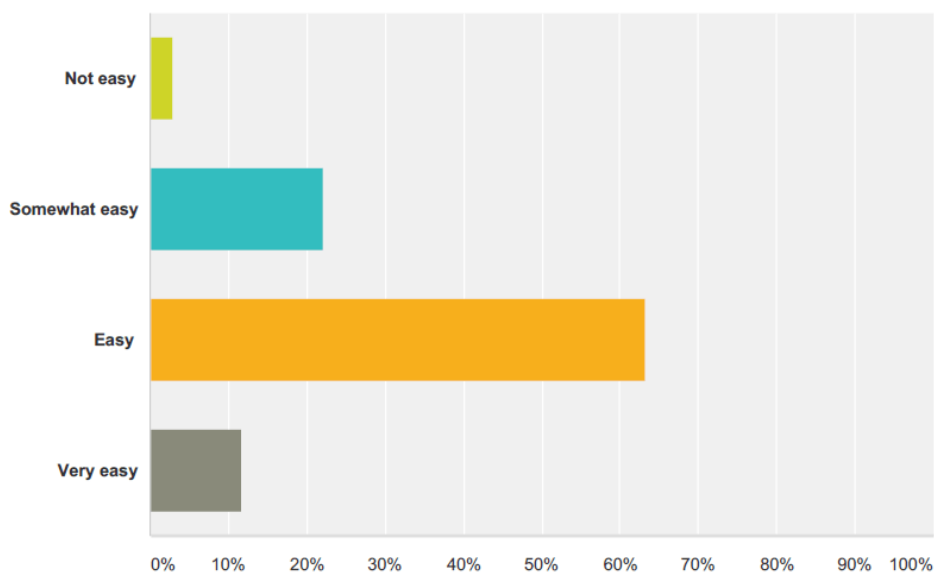
Table 8 shows survey results related to technical issues with the OCT platform. The vast majority of respondents did not come across any technical issues, but those that did were able to resolve them to their satisfaction with the exception of 3 respondents. Seventy-five percent of respondents thought that the OCT was “easy” or “very easy” to use. Regarding visual and audio quality of the videos, there did not appear to be any major issues. However, some participants did report that there were volume and noise issues with some of the videos. Two participants would have liked to see different video angles to be able see students’ faces/expressions. One participant said he/she would have liked to watch “more clips in some instances prior to responding.” In addition, it was noted by several respondents that discussing the video lessons afterwards in a collaborative setting was very helpful.

**TABLE 8. SURVEY RESULTS: TECHNICAL ISSUES WITH OCT PLATFORM**

Question	Response	Frequency	Percent
<b>Did you experience any technical issues with the OCT platform?</b>	No	51	75.0%
	Yes, but resolved the issue on my own	9	13.2%
	Yes, but resolved the issue with help from another person	5	7.4%
	Yes, but the issue was not resolved to my satisfaction	3	4.4%

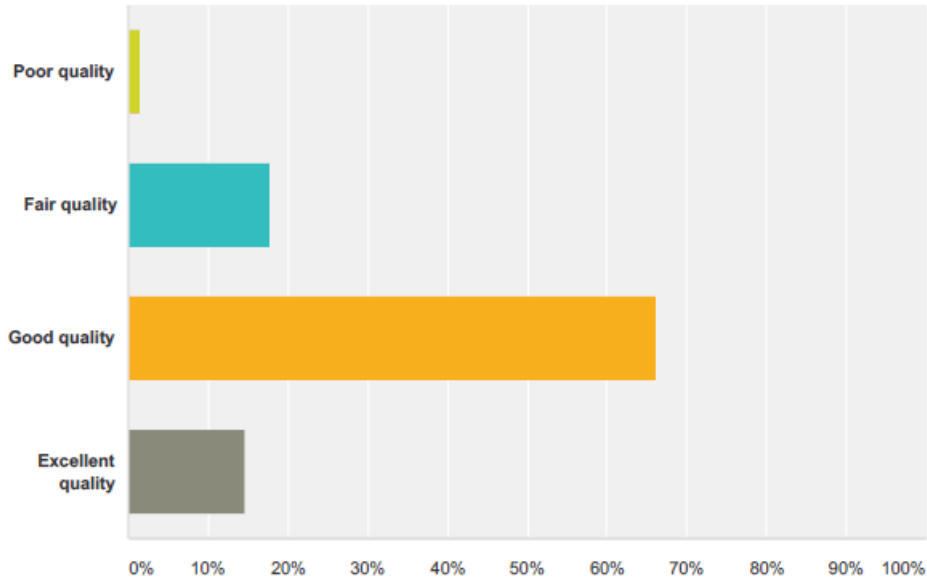


<b>How would you rate the OCT's ease of use?</b>	Not easy	2	2.9%
	Somewhat easy	15	22.1%
	Easy	43	63.2%
	Very easy	8	11.8%



**TABLE 8. SURVEY RESULTS: TECHNICAL ISSUES WITH OCT PLATFORM**

<b>Of the videos you watched, how would you rate the quality of the picture?</b>	Poor quality	1	1.5%
	Fair quality	12	17.6%
	Good quality	45	66.2%
	Excellent quality	10	14.7%



<b>Of the videos you watched, how would you rate the quality of the audio?</b>	Poor quality	1	1.5%
	Fair quality	14	20.9%
	Good quality	47	70.2%
	Excellent quality	5	7.5%

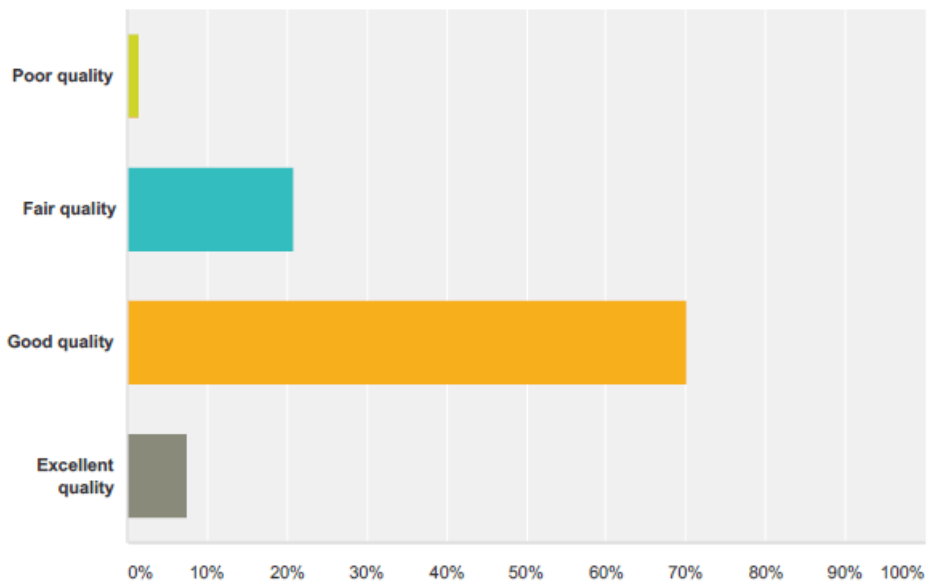
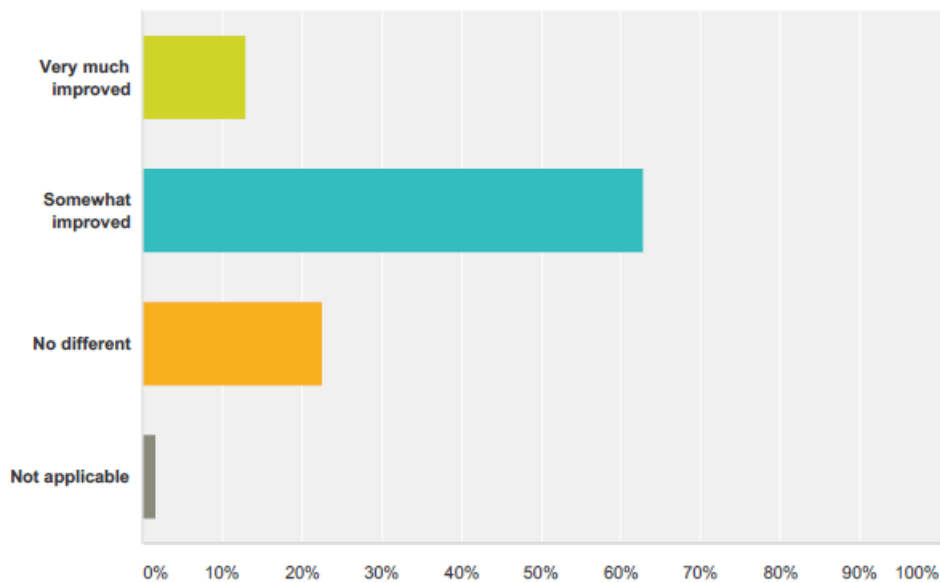


Table 9 contains feedback on the Scoring Studies. Between SS1 and SS2, 78% of respondents felt that their application of the NCEES rubric improved. The bottom section of the table lists a selection of comments related to Scoring Studies. The participants expressed that they learned a lot from the experience and appreciated having justifications for the scores. A few participants were surprised at the target scores and one participant thought some of the videos did not provide good examples of the specific elements; some expressed the desire to have more variety in teaching levels (i.e. more exemplary videos). Again, those participants who had the opportunity to discuss the videos with colleagues found it very helpful; those who didn't noted that group discussions would have been helpful.

**TABLE 9. SURVEY RESULTS: FEEDBACK ON SCORING STUDIES**

Question	Response	Frequency	Percent
<b>To what extent do you feel your application of the NCEES rubric improved between Scoring Study 1 and Scoring Study 2?</b>	Very much improved	8	12.9%
	Somewhat improved	39	62.9%
	No different	14	22.6%
	Not applicable	1	1.6%





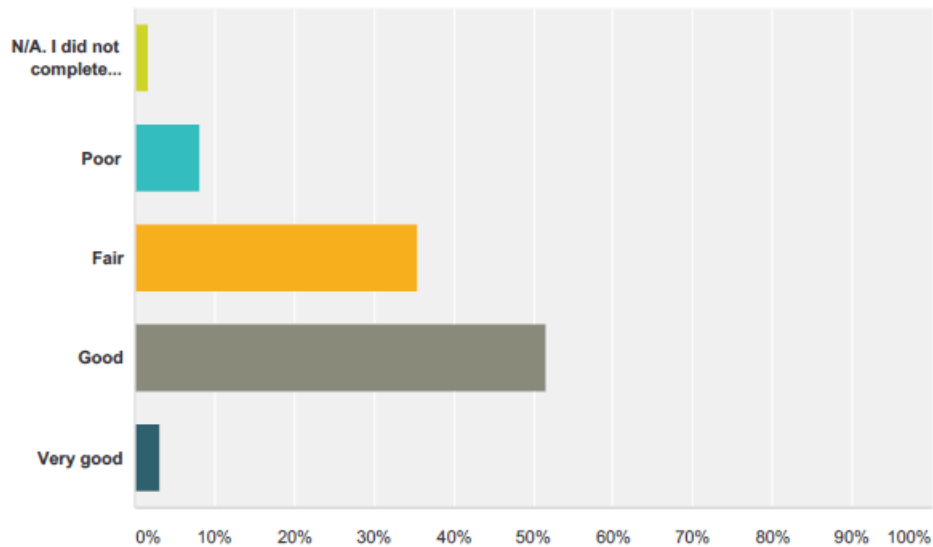
**TABLE 9. SURVEY RESULTS: FEEDBACK ON SCORING STUDIES CONT.**

<p><b>Please provide any additional feedback you have regarding your experience with the Scoring Studies, including comparing it to other forms of training you have participated in.</b></p>	<p>"The group discussions on the observations were the most helpful."</p> <p>"I think the feedback will help me become a better administrator."</p> <p>"It is not useful to participate as an individual. Some of the videos were not great examples of the elements."</p> <p>"In other forms of training on NCEES that I have been a part of did not have the videos of actual classrooms to observe. It was always hypothetical situations. This was much more realistic and helpful in learning what to look for."</p> <p>"It would be nice to be able to hold discussions on the scoring studies to hear other administrators' observations on what they were seeing."</p> <p>"Viewing the videos was great and they could be used for teachers to view. Having no knowledge of the people on the videos was hard to evaluate at first."</p> <p>"I appreciate the specific feedback on the Scoring Studies and why a particular score was selected on the rubric."</p> <p>"I found the scoring studies to be very helpful- I particularly liked the feedback that was provided after the scoring was complete."</p> <p>"I have found that indicators of the evaluation system can differ from school to school site and county to county. This was a great way to attempt to be better aligned."</p> <p>"These were way too long and we didn't have enough background knowledge to really fill in Standards II, III, and IV properly."</p> <p>"It was great training. I was surprised by some of the areas that were rated low."</p> <p>"I would try to incorporate more variety in the examples as far as scoring - most of these appeared to fall in the Not Observed or 1."</p> <p>"I question myself more, because throughout the studies and the lessons I was often shocked at the ratings that I'd give vs. the ratings that the OCT system suggested was correct."</p>
---	--

Table 10 presents feedback on OCT Lessons. Although there was some variability in what respondents thought of Lessons, most thought that the target score feedback was either fair or good. However, several participants noted that they were surprised by or disagreed with the target scores for the videos, and/or thought more explanation would have been helpful, particularly for the lower scores. The comments in the bottom half of the table help elucidate both strengths and weaknesses of the Lessons.

**TABLE 10. SURVEY RESULTS: FEEDBACK ON LESSONS**

Question	Response	Frequency	Percent
<b>In the Lessons you completed, how would you rate the quality and usefulness of the target scores and justifications?</b>	I did not complete Lessons	1	1.6%
	Poor	5	8.1%
	Fair	22	35.5%
	Good	32	51.6%
	Very good	2	3.2%



Question	Response
<b>Please provide any additional feedback you have regarding the use of Lessons, including comparing it to other forms of training you have participated in.</b>	<p>"My scoring study aligned closely with the rates scores; however the more lessons I completed, the less comparable my scores were to the master raters--however, my observation comments were almost the same as the raters."</p> <p>"Having an actual lesson to view provides more authentic evidence to understanding the standards."</p> <p>"Once again it would be nice to be able to hold discussions with other administrators."</p>

**Please provide any additional feedback you have regarding the use of Lessons, including comparing it to other forms of training you have participated in.**

"Much more valuable when you have colleagues to discuss scoring."

"I could tell I had difficulty determining the difference between accomplished and distinguished."

"Great way to become uniformed across the state."

"I think that it is no true fault of the OCT, but several elements based on the developing/proficient levels were very similar. However, the scoring fluctuated at times and the argument could have been made for 1s or 2s."

"This was very difficult because we didn't have all the knowledge we would have going in to a real observation. In addition, we did many of these together as a group and disagreed with multiple of the target scores and found the justifications didn't match the scores at times."

"... some of the justifications did not fit with the target score provided. "

"I thought in many instances the teachers performed better than the score given."

"Better than other forms of training. Would have liked a little more information on some of the areas that were rated low."

"In some of the justifications there was no or little explanation why the score was scored as such-ie. why wasn't the rating above or below."

The survey also solicited suggestions from participants about what additional supports or structures would be helpful in implementing the OCT. Table 11 displays the comments from participants. Once again, the importance of having facilitated and collaborative group discussions around the videos was stressed.

**TABLE 11. SURVEY RESULTS: SUPPORTS NEEDED FOR SUCCESSFUL IMPLEMENTATION**

Question	Response
<p><b>Going forward, what additional supports or structures would be helpful in implementing the OCT at your LEA or school site?</b></p>	<p>"I felt our LEA did an outstanding job using your resources to help all of us grow in our knowledge of the observation tool."</p> <p>"More facilitator resources and ideas, especially when administrators from different schools are participating and face-to-face discussions are not convenient."</p> <p>"Additional discussion relating to the scoring."</p> <p>"Additional videos where the teacher achieved an accomplished rating"</p> <p>"I think the time of the year that the OCT was done in our district wasn't good as it was the busiest time of the year. I think that our participation would have been better if the timing could have been different."</p> <p>"I would like for my LEA to work collaboratively to better align our use of NCEES for teacher evaluations."</p> <p>"interaction and conversation with other teachers/administrators"</p> <p>"I think commentary during a second watch of the video with specifics about the elements that were being targeted would be necessary to help the evaluator see what was missed if she scored it differently."</p> <p>"More videos for elective classes (foreign languages, PE, Music, Art, CTE etc...)"</p> <p>"Worked good to have a facilitator leading discussion"</p>

Note. Not all survey respondents answered every question.

## Discussion

The 2015-16 OCT implementation was a voluntary participation opportunity that included evaluators in 33 of the 267 North Carolina LEAs. Despite the lower than expected numbers of participation, the results of this program were consistently positive. Consistent with last year's pilot results, Observers showed statistically significant improvement on scoring from the start of the program to the end. That improvement was directly tied to the extent to which they utilized the available online resources. Engagement with the BloomBoard Marketplace also increased significantly this year. Participants expressed that the OCT platform was easy to use, had few technical issues, and helped them improve their application of the NCEES rubric.

As evidenced by participant comments, improvements can be made to the program moving forward. The target score justifications could be expanded in some cases, and the available library of videos could also be expanded to include teachers of more varying quality. It is also evident that encouraging more collaborative group activities either locally or regionally would maximize the benefit of the tool.

We continue to learn more about how the program can be best implemented to help participants increase their understanding of the NCEES and accurately apply it in practice during live classroom observations. A concerted effort will be made in future implementations to engage LEAs who were unable to participate due to scheduling and bandwidth concerns. A stronger facilitator presence will also be encouraged to help foster collaborative environments and group learning opportunities.

## Appendix A. North Carolina Educator Evaluation System Rubric

**TABLE A1. NCEES TEACHER RUBRIC OBSERVABLE ELEMENTS\***

Standard	Element number	Element description
<b>Standard 1: Teachers demonstrate leadership</b>	1a	Teachers lead in their classrooms
	2a	Teachers provide an environment in which each child has a positive, nurturing relationship with caring adults
<b>Standard 2: Teachers establish a respectful environment for a diverse population of students</b>	2b	Teachers embrace diversity in the school community and in the world
	2c	Teachers treat students as individuals
	2d	Teachers adapt their teaching for the benefit of students with special needs
	<b>Standard 3: Teachers know the content they teach</b>	3a
3b		Teachers know the content appropriate to their teaching specialty
3c		Teachers recognize the interconnectedness of content areas/disciplines
3d		Teachers make instruction relevant to students
<b>Standard 4: Teachers facilitate learning for their students</b>	4a	Teachers know the ways in which learning takes place, and they know the appropriate levels of intellectual, physical, social, and emotional development of their students
	4b	Teachers plan instruction appropriate for their students.
	4c	Teachers use a variety of instructional methods
	4d	Teachers integrate and utilize technology in their instruction
	4e	Teachers help students develop critical-thinking and problem-solving skills
	4f	Teachers help students work in teams and develop leadership qualities
	4g	Teachers communicate effectively
	4h	Teachers use a variety of methods to assess what each student has learned

\*The elements listed in this table only represent the elements that are directly observable through video observation.