

Observation Calibration Training 2014-15 Pilot Case Study: Newton-Conover City Schools

Background

OVERVIEW OF THE OCT

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. The online platform combines BloomBoard's Professional Development resources with Empirical Education's observer training and calibration tool, Observation Engine. 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*. The OCT aims to improve observation skills, increase rater agreement, and to provide a common experience for LEAs to host collaborative conversations to improve instructional leadership skills.

OCT PILOT

During the 2014-15 school year, the North Carolina Department of Public Instruction (NCDPI), BloomBoard, and Empirical Education initiated a pilot implementation of the OCT. This included voluntary participation from administrators in 32 LEAs. Participants were provided with an introduction to the project and available resources via webinar in November 2014, and were given access to written instructional materials and a short demonstration video. As a part of the pilot, participants were to complete the following tasks in Observation Engine:

- Scoring Study 1 (SS1): a one-video observation where participants are to rate all observable NCEES elements
- Element-specific lessons: short video clips focused on one specific NCEES element (17 lessons available with 2 clips for each observable element)
- Scoring Study 2 (SS2): after completing SS1 and engaging in rating practice through the element-specific lessons, observers were to complete a second scoring study with a different full-length video and score all observable NCEES elements

The purpose of the first year pilot was to evaluate the effectiveness of the OCT resources and gather feedback from administrators. This case study describes the pilot activity and quantitative results of one of the participating LEAs, Newton-Conover City Schools (NCCS). NCCS provides an excellent

example of how the OCT resources can be leveraged to facilitate collaboration and customized professional development opportunities.

NEWTON-CONOVER CITY SCHOOLS & THEIR APPROACH TO THE PILOT

Newton-Conover City Schools (NCCS) is a suburban public school district in North Carolina located one hour outside of Charlotte. NCCS serves over 3,000 students across 7 schools.¹ The 12 participating administrators of NCCS took a highly collaborative approach to the 2014-2015 OCT pilot. Heather Mullins, NCCS’s Chief Academic Officer, integrated OCT resources into the district’s biweekly principal meetings and customized a detailed schedule for completing activities in the platform, both as a group during the principal meetings and individually between meetings. During an OCT webinar in March, Heather discussed their approach to the project: “We wanted to take a systemic approach to the OCT pilot, and really dig into this with our principals, understanding that their time is valuable.” Table 1 below shows their schedule of activities.

TABLE 1. NCCS OCT SCHEDULE

Date or Completion Window	OCT Task	Together or Individually
11/18/14	OCT Kick-Off & Scoring Study 1	together
12/3/14	Element-specific lesson 1a	together
12/3/14 – 1/7/15	Element-specific lessons 2a & 2b	individually
1/7/15	Element-specific lesson 2c	together
1/7/14 – 1/12/15	Element-specific lessons 2d & 3a	individually
1/12/15	Element-specific lesson 3b	together
1/12/15 – 2/4/15	Element-specific lessons 3c & 3d	individually
2/4/15	Element-specific lesson 4a	together
2/4/15 – 2/17/15	Element-specific lessons 4b & 4c	individually
2/17/15	Element-specific lesson 4d	together
2/17/15 – 3/4/15	Element-specific lessons 4e & 4f	individually
3/4/15	Element-specific lesson 4g	together
3/4/15 – 3/17/15	Element-specific lessons 4h	individually
3/17/15	Scoring Study 2	together

*NCCS plans to complete additional activities in the OCT platform beyond those listed above, but those activities go beyond the scope of this initial pilot.

¹ Common Core of Data Public school district data for the 2012-2013 school year.

For the activities completed together, participants watched the video clip as a group, rated silently and individually, and then looked at the target scores as a group. They discussed misconceptions and deconstructed and analyzed the language of the NCEES rubric. Between meetings, participants were to complete element-specific lessons on their own.

Although NCCS had a relatively small number of participants, their concentrated efforts and collaboration provide an excellent example for other LEAs to follow in utilizing the OCT. In the following section, usage and scoring data from Observation Engine are used to present participation and performance results are presented. In addition, a focus group discussion conducted at the end of the pilot is summarized. To preserve confidentiality, results and opinions of individual participants are not reported; all performance results are aggregated at the group level.

Results

PARTICIPATION

All 12 participants completed two scoring studies: Scoring Study 1 (SS1) at the beginning of the pilot and Scoring Study 2 (SS2) at the end of the pilot. A majority of participants (nine of 12) completed all 17 available element-specific lessons, and all the remaining participants completed a majority of the lessons (at least nine).

COMPARISON OF SCORING STUDY 1 AND SCORING STUDY 2

Participants' performance improved on all three measures on which they are assessed: percent target agreement, percent discrepant, and scoring bias. Target agreement increased by more than five percentage points, percent discrepant declined from eight to three, and the proportion of participants exhibiting some scoring bias declined substantially (only two participants in SS2 compared to seven in SS1). The decline in the proportions of discrepant scores and participants with scoring bias was statistically significant (at *p level* of 0.05 or better), which is a particularly strong result since the likelihood of finding any statistically significant differences in a sample of 12 subjects is generally very low. Moreover, analysis of score discrepancy by element showed that large deviations from target scores were practically limited to element 2b, which apparently presented a particular difficulty for all participants. By contrast, discrepant scores were spread across all elements. This suggests that the participants developed an essential understanding of the rubric after completing the lessons.

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

Metric	Scoring Study 1	Scoring Study 2
Mean Percent Target Agreement	49.6%	54.9%
Mean Target Discrepant	8%	3%*
# Participants w/ Scoring Bias	7	2*

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.

*Difference from SS1 results statistically significant at $p < .05$

Findings from Focus Group

On May 19th, Empirical Education and BloomBoard hosted a focus group discussion to gather qualitative feedback from NCCS participants on their experience with the OCT pilot. Participants shared that the OCT provided an opportunity to “dig deeply” into the NCEES standards as a group. The conversations that ensued around a particular event or instructional practice seen in the video were highly valuable, allowing for understanding of varying perspectives and interpretations. They looked specifically at semantics and interpretations of wording around particular elements in the NCEES framework. These collaborative activities brought far more value to the OCT process than completing the OCT video observations independently. However, the Newton-Conover group acknowledged that they had much more experience with the NCEES and classroom observation than newer evaluators, who would definitely benefit from the OCT’s self-paced activities.

Prior to the OCT, NCCS conducted NCEES work by discussing the specific standards and generating look-fors for each element. They were not able to refer to videos of classroom lessons to discuss specific examples. The OCT’s video observation capabilities gave them the opportunity to look more in-depth and apply the standards as a group.

The participants expressed that the OCT videos did have some limitations. They believe that some of the target scores were slightly inflated and that there was not sufficient evidence given for some of the target scores. In addition, it would have been helpful to see some examples of distinguished teaching in the video set as well as more diverse grade levels and subject areas. They also provided suggestions for improvements to the rating page interface, including the ability to toggle between elements, so that users could see all element specific information on one page. When asked if they would like to have access to the OCT next year, they said that they would certainly benefit from additional collaborative work with videos they had not yet seen.

In the end, they indicated that they learned a lot from this process. They gained a better understanding of how different evaluators go through the rating process and felt that the collaboration the OCT encourages improved their understanding of the NCEES framework.

Discussion

These results suggest that the OCT provides resources that can help administrators become better evaluators. It is likely that the face-to-face collaboration augmented the experience for participants and set them up for success. Heather Mullins has shared that the overall experience has been positive for NCCS: “As a curriculum leader, it’s been really important to see my principals emerge as leaders in this process. This gives them a real-life opportunity in real-time to come back and share what they’ve learned. We feel that the OCT pilot has really helped us and we’re learning together.”

On April 28th, NCCS will review performance on both scoring studies and select the four most challenging elements for continued, concentrated work. Their review will be based on the automatically generated report from the OCT platform that provides performance metrics and agreement trends by individual element. They plan to split into two groups and conduct live classroom observations. They will then discuss the four elements and how the teacher should be rated. Lastly, they will consider how they would coach that teacher, and select a few key points for feedback. The OCT provides a springboard for these kinds of activities, assisting evaluators with helpful data and learning material for continuous improvement.

Appendix A – North Carolina Educator Evaluation System Teacher Rubric

TABLE A1. NCEES TEACHER RUBRIC OBSERVABLE ELEMENTS*

Standard	Element Number	Element Description
Standard 1: Teachers demonstrate leadership	1a	Teachers lead in their classrooms
	2a	Teachers provide an environment in which each child has a positive, nurturing relationship with caring adults
Standard 2: Teachers establish a respectful environment for a diverse population of students	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
	3a	Teachers align their instruction with the North Carolina Standard Course of Study
Standard 3: Teachers know the content they teach	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
	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
Standard 4: Teachers facilitate learning for 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.