

2023 ELA and Mathematics NSCAS Growth Standard Setting Report for the Nebraska Department of Education

Submitted By: sdavisbecker@acsventures.com

Contents

1.	Study Overview and Preparation	3
	 Assessments Methodology Project Participants and Staff Security 	3 4 6 7
2.	Standard Setting Process	8
	 2.1. General Orientation 2.2. Panel Orientation 2.3. Panel Training 2.4. Operational Standard Setting 2.5. Articulation 2.6. Reporting 	
3.	Results	14
	3.1. Cut Score Recommendations	14 26
4.	References	31
5.	Appendices	



1. Study Overview and Preparation

In 2023, NWEA contacted with ACS Ventures, LLC (ACS) to conduct a standard setting for the Nebraska Student-Centered Assessment System (NSCAS) for grades 3 to 8 in mathematics and English Language Arts (ELA).

The plan for the standard setting activities was developed through collaborative efforts between ACS, NWEA, the statewide assessment team at the Nebraska Department of Education (NDE), and the Nebraska Technical Advisory Committee (TAC).

ACS presented an overview of the standard setting plan to the Nebraska Department of Education (NDE) and This report documents the preparation for, execution of, and results from the standard setting activities.

1.1. NSCAS Assessments

The NSCAS in ELA and mathematics are designed to measure and report Nebraska students' depth of achievement regarding the Nebraska College and Career Ready Standards (NCCRS) across three defined achievement levels: *Developing*, *On-Track*, and *Advanced*. (see NDE, 2022)¹.

- *Developing* Developing learners do not yet demonstrate proficiency in the knowledge and skills necessary at this grade level, as specified in the assessed Nebraska College and Career Ready Standards. These results provide evidence that the student may need additional support for academic success at the next grade level.
- On Track On Track learners demonstrate proficiency in the knowledge and skills necessary at this grade level, as specified in the assessed Nebraska College and Career Ready Standards. These results provide evidence that the student will likely be ready for academic success at the next grade level.
- Advanced Advanced learners demonstrate advanced proficiency in the knowledge and skills necessary at this grade level, as specified in the assessed Nebraska College and Career Ready Standards. These results provide evidence that the student will likely be ready for academic success at the next grade level.

Each NCCRS is further articulated by achievement level (i.e., range achievement level descriptors; RALDs) – see Appendix A. These assessments are administered online using an adaptive design and student final scores are reported as composite scale scores along with the associated achievement levels.

NSCAS assessments in ELA and mathematics consist of operational (OP), diagnostic operational (DO), and field test (FT) items from a variety of item types, such as multiple-choice, multiselect, and technology-enhanced items (i.e., hot text, text entry, composite, drag & drop, gap match, and graphic gap match). Table 1 shows a summary of the items included on the assessments for each grade level in ELA and mathematics and a description of each item type is shown Table 2.

Table 1. NSCAS Growth Assessments in ELA and Mathematics

NSCAS Growth Spring 2023 Adaptive Form								
Content / Total FT DO #Items for NSCAS #Items for NSCAS RIT Sc Grade Items Items Items Scale Score RIT Score Points Points						RIT Score Points		
ELA	ELA							

¹ 2022-NSCAS-Growth-Technical-Report.pdf (ne.gov)



Grade 3-8	de 3-8 45 7 8-11 27-30 31-32				31-32	27-38	31-42		
Math									
Grade 3-8	45	45 1 17 27		44	31-35	48-52			

Table 2. Description of Item Types

	Online Item Types						
Item Type	Description						
Multiple-Choice (Choice)	Students select one response from multiple options.						
Multiselect	Students select two or more responses from multiple options. Some						
(Choice Multiple)	multiselect items are also two-point items for which students can earn partial credit.						
Hot Text	Students select a response from within a piece of text or a table of						
	information (e.g., word, section of a passage, number, symbol, or						
	equation), which highlights the selected text. Some hot text items are						
	also two-point items for which students can earn partial credit.						
Text Entry	Students input answers using a keyboard.						
Composite	Students interact with multiple interaction types included within a single item. Students may receive partial credit for composite items.						
Drag & Drop	Students select an option or options in an area called the toolbar and move or "drag" these options (e.g., words, phrases, symbols, numbers, or graphic elements) to designated containers on the screen. Drag-and- drop items can include a click and click functionality in which students select the option and select the container it goes into instead of physically dragging it.						
Gap Match	A type of drag-and-drop item in which students select one or more answer options from the item toolbox and populate a defined area, or "gap."						
Graphic Gap Match	A type of drag-and-drop item in which students move one or more answer options from the toolbox and populate a defined area, or "gap," that has been embedded within an image in the item response area.						

1.2. Methodology

ACS used the procedures described within the Bookmark Method of standard setting (Lewis et al, 2012) to guide panels of Nebraska educators through the process of recommending two cut scores to be used to distinguish the three achievement levels (i.e., *Developing*, *On-Track*, *Advanced*):

- 1. The cut score that differentiates *On-Track* performance from *Developing* performance
- 2. The cut score that differentiates Advanced performance from On-Track performance

A key feature of the Bookmark method is presentation of the assessment items in an ordered item booklet (OIB). Specifically, the assessment content is presented in order of difficulty starting with the easiest item and progressing through more and more difficult items. Expert panelists are instructed to review the OIB and identify the expected level of performance for a student who is just barely within the *On-Track* achievement level and then the expected level of performance for a student who is just barely within the *Advanced* achievement level.



ACS collaborated with NWEA to prepare the OIB as follows:

- 1. NWEA analyzed the spring 2023 NSCAS results and estimate item parameters (theta scale) for each item/score point (i.e., if an item has possible scores beyond 0 and 1, parameters were provided for each score point).
- 2. NWEA provided ACS with information about each item including difficulty parameters from step 1, item type, item content alignment. In addition, NWEA provided information about student performance on the theta scale to calculate impact values.
- 3. ACS estimated the RP67 values for each item/score point.
- 4. ACS ordered the full set of items based on the RP67 values and selected a sample of items/score points for the OIB that resembled a test form based on content representation, item type, and difficulty range. ACS attempted to keep the OIBs in close resemblance to the operational test experience, but some OIBs deviated from the length of the operational test forms to ensure there was adequate content representation. Any included items that have scores beyond 0 and 1 were represented in the OIB multiple times.
- 5. ACS calculated the impact (% of students scoring at/above) each location within the OIB based on the associated theta value.
- 6. ACS worked with NWEA to validate the results of steps 3, 4, and 5.
- 7. From those items NOT selected for the standard setting OIB, ACS selected a small sample (~ 10) to construct a practice OIB, which panelists used to practice evaluating items.

Since the spring 2023 assessments contained both dichotomous (i.e., multiple-choice) and polytomous (i.e., constructed response) items, the OIB was created to include both item types. Dichotomous items (i.e., score point of 0 and 1) appeared only once in the OIB. Polytomous items (i.e., maximum score points of 2 or higher) appeared multiple times in the OIB, once for each score point. For these items, the item prompt and scoring rubric for the particular score point was provided to panelists. ACS used the *NSCAS Table of Specifications* to set the targeted distribution of assessment content within the OIBs. These targeted distributions are shown for ELA and mathematics in Table 3 and Table 4 below.

	ELA Table of Specification Target Content Distributions							
Grade	Reading Prose and Poetry	Reading Informational Text	Vocabulary	Writing				
Grade 3	28% - 33%	28% - 33%	15% - 20%	23% - 28%				
Grade 4	28% - 33%	28% - 33%	15% - 20%	23% - 28%				
Grade 5	28% - 33%	28% - 33%	15% - 20%	23% - 28%				
Grade 6	28% - 33%	28% - 33%	15% - 20%	23% - 28%				
Grade 7	25% - 30%	30% - 35%	15% - 20%	23% - 28%				
Grade 8	25% - 30%	30% - 35%	15% - 20%	23% - 28%				

Table 3. Target Content Distributions: ELA

Table 4. Target Content Distributions: Math

Math Table of Specification Target Content Distributions							
Grade	Number	Algebra	Geometry	Data			
Grade 3	35% - 45%	15% - 20%	20% - 30%	15% - 20%			
Grade 4	35% - 45%	20% - 30%	15% - 25%	10% - 20%			
Grade 5	35% - 45%	20% - 30%	15% - 25%	10% - 20%			
Grade 6	25% - 35%	30% - 40%	15% - 25%	10% - 20%			



Grade 7	15% - 25%	30% - 40%	15% - 25%	20% - 30%
Grade 8	20% - 30%	25% - 35%	25% - 35%	10% - 20%

1.3. Project Participants and Staff

1.3.1. ACS Staff

ACS staff led the preparation activities for the standard setting, led the facilitation of the standard setting, and prepared the results in the form of presentations and technical reports.

ACS provided a total of six facilitators to lead each panel, as well as one lead facilitator to conduct the general sessions and serve as a floating facilitator during the study. The facilitators were experienced with the Bookmark methodology and with leading panelists through the standard setting process. Within each panel, the facilitator served the following functions:

- Guided the panelists through the standard setting process.
- Provided feedback and answered questions from panelists.
- Analyzed data at the end of each round to prepare for the next round.
- Provided feedback data to panelists and facilitated discussions after each round.

All facilitators used the same study resources to ensure all panels received the same instructions and followed the same processes.

1.3.2. NWEA Staff

NWEA staff were involved in the preparation and execution of the standard setting activities that included:

- Managing the logistics for the study (e.g., meeting space, technology).
- Providing ACS with information about each assessment (detailed in next section).
- Providing access for panelists to the online system for accessing the OIBs.
- Being available for technical assistance and content questions during the standard setting.

1.3.3. NDE Staff

NDE staff were involved in the preparation and execution of the standard setting activities that included:

- Recruitment and selection of panelists.
- Providing ACS with information about the assessment programs as needed.
- Answering policy questions related to the standard setting.

1.3.4. Nebraska Educators

Six standard setting panels were recruited by NWEA and NDE as shown in Table 5. Panelists were recruited by NDE to ensure that the sample represents the diversity of Nebraska educators and possess the content area expertise necessary to provide the cut score recommendations.

Panel #	Content Area	Grades	# Panelists	Facilitator
1	ELA	3 & 4	10	Sarah Ahlamadi
2	ELA	5&6	10	Melia Franklin
3	ELA	7&8	9	Jaime Kavanaugh
4	Math	3 & 4	7	Unber Ahmad

Table 5. Standard Setting Panels

5	Math	5&6	11	Ross Markle
6	Math	7&8	9	Kelley Stethen

Prior to the workshop, all panelists received a brief participant advance document. This document included a description of the goals and high-level agenda of the workshop, the role of the panel in the overall process of setting cut scores, and the expectations (time, engagement, security) for each panelist. Panelists were also asked to complete a demographic form in advance of the meeting. A full summary of the demographic information collected in this form can be found in Appendix B.

Within each panel, individual panelists were assigned to smaller groups of 3-4 panelists for focused discussions. Each small group included representatives from different parts of the state and one panelist from each group was asked to serve as the leader and take notes during the discussions.

1.4. Security

There were several necessary security measures involved in the study given the need to maintain the confidentiality of the test content, the standard setting discussions, and the standard setting results.

NWEA created non-disclosure agreements for panelists to sign before participating in the workshop. ACS facilitators continuously reminded panelists about the security policies throughout the meeting and emphasized that the security of testing materials should always be maintained.

Panelists were permitted to access test material on their own device but were reminded of the security requirements throughout the study and were excluded from using cell phones near the test materials. All paper notes and documentation were collected at the end of the study.



2. Standard Setting Process

The standard setting study occurred over three days (07/25/23 - 07/28/23). The primary standard setting activities (large group orientation and training, panel-specific training, iterative judgmental process) were conducted during the first two days and half of the third. A subset of each panel was then asked to participate in an articulation meeting on the third afternoon to review the results across grades and subject areas. Table 6 shows the detailed study agenda; each step of the process is described in the following sections.

Table 6. Study Agenda

Day/Time	Key Activities	Materials
Tuesday Ju	ly 25	
8-9	General Orientation	Training Presentation
9-11	 Panel Orientation (Break included) Review assessment (purpose, blueprint, and sample form) Review range ALDs and develop threshold ALDs Make Initial Estimates 	 Training Presentation Panelist Resource List Sample Assessment form (via CRT system) Range ALDs Threshold ALD template Initial Estimate Form
11-2	 Panel Training (Lunch break included) Review of Bookmark process Bookmark practice Evaluation #1 	 Training Presentation Panelist Resource List Threshold ALDs Practice OIB Practice Judgment Form Readiness Evaluation
2-4	 Lower Grade Level Operational Standard Setting Activities Round 1 judgments 	 Training Presentation Panelist Resource List Threshold ALDs Operational OIB Operational Form
Wednesda	y July 26	
8-1	Lower Grade Level Operational Standard Setting Activities (Break + Lunch included) • Round 1 feedback • Presentation of results • Discussion within table groups • Discussion across full panel • Round 2 instructions • Round 2 judgments • Round 2 feedback	 Training Presentation Panelist Resource List Threshold ALDs Operational OIB Operational Judgment Form Results Evaluation



1-4	 Discussion within table groups Discussion across full panel Round 3 instructions Round 3 judgments Evaluation #2 Upper Grade Operational Standard Setting Activities (Break included) Review assessment (purpose, blueprint, and sample form) Review range ALDs and develop threshold ALDs Round 1 judgments 	 Training Presentation Panelist Resource List Sample Assessment form (via CRT system) Range ALDs Threshold ALD template Threshold ALDs Initial Estimate Form
		 Operational OIB Operational Judgment Form
Thursday J	uly 27	
8-12	Upper Grade Operational Standard Setting Activities (Break included) Round 1 feedback	 Training Presentation Panelist Resource List Threshold ALDs Operational OIB Operational Judgment Form Results Evaluation
12-1	Lunch	
1-3	 Articulation Review Presentation of results Standards Articulation judgments Evaluation #4 	 Articulation Presentation Articulation Initial Judgments Form Articulation Evaluation

2.1. General Orientation

The first part of the meeting served as an introduction to the general standard setting process. It began with a large-group general session that included a welcome and introductions from NDE. Then, ACS lead facilitators



provided a high-level orientation and training on the standard setting process and methodology that was to be followed. The overview also included a brief review of the format of the assessments, the RALDs, and how the panelists were to make their judgments.

2.2. Panel Orientation

After the general orientation session, all panelists began their work within the grade-level panels. Each panel began with the lower of the two assigned grades (e.g., ELA grades 3 and 4 panel began with the grade 3 test form). All panel activities were facilitated by the ACS staff and NDE and NWEA staff were available to address questions that arose which were specific to content, administration, or assessment policy.

2.2.1. Review Assessment

Following the general orientation, the facilitator reviewed the assessment's purpose, format (e.g., item types), and blueprint. After this introduction, the panelists were instructed to review sample items from the assessment from the *NSCAS Item Type Samplers*. The purpose of this review was to understand a student's experience interacting with the assessment itself. This review was limited to 30 minutes so that the panelists did not focus on determining each correct answer but rather on getting a general sense of the assessment from the student experience.

2.2.2. Develop Threshold ALDs

Panelists were then asked to review the RALDs which describe the knowledge, skills, and ability (KSA) expectations for each achievement level that are tied to the grade-level content standards for an assessment. After the review, the panelists collaborated within their table groups to identify which KSAs they expected a student to perform if they were at the *threshold* of each achievement level. Each small group was assigned a specific domain to focus on when defining the threshold ALDs.

Once all domains were covered, the facilitator reviewed the results with the whole panel and guided them through a discussion focused on refining the document until a final consensus was reached. The outcome of this activity resulted in a working document that describes the expectations for students who are just at the transition point from one achievement level to the next. Copies of this document were made available to all panelists before beginning the standard setting process.

After the panelists discussed the expectations for students at each achievement level, they were instructed to make their initial estimates as to what percent of all Nebraska students at this grade level are at each achievement level. This information was collected for internal research purposes and the results were not presented to the panelists.

2.3. Panel Training

2.3.1. Bookmark Training

Next, the facilitator provided additional training on the Bookmark method. This training began with a conceptual review of how the panelists were to translate the expectations outlined in the threshold ALDs into progress within the OIB. The facilitators described the OIB and how it was created to consist of a set of items placed in order from least to most difficult using a response probability of two-thirds (0.67) and used data from the spring 2023 (and previous) assessments to determine the item level values. The presentation used for the general orientation, as well as the panel-level training can be found in Appendix C.



The panelists were instructed how to access the OIBs through the online NWEA assessment system and the facilitator reviewed how dichotomous and polytomous items were to be presented in the OIB, as well as how to access the scoring rubrics that were provided for reference.

Following the training, the panelists then had the opportunity to practice the Bookmark method by applying the description for the threshold *On-Track* student to a shorter practice OIB. Once the panelists had completed making their practice ratings, the facilitator led the entire panel in a discussion of the results and the panelists were allotted time to ask any questions that might have come up during the practice.

2.3.2. Readiness Evaluation

After panelists have completed the Bookmark method practice and felt ready to complete the operational judgments, they were asked to complete the *Readiness Evaluation* form which asked them to indicate how ready they feel to proceed with the operational standard setting judgments. Prior to moving on to the next activity, the facilitator responded to any remaining questions panelists might have had regarding the process. All (100%) of panelists indicated they were ready to begin providing operational judgments. A copy of this form can be found in Appendix C.

2.4. Operational Standard Setting

2.4.1. Lower Grade Level

Following the confirmation that all panelists understand the procedures and are prepared to make their operational judgments, panelists were instructed to begin reviewing the OIB for the lower grade level and making their Round 1 Bookmark judgments. The panelists were reminded that the process of making judgments is an individual activity, but that they would be provided ample time to discuss items after all Round 1 judgments are completed.

After Round 1, the panelists were provided feedback that included summary statistics of the panel recommendations and a graphical depiction of the individual recommendations within the panel. The panelists were first instructed to discuss their reaction to the feedback within their small groups and then asked to share their small group's key discussion points with the whole panel. Throughout the discussion, the panelists were prompted to consider whether they were grounding their Round 1 judgments in how they expect a student should be able to perform or in the expectations defined in the threshold ALDs of how a *threshold* student is likely to perform. Throughout the discussion, the facilitator displayed specific items and asked panelists to discuss how they reached a judgment using the expectations defined in the threshold ALDs. After the Round 1 discussion, the facilitator reviewed instructions for making the Bookmark judgments and instructed the panelists to consider their initial judgments, the Round 1 results, and the Round 1 discussions when making their Round 2 judgments.

Following Round 2, the panel was provided with the same type of feedback from Round 1 was provided and impact data or the percentage of students that would be classified into each achievement level using the median Round 2 recommendation. After reviewing the feedback, panelists once again discussed their reactions to the feedback in their table groups and then in the whole group setting. Panelists discussed whether they thought the presented was an accurate depiction of all Nebraska students.

Panelists were then given review of how to provide Round 3 judgments which were their final judgments for that grade level. Following the completion of Round 3, the results (same feedback data as what was provided in Round 2) were shared with the panelists for review.



After Round 3 judgments have been reviewed, panelists were asked to complete the *Recommendation Evaluation* form. This questionnaire was designed to collect information on how panelists felt about the overall recommendation, the threshold ALDs that were developed, and their understanding of the overall process. A copy of this form can be found in Appendix C.

2.4.2. Upper Grade Level

After completing the standard setting process for the lower grade level, the panelists continued with the process for the upper grade level assessment. The procedures for the upper grade level were the same as what was completed for the first with just a few minor differences such as:

- The panel training was reduced to a quick review of the purpose of the task and the process of making standard setting judgments.
- The panel did not have to practice their OIB judgments for the upper grade level given their familiarity with the process.
- When the Round 2 feedback was delivered, the panels had the opportunity to see the impact of their recommended cut scores alongside the 1) impact of the cut scores they set for the grade below, and 2) the impact for the cut scores set by the panel working on the grade above if available (i.e., the grade 3/4 panel were able to review the impact of their recommendations from grade 3 as well as what was recommended for grade 5 while working with grade 4).

2.5. Articulation

After the final round of standard setting, three representatives from each panel were invited to participate in a standards articulation process. During the articulation process, the panelists evaluated whether the cross-grade and cross-subject impact represented a reasonable and coherent set of results. The articulation process anchored on two underlying principles:

- Achievement level expectations should be coherent across grades and subjects. This does not mean they need to match or follow a specific pattern but rather they should be reasonable.
- The judgments of the standard setting panels should be honored, unless doing so would clearly violate the above principle.

The articulation began with the facilitator explaining the purpose and process of articulation to the panelists. After the orientation, panelists were asked to describe how they anticipate students would be classified across grades. The facilitator guided the discussion to help ensure that expectations were grounded in the RALDs, and in the knowledge and skills required to complete the test items. Panelists were then presented with a high-level summary of the impact of the panel-level recommendations following the Round 3 judgments for all grade levels in both ELA and mathematics. Panelists were allotted time to independently review the results and then were asked to respond to an *Articulation Initial Judgment Form* (Appendix C) to indicate which grade-level comparisons they were interested in discussing. Responses to this form were solely used to organize the discussion and all grades were discussed during the articulation.

During the discussion, panelists were asked to compare the results across grades and subject areas to evaluate the trends and make recommendations for any changes that seem appropriate. The facilitator guided the discussion by asking panelists to explain why they might have indicated they were interested in discussing a particular grade level. Representatives from that grade were then given the opportunity to describe the



ACS Ventures – Bridging Theory & Practice

conversations that occurred within the panel. Discussion topics varied slightly from grade to grade, but some common themes included:

- the level of agreement within the panel
- the novelty of the content being assessed
- the expected difficulty of the content being assessed

The primary question for the panelists to consider was whether the magnitude and pattern of the impact data match the magnitude of the shifts and expectations from a content/standards perspective.

During the session, panelists were able to make real-time recommendations for changes in the cut scores, and immediately see the revised impact numbers for their recommendations. As a reference point, the panelists were also provided with the 2022 results of the ACT assessment in ELA and mathematics at high school, as well as the 2022 NAEP results for grade 4 and grade 8 but were cautioned to consider this information carefully given the differences in the assessments.

2.6. Reporting

Immediately following the standard setting meeting, ACS presented the results to the AAAC who discussed the impact of the recommended cut scores and then made additional recommendations for final adjustments to improve coherence across the grades. The recommendations from the standard setting panels, the standards articulation panel, and the AAAC committee are included in the following section of this report.



3. Results

This section presents the resulting cut score recommendations that were generated from this standard setting study, as well as an evaluation of the standard setting activities and results according to the sources of evidence described within Kane's (1994; 2001) framework.

3.1. Cut Score Recommendations

3.1.1. Panel Results

Prior to making any operational judgments, the panelists were provided the opportunity to practice making cut score recommendations using the Bookmark methodology and then were asked to respond to a *Readiness Evaluation* form (Appendix C) to provide feedback on the training and to report whether they felt ready to proceed with the process. The results of the evaluations are shown in Table 7 and Table 8 below.

Readiness Evaluation	E	LA 3&4	El	A 5&6	ELA 7&8	
General Session	N	%	N	%	N	%
Very good	5	45.5%	9	90.0%	5	55.6%
Good	6	54.5%	1	10.0%	4	44.4%
Inadequate	0	0.0%	0	0.0%	0	0.0%
Developing Threshold ALDs	Ν	%	Ν	%	Ν	%
Very good	6	54.5%	8	80.0%	6	66.7%
Good	5	45.5%	2	20.0%	3	33.3%
Inadequate	0	0.0%	0	0.0%	0	0.0%
Bookmark Judgments		%	N	%	N	%
Very good	9	81.8%	9	90.0%	5	55.6%
Good	2	18.2%	1	10.0%	4	44.4%
Inadequate	0	0.0%	0	0.0%	0	0.0%
Practice Bookmark	N	%	N	%	N	%
Very good	9	81.8%	8	80.0%	7	77.8%
Good	2	18.2%	2	20.0%	2	22.2%
Inadequate	0	0.0%	0	0.0%	0	0.0%
Prepared for Judgements		%	N	%	N	%
Yes	11	100.0%	10	100.0%	9	100.0%
No	0	0.0%	0	0.0%	0	0.0%

Table 7. Readiness Evaluation Results: ELA

Table 8. Readiness Evaluation Results: Math

Readiness Evaluation		Math 3&4		Math 5&6		Math 7&8	
General Session		%	N	%	N	%	
Very good	8	72.7%	5	50.0%	6	66.7%	
Good	3	27.3%	5	50.0%	3	33.3%	
Inadequate	0	0.0%	0	0.0%	0	0.0%	
Developing Threshold ALDs	N	%	N	%	N	%	
Very good	8	72.7%	6	60.0%	7	77.8%	
Good	3	27.3%	4	40.0%	2	22.2%	





Inadequate	0	0.0%	0	0.0%	0	0.0%
Bookmark Judgments	N	%	N	%	N	%
Very good	9	81.8%	6	60.0%	6	66.7%
Good	2	18.2%	4	40.0%	3	33.3%
Inadequate	0	0.0%	0	0.0%	0	0.0%
Practice Bookmark		%	N	%	N	%
Very good	8	72.7%	6	60.0%	5	55.6%
Good	3	27.3%	4	40.0%	4	44.4%
Inadequate	0	0.0%	0	0.0%	0	0.0%
Prepared for Judgements		%	N	%	N	%
Yes	11	100.0%	10	100.0%	9	100.0%
No	0	0.0%	0	0.0%	0	0.0%

The results of the evaluation demonstrate that the panelists across all panels felt that the training and preparation was appropriate to prepare them for making judgements. Evidence to support this includes all panelists rating each training component as either "Very good" or "Good" and by all panelists responding "Yes" to the question asking if they felt prepared to make operational judgments.

The results of the panel-level cut score recommendations for all three rounds of the operational judgments are summarized in tables 9-12 below. The cut scores in the tables are expressed as median page numbers from the OIBs as well as on the theta scale (RP 67 values). These values indicate the minimum performance necessary for each achievement level (*On-Track, Advanced*).

ELA On Track Cut Score							
Grade / Round	N Ratings	Min (OIB Pg)	Max (OIB Pg)	Median (OIB Pg)	SE _{Median} (OIB Pg)	Median (Theta)	
Grade 3							
Round 1	9	5	15	10	1.74	-0.791	
Round 2	8	5	15	10	1.76	-0.791	
Round 3	8	5	15	10	1.59	-0.791	
Grade 4							
Round 1	8	8	23	15	2.34	-0.217	
Round 2	9	5	20	10	1.72	-0.688	
Round 3	9	7	14	10	1.05	-0.688	
Grade 5							
Round 1	9	9	35	25	3.58	0.608	
Round 2	9	18	25	19	0.86	0.250	
Round 3	9	17	19	19	0.37	0.250	
Grade 6							
Round 1	10	7	27	24	2.54	0.889	
Round 2	10	15	18	17	0.35	0.527	
Round 3	10	15	17	17	0.27	0.527	
Grade 7							

Table 9. ELA Cut Score Recommendations: On Track



Round 1	9	9	14	12	0.69	0.373
Round 2	9	9	13	13	0.64	0.425
Round 3	9	10	13	12	0.54	0.373
Grade 8						
Round 1	9	10	15	13	0.89	-0.057
Round 2	9	10	15	13	0.89	-0.057
Round 3	9	10	15	13	0.86	-0.057

Table 10. ELA Cut Score Recommendations: Advanced

ELA Advanced Cut Score								
Grade / Round	N Ratings	Min (OIB Pg)	Max (OIB Pg)	Median (OIB Pg)	SE _{Median} (OIB Pg)	Median (Theta)		
Grade 3								
Round 1	9	19	35	29	2.08	0.503		
Round 2	8	29	29	29	0.00	0.503		
Round 3	8	29	34	29	0.78	0.503		
Grade 4								
Round 1	8	25	36	32	1.77	1.166		
Round 2	9	24	34	29	1.45	0.820		
Round 3	9	27	34	29	1.00	0.820		
Grade 5								
Round 1	9	30	40	36	1.15	1.211		
Round 2	9	35	36	35	0.14	1.141		
Round 3	9	35	35	35	0.00	1.141		
Grade 6								
Round 1	10	32	37	36	0.80	1.719		
Round 2	10	32	35	35	0.38	1.572		
Round 3	10	34	37	35	0.36	1.572		
Grade 7								
Round 1	9	31	40	39	1.35	1.740		
Round 2	9	34	40	36	1.00	1.595		
Round 3	9	34	40	36	0.94	1.595		
Grade 8								
Round 1	9	31	40	36	1.34	1.431		
Round 2	9	33	38	38	0.71	1.541		
Round 3	9	33	38	38	0.71	1.541		

Table 11. Math Cut Score Recommendations: On Track

Mathematics On Track Cut Score							
Grade / N Min Max Median SE _{Median} Median							
Round	Ratings	(OIB Pg)	(OIB Pg)	(OIB Pg)	(OIB Pg)	(Theta)	
Grade 3							
Round 1	7	13	29	21	2.79	0.241	

Round 2	7	15	26	21	1.77	0.241
Round 3	7	10	16	14	1.00	-0.440
Grade 4						
Round 1	7	10	24	20	2.26	0.614
Round 2	6	15	20	16	0.99	0.274
Round 3	6	12	17	14	0.92	0.154
Grade 5						
Round 1	10	4	30	11	2.72	0.182
Round 2	10	10	12	10	0.27	0.136
Round 3	10	10	12	10	0.27	0.136
Grade 6						
Round 1	10	14	27	16	1.98	0.623
Round 2	9	15	17	15	0.41	0.507
Round 3	9	15	17	15	0.41	0.507
Grade 7						
Round 1	9	16	27	25	2.09	1.483
Round 2	9	11	27	21	1.90	1.134
Round 3	9	10	23	11	1.81	0.219
Grade 8						
Round 1	9	11	24	16	1.72	0.986
Round 2	8	12	18	14	0.95	0.769
Round 3	8	12	16	12	0.73	0.563

Table 12. Math Cut Score Recommendations: Advanced

Mathematics Advanced Cut Score							
Grade / Round	N Ratings	Min (OIB Pg)	Max (OIB Pg)	Median (OIB Pg)	SE _{Median} (OIB Pg)	Median (Theta)	
Grade 3							
Round 1	7	34	38	35	0.90	1.921	
Round 2	7	33	37	35	0.66	1.921	
Round 3	7	31	37	33	1.02	1.770	
Grade 4							
Round 1	7	31	40	38	6.86	2.386	
Round 2	6	27	36	36	1.81	1.944	
Round 3	6	31	40	38	1.65	2.386	
Grade 5							
Round 1	10	28	37	32	1.28	2.179	
Round 2	10	29	32	32	0.46	2.179	
Round 3	10	29	32	32	0.46	2.179	
Grade 6							
Round 1	10	25	39	32	1.84	2.031	
Round 2	9	31	35	35	0.54	2.200	
Round 3	9	31	35	35	0.54	2.200	



Grade 7						
Round 1	9	35	42	38	0.75	2.842
Round 2	9	25	38	38	1.76	2.842
Round 3	9	25	38	33	2.11	2.084
Grade 8						
Round 1	9	28	42	33	2.14	2.273
Round 2	8	26	33	29	1.06	2.006
Round 3	8	28	34	29	0.85	2.006

The median thetas (RP 67 values) were used to calculate the impact of the recommended cut scores and were shown to panelists as a part of the feedback data for Round 2 and Round 3 of the standard setting activities. The Round 2 and Round 3 impact percentages are shown in Table 13 and Table 14 below.

	ELA Cut Sco	re Impact	
Grade /	%	%	%
Round	Developing	On Track	Advanced
Grade 3			
Round 1	37%	42%	21%
Round 2	37%	42%	21%
Round 3	37%	42%	21%
Grade 4			
Round 1	41%	44%	15%
Round 2	28%	47%	25%
Round 3	28%	47%	25%
Grade 5			
Round 1	62%	19%	19%
Round 2	49%	30%	21%
Round 3	49%	30%	21%
Grade 6			
Round 1	69%	23%	8%
Round 2	55%	34%	11%
Round 3	55%	34%	11%
Grade 7	·		
Round 1	46%	44%	10%
Round 2	47%	40%	13%
Round 3	46%	41%	13%
Grade 8			
Round 1	25%	52%	23%
Round 2	25%	55%	20%
Round 3	25%	55%	20%

Table 13. Cut Score Impact: ELA



Table 14. Cut Score Impact: Math

Mathematics Cut Score Impact								
Grade /	%	%	%					
Round	Developing	On Track	Advanced					
Grade 3								
Round 1	60%	28%	12%					
Round 2	60%	28%	12%					
Round 3	41%	46%	13%					
Grade 4								
Round 1	54%	34%	12%					
Round 2	45%	37%	18%					
Round 3	42%	46%	12%					
Grade 5								
Round 1	36%	45%	19%					
Round 2	35%	46%	19%					
Round 3	35%	46%	19%					
Grade 6								
Round 1	45%	35%	20%					
Round 2	42%	40%	18%					
Round 3	42%	40%	18%					
Grade 7								
Round 1	67%	23%	10%					
Round 2	59%	31%	10%					
Round 3	34%	46%	20%					
Grade 8								
Round 1	49%	30%	21%					
Round 2	43%	31%	26%					
Round 3	38%	36%	26%					

The standard error of the median (SEM) was calculated from the individual panelist ratings and used to provide a range of recommended cut scores (as represented by theta values) that reflect both the variation and sample size of the panels. These final panel-level recommended ranges are shown in Tables 15 and 16 below.

Table 15.	Recommended	Ranges: ELA
-----------	-------------	--------------------

	ELA Panel Recommendations												
Grade / Cut	Cut - 2 SEM (Theta)	Cut Score (Theta)	Cut + 2 SEM (Theta)										
Grade 3													
On-Track	-1.261	-0.791	-0.320										



Advanced	-0.166	0.503	1.173
Grade 4			
On-Track	-1.143	-0.688	-0.232
Advanced	0.353	0.820	1.287
Grade 5			
On-Track	0.132	0.250	0.368
Advanced	1.141	1.141	1.141
Grade 6			
On-Track	0.332	0.527	0.722
Advanced	1.405	1.572	1.740
Grade 7			
On-Track	0.180	0.373	0.565
Advanced	1.448	1.595	1.742
Grade 8			
On-Track	-0.394	-0.057	0.279
Advanced	1.170	1.541	1.913

Table 16. Recommended Ranges: Math

N	lathematics Pa	nel Recommen	dations
Grade / Cut	Cut - 2 SEM (Theta)	Cut Score (Theta)	Cut + 2 SEM (Theta)
Grade 3			
On-Track	-0.469	0.241	0.950
Advanced	0.668	1.921	3.175
Grade 4			
On-Track	0.102	0.614	1.126
Advanced	2.092	2.386	2.681
Grade 5			
On-Track	-0.460	0.182	0.825
Advanced	1.242	2.179	3.117
Grade 6			
On-Track	0.135	0.623	1.111
Advanced	1.524	2.031	2.538
Grade 7			
On-Track	1.096	1.483	1.869
Advanced	2.198	2.842	3.485
Grade 8			
On-Track	0.534	0.986	1.439
Advanced	1.868	2.273	2.679

Following the standard setting process for each grade level, the panelists completed a *Results Evaluation* (Appendix C) where they were asked to provide feedback on the time allocated to each component of the



standard setting process, as well as their confidence in the recommended cut scores. The results of these forms are summarized for all assessments in Tables 17 and 18 below.

ELA Results Evaluation	G	irade 3	G	rade 4	G	rade 5	G	rade 6	G	rade 7	G	irade 8
Training	Ν	%	Ν	%	Ν	%	Ν	%	Ν	%	Ν	%
More than enough time	4	50.0%	3	33.3%	3	33.3%	5	50.0%	1	11.1%	1	14.3%
Sufficient time	4	50.0%	6	66.7%	6	66.7%	5	50.0%	8	88.9%	6	85.7%
Not enough time	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Round 1 Judgments	Ν	%	Ν	%	Ν	%	Ν	%	Ν	%	Ν	%
More than enough time	5	62.5%	4	44.4%	5	55.6%	4	40.0%	1	11.1%	4	57.1%
Sufficient time	3	37.5%	5	55.6%	4	44.4%	6	60.0%	8	88.9%	3	42.9%
Not enough time	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Discussion	Ν	%	Ν	%	Ν	%	Ν	%	Ν	%	Ν	%
More than enough time	4	50.0%	4	44.4%	4	44.4%	5	50.0%	2	22.2%	3	42.9%
Sufficient time	4	50.0%	5	55.6%	5	55.6%	5	50.0%	7	77.8%	4	57.1%
Not enough time	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Round 2 Judgments	Ν	%	Ν	%	Ν	%	Ν	%	Ν	%	Ν	%
More than enough time	5	62.5%	4	44.4%	5	55.6%	4	40.0%	5	55.6%	2	28.6%
Sufficient time	3	37.5%	5	55.6%	4	44.4%	6	60.0%	4	44.4%	5	71.4%
Not enough time	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
On Track Confidence	Ν	%	Ν	%	N	%	Ν	%	Ν	%	Ν	%
Confident	6	75.0%	3	33.3%	4	44.4%	5	50.0%	6	66.7%	6	85.7%
Somewhat Confident	0	0.0%	2	22.2%	4	44.4%	3	30.0%	3	33.3%	1	14.3%
Somewhat not confident	2	25.0%	4	44.4%	1	11.1%	2	20.0%	0	0.0%	0	0.0%
Not Confident	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Advanced Confidence	Ν	%	Ν	%	N	%	Ν	%	Ν	%	Ν	%
Confident	8	100.0%	6	66.7%	6	66.7%	5	50.0%	8	88.9%	7	100.0%
Somewhat Confident	0	0.0%	3	33.3%	3	33.3%	4	40.0%	1	11.1%	0	0.0%
Somewhat not confident	0	0.0%	0	0.0%	0	0.0%	1	10.0%	0	0.0%	0	0.0%
Not Confident	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%

Table 17. Results Evaluation: ELA

Table 18. Results Evaluation: Math

Math Results Evaluation	G	rade 3	G	irade 4	G	rade 5	Grade 6		Grade 7		Grade 8	
Training	Ν	%	Ν	%	Ν	%	N	%	Ν	%	Ν	%
More than enough time	2	33.3%	1	16.7%	4	40.0%	9	100.0%	1	11.1%	1	11.1%
Sufficient time	4	66.7%	5	83.3%	6	60.0%	0	0.0%	8	88.9%	8	88.9%
Not enough time	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Round 1 Judgments	Ν	%	Ν	%	Ν	%	N	%	Ν	%	Ν	%
More than enough time	1	16.7%	0	0.0%	5	50.0%	8	88.9%	1	11.1%	1	11.1%
Sufficient time	5	83.3%	6	100.0%	5	50.0%	1	11.1%	8	88.9%	8	88.9%
Not enough time	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Discussion	N	%	Ν	%	N	%	N	%	N	%	Ν	%

ACS Ventures – Bridging Theory & Practice



More than enough time	1	16.7%	0	0.0%	5	50.0%	8	88.9%	2	22.2%	2	22.2%
Sufficient time	5	83.3%	6	100.0%	5	50.0%	1	11.1%	7	77.8%	7	77.8%
Not enough time	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Round 2 Judgments	Ν	%	Ν	%	N	%	Ν	%	Ν	%	Ν	%
More than enough time	1	16.7%	0	0.0%	5	50.0%	8	88.9%	1	11.1%	2	22.2%
Sufficient time	5	83.3%	6	100.0%	5	50.0%	1	11.1%	8	88.9%	7	77.8%
Not enough time	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
On Track Confidence	Ν	%	Ν	%	N	%	Ν	%	Ν	%	Ν	%
Confident	5	83.3%	4	66.7%	10	100.0%	8	88.9%	7	77.8%	7	77.8%
Somewhat Confident	1	16.7%	2	33.3%	0	0.0%	1	11.1%	2	22.2%	2	22.2%
Somewhat not confident	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Not Confident	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Advanced Confidence	Ν	%	Ν	%	Ν	%	Ν	%	Ν	%	Ν	%
Confident	5	83.3%	6	100.0%	9	90.0%	9	100.0%	3	33.3%	8	88.9%
Somewhat Confident	1	16.7%	0	0.0%	1	10.0%	0	0.0%	6	66.7%	1	11.1%
Somewhat not confident	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Not Confident	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%

The results of these evaluations demonstrate that the panelists felt there was sufficient time allocated to each component of the standard setting process and that the panelists had confidence in the cut score recommendations. This is evident by having no panelist ratings of "Not Confident" across all panels. However, there were some panelists who provided ratings of "Somewhat not confident" in the recommended cut scores in the ELA panels for grades 3-6. The highest incident of this rating occurred for the recommended ELA Grade 4 *On-Track* cut where four panelists (44%) provided this rating.



3.1.2. Articulation Panel Results

The results shown in Table 19 below represent the recommendations following the standards articulation review. The panel recommended adjusting the *On-Track* cuts for the ELA grade 4, grade 5, and grade 6 assessments and recommended adjusting the *Advanced* cut for the ELA grade 4 assessment. The cuts that were recommended to be adjusted are highlighted in the table. The articulation panel did not recommend any adjustments for mathematics. The impact of these recommendations is shown in Figures 1 and 2 that follow.

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
ELA						
On-Track	-0.791	<mark>-0.096*</mark>	<mark>0.060*</mark>	<mark>0.434</mark>	0.373	-0.057
Advanced	0.503	<mark>0.921</mark>	1.141	1.572	1.595	1.541
Math						
On-Track	-0.440	0.154	0.136	0.507	0.219	0.563
Advanced	1.770	2.386	2.179	2.200	2.084	2.006

Table 19. Cut Score Recommendations: Articulation Panel Resul

*Notes adjustments that were outside recommended ranges shown in tables 15 & 16.



Figure 1. Articulation Recommendation Impact: ELA



The standards articulation activity concluded with the panelists completing a *Standards Articulation Evaluation* (Appendix C) in which they were asked to provide feedback on their evaluation of the time allocated for each activity, as well as provide feedback on their confidence in the cut scores being recommended for their panel. The results of the evaluation are shown in Tables 20 and 21 below.

ELA Articulation Evaluation	G	Grades 3&4	Grades 5&6		G	rades 7&8
Presentation of Results	N	%	N	%	N	%
More than enough time	2	66.7%	2	66.7%	2	66.7%
Sufficient time	1	33.3%	1	33.3%	1	33.3%
Not enough time	0	0.0%	0	0.0%	0	0.0%
Discussion	N	%	N	%	N	%
More than enough time	2	66.7%	1	33.3%	2	66.7%
Sufficient time	1	33.3%	2	66.7%	1	33.3%
Not enough time	0	0.0%	0	0.0%	0	0.0%

Table 20. Standards Articulation Evaluation: ELA

Recommendations	Ν	%	N	%	Ν	%
More than enough time	2	66.7%	2	66.7%	1	33.3%
Sufficient time	1	33.3%	1	33.3%	2	66.7%
Not enough time	0	0.0%	0	0.0%	0	0.0%
On-Track Cut Confidence	N	%	N	%	N	%
Confident	1	33.3%	2	66.7%	2	66.7%
Somewhat Confident	1	33.3%	1	33.3%	1	33.3%
Somewhat not Confident	1	33.3%	0	0.0%	0	0.0%
Not Confident	0	0.0%	0	0.0%	0	0.0%
Advanced Cut Confidence	N	%	N	%	N	%
Confident	2	66.7%	3	100.0%	3	100.0%
Somewhat Confident	0	0.0%	0	0.0%	0	0.0%
Somewhat not Confident	1	33.3%	0	0.0%	0	0.0%
Not Confident	0	0.0%	0	0.0%	0	0.0%

Table 21. Standards Articulation Evaluation: Math

Math Articulation Evaluation		Grades 3&4		Grades 5&6		Grades 7&8
Presentation of Results	N	%	N	%	N	%
More than enough time	0	0.0%	1	50.0%	1	33.3%
Sufficient time	3	100.0%	1	50.0%	2	66.7%
Not enough time	0	0.0%	0	0.0%	0	0.0%
Discussion	N	%	N	%	N	%
More than enough time	0	0.0%	1	50.0%	1	33.3%
Sufficient time	3	100.0%	1	50.0%	2	66.7%
Not enough time	0	0.0%	0	0.0%	0	0.0%
Recommendations	N	%	N	%	N	%
More than enough time	0	0.0%	1	50.0%	1	33.3%
Sufficient time	3	100.0%	1	50.0%	2	66.7%
Not enough time	0	0.0%	0	0.0%	0	0.0%
On-Track Cut Confidence	Ν	%	N	%	Ν	%
Confident	2	66.7%	2	100.0%	3	100.0%
Somewhat Confident	1	33.3%	0	0.0%	0	0.0%
Somewhat not Confident	0	0.0%	0	0.0%	0	0.0%
Not Confident	0	0.0%	0	0.0%	0	0.0%
Advanced Cut Confidence	N	%	N	%	N	%
Confident	3	100.0%	2	100.0%	3	100.0%
Somewhat Confident	0	0.0%	0	0.0%	0	0.0%
Somewhat not Confident	0	0.0%	0	0.0%	0	0.0%
Not Confident	0	0.0%	0	0.0%	0	0.0%

The results of these evaluations demonstrate that the panelists felt there was sufficient time allocated to each component of the standard setting process and they had confidence in the cut score recommendations. There



was one panelist (33.3%) who provided ratings of "Somewhat not confident" in both recommended cut scores in the ELA grades 3&4 panel.

3.1.3. AAAC Results

The results shown in Table 20 below represent the recommendations from the AAAC policy review meeting. For ELA, the AAAC policy committee recommended adjusting *On-Track* cuts for grade 6, grade 7, and grade 8 assessments, as well as the *Advanced* cuts for grade 6, grade 7, and grade 8 assessments. For mathematics, the AAAC policy committee only recommended adjusting the *Advanced* cut for the grade 8 assessment. The cuts that were recommended to be adjusted are highlighted in the table. The impact of these recommendations is shown in Figures 3 and 4 that follow.

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
ELA						
On-Track	-0.791	-0.096*	0.060*	<mark>0.242*</mark>	0.373	<mark>0.326*</mark>
Advanced	0.503	0.921	1.141	<mark>1.303*</mark>	<mark>1.510</mark>	<mark>1.707</mark>
Math						
On-Track	-0.440	0.154	0.136	0.507	0.219	0.563
Advanced	1.770	2.386	2.179	2.200	2.084	<mark>2.158</mark>

Table 22. Cut Score Recommendations: AAAC Results

*Notes adjustments that were outside recommended ranges shown in tables 15 & 16.





Figure 4. AAAC Recommendation Impact: Math





3.2. Evaluation of Standard Setting Activities

The standard setting activities and resulting cut score recommendations that were produced from this study were evaluated using the following three sources of validity evidence as recommended by Kane (1994; 2001):

- Procedural evidence
- Internal evidence
- External evidence

The following sections summarize the results of this evaluation according to each source of validity evidence.

3.2.1. Procedural Evidence

Procedural validity is evaluated based on evidence related to the selection and execution of the standard setting methodology used in the study (Kane, 1994; 2001). The Bookmark method is well suited for adaptive measures as the extensive item banks support creation of an OIB for each subject and grade.

Evidence supporting the execution of the method is demonstrated through the following sources (Kane, 1994; 2001):

- 1. Definition of goals for the decision procedure
- 2. Selection of panelists
- 3. Training of panelists
- 4. Definition of performance standard
- 5. Data collection procedures

Table 23 shows evidence to support the procedural validity of the study organized by the sources listed above.

	Source	Evaluation Criteria	Evidence Presented
		(Kane, 1994; 2001)	
1.	Definition of goals for	The general purpose of the use of	Policy ALDs for both ELA and
	the decision procedure	the passing score should be	mathematics were created and
	-	defined before the standard	published in 2021 (see Appendix A).
		setting process begins.	
2.	Selection of judges	The panelists selected should be	The average years of experience in the
		both technical experts in the	subject was 12.4 years in the ELA
		material and familiar with the	panels and 15.1 years in the math
		population.	panels. The largest category of current
			job role was classroom teacher in all
			panels (see Appendix B).
		The number of panelists should be	The panel size in this study ranged from
		large enough to achieve an	7 to 11 panelists (see Table 5) and
		acceptable standard error of	resulted in Round 3 standard errors
		measurement for the	ranging from 0.00 to 2.11 across all
		recommended cut score.	panels (see tables 9-12).
3.	Training of judges	The panelists should be oriented	In the Readiness Evaluation, 100% of
		to the goals of the study, be	panelists across both panels rated all
		trained on the steps of the rating	components of the training as either



		process, and have an opportunity	being "Very good" or "Good".
		to practice the steps before	Additionally, 100% of panelists also
		making operation ratings.	reported feeling prepared to make
			operational Bookmark judgments (see
			Table 7 and Table 8).
4.	Definition of	The panelists were given the	A large portion of the study
	performance standard	opportunity to develop a	preparation was dedicated to drafting
		definition of the standard of	threshold ALDs from the existing Range
		performance they consider	ALDs (see Appendix A). The whole
		adequate for the intended	panel had to approve each set of
		purposes of the decision process.	threshold ALDs before they were to be
			used in the rating process. Additionally,
			in the Readiness Evaluation 100% of
			panelists across both panels rated the
			training on threshold ALD development
			as either being "Very good" or "Good"
			(see Table 7 and Table 8).
5.	Data collection	The procedures to collect data	The panelists provided judgments over
	procedures	allow panelists multiple	three rounds and were allowed to
		opportunities to review their	review their individual item ratings
		decisions before the passing score	prior to and during the subsequent
		is finalized.	rounds (see Agenda).
		The panelists were allotted ample	The panelists were allotted time
		time to discuss ratings and results.	dedicated to discussing the individual
			item ratings following Round 1 and
			Round 2 of the study (see Agenda).
			Additionally, the panelists were asked
			rate the amount of time allocated to
			each activity in the Results Evaluation
			and 100% of all panelists rated having
			either "Sufficient time" or "More than
			enough time" for discussion (see Table
			17 and Table 18).

3.2.2. Internal Evidence

Internal validity is evaluated based on evidence related to the consistency of the panelist judgments and the convergence of the resulting cut score recommendations (Kane, 1994; 2001).

Evidence supporting the claim that the panelist judgments are consistent and that the cut score recommendations are converging can be demonstrated by the overall low standard errors as well as a reduction in the standard error of the recommended cuts across rounds.

Table 24 and Table 25 show evidence supporting the internal validity of the standard setting activities because 21 of the 24 recommended cuts across all panels resulted in Round 3 standard errors that were the less than or equal to the standard errors of the Round 1 recommendations. Additionally, all of the cuts that did have increasing standard error (noted by an asterisk below) only increased by less than 0.1 from the next largest



value. These likely indicated that some panelists had a different reaction to the Round 2 results including the impact.

Table 24. Cut Score Standard Errors: ELA

ELA Cut Score Standard Errors							
Grade / Round	On-Track SEM	Advanced SEM					
Grade 3							
Round 1	0.24	0.22					
Round 2	0.26	0.00					
Round 3	0.24	0.33*					
Grade 4							
Round 1	0.26	0.30					
Round 2	0.30	0.32					
Round 3	0.23	0.23					
Grade 5							
Round 1	0.27	0.47					
Round 2	0.10	0.26					
Round 3	0.06	0.00					
Grade 6							
Round 1	0.25	0.45					
Round 2	0.10	0.10					
Round 3	0.10	0.08					
Grade 7							
Round 1	0.12	0.09					
Round 2	0.13	0.11					
Round 3	0.10	0.07					
Grade 8							
Round 1	0.17	0.14					
Round 2	0.17	0.19					
Round 3	0.17	0.19*					

Table 25. Cut Score Standard Errors: Math

Math Cut Score Standard Errors								
Grade / Round On-Track SEM Advanced SEM								
Grade 3								
Round 1 0.35 0.63								
Round 2	0.30	0.48						
Round 3	0.12	0.11						
Grade 4								
Round 1 0.26 0.15								
Round 2 0.16 0.12								

Round 3	0.14	0.10					
Grade 5							
Round 1	0.32	0.47					
Round 2	0.07	0.05					
Round 3	0.07	0.05					
Grade 6							
Round 1	0.24	0.25					
Round 2	0.08	0.07					
Round 3	0.08	0.07					
Grade 7							
Round 1	0.19	0.32					
Round 2	0.21	0.22					
Round 3	0.25*	0.23					
Grade 8							
Round 1	0.23	0.20					
Round 2	0.17	0.16					
Round 3	0.17	0.16					

3.2.3. External Evidence

External validity is the most difficult to evaluate and is based on evidence that comes from triangulating the results of the standard setting process with some other indicator of examinee performance that is related but external to the process (Kane, 1994; 2001).

Evidence supporting external validity was collected through multiple cross-panel validations and policy review committee feedback. Results from all three sources should be considered in the final recommendation.

Table 26 shows the evidence organized by the method of collection.

Source	Evidence Presented						
Adjacent Grade	After the panelists made Round 2 judgments for the second assessment, they						
Validation	were presented with the results (impact) of the upper grade level panel's first						
	assessment to review and to consider whether the results were in line with						
	their expectations of student performance (see Training slides). The facilitators						
	recorded notes to discuss any concerns in the standards articulation panel.						
Standards Articulation	During the standards articulation panel, the representatives from each panel						
Panel	were presented with the results of all grade-levels across the ELA and math						
	panels. Panelists were able to ask questions about results that differed from						
	expectations and representatives from the grade level in question were given						
	the opportunity to provide a content-based rationale that explained the result.						
	When the representatives did not have a justification and/or expressed their						
	own scrutiny, they were given an opportunity to recommend adjustments to						
	better reflect expectations.						
	Panelists were also provided the opportunity to compare the impact data of the						
	results to that of the ACT in high school, as well as the NAEP results for grades 4						

 Table 26. External Validity Evidence



	and 8. The panelists were provided the opportunity to discuss whether their					
	recommendations reflected comparable expectations and overall, the result					
	did not affect their recommendations.					
Policy Review	Following the standards articulation, a policy review committee was presented					
Committee	with a summary of the study procedures and results. This committee reviewed					
	the cut scores and impact data that resulted from the standard setting activities					
	and recommended adjustments to cut scores when they believed the results					
	did not accurately reflect the student expectations.					



4. References

- Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64 (3), 425-461.
- Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. In G.
- Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, E. M. (2012). The Bookmark standard setting procedure. In G. J. Cizek (Ed.) *Setting Performance Standards: Foundations, Methods, and Innovations* (2nd Ed.). New York, NY: Routledge.
- Nebraska Department of Education (NDE). (2022). Spring 2022 NSCAS Growth ELA, Mathematics, and Science Technical Report. <u>https://www.education.ne.gov/wp-content/uploads/2022/12/2022-NSCAS-Growth-Technical-Report.pdf</u>



Appendices 5.

5.1.1. Appendix A – Content Standards and RALDs

ELA Standards



Mathematics Standards



ege_and_Career_Stan

ELA Range ALDs



ELA-2021-Range-ALD s_June-2022-2.xlsx

Mathematics Range ALDs



8m - Range ALDs (grade 4).pdf

8m - Range ALDs (grade 5).pdf



8m - Range ALDs (grade 7).pdf



(grade 8).pdf



8m - Range ALDs (grade 3).pdf

ELA Threshold ALDs



ALDs.docx

w

ALDs.docx

Math Threshold ALDs

ELA G4 Threshold

ALDs.docx

ALDs.docx



Math G3 Threshold Math G4 Threshold Math G5 Threshold Math G6 Threshold Math G7 Threshold

ALDs.docx

ELA G5 Threshold ALDs.docx

ELA G6 Threshold ALDs.docx

ALDs.docx

ELA G7 Threshold ALDs.docx

ALDs.docx

ELA G8 Threshold ALDs.docx



Math G8 Threshold ALDs.docx



2023 NDE ELA and Math Standard Setting Report

ACS Ventures - Bridging Theory & Practice

5.1.2. Appendix B – Panelist Demographic Information

ELA Panel Demographics	LA Panel Demographics ELA 3&4 ELA 5&6		LA 5&6	ELA 7&8		
Current Job Title	N	%	N	%	N	%
Classroom teacher	10	100%	8	80%	7	78%
Curriculum/Instructional coach	0	0%	2	20%	1	11%
Special Education professional	0	0%	0	0%	1	11%
Teaching and Learning Specialist	0	0%	0	0%	0	0%
District	0	0%	0	0%	0	0%
Educational Service Unit	N	%	N	%	N	%
ESU 1 - Wakefield	2	20%	0	0%	0	0%
ESU 2 - Fremont	0	0%	0	0%	1	11%
ESU 3 - Lavista	0	0%	1	10%	1	11%
ESU 4 - Auburn	0	0%	0	0%	0	0%
ESU 5 - Beatrice	0	0%	0	0%	2	22%
ESU 6 - Milford	0	0%	1	10%	0	0%
ESU 7 - Columbus	0	0%	1	10%	1	11%
ESU 8 - Neligh	0	0%	1	10%	1	11%
ESU 9 - Hastings	0	0%	0	0%	0	0%
ESU 10 - Kearney	0	0%	0	0%	0	0%
ESU 11 - Holdrege	0	0%	0	0%	0	0%
ESU 13 - Scottsbluff	2	20%	2	20%	0	0%
ESU 15 - Trenton	0	0%	0	0%	1	11%
ESU 16 - Ogallala	0	0%	0	0%	0	0%
ESU 17 - Ainsworth	1	10%	1	10%	0	0%
ESU 18 - Lincoln	2	20%	1	10%	1	11%
ESU 19 - Omaha	3	30%	1	10%	1	11%
None / Unsure of ESU	0	0%	1	10%	0	0%
Highest Level of Education	N	%	N	%	N	%
Bachelor's degree	4	40%	1	10%	4	44%
Master's degree	6	60%	8	80%	5	56%
Doctoral degree	0	0%	1	10%	0	0%
Years of Experience in ELA	N	%	N	%	N	%
0 to 2 years	2	20%	2	20%	0	0%
3 to 5 years	0	0%	1	10%	4	44%
6 to 9 years	2	20%	1	10%	0	0%
10 to 14 years	2	20%	2	20%	1	11%
15 to 19 years	2	20%	1	10%	2	22%
20 or more years	2	20%	3	30%	2	22%

Table A. 1 ELA Panel Demographic Information



Math Panel Demographics	nel Demographics Math 3&4 Math 5&6 Math		ath 7&8			
Current Job Title	N	%	N	%	N	%
Classroom teacher	4	40%	9	90%	7	78%
Curriculum/Instructional coach	2	20%	1	10%	2	22%
Special Education professional	0	0%	1	10%	0	0%
Teaching and Learning Specialist	0	0%	0	0%	0	0%
District	1	10%	0	0%	0	0%
Educational Service Unit	N	%	N	%	N	%
ESU 1 - Wakefield	0	0%	4	40%	0	0%
ESU 2 - Fremont	0	0%	0	0%	2	22%
ESU 3 - Lavista	1	10%	1	10%	1	11%
ESU 4 - Auburn	0	0%	0	0%	0	0%
ESU 5 - Beatrice	0	0%	1	10%	0	0%
ESU 6 - Milford	0	0%	1	10%	0	0%
ESU 7 - Columbus	1	10%	0	0%	1	11%
ESU 8 - Neligh	0	0%	0	0%	0	0%
ESU 9 - Hastings	0	0%	0	0%	0	0%
ESU 10 - Kearney	1	10%	1	10%	3	33%
ESU 11 - Holdrege	0	0%	0	0%	0	0%
ESU 13 - Scottsbluff	0	0%	0	0%	0	0%
ESU 15 - Trenton	0	0%	1	10%	0	0%
ESU 16 - Ogallala	1	10%	0	0%	0	0%
ESU 17 - Ainsworth	0	0%	0	0%	1	11%
ESU 18 - Lincoln	1	10%	2	20%	0	0%
ESU 19 - Omaha	2	20%	0	0%	1	11%
None / Unsure of ESU	0	0%	0	0%	0	0%
Highest Level of Education	N	%	N	%	N	%
Bachelor's degree	0	0%	2	20%	1	11%
Master's degree	6	60%	9	90%	8	89%
Doctoral degree	1	10%	0	0%	0	0%
Years of Experience in Math	N	%	N	%	N	%
0 to 2 years	1	10%	3	30%	8	89%
3 to 5 years	0	0%	2	20%	0	0%
6 to 9 years	1	10%	1	10%	0	0%
10 to 14 years	4	40%	1	10%	0	0%
15 to 19 years	0	0%	1	10%	0	0%
20 or more years	1	10%	3	30%	1	11%



5.1.3. Appendix C – Study Materials





Evaluation.docx



Evaluation.docx





Articulation Initial Judgments.docx



Vertical Articulation Evaluation.docx



