

**Observation and Report on**

**NCEXTEND1 Reading (Grade 8, English II) Achievement Level Descriptor  
Development; NCEXTEND1 Science (Grades 5 and 8, and Biology) Standard  
Setting;**

**North Carolina Reading End-of-Grade (Grades 3-8) Standard Setting; and  
NCEXTEND1 Reading (Grades 3-8 and English II) Standard Setting**

July 12-16, 2021

Marriott Crabtree Valley  
Raleigh, NC

Prepared by:

Gregory J. Cizek, PhD

July 24, 2021

## Executive Summary

The North Carolina Department of Public Instruction (NCDPI) contracted with Data Recognition Corporation (DRC) to provide services for three standard setting related activities: 1) development of Achievement Level Descriptors (ALDs) for the grade 8 and English II NCEXTEND1 assessments in reading; 2) setting performance standards (i.e. “academic achievement levels,” or “cut scores”) for NCEXTEND1 assessments in science (Grades 5, 8 and Biology; and 3) setting performance standards for the NC End-of-Grade reading assessments for grades 3-8 and for the NCEXTEND1 reading assessments for grades 3-8 and the English II assessment. It should be noted that, because the North Carolina State Board of Education is the entity with the authority to actually set the performance standards on NC state assessments, the processes of the workshops described in this report are most accurately viewed as standards *recommendations*; the participants’ recommendations are subsequently reviewed and potentially approved by the relevant state personnel.

All activities were conducted in-person during the week of July 12-16, 2021. A large (total n = 86) and diverse groups of participants—namely, North Carolina classroom educators—were empaneled to perform the standard setting procedures across the week of July 12-16, 2021 at the Marriott Crabtree Valley hotel in Raleigh, North Carolina.

The sessions for each assessment program (i.e., general assessment, alternate assessment) were implemented using different standard-setting methods. The Bookmark procedure (Lewis, Mitzel, Mercado, & Schulz, 2012) was used for the general assessments; the Yes/No Angoff procedure (Plake & Cizek, 2012) was used for the NCEXTEND1 assessments.

Qualified educators from North Carolina were trained in the methods and led through the standard setting procedures by content and process specialists. The participants’ judgments were

solicited in two ways: they first generated exclusively content-based judgments and cut scores across three rounds of judgments; following these activities, participants who had served as table leaders in the individual sessions met to consider adjustments to the system of recommended cut scores in processes called “vertical articulation.”

The author of this report was contracted by NCDPI to provide an independent, external observation of the standard setting sessions and to submit a report of observations and findings. The author has expertise and extensive experience in the area of setting performance standards (see, e.g., Cizek, 2001, 2012; Cizek & Bunch, 2007), including designing, facilitating, and observing standard setting procedures. The author has served previously as an independent observer of previous standard setting workshops for NCDPI; the author also serves as a technical advisor for NCDPI assessment programs.

Overall, the workshops followed best practices in the area of standard setting; the procedures as implemented followed the plan that had been reviewed and approved by the state’s technical advisors; and the activities produced cut score recommendations that reliably and accurately reflect the intended performance expectations for North Carolina students and the expert content judgments of North Carolina educators. Unless the panelists’ evaluations (not available as of the date this report was submitted) indicate otherwise, policy makers can have confidence that the recommendations from the standard setting activity are based on sound procedures, producing credible, defensible, and educationally useful results.

The report is organized into four sections: 1) Executive Summary 2) Standard Setting Workshop Observations; 3) Summary and Conclusions; and 4) References. The remainder of this report provides a description of the standard setting activities, some recommendations, and a summary evaluation.

## II. Workshop Observations

### *Monday, July 12: Morning Activities*

Two standard setting activities for NCEXTEND1 assessments began on the morning of Monday, July 12, 2021. The sessions began at approximately 8:30 a.m. In one of the larger conference rooms, participants in the NCEXTEND1 science standard setting were welcomed by DRC staff and provided with an orientation to their tasks for the workshop. In another conference room, participants were welcomed and introduced to their task of developing ALDs for the grade 8 and English II NCEXTEND1 reading assessments. Each of the NCEXTEND1 assessments uses a system comprising three achievement levels: *Not Proficient*, *Level 3*, and *Level 4*. Two cut scores are used to create the three performance categories.

### NCEXTEND1 Reading ALDs (Day 1 of 1)

Because a previously-scheduled workshop was interrupted by severe weather in summer 2020, a session was conducted to complete ALDs for NCEXTEND1 in reading for grade 8 and English II. A total of 6 North Carolina educators participated in the reading ALD development workshop. Drafts of ALDs were developed for three achievement levels for the grade 8 and English II assessments; the participants' task was to review and revise (if needed) the drafts. All work was begun and completed on the morning of Day 1. Participants were welcomed to the session by DRC staff, reminded of the responsibilities related to confidentiality and security of materials, and provided with a review of their task. Participants reviewed the key components of ALDs, which included attention to the content of the relevant standards, the cognitive complexity increases across achievement levels, and the contextual aspects of student performance (e.g., scaffolding) across levels.

Participants began their work with a review of the draft ALDs for grade 8; they then proceeded to review the draft ALDs for English II. This activity concluded by noon on the first day with the group considering *vertical articulation* of the ALDs; that is, the group reviewed the grade 8 and English II ALDs they had drafted and considered whether they reflected a cohesive, steady increase in performance expectations across grade levels; overall, few changes were made. Participants were again thanked for their work on the ALDs, and all panelists were dismissed at approximately 12:00 p.m.

#### NCEXTEND1 Science Standard Setting (Day 1 of 2)

Standard setting for the NCEXTEND1 science assessments (grades 5, 8, and biology) were scheduled to be completed over the first two days of the week. Before the formal opening on Monday of the NCEXTEND1 standard setting workshop for science, DRC workshop facilitator, Dr. Ricardo Mercado asked all participants to complete a preworkshop survey. Surveys were completed on individual laptop computers that were preloaded with meeting materials and the application (the “Hub”) that participants would use for completing their work during the workshop. Finally, Dr. Mercado directed participants to complete hard copies of security and non-disclosure agreements before providing access to any secure assessment materials used in the workshop.

Upon completion of the preworkshop survey, Dr. Tammy Howard (Director of Accountability Services) welcomed all participants, thanked them in advance for their willingness to serve, introduced NCDPI staff, and provided with a general orientation to the NC assessments, the standard setting development process, and the standard setting activities that they would be engaging in over the next days. Following the orientation, a brief informational session was presented related to participants’ CEU credits, honoraria, travel reimbursement, and other

housekeeping details.

The session facilitator, Dr. Mercado, then introduced other DRC staff and facilitated introductions of all participants. A total of 24 North Carolina educators participated in the general orientation session on Day 1. Dr. Mercado then provided a more detailed orientation to the NCEXTEND1 assessment and the standard setting activities. The orientation included information on:

- \* the NCEXTEND1 items and scoring, the achievement levels and ALDs;
- \* the concept of a “threshold student” (i.e., a student just on the threshold of a performance level);
- \* the item maps that participants would be using to aid in making their judgments; and
- \* the specific standard setting procedure that would be used (Yes/No Angoff method).

The first part of the morning orientation session concluded at 10:15 a.m. and participants were given a break until activities resumed at approximately 10:30 a.m.

On return from the morning break, Ms. Bonnie Wright (DRC reading alternate assessment specialist) provided participants with additional information on the NCEXTEND1 assessment items and scoring to provide them with greater familiarity with the actual assessment and its administration. Ms. Wright then facilitated a discussion on the student population eligible to take the NCEXTEND1 assessment and the variability in ability and educational experiences within the NCEXTEND1 population.

Mr. David Durette (DRC science content lead) provided an introduction to the first specific task that participants would complete on Day 1—a review of the ALDs and consideration of the knowledge, skills, and abilities of students just on the threshold of each achievement level. He emphasized three aspects of ALDs: 1) the grounding of ALDs in the North Carolina extended

content standards; 2) that ALDs reflect the knowledge and skills that students possess at each level, as opposed to knowledge and skills that they lack; and 3) the need for participants to reflect on relevant students generally, as opposed to their particular students only. Mr. Durette then facilitated consideration of threshold students, beginning with those at the threshold of the Level 3 performance category for the grade 8 assessment. Participants were seated at six tables of four participants each. Each table was assigned 1-2 specific content standard and discussions of the Level 3 threshold students occurred at each of the tables. Participants were provided with the Range ALDs corresponding to the content standard(s) from which to develop drafts of the threshold ALD descriptions. Each table had one participant charged with notetaking on the discussions for later presentation to the full group. Table discussions began at approximately 11:15 a.m. Table leaders began reporting on their group discussions regarding Level 3 threshold students until approximately noon, when participants were dismissed for lunch until 12:30 p.m.

#### *Monday, July 12: Afternoon Activities*

Following lunch, participants then repeated a threshold ALD activity for the Level 4 threshold. Table leaders reported on their groups' work and discussion of each group's draft Level 4 threshold descriptions followed. Using the "Hub" application, participants were then directed to examine in detail 25 actual items from a released form of grade 8 NCEXTEND1 assessment and to begin work developing their item maps by analyzing the knowledge and skills threshold students would need to answer each of the 25 items correctly. Approximately one hour was allocated for this activity. Participants began their work at 1:15 p.m. and concluded at approximately 2:00 p.m.

Following additional presentation of information on the threshold students, participants were directed to engage in the Yes/No Angoff standard setting task in three rounds of judgments.

The specific task for Round 1 was to estimate, for each item, how many points (0, 1, or 2) that the threshold student would be expected to achieve. For Round 1, participants were directed to start with a review and to make independent judgments for all items with reference to threshold Level 3 students and to indicate their judgments on the item map (found on the “Hub”). Before beginning their work, participants completed and discussed a survey to gauge their understanding of the Yes/No Angoff procedure and to ensure that they were ready to proceed with making their judgments. Participants could enter any lingering questions or ask for clarifications about the process in the survey so that these could be addressed before continuing. The survey and discussion concluded at approximately 2:45 p.m. at which point participants began making their Round 1 judgments on all items for grade 8 threshold Level 3, followed by judgments on the same items for threshold Level 4.

When all judgments were completed, participants’ overall values were analyzed by DRC and a briefing session with NCDPI staff was held at approximately 3:45 p.m. Each participant’s overall judgment was computed by summing their individual judgments (0, 1, 2) for each item. The Round 1 group recommendation was obtained by computing the median of the individual results. Round 1 results were discussed and a graphical display of results was prepared for presentation to participants. At 4:15 p.m., results for the first round of ratings were presented to panelists in a whole group setting. Round 1 results consisted of histograms showing the locations of each participant’s recommended cut score, group median cut scores, and the percentages of students that would be classified into each achievement level if the Round 1 median cut scores were applied to the actual test performance; (this information is called “impact data”).

Panelists were given time to contemplate the results, engage in discussion, and were briefed on the process for generating their Round 2 judgments. Additional information provided for Round



2 included the difficulty levels for each item (that is, the percentages of students earning at least one point and two points on each item). Round 2 provides an opportunity for participants to revise their judgments; however, it was stressed that no changes were required: the purpose of Round 2 was only to allow participants to reconsider their judgments, to consider the additional information, and to generate a second set of judgments in consultation with colleagues at their tables. The discussion of Round 1 judgments and directions for completing Round 2 judgments ended at approximately 4:30 p.m. At that time, it was determined that there would not be enough time to complete a second round of judgments by the scheduled end time for Day 1 (5:00 p.m.), so participants were dismissed for the day and the workshop schedule was adjusted to allow the Round 2 judgements to be completed at the beginning of Day 2. At the end of Day 1, all secure materials were collected.

A debriefing session with DRC and NCDPI staff was held at the end of Day 1. Facilitators from the ALD development group and the NCEXTEND1 science standard setting group summarized the events of the day; no major concerns were identified. Some discussion centered on how to incorporate additional information into the Day 2 opening presentation to ensure that the science panelists were correctly interpreting and applying the policy ALDs at the level of rigor intended. The debriefing session concluded at approximately 6:00 p.m.

### *Tuesday, July 13: Morning Activities*

On the second day of workshops, standard setting for the NCEXTEND1 science assessments continued and a new standard setting activity was initiated: standard setting for the NCEXTEND1 reading assessments for grades 3-8 and English II.

## NCEXTEND1 Reading/English II Standard Setting (Day 1 of 3)

Standard setting for the NCEXTEND1 reading and English II assessments were scheduled to be completed over three days (Tuesday, Wednesday, Thursday). Similar to the standard setting for NCEXTEND1 science assessments, DRC workshop facilitator, Dr. Ricardo Mercado welcomed participants and directed them to complete a preworkshop survey. Surveys were completed on individual laptop computers that were preloaded with meeting materials and the application (the “Hub”) that participants would use for completing their work during the workshop. Participants also completed security and non-disclosure agreements before they were permitted access to any secure assessment materials used in the workshop.

Upon completion of the preworkshop survey, Dr. Tammy Howard (Director of Accountability Services) welcomed all participants, thanked them in advance for their willingness to serve, introduced NCDPI staff, and provided with a general orientation to the NC assessments, the test development process used for the NCEXTEND1 assessments, the standard setting development process, and the standard setting activities that they would be engaging in over the next days. Following the orientation, a brief informational session was presented related to participants’ CEU credits, honoraria, travel reimbursement, and other housekeeping details.

The session facilitator, Dr. Mercado, then introduced other DRC staff and facilitated introductions of all participants. A total of 24 North Carolina educators participated in the general orientation session on Day 1. All participants had experience in the subject area of reading/English language arts.

Parallel to the opening session for the NCEXTEND1 science standard setting, Dr. Mercado then provided a more detailed orientation to the NCEXTEND1 assessment and the standard setting activities, including:

- \* the NCEXTEND1 items and scoring, the achievement levels and ALDs;
- \* the concept of a “threshold student” (i.e., a student just on the threshold of a performance level);
- \* the item maps that participants would be using to aid in making their judgments; and
- \* the specific standard setting procedure that would be used (Yes/No Angoff method).

The first part of the morning orientation session concluded at 10:15 a.m. and participants were given a break until activities resumed at approximately 10:30 a.m. Following the break, participants began work on their main task for the morning of Day 1—development of threshold ALDs for the grade 6 assessment. The threshold ALDs describe the knowledge and skills of the hypothetical grade 6 student who is *just* barely into each performance level. For example, the threshold *Level 4* student is a student who possess just enough knowledge and skill to be classified as higher than *Level 3* and only enough to be just barely into the *Level 4* category. With three achievement levels, two threshold descriptions are needed—a description of the threshold *Level 3* student and a description of the threshold *Level 4* student. Work on the threshold ALDs for grade 6 continued up to a scheduled lunch break at noon and continued after the lunch break.

#### NCEXTEND1 Science Standard Setting (Day 2 of 2)

The second day of NCEXTEND1 science standard setting began with meeting facilitators leading a review of the graphic presented previously showing the Round 1 ratings. Also presented were: a review of the impact on student performance that would result if the Round 1 median cut scores were applied; a review of the intended rigor of the policy ALDs around college and career readiness; a review of the actual mean level of student performance from the spring administration; and a review, for comparison, of the 2019 NCEXTEND1 mathematics assessment results. A whole

group discussion ensued and, at approximately 9:00 a.m., panelists were directed to begin their Round 2 judgments. Upon completing their Round 2 judgments, participants completed an evaluation survey and their judgments were submitted for analysis by DRC. All panelists completed their judgments and surveys by approximately 10:00 a.m.

After analysis, NCDPI and DRC staff reviewed Round 2 results. In reviewing the Round 2 results, it appeared that participants may have used the information provided at the end of Round 1 in a way that was inconsistent with what the information was communicating. That is, instead of interpreting the information as evidence that the cut score rigor should be more aligned with the rigor implied by the policy ALDs, participants appeared to make cut score adjustments that reduced the rigor of the expectations. In the meeting of NCDPI and DRC staff, considerable discussion centered on how to address the misalignment. It was decided to explain the issue first to table leaders, then to the whole group. It was further decided to alter the planned schedule of activities to delay making Round 3 judgments for grade 8 science, and to first ensure understanding of the process and the intended level of rigor of the performance standards. Then, participants would break into grade level groups, complete the three rounds of judgments for grade 5 and biology, and return to complete Round 3 for the grade 8 science assessment at the conclusion of the day.

This plan was then reviewed first with table leaders, then presented to the full group. All participants appeared to understand the concern and were agreeable with proceeding with the plan as revised. The full group discussion concluded at approximately 11:20 a.m. At that time, participants separated into grade level groups to begin working on threshold ALDs for their respective grade level assessments (i.e., grade 5, biology).

*Tuesday July 13: Afternoon Activities*

NCEXTEND1 Reading/English II Standard Setting (Day 1 of 4)

After returning from a lunch break, participants continued work on the threshold descriptions task in small table groups (n = approximately 4 per table); they completed this task at 1:00 p.m. At 1:00 p.m., participants were assembled for a whole group discussion of the threshold characteristics that they had developed. At the conclusion of this discussion, the workshop facilitator provided more detailed information on how to make the Yes/No Angoff judgments for each item. This portion of the afternoon activities concluded at approximately 3:30 p.m. with participants completing a mid-procedure evaluation survey.

When all surveys had been completed, participants were then directed to begin making their Round 1 judgments. Round 1 judgments were completed by approximately 4:15 p.m. The resulting data were collected and analyzed by DRC staff and provided to NCDPI at the end of Day 2, with the plan to present a summary of Round 1 results to participants the following morning. Participants were encouraged to begin discussing their perceptions of the day's activities and impressions of their Round 1 judgments they had just completed. To close the day's activities, the facilitator informed the whole group that each table should choose a table leader to help guide subsequent discussions at the table in the coming days.

NCEXTEND1 Science Standard Setting (Day 2 of 2)

After a lunch break, NCEXTEND1 standard setting participants began making their Round 1 judgments within their grade level groups (i.e., grade 5, biology). When all Round 1 judgments were completed, DRC staff analyzed the data and results were presented to participants for grade 5 and biology. Participants discussed their judgments and, at approximately 1:30 p.m., participants

were instructed to begin their Round 2 judgments. To conclude the days activities for the majority of participants, Round 2 judgments were reviewed and discussed; Round 3 judgments for grade 5 and biology were generated by participants, analyzed, presented and discussed. Finally, the grade 5 and biology groups were brought back together as a whole group. The whole group was presented with the complete set of results for grade 5 and biology and asked to return to their task of providing Round 3 judgments for grade 8 science.

When all Round 3 judgments for grade 8 science were submitted, participants completed a final evaluation, materials were collected, and participants were informed of next steps and thanked for their service by Dr. Tammy Howard of NCDPI. Table leaders were requested to remain for an articulation process. (Any participant who wanted to observe the articulation were also welcomed to remain.)

The articulation process consisted of presenting the results for all three levels (grades 5, 5, and biology). The purpose of articulation is to review and promote consistency and coherence of results across grade levels within a subject area. Results for NCEXTEND1 science were presented in terms of percentages of students that would be classified in each of the three performance categories if the group's final Round 3 median cut score recommendations were applied. The table leaders considered the overall results, engaged in thoughtful discussion and ultimately concurred that no changes to any of the recommended Round 3 cut scores.

At the end of Day 2, a debriefing session among NCDPI and DRC staff again took place beginning at approximately 5:00 p.m. A review of the day's activities took place, comprising a summary of results for NCEXTEND1 science and an update on the progress of the NCEXTEND1 reading activities. The work of the NCEXTEND1 science standard setting was complete and the work of NCEXTEND1 reading would continue on Day 3. It was decided that no changes to the

planned activities or schedule for the next day were needed.

### *Wednesday July 14: Morning Activities*

The third day of standard setting activities comprised the continuation of standard setting for the NCEXTEND1 Reading/English II assessments and the beginning of standard setting for the NC End-of-Grade Reading Assessments. Activities for both standard setting efforts began at 8:30 a.m.

### NCEXTEND1 Reading/English II Standard Setting (Day 2 of 4)

The morning of the second day of standard setting for the NCEXTEND1 Reading/English II standard setting activities continued the focus on the standard setting judgments for the grade 6 assessment. The morning began with a review of the purpose of the standard setting work, information on what participants could expect when their Round 1 results were presented, and how they would use those results. Round 1 results (consisting of the median raw cut scores defining *Level 3* and *Level 4* and percentages of students that would be classified into the three achievement levels based on those cut scores) were then presented, discussed, and directions for completing Round 2 judgments were provided. Participants then began generating their Round 2 judgments at approximately 9:15 a.m. Upon finishing their Round 2 judgments, participants completed an evaluation survey and began a morning break while the Round 2 data was analyzed and prepared for presentation.

Before presenting Round 2 results, meeting facilitators decided to gather the table leaders for a special briefing. Similar to what had occurred the previous day for NCEXTEND1 science, it was decided that participants may not be interpreting performance of the threshold students

correctly, so table leaders were provided an advance briefing on results and an extensive discussion occurred regarding how to translate the rigorous expectations of the policy ALDs into Yes/No Angoff judgments. The table leader briefing concluded at approximately 11:15 a.m. and Round 2 results (histogram of judgments for each level, median raw cut scores, impacts) were presented and discussed by all participants. Table leaders led discussions at their tables regarding appropriate interpretation of the policy ALDs and provided guidance on making Round 3 judgments. When discussions at each table concluded, participants then began making their Round 3 judgments; all Round 3 judgments for the grade 6 assessment were completed before noon when participants were dismissed for the scheduled lunch break.

#### NC End-of-Grade Reading Assessments (Day 1 of 3)

Standard setting for the NC End-of-Grade reading assessments (grades 3 – 8) began on Wednesday, July 14 at 8:30 a.m. as a whole group activity. A total of 38 participants were empaneled, all of whom were North Carolina educators with experience in the subject area of reading/English language arts. The session opened with a welcome from Ms. Julie Pointer, DRC, who instructed participants to sign the hard copies of the security agreements that had been distributed to each table in advance, then provided instructions for logging in to the provided laptops that would be used for the standard setting activities. The first activity completed using the standard setting application on the laptops was completion of a first workshop survey. All participants completed the survey by approximately 8:50 a.m. and a formal orientation portion to the workshop began. Participants were welcomed by Dr. Tammy Howard of NCDPI, who also introduced other NCDPI staff. Dr. Howard then provided an overview of the North Carolina assessment development process, the criterion-referenced focus of the assessments, why standard



setting for assessments is necessary, how participants were selected for the standard setting workshop, and the nature and use of policy achievement level descriptors (ALDs). At the conclusion of Dr. Howard's presentation, information on logistic details (honoraria, travel reimbursement, CEUs) was presented by Ms. Beth Nash of NCDPI.

The next portion of the morning session was facilitated by Dr. Ricardo Mercado of DRC who provided an overview of the workshop activities. Dr. Mercado provided a more detailed orientation to the standard setting activities, including information on:

- \* the achievement levels relevant to the end-of-grade assessments (four levels: *Not Proficient, Level 3, Level 4, and Level 5*), the achievement level descriptions (ALDs) and their relationship to cut scores;
- \* the concept of a "threshold student" (i.e., a student just on the threshold of an achievement level);
- \* the item maps that participants would be using to aid in making their judgments;
- \* the specific standard setting procedure that would be used (the "Bookmark" method), the ordered item booklets (OIBs; an OIB consists of all items representing a specific assessment, presented one item per page, in order from the easiest to most difficult item), and the rounds of judgments that participants would be making;
- \* the general plan for the workshop (i.e., first discussing threshold descriptions for grades 3, 5, and 7 and making cut score recommendations for those grades, then developing threshold descriptions for grades 4, 6, and 8, and making cut score recommendations for those grades; and
- \* additional information on the DRC application, the "Hub," that provides participants with a website containing workshop materials.

Detailed information on the item map and threshold students was then presented in conjunction with an abbreviated OIB that would be used for practice with the Bookmark method. Participants used the abbreviated OIB to gain experience with the ordering of items according to difficulty, and to practice analyzing the knowledge and skills needed to answer test questions. At the conclusion of this activity, participants were dismissed for a morning break at approximately 10:30 a.m. During the break, the whole-group room was divided into three separate rooms so that the next activity could be conducted in grade-level groups.

Participants returned from the morning break at approximately 11:00 a.m. into separate rooms for grades 3, 5, and 7 to begin discussing threshold descriptions of each achievement level. These discussions continued until the scheduled lunch break at noon.

#### *Wednesday July 14: Afternoon Activities*

##### NCEXTEND1 Reading/English II Standard Setting (Day 2 of 4)

After the lunch break, participants in the NCEXTEND1 reading/English II standard setting were split into two grade level groups to begin work on reviewing the ALDs for grades 5 and 7. The ALDs were used to aid in the development of threshold descriptions for the two achievement levels that would be determined. The threshold discussions occurred within grades at the table level, with table leaders facilitating discussion of the knowledge and skills that characterize the just barely *Level 3* and just barely *Level 4* students. At the conclusion of the discussions, whole-group, within-grade discussions were facilitated to promote consistency of threshold conceptualizations across all participants. As a final activity to provide context and background information for making their standard setting judgments, all participants completed a review of actual test items for their respective grade levels. At the conclusion of this activity, participants

completed an evaluation/readiness survey and were given instructions from the room facilitators on how to make their Round 1 Yes/No Angoff judgments. Grade level groups completed their Round 1 judgments at slightly different times, with the Grade 7 group completing their judgments by approximately 2:30 p.m.; the grade 5 group completed their Round 1 judgments at approximately 3:00 p.m. Both grade level groups took afternoon breaks while results were analyzed and NCDPI staff were briefed.

Following their breaks, participants in each grade level group were presented with their Round 1 results. As in all previous Yes/No Angoff procedures during the week, the results consisted of histograms of individual raw cut scores, group median cut score information, and anticipated impacts (i.e., percents of students that would be classified into each achievement level using the median cut scores from that round). After discussion of the Round 1 results, both groups began making their Round 2 judgments. The groups completed their Round 2 judgments by the scheduled end of the second day.

### NC End-of-Grade Reading Assessments (Day 1 of 3)

Following the lunch break, participants in grade level breakout rooms returned to the task of reviewing the ALDs and considering the characteristics of threshold students for each achievement level. Participants then began a review of test items relevant to their grade levels. The major task of the afternoon was to examine, in grade level groups, the ordered item booklets (OIBs) for each assessment. In each grade level breakout room (grades 3, 5, 7), participants reviewed items in the OIBs and recorded notes regarding what knowledge and skills a student would need to answer each item correctly and what aspects of each item contributed to its difficulty. Participants took an afternoon break at approximately 2:30p.m. and returned to complete their

reviews.

Because of a technical issue in accessing the OIBs for one of the groups, the remainder of the afternoon proceeded differently for the grade level groups. Two of the groups (grades 3 and 7) completed their OIB reviews and then received a presentation with more detailed instructions on how to make their Round 1 Bookmark judgments. Access to the OIBs was delayed for the grade 5 group, so the tasks for that group were adjusted slightly to have them begin work on the threshold descriptions for grade 6. Beginning at approximately 4:00 p.m., two of the three grade-level groups (grades 3 and 7) received final instructions for using the “Hub” application to record their Bookmark judgments and participants. Because it was near the end of the day, it was decided to delay beginning the task of independently placing their Round 1 bookmarks until the beginning of Day 2 and participants were dismissed for the day.

As had been done at the end of each of the previous days, NCDPI and DRC staff gathered for a debriefing and planning session following the dismissal of all participants. Discussion focused on two topics: 1) schedule changes that would be necessary to accommodate the different sequences of activities for the end-of-grade reading assessment groups for grades 3 and 7 and for grade 5; and 2) whether additional training would be required for the NCEXTEND1 reading assessment before they would begin work on their final Round 3 judgments. Minor adjustments were made to the end-of-grade schedule for the grade 5 and 6 group; for NCEXTEND1 reading, it was decided to reiterate the information panelists had been provided for Round 2 before they would make their Round 3 judgments.

*Thursday July 15: Morning Activities*

NCEXTEND1 Reading/English II Standard Setting (Day 3 of 4)

The third day of standard setting for the NCEXTEND1 Reading/English II assessments began at approximately 8:30 a.m. with a review of the previous day's Round 2 results in grade level groups for grades 5 and 7. After discussion of the results, both groups proceeded to make the final (Round 3) judgments. Both groups finished this work by 9:15 a.m. An update on results was presented by DRC to NCDPI staff and, afterward, to the participants.

Upon completing their work making judgments for grades 5 and 7, the group began work developing threshold ALDs for adjacent grades (i.e., one group worked on grade 4; the other group worked on grade 8). Before the scheduled lunch break, the groups were able to also complete their Round 1 judgments for those grades.

#### NC End-of-Grade Reading Assessments (Day 2 of 3)

The second day of standard setting for the NC end-of-grade reading assessments also began at approximately 8:30 a.m. The activities were slightly different depending on grade level. Groups assigned to the Grade 3 and 7 assessments received a refresher on how to set their bookmarks and began the work of making their independent Round 1 bookmark recommendations. Groups were reminded to begin with the Level 4 bookmark placement, followed by the Level 5 bookmark placement, and finally the Level 3 bookmark. Both groups completed this work by 10:00 a.m. Data from this round was analyzed, reviewed by NCDPI and DRC staff, then presented to participants. The summary information presented to participants included a histogram showing each participants' individual bookmark page locations, the median bookmark page placement for each level, and "benchmark" page locations for *Level 3* and *Level 5*. Benchmark locations refer to the approximate location within the current OIBs where the 2019 cut scores are located; this benchmark information can be used by participants to gauge the extent to which their

recommendations are aligned to the achievement standards previously in place for the end-of-grade assessments.

After presentation of the Round 1 feedback to participants, the grade level groups began working on their Round 2 bookmark placements until the scheduled lunch break.

#### *Thursday July 15: Afternoon Activities*

##### NCEXTEND1 Reading/English II Standard Setting (Day 3 of 4)

After a lunch break, participants in the grade 4 and Grade 8 reviewed the results of their Round 1 judgments, were presented with feedback, discussed the results and feedback, and received directions for beginning their Round 2 judgments for the grade 4 and grade 6 assessments. Round 2 judgments were completed, analyzed, reported, and discussed by approximately 1:30 p.m. The participants then proceeded to make their Round 3 (final round) of judgments. Round 3 judgments were completed for grades 4 and 8 of the NCEXTEND1 reading assessments by 2:00 p.m. Following an afternoon break, NCEXTEND1 panelists began working on threshold ALDs for grades 3 and English II. Panelists continued this work and reviewed items in their OIBs until the end of the day; Round 1 judgments for grades 3 and English II would be the first task on the morning of the next day. Some participants finished their OIB reviews as early as 3:30 p.m. All participants completed their work and were dismissed at approximately 4:00 p.m.

##### NC End-of-Grade Reading Assessments (Day 2 of 3)

After completing Round 2 judgments, the grade 3 and 7 groups received the same kinds of feedback on their judgments as they had for Round 1. The groups discussed the results as a whole group, received reminders on how to make their Round 3 bookmark judgments, and began work

on their Round 3 bookmark placements. On completing those judgments, the grade 3 and 7 groups began work on threshold descriptions for grades 4 and 8. The remainder of the afternoon for the grade 4 and grade 8 panelists was spent finishing threshold descriptions and reviewing the OIBs for those grades, with Round 1 judgments for those grades to be made as the first activity of the next morning. Participants completed their work for Day 2 by approximately 4:30 p.m.

The grade 5 panelists completed their Round 1 and Round 2 judgments, received feedback on those judgments, and began work on their final (Round 3) bookmark placements. After receiving and discussing feedback on their Round 3 bookmark placements, panelists in this group were also dismissed for the day; they had previously completed threshold descriptions for both grade 5 and grade 6 assessments, so it was planned for them to also make their Round 1 bookmark placements (for the grade 6 assessment) as the first activity the next day.

As was the case for each preceding day of the week, a debriefing activity with NCDPI and DRC staff took place at the end of the day. Workshop facilitators for each group reported that all activities had gone smoothly for the day. No issues requiring attention or schedule adjustments for the final day were identified.

#### *Friday, July 15: Morning Activities*

##### NCEXTEND1 Reading/English II Standard Setting (Day 4 of 4)

The final day of standard setting for the NCEXTEND1 Reading/English II assessments began as scheduled at 8:30 a.m. In grade level breakout rooms, groups began the day by making their Round 1 Yes/No Angoff judgments for the grade 3 and English II assessments. Those judgments were completed by approximately 9:30 a.m. As was done for the other NCEXTEND1 grade levels, the data were submitted for analysis by DRC and presented to NCDPI staff. At 10:00

a.m., the two grade level groups were presented with the Round 1 results (histograms of individual participants' judgments, median raw cut recommendations for each level, impact) and this information was discussed in the grade level groups.

The groups then began making their Round 2 judgments. Both groups completed Round 2 judgments by 10:30 a.m. The data were analyzed and presented to NCDPI staff; however, in order to make best use of participants' time, it was decided to simply present the Round 2 results to them and delay discussion of Round 2 results until after the lunch break. Thus, participants were dismissed after that presentation for an extended break so that they could check out of hotel rooms, submit reimbursement information, etc.

#### NC End-of-Grade Reading Assessments (Day 3 of 3)

The final day of standard setting for the NC end-of-grade reading assessments began as scheduled at 8:30 a.m. Participants met in a single large meeting room to start the day. They began working on a review of their grade level OIBs for grades 4, 6, and 8. At approximately 9:00 a.m., because it was the final day of standard setting for all participants, Ms. Beth Nash of NCDPI addressed the whole group to provide reminders to the whole group on reimbursements and other final logistic details. At the conclusion of those remarks, and while still assembled as a whole group, participants were provided with additional context information by the workshop facilitator, Dr. Mercado. Information included percentages of students that were in each performance category for all grades (grade 3 through English II) that resulted from the 2019 administration of the EOG assessments. Because the new content standards did not differ significantly from the previous version, such comparison information is relevant context for standard setting in the current year; Dr. Mercado also reminded participants regarding variability in student participation, characteristics of the population of students who participated in the 2021 EOG administration, and



the intended policy interpretation of the highest performance level (*Level 5*) which emphasizes that achievement at that level represents performance of students who are prepared for challenging, advanced level coursework. He invited participants to comment on any of the information that had been presented and outlined the tasks they would be for the rest of the morning.

*Friday, July 15: Afternoon Activities*

NCEXTEND1 Reading/English II Standard Setting (Day 4 of 4)

After a lunch break, participants in the NCEXTEND1 reading/English II standard setting completed their Round 3 judgments for the grade 3 and English II assessments by 1:00 p.m. Data from those judgments was analyzed and presented to NCDPI staff; participants' overall recommendations for Round 3 were unchanged from their Round 2 results. DRC staff presented the final results across all grades to their rooms and facilitated discussion. At the conclusion of the discussions, a final evaluation was completed and all secure materials were collected. Most panelists were dismissed at this time; however, table leaders (n = 6) remained on the afternoon of the final day to participate in a vertical articulation discussions facilitated by Dr. David Chayer of DRC. *Vertical articulation* is a term used to describe review of multiple/cross grade results as a check on the coherence and consistency of the entire system of achievement levels. For this activity, table leaders were presented with the full results for all grades of the percentages of students that would be classified into each of the three achievement levels if their final Round 3 recommendations were applied to the students' test performances.

Table leaders were invited to speak individually about the process and about their perceptions of the full system of results. The table leaders voiced strong support and appreciation for the process. They indicated that they had a good grasp of their tasks, that all members of their

group approached the tasks conscientiously, and that all participants were able to articulate their perspectives. Uniformly, the table leaders strongly endorsed the final results without changes, noting that they were proud of the work they had accomplished and very confident that they had landed on appropriate achievement expectations for the NCEXTEND1 students. The facilitator thanked the panelists for their participation and, as their final task, the table leaders were asked to complete an evaluation of the articulation activity. All surveys were completed and all NCEXTEND1 table leaders who participated in the vertical articulation activity were dismissed at 2:00 p.m.

#### NC End-of-Grade Reading Assessments (Day 3 of 3)

The afternoon of the final day for the NC end-of-grade reading assessments completed the rounds of review for grade 4, 6, and 8. For each round, participants reviewed their OIBs and the threshold ALDs, indicated their bookmark judgments, received feedback on those judgments (i.e., histograms showing individual bookmark placements, median bookmark placements, and impact data), discussed the feedback information and proceeded to the next round. Grade level groups worked at somewhat different rates, with the grade 6 and grade 8 groups finishing all three of their rounds of judgments by 3:00 p.m., and the grade 4 group completing their three rounds of judgments by 3:40 p.m. Round 3 feedback for all groups consisted of showing all participants the final impact percentages for all grade levels (i.e., the grade levels they had worked on in their group plus the results for the other grades); room facilitators encouraged discussion of the results and note-taking on those discussions by table leaders to inform the contributions to a vertical articulation discussions.

Mirroring the procedures used for the NCEXTEND1 assessments, after completing a final

evaluation, all secure materials were collected and most NC EOG panelists were dismissed but table leaders (n = 9) remained to participate in vertical articulation facilitated by Dr. Mercado of DRC. Vertical articulation for grade 3-8 NC EOG reading assessments was facilitated by Dr. Mercado of DRC and began at approximately 4:25 p.m. on the final day. Dr. Mercado detailed the next steps in the standard setting, including review by NCDPI and presentation to the NC State Board of Education. He then projected the full system of median impact results across all grade levels and invited table leaders to offer any comments on the work of their groups and their perceptions of the full system of results.

Overall, the table leaders indicated that they had a good grasp of the process and understanding of their tasks. They indicated that their groups had deep, thoughtful discussions that incorporated and valued the perceptions of each group member. They expressed their opinion that the standard setting work had been a highly valuable professional development activity for them, and they hoped that their colleagues could be involved in such activities in the future so that there might be broader knowledge of the rigorous procedures used to set achievement levels on North Carolina statewide assessments. They expressed appreciation for the process and strong confidence in the results.

After discussion of the full results for the NCEOG reading assessments, the table leaders unanimously voiced endorsement of their recommended cut score ranges without changes, noting that they were proud of the work they had accomplished and very confident that they had landed on appropriate reading achievement expectations for the NC EOG students. To conclude the activities, panelists were asked to complete a final evaluation survey regarding the vertical articulation portion of the workshop. The facilitator and Dr. Tammy Howard of NCDPI staff again thanked the panelists for their participation and reminded them that although they could freely

Speak about any aspects of the standard setting process, they should not disclose any results, as the results could only be considered as final upon action of the North Carolina State Board of Education. All surveys were completed and the table leaders were dismissed by 5:15 p.m.

A final debriefing session was held at the conclusion of the vertical articulation involving NCDPI and DRC staff. The purpose of the session was to confirm information needs, next steps in processing the data, and critical due dates for reports. All staff were thanked by Ms. Kristen Maxey-Moore<sup>1</sup> of NCDPI and the meeting was adjourned.

---

<sup>1</sup> Ms. Maxey-Moore is Section Chief, Test Development, at NCDPI. Due to illness, Ms. Moore was not able to attend the standard setting activities in person. However, arrangements were made for her to virtually observe workshop sessions, participate in daily debriefings, provide overall oversight of the activities each day, and extend her appreciation remotely.

### **III. Summary and Conclusions**

Based on my observations of the procedures, materials, and processes used to obtain recommended performance standards, it is my opinion that the standard setting activities implemented for the North Carolina general and alternate (NCEXTEND1) assessments were conducted in skilled and professional manner consistent with sound psychometric practices. The procedures as implemented followed the plan that had been reviewed and approved by the state's technical advisors. The activities produced cut score recommendations that can be considered as reliable and accurate reflections of the intended performance expectations for North Carolina students and the expert content judgments of North Carolina educators. Unless the panelists' evaluations (not available as of the date this report was submitted) indicate otherwise, policy makers can have confidence that the recommendations from the standard setting activity are based on sound procedures, producing defensible and educationally useful results.

Overall, the process was characterized by a number of strengths; few concerns arose during the course of the standard setting. In the following sections some specific strengths and general conclusions are presented.

#### *Strengths*

- 1) The contractor for setting performance standards on the North Carolina general and NCEXTEND1 assessments developed appropriate and specific plans for implementing accepted standard setting methods (i.e., the Bookmark and Yes/No Angoff methods).
- 2) The state's technical advisors reviewed key elements of the plans in advance and judged them to be sound and defensible.
- 3) The approved plans were implemented faithfully with only minor deviations in schedule

and procedures. When such deviations were deemed necessary, DRC and NCDPI staff conferred and made appropriate, well-reasoned adjustments that appropriately kept the workshop on schedule and safeguarded the integrity of the process.

4) The contractor provided adequate resources and personnel to ensure that the standard setting was conducted professionally and paced appropriately. The contractor staff were qualified to conduct the standard setting activities, representing both psychometric and relevant content area expertise. The content specialists who supported the workshops were knowledgeable about the North Carolina content standards and ALDs. Overall, the contractor facilitated the workshops in a skillful, professional manner.

5) All materials for the workshops appeared to be complete, accurate, and easy for participants to use. The online standard setting utility appeared to work well, as did all other technology used in the workshops.

6) Participants in the standard setting activities had relevant qualifications for making the judgments they were asked to make. Participants consisted of North Carolina educators with experience teaching reading (or science) at relevant grade levels, and educators having experience with either general population students, special needs students, or both. All participants appeared to be motivated to complete their work conscientiously, and they worked attentively and thoughtfully.

7) Participants appeared to understand the standard setting tasks they were to perform, and the nature of the feedback provided to them (i.e., normative and impact information). Participants who were identified as table leaders functioned well in their roles.

8) The meeting arrangements, food service, and other logistics appeared to meet the needs of the participants. Contractor staff were highly attentive to ensuring that meeting rooms

were comfortable and conducive to supporting participants' work.

9) There was appropriate concern for and diligence regarding confidentiality and security of materials and results.

10) NCDPI staff observing the workshops provided support for DRC staff, but otherwise maintained an unobtrusive presence and did not interact with participants related to the standard setting activities. The only visible role of NCDPI staff was to welcome participants at the beginning of the sessions and to thank them for their participation at the conclusion of the workshops.

11) The contractor thoroughly evaluated the standard setting activities to ensure that all panelists understood their tasks and to solicit information about various aspects of the standard setting and the workshop logistics. Evaluations included a pre-workshop survey, mid-process/readiness and end-of-round surveys, and post-workshop summary evaluations including post-vertical articulation surveys. When available, results of these surveys should provide an additional, strong source of validity evidence.

### *Conclusions*

Four conclusions are warranted based on observations conducted of the current standard setting workshop:

1) Performance standards for the NC general and alternate mathematics assessments were recommended by qualified, engaged, and thoughtful groups of North Carolina educators. The standard setting plan was developed and implemented by qualified and conscientious contractor staff. The entire endeavor was overseen by qualified, attentive, and experienced NCDPI staff.

- 2) Overall, no issues of concern arose during the standard setting process for the North Carolina assessments. Minor issues involving daily agendas and procedures did not appear to have any immediate or discernible effects on the procedures or results. Adjustments to the daily agendas and intended procedures were reasonable, appropriate, and supported the integrity of the standard setting process.
- 3) The procedures implemented for recommending performance standards on the North Carolina assessments followed best practices for standard setting and were generally faithful to the specific methodological procedures intended (i.e., Yes/No Angoff, Bookmark).
- 4) The vertical articulation procedures provided an effective mechanism for participants (through their table leaders) to review the overall coherence and consistency of recommended performance standards across grade levels. In each case, the vertical articulation panels recommended no changes to the final results.

The preceding four points support the overall conclusion that the participants' cut score recommendations can be considered to be thoughtful, dependable, and accurate estimates of the cut scores for the relevant assessments. Unless the panelists' evaluations indicate otherwise, policy makers can have confidence that the recommendations from the standard setting activity are based on sound procedures, producing credible, defensible, and educationally useful results.



## IV. References

- Cizek, G. J. (Ed.) (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum.
- Cizek, G. J., & Bunch, M. (2007). *Standard setting: A practitioner=s guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: SAGE.
- Cizek, G. J. (Ed.) (2012). *Setting performance standards: Foundations, methods, and innovations*. New York: Taylor and Francis.
- Lewis, D. M., Mitzel, H. C., Mercado, R., Schulz, E. M. (2012) The Bookmark standard setting procedure. In G. J. Cizek (Ed.) *Setting performance standards: Foundations, methods, and innovations* (pp. 225-254). New York: Taylor and Francis.
- Plake, B. S., & Cizek, G. J. (2012). Variations on a theme: The modified Angoff, extended Angoff, and yes/no standard setting methods. In G. J. Cizek (Ed.) *Setting performance standards: Foundations, methods, and innovations* (pp. 181-200) New York: Routledge.