North Carolina Testing Program

Standard Setting
for the
End-of-Course and End-of-Grade Assessments

Executive Summary

J. Michael Clark III, Ph.D.
Stephen T. Murphy, Ph.D.

August 24, 2013

ALWAYS LEARNING PEARSON

Executive Summary

Introduction

Nine committees of North Carolina educators convened to make cut score recommendations for the End-of-Grade (EOG) assessments for Grades 3-8 Mathematics, Grades 3-8 Reading, and Grades 5 and 8 Science; and for the End-of-Course (EOC) assessments for Biology, English II, and Mathematics I. A total of 164 North Carolina educators convened in Chapel Hill, North Carolina between July 22 and July 26, 2013, using the item mapping method to make content-oriented recommendations for cut scores. A brief summary of the outcomes of this workshop are provided in this executive summary, and a more detailed account of the workshop is provided in the full standard setting technical report.

Panelists

All panelists were asked to provide voluntary demographic information. A brief summary of panelist characteristics is provided in this executive summary. Complete panelist demographics are provided in the full standard setting technical report.

The panelists' years of experience as educators are summarized in Table 1. As illustrated by this table, participants in this standard setting had a wide range of teaching experience.

Table 1. Panelist Experience

Panel	N	Years in Current Position					
		1-5	6-10	11-15	16-20	21+	NR
Mathematics 3-5	20	1	4	8	2	4	1
Mathematics 6-8	16	2	3	4	5	2	0
Reading 3-5	18	1	3	5	1	8	0
Reading 6-8	19	2	2	6	6	3	0
Science 5	16	1	5	5	5	0	0
Science 8	17	3	6	5	1	2	0
Biology	20	2	5	6	4	3	0
English II	17	3	5	5	2	1	1
Mathematics I	21	4	3	5	2	7	0

Note: NR = No Response.

The panelists' professional backgrounds are summarized in Table 2 and Table 3. As will be described in greater detail in a subsequent section of this executive summary, panelists summarized in Table 2 made cut score recommendations for three grade levels within a particular subject area. Individuals reported as teaching in lower, middle, or upper grades are reported in the context of their committee. For example, a lower-grade panelist in the Mathematics 3-5 panel teaches Grade 3 Mathematics, while a lower-grade panelist in the Reading 6-8 panel teaches Grade 6 Reading. Panelists who reported teaching more than one grade level within the subject area are listed under the multiple grades column, and panelists who primarily teach a grade level outside of the panel's range (e.g., a Grade 2 teacher who participated in the Mathematics 3-5 panel) are listed in the off-grade column. Finally, other groups of educators are summarized in the remaining columns of this table. As shown in this table, all grade levels were represented on these panels, and a variety of professional backgrounds was represented on these panels.

Table 2. Panelist Professional Background: Three-Grade Panels

Panel	LOW	MID	UP	MUL	OFF	SED	SPE	COA	GNS	OTH
Mathematics 3-5	3	6	5	2	1	0	2	1	0	0
Mathematics 6-8	7	3	3	1	0	1	1	0	0	0
Reading 3-5	3	1	4	3	1	0	4	0	1	1
Reading 6-8	4	5	3	2	0	3	0	0	0	2

Note: LOW = lower grade, MID = middle grade, UP = upper grade, MUL = multiple grades, OFF = off-grade, SED = special education, SPE = specialist, COA = coach, GNS = grade level not specified, OTH = other.

Panelists summarized in Table 3 recommended cut scores for a single grade and/or subject. Panelists listed in the on-grade column actively teach in the grade/subject for which standards are being set. Panelists summarized in the off-grade column teach in a related subject area, but at a different grade level. Other types of professional backgrounds are summarized to the right of these columns in the table. As shown in this table, the majority of each panel was comprised of individuals who teach the grade/subject of interest, but each showed diversity in panelists' professional backgrounds as well.

Table 3. Panelist Professional Background: Single-Grade Panels

Panel	ON	OFF	SED	SPE	COA	HED	OTH	RET	NR
Science 5	7	4	0	1	0	0	4	0	0
Science 8	11	4	1	0	0	0	0	0	1
Biology	17	0	0	1	0	1	0	1	0
English II	11	2	1	0	0	2	1	0	0
Mathematics I	15	2	0	1	1	0	1	0	1

Note: ON = on-grade, OFF = off-grade, SED = special education, SPE = specialist, COA = coach, HED = higher education, OTH = other, RET = retired, NR = no response.

Table 4 contains a summary of panelists' gender and ethnicity, and Table 5 summarizes panelists' geographic regions within the state. As these two tables illustrate, panels generally were representatively diverse in terms of gender, ethnicity, and geographic region.

Table 4. Panelist Gender and Ethnicity

	Gender			Ethnicity						
Panel	F	M	NR	AA	AS	HI	NA	WH	MU	NR
Mathematics 3-5	18	2	0	7	0	0	0	12	0	1
Mathematics 6-8	11	5	0	3	0	1	0	12	0	0
Reading 3-5	17	1	0	7	1	1	1	6	2	0
Reading 6-8	18	1	0	4	0	0	1	14	0	0
Science 5	16	0	0	4	0	0	0	12	0	0
Science 8	13	4	0	0	1	1	1	13	1	0
Biology	17	3	0	1	0	1	0	18	0	0
English II	14	3	0	1	0	2	0	14	0	0
Mathematics I	20	1	0	3	0	1	0	17	0	0

Note: F = female, M = male, NR = no response, AA = African American, AS = Asian, HI = Hispanic, NA = Native American, WH = white, MU = multiple responses.

Table 5. Panelist Geographic Region

Panel	C	NC	NE	NW	SC	SE	SW	W	MU	NR
Mathematics 3-5	4	1	0	1	4	4	5	1	0	0
Mathematics 6-8	1	2	1	1	2	3	4	2	0	0
Reading 3-5	2	1	1	0	4	3	4	2	0	1
Reading 6-8	0	1	1	4	2	5	5	0	1	0
Science 5	4	2	0	0	2	1	5	2	0	0
Science 8	5	1	1	1	4	2	2	1	0	0
Biology	3	4	1	3	1	2	5	0	0	1
English II	4	0	1	3	4	2	2	1	0	0
Mathematics I	6	2	0	3	4	0	6	0	0	0

Note: C = central, NC = north central, NE = northeastern, NW = northwestern, SC = south central, SE = southeastern, SW = southwestern, W = western, NR = no response.

Method and Procedure

A total of nine panels set standards for 17 grades and subjects. Panels were divided into two groups. Panelists setting standards for Mathematics or Reading for grades 3-8 each worked on three adjacent grade levels (3-5 or 6-8). These panels are referred to in this executive summary as three-grade panels. For the remaining grades and subjects—Grades 5 and 8 Science, Biology, English II, and Mathematics I—panelists set standards for a single grade/subject. These are referred to as single-grade panels. Although all nine panels used a similar methodology for panelists to render their judgments, the scope of activities varied across panel types. The three-grade panels convened between July 22-26, 2013, while the single-grade panels convened between July 24-25, 2013. The agenda for the single-grade panels is provided in Appendix A, and the agenda for the three-grade panels is provided in Appendix B.

On the morning of Monday, July 22, prior to the standard setting workshop, training was held for table leaders for the three-grade panels. For the single-grade panels, table leader training was held during the morning of Wednesday, July 24. During this training session, table leaders were introduced to the standard setting facilitators, trained on their role in the standard setting process, and received a general introduction and instruction on the item mapping process. Following table leader training, representatives of the North Carolina Department of Public Instruction and Pearson presented an opening session to all panelists. The three-grade panel opening session occurred on July 22, and the single-grade opening session occurred on July 24. After the conclusion of the opening session, panelists dispersed to their breakout session meeting rooms. Each panel convened in a separate breakout session room to complete the required standard setting activities.

Following committee introductions, the three-grade panels spent the remainder of Monday, July 22 writing and discussing achievement level descriptors (ALDs), which serve as content-oriented statements describing expectations of student performance at each achievement level, for the three grade levels assigned to their panels. For the single-grade panels, a portion of July 24 was devoted to ALD writing for their single assigned assessment, and then the single-grade panels moved on to other standard setting activities that day.

Following ALD writing activities, panelists performed tasks to set standards for their assigned subject area and grade(s). Panelists began by writing “just barely” level descriptors: statements describing performance expectations for students who are *just barely* at the three cut points separating the four achievement levels. Next, panelists reviewed the ordered item book (OIB), which contains items from the previous administration’s assessment as well as additional supplemental items selected from the item

pool, ordered in ascending empirical difficulty as estimated from actual student performance, and presented such that each page of the booklet contains a single item.

The item mapping procedure (Lewis, Green, Mitzel, Baum, & Patz, 1998; Mitzel, Lewis, Patz, & Green, 2001) is the judgmental process that was used in this standard setting. According to this procedure, panelists are asked to identify the item in the ordered item book that is the last item that a student who is just barely at a given achievement level should be able to answer correctly more often than not. The locations for the items in the ordered item book were established using a guess-adjusted response probability of two-thirds (or 2/3), representing the point on the item characteristic curve at which the probability of a correct response is two-thirds of the way between the curve's lower asymptote and 1.0. Following item mapping methodology training by a Pearson breakout session facilitator and a practice round of judgment, panelists began the standard setting process. For the three-grade panels, standard setting activities began at the lower grade (i.e., grade 3 for the panels assigned to grades 3-5, grade 6 for panels assigned to grades 6-8). Panelists set three recommended cut scores, which separate student performance into four distinct achievement level categories.

The standard setting process consisted of three rounds of judgment. Panelists were provided with feedback data, which was intended to inform panelists' decisions, to consider and discuss between each round. Following Round 1, panelists broke up into small groups of 5 to 7 and discussed their cut scores and associated rationales. Following small-group discussions, the entire panel shared their cut scores. For both discussions, panelists were instructed that reaching consensus was not the goal of these discussions, but rather, they should share their perspectives that led to their chosen cut scores.

In addition to the Round 1 cut score agreement data, panelists were shown external data to further inform their judgments in subsequent rounds of judgment. Panelists were provided with empirical item difficulty data showing the proportion of all test-takers from the spring 2013 administration who correctly answered each item (i.e., item *p*-values). The standard setting facilitator also shared with panelists the ACT Explore® cut score, which was linked to the North Carolina assessment by NCDPI, representing the score point at which students are on-track to be college and career-ready. Finally, the facilitator shared with panelists the expected cut scores obtained by NCDPI from a recent survey of North Carolina educators. Following discussion of Round 1 cut scores and the provided feedback data, panelists proceeded to the second round of judgment.

Following Round 2, panelists received updated cut score agreement data and engaged in discussions in both small groups and across the entire panel. Additionally, panelists were shown a graphical display of student impact data. The impact data displayed the percentages of spring 2013 test-takers who would be classified into the four achievement levels based on the panel's median cut score recommendation. Impact was shown for the overall North Carolina test-taking population, and impact was also broken down by gender and ethnicity subgroups. Panelists were given an opportunity to discuss the appropriateness of their cut scores given the current impact data. Following discussion of the Round 2 feedback data, panelists proceeded to the third and final round of judgment.

Following Round 3, panelists were shown their final recommended cut scores, which were based on their median cut score judgments from this final round of ratings. Panelists were shown impact data, again illustrating overall impact as well as impact broken down by gender and ethnicity. After reviewing and discussing the Round 3 impact data, panelists completed an evaluation survey capturing their reactions to the final cut score

recommendations and associated impact data. The results of the evaluation survey are documented in the full standard setting technical report.

The standard setting workshop activities concluded at this point for the single-grade committees. For the three-grade committees, the breakout session facilitator guided panelists through the same process for the middle and upper grades. Following the conclusion of standard setting activities, all panelists were dismissed with the exception of table leaders, who attended the vertical articulation session on Friday, July 26.

Table leaders from each committee convened in a single room to participate in the vertical articulation session. During this session, impact data were compared across grade levels within subject areas (e.g., Grades 3-8 Reading) and also across subjects. Panelists were asked to evaluate, from a policy perspective, the reasonableness of the committees' content-oriented cut score recommendations and the impact of imposing these achievement expectations on student test scores. Panelists were guided through a process whereby they evaluated the reasonableness of impact for particular grades/subjects, both in isolation and in contrast to other grades and subject areas. Table leaders from each committee were present in the vertical articulation meeting, which allowed them an opportunity to share with the entire group their recollection of the process and discussions that occurred within their committees. Following group discussion, each participant on the vertical articulation panel considered the recommended cut scores and their impact data as well as other potential cut scores and the changes in impact data associated with other potential cut scores. Each member of the vertical articulation committee provided a unique recommendation to keep or change the final cut scores. Prior to rendering judgments, the lead facilitator impressed upon the vertical articulation panel that their holistic, policy-oriented cut score recommendations would supplement, not overwrite, the content-oriented cut recommendations provided by the standard setting panels and would provide the North Carolina State Board of Education with additional information to consider when deciding which cut scores to adopt.

Results

The standard setting panels' final recommended cut scores, obtained prior to the vertical articulation session, are presented in Table 6. The reader should note that these cut scores are reported as page numbers within the ordered item book, not raw scores. NCDPI will translate these page cuts into the final reporting scale in a future study. The figures following Table 6 display impact data for the Mathematics, Reading, Science, and End-of-Course, respectively, based upon these cut score recommendations.

Table 6. Pre-Vertical Articulation Page Cuts

Assessment	Level 2	Level 3	Level 4
Mathematics 3	16	41	69
Mathematics 4	15	34	70
Mathematics 5	9	33	65
Mathematics 6	10	32	67
Mathematics 7	9	28	59
Mathematics 8	10	30	70
Reading 3	26	55	74
Reading 4	25	58	75
Reading 5	23	55	71
Reading 6	15	46	69
Reading 7	15	45	70
Reading 8	16	42	70
Science 5	12	45	69
Science 8	6	20	64
Biology	20	47	68
English II	9	34	79
Math I	9	29	60

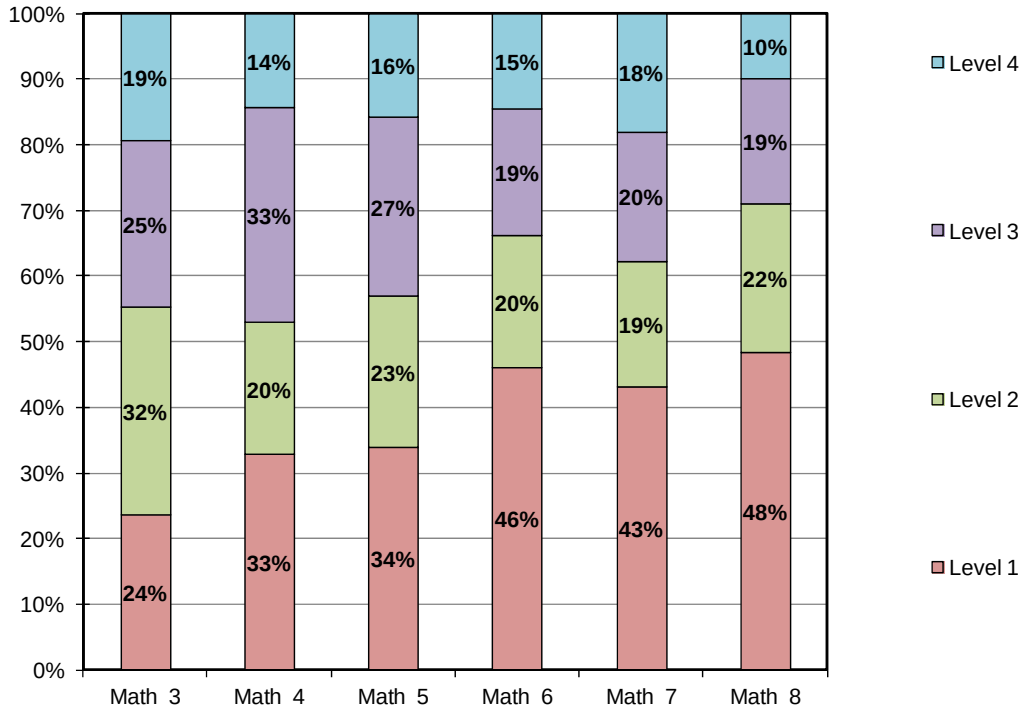

Figure 1. Pre-Vertical Articulation Impact Data: Mathematics 3-8

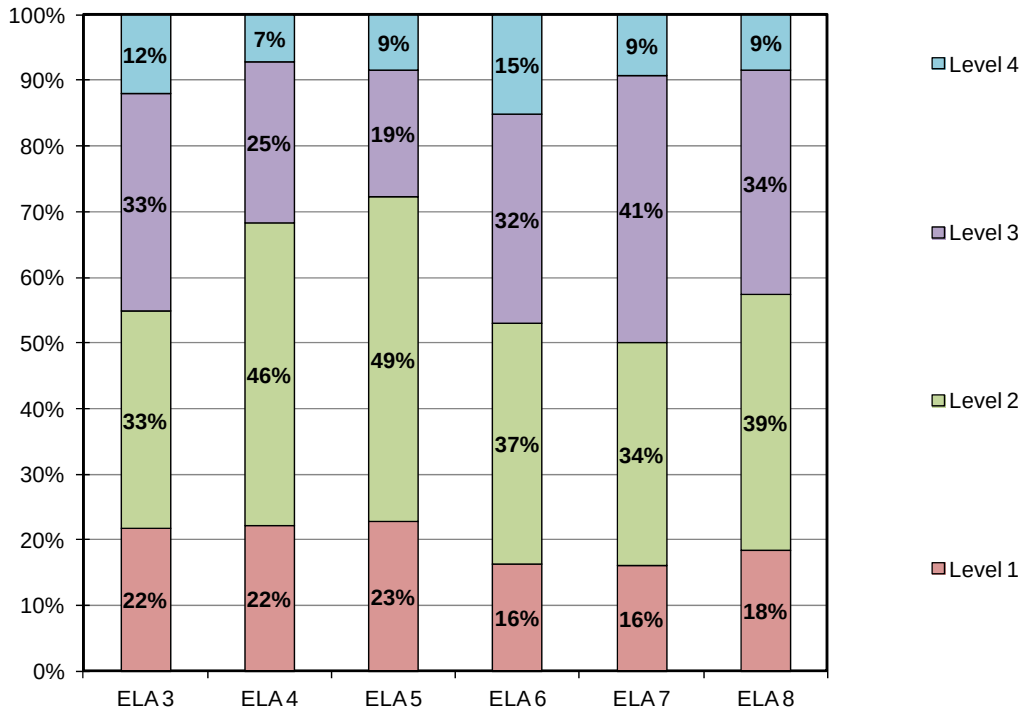

Figure 2. Pre-Vertical Articulation Impact Data: Reading 3-8

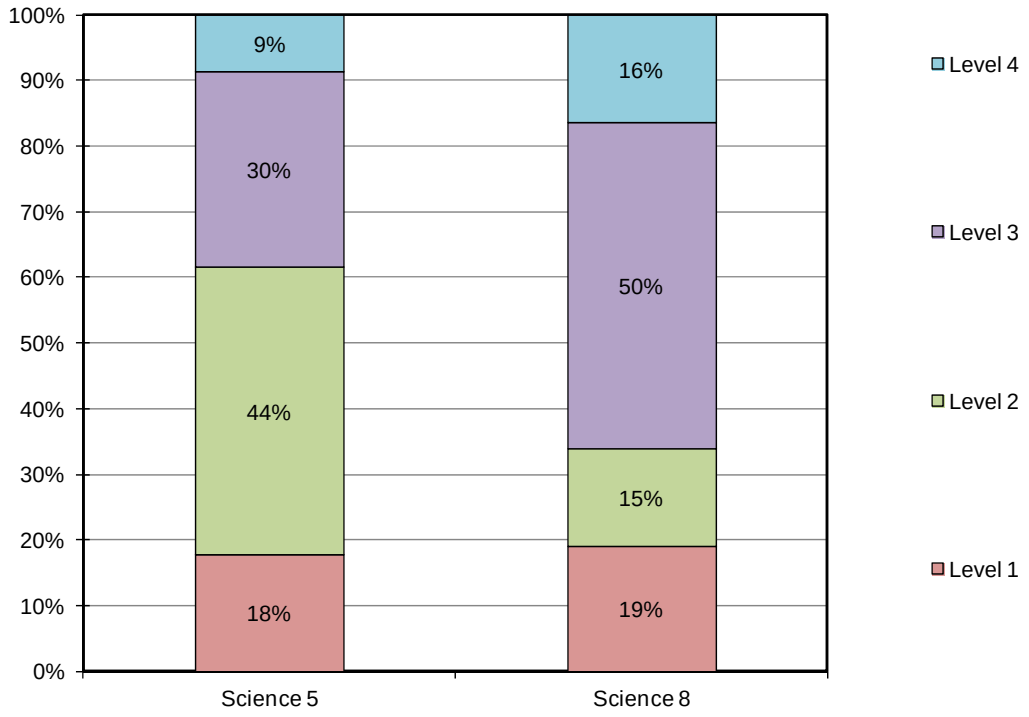

Figure 3. Pre-Vertical Articulation Impact Data: Science 5 and 8

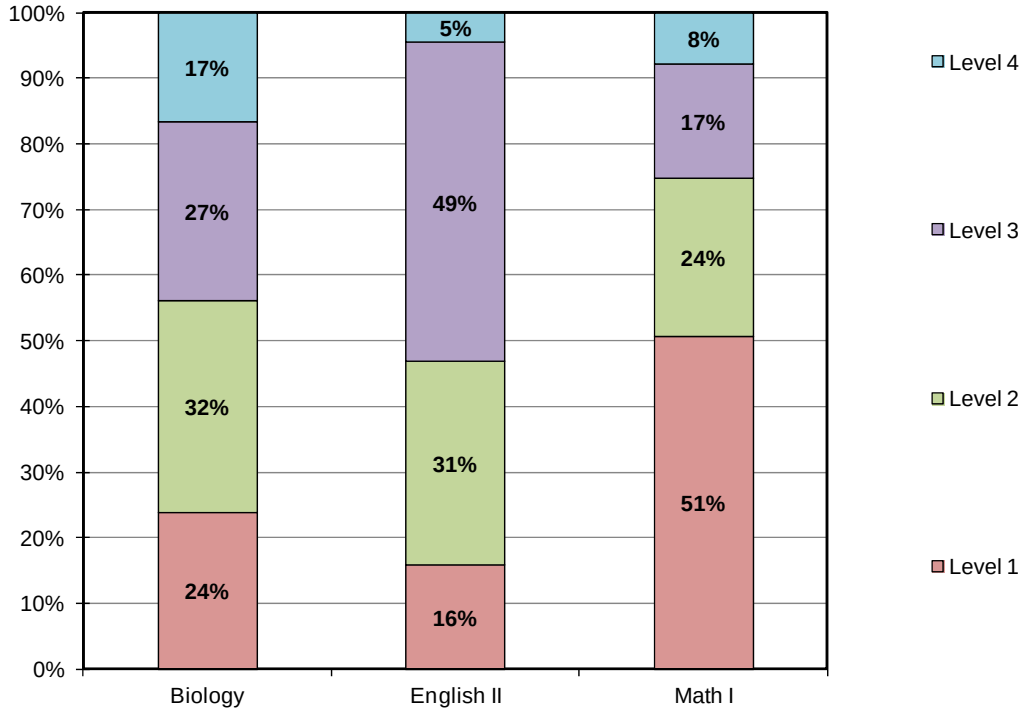

Figure 4. Pre-Vertical Articulation Impact Data: EOC

Cut scores obtained following the vertical articulation session are shown in Table 7, and impact data associated with these recommended cut scores are displayed in the subsequent figures.

Table 7. Post-Vertical Articulation Page Cuts

Assessment	Level 2	Level 3	Level 4
Mathematics 3	16	38	73
Mathematics 4	10	34	70
Mathematics 5	7	30	65
Mathematics 6	4	24	67
Mathematics 7	6	28	65
Mathematics 8	5	25	70
Reading 3	26	55	74
Reading 4	25	50	75
Reading 5	23	46	71
Reading 6	15	46	73
Reading 7	15	47	70
Reading 8	16	42	70
Science 5	12	40	69
Science 8	6	25	64
Biology	20	47	71
English II	9	36	79
Math I	2	20	60

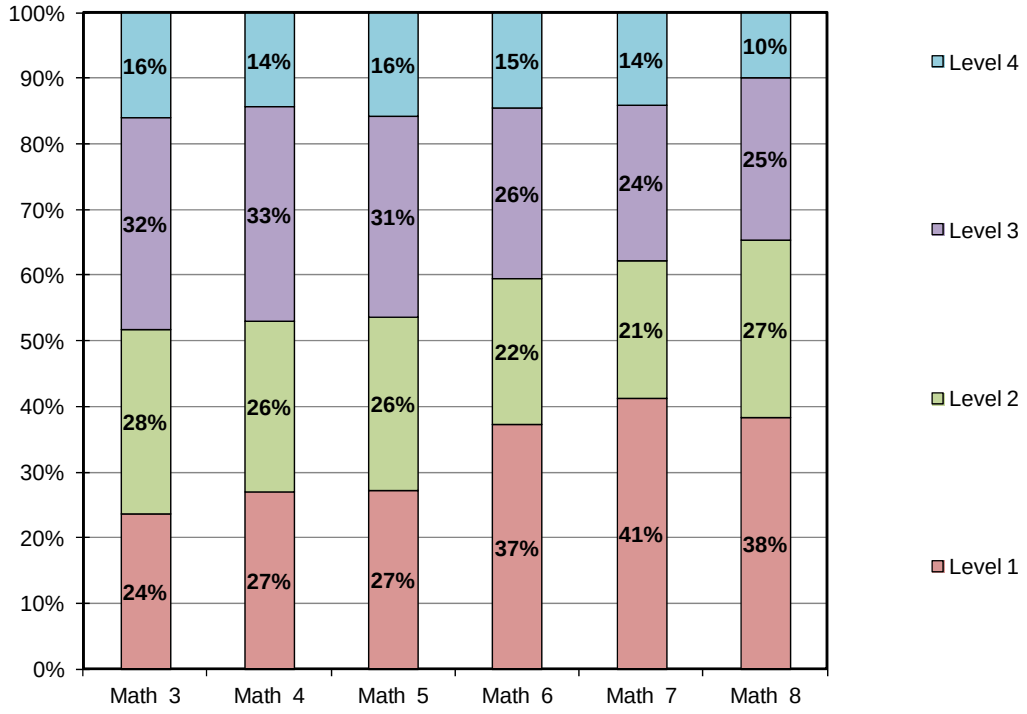

Figure 5. Post-Vertical Articulation Impact Data: Mathematics 3-8

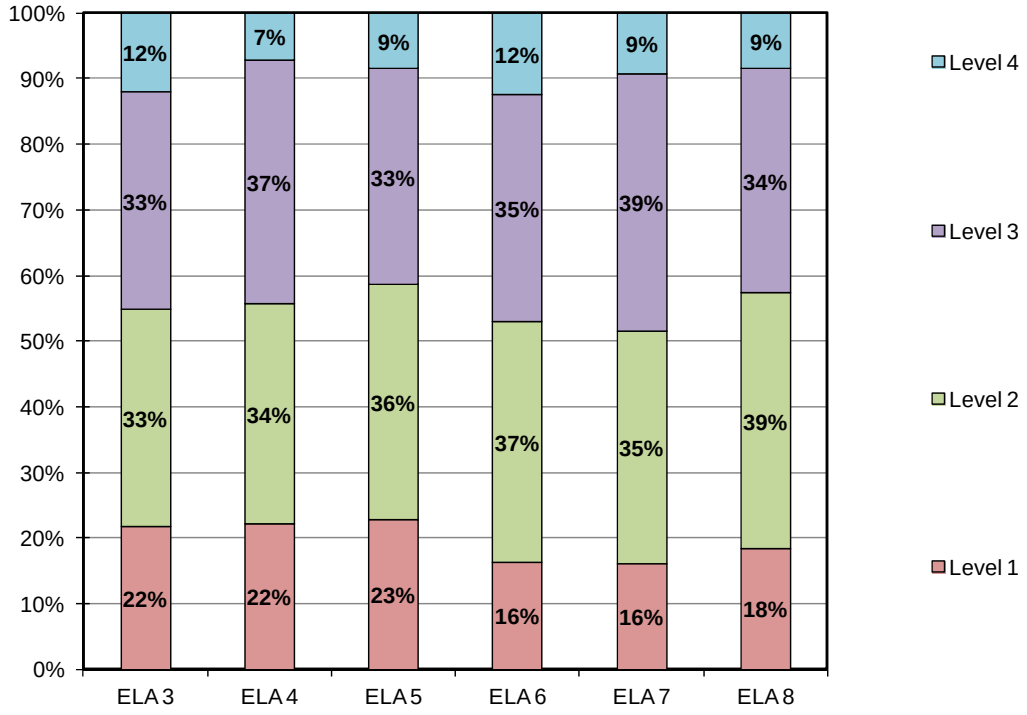

Figure 6. Post -Vertical Articulation Impact Data: Reading 3-8

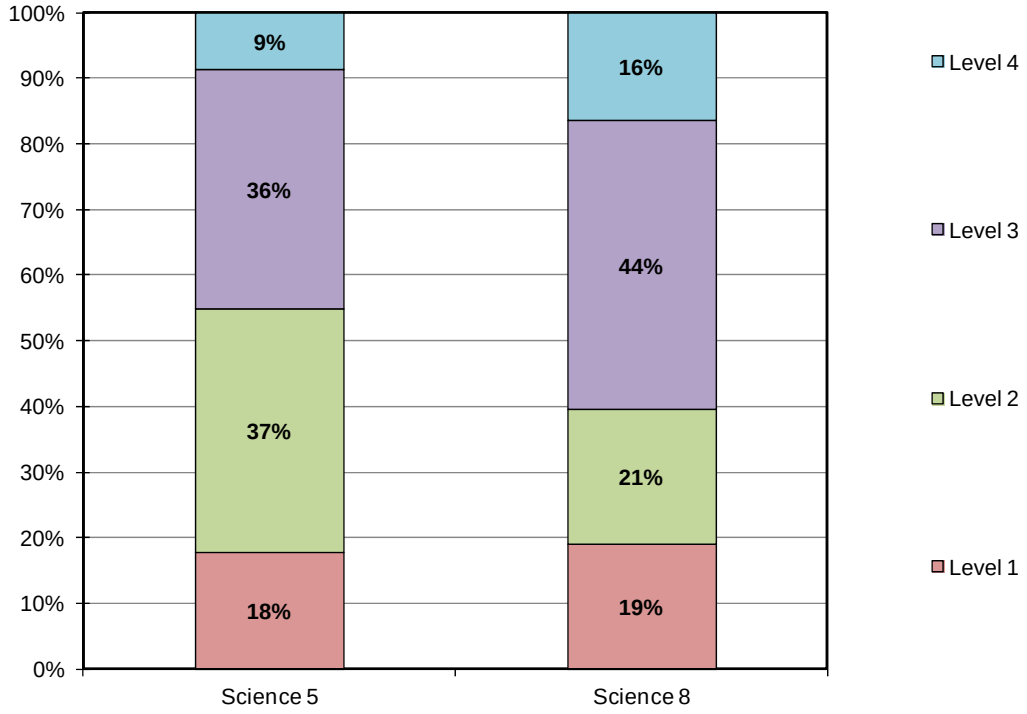

Figure 7. Post -Vertical Articulation Impact Data: Science 5 and 8

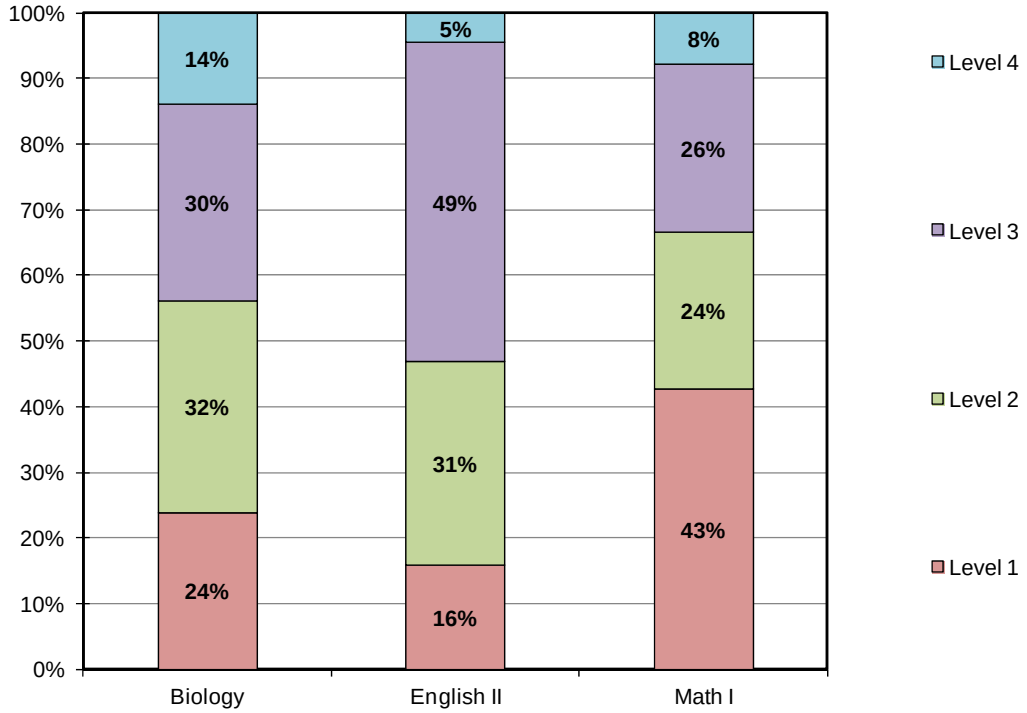

Figure 8. Post -Vertical Articulation Impact Data: EOC

References

Lewis, D. M., Green, D. R., Mitzel, H.C., Baum, K. & Patz, R.J. (1998). The Bookmark standard setting procedure: Methodology and recent implementations. Paper presented at the annual meeting of the National Council on Measurement in Education. San Diego, CA.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The Bookmark procedure: Psychological perspectives. In G. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Erlbaum.

North Carolina Testing Program
EOC/EOG 2013 Standard Setting

Agenda: Single-Grade Panels

Day 1: Wednesday, July 24

Activity	Time
Table leader training (*Table leaders only*)	8:00 – 8:45 AM
Large group kick-off meeting	9:00 – 9:30 AM
Break	9:30 – 9:45 AM
Committee introductions	9:45 – 10:00 AM
Achievement level descriptor revision training	10:00 – 10:15 AM
Achievement level descriptor revisions	10:15 AM – 12:15 PM
Lunch	12:15 – 1:00 PM
"Just barely" level descriptions	1:00 – 2:15 PM
Ordered item booklet review	2:15 – 3:15 PM
Break	3:15 – 3:30 PM
Standard setting training and practice round	3:30 – 4:15 PM
Round 1	4:15 – 5:30 PM

Day 2: Thursday, July 25

Activity	Time
Round 1 feedback and discussion	8:00 – 9:15 AM
Round 2	9:15 – 10:15 AM
Break	10:15 – 10:30 AM
Write behavioral descriptions	10:30 – 11:15 AM
Round 2 feedback and discussion	11:15 AM – 12:15 PM
Lunch	12:15 – 1:00 PM
Round 3	1:00 – 1:30 PM
Break/Collect secure materials	1:30 – 2:30 PM
Round 3 feedback and discussion	2:30 – 3:00 PM
Wrap-up and evaluations	3:00 – 3:15 PM

Day 3: Friday, July 26

Activity	Time
Vertical articulation (*Table leaders only*)	1:00 – 3:30 PM

North Carolina Testing Program
EOC/EOG 2013 Standard Setting

Agenda: Three-Grade Panels

Day 1: Monday, July 22

Activity	Time
Table leader training (*Table leaders only*)	8:00 – 8:45 AM
Large group kick-off meeting	9:00 – 9:30 AM
Break	9:30 – 9:45 AM
Committee introductions	9:45 – 10:00 AM
Achievement level descriptor revision training	10:00 – 10:15 AM
Achievement level descriptor revisions – LOWER GRADE	10:15 AM – 12:15 PM
Lunch	12:15 – 1:15 PM
Achievement level descriptor revisions – MIDDLE GRADE	1:15 – 3:15 PM
Break	3:15 – 3:30 PM
Achievement level descriptor revisions – UPPER GRADE	3:30 – 5:30 PM
NCDPI/Pearson debrief meeting	5:45 – 6:15 PM

Day 2: Tuesday, July 23

Activity	Time
"Just barely" level descriptions – LOWER GRADE	8:00 – 9:15 AM
Ordered item booklet review – LOWER GRADE	9:15 – 10:15 AM
Break	10:15 – 10:30 AM
Standard setting training and practice round	10:30 – 11:15 AM
Round 1 – LOWER GRADE	11:15 AM – 12:30 PM
Lunch	12:30 PM – 1:30 PM
Write behavioral descriptions – LOWER GRADE	1:30 – 2:15 PM
Round 1 feedback and discussion – LOWER GRADE	2:15 – 3:30 PM
Break	3:30 – 3:45 PM
Round 2 – LOWER GRADE	3:45 – 4:45 PM
NCDPI/Pearson debrief meeting	5:00 – 5:30 PM

Day 3: Wednesday, July 24

Activity	Time
Round 2 feedback and discussion – LOWER GRADE	8:00 – 9:00 AM
Round 3 – LOWER GRADE	9:00 – 9:30 AM
"Just barely" level descriptions – MIDDLE GRADE	9:30 – 10:45 AM
Round 3 feedback and discussion – LOWER GRADE	10:45 – 11:15 AM
Ordered item booklet review – MIDDLE GRADE	11:15 AM – 12:15 PM
Lunch	12:15 – 1:00 PM
Round 1 – MIDDLE GRADE	1:00 – 2:15 PM
Write behavioral descriptions – MIDDLE GRADE	2:15 – 3:00 PM
Break	3:00 – 3:15 PM
Round 1 feedback and discussion – MIDDLE GRADE	3:15 – 4:30 PM
Round 2 – MIDDLE GRADE	4:30 – 5:30 PM
NCDPI/Pearson debrief meeting	5:45 – 6:15 PM

Appendix B

Day 4: Thursday, July 25

Activity	Time
Round 2 feedback and discussion – MIDDLE GRADE	8:00 – 9:00 AM
Round 3 – MIDDLE GRADE	9:00 – 9:30 AM
"Just barely" level descriptions – UPPER GRADE	9:30 – 10:45 AM
Round 3 feedback and discussion – MIDDLE GRADE	10:45 – 11:15 AM
Ordered item booklet review – UPPER GRADE	11:15 AM – 12:15 PM
Lunch	12:15 – 1:00 PM
Round 1 – UPPER GRADE	1:00 – 2:15 PM
Write behavioral descriptions – UPPER GRADE	2:15 – 3:00 PM
Break	3:00 – 3:15 PM
Round 1 feedback and discussion – UPPER GRADE	3:15 – 4:30 PM
Round 2 – UPPER GRADE	4:30 – 5:30 PM
NCDPI/Pearson debrief meeting	5:45 – 6:15 PM

Day 5: Friday, July 26

Activity	Time
Round 2 feedback and discussion – UPPER GRADE	8:00 – 9:00 AM
Round 3 – UPPER GRADE	9:00 – 9:30 AM
Break/Collect secure materials	9:30 – 10:30 AM
Round 3 feedback and discussion – UPPER GRADE	10:30 – 11:00 AM
Wrap-up and evaluations	11:00 – 11:15 AM
Lunch	11:15 AM – 1:00 PM
Vertical articulation (*Table leaders only*)	1:00 – 3:30 PM