Appendix A: Meeting Reports

1. GAA 2.0 Content, Bias, and Data Review Meeting Report 2
2. GAA 2.0 High School Standards Validation Meeting Report 47

121 NW Everett Street
Portland, Oregon 97209
www.nwea.org

Georgia Alternate Assessment 2.0
Content, Bias, and Data Review Meeting Report
June 14, 2022–June 16, 2022

Executive Summary

A Content, Bias, and Data Review meeting was conducted for the Georgia Alternate Assessment (GAA) 2.0. The GAA 2.0 is an assessment of the Extended Content Standards (ECS), which are aligned to Georgia's state academic content standards and is administered to students with the most significant cognitive disabilities as part of the Georgia Student Assessment Program. The GAA 2.0 was developed to assess achievement of the knowledge and skills inherent in the ECS, Georgia's alternate academic content standards, and was administered operationally for the first time in Spring 2019. Content experts from throughout the state met in Atlanta, Georgia June 14–June 16, 2022, to review GAA 2.0 tasks for English language arts (ELA), mathematics, science, and social studies.

Panels

Georgia Department of Education (GaDOE) staff recruited 55 committee members to participate in the Content, Bias, and Data Review Meeting. Newly developed assessment tasks were reviewed for content and bias, and data was reviewed for the field test tasks administered in the Spring. Each committee member had access to the newly developed and field test tasks, the ECS, statistical data cards, content-specific vocabulary lists, and other resources during their review of the tasks. Committee members were identified for each content area and were divided into the following grade-level groups:

ELA

- Grade 3, Grade 4, Grade 5
- Grade 6, Grade 7
- Grade 8, High School

Mathematics

- Grade 3, Grade 4, Grade 5
- Grade 6, Grade 7
- Grade 8, High School

Science

- Grade 5, Grade 8, High School

Social Studies

- Grade 8, High School

Workshop

The objective of the GAA 2.0 is to ensure that students with significant cognitive disabilities are given the opportunity to demonstrate achievement of the knowledge, concepts, and skills inherent in the grade-level general education standards—the GSE. Like Georgia's general statewide assessments, the GAA 2.0 assesses ELA and mathematics for students in grades 3–8 and high school, science for students in grades 5, 8, and high school, and social studies for students in grade 8 and high school.

Committee members began their work with a general session overview and PowerPoint presentation that detailed the expectations for the three-day meeting. After the general session, committee members joined their grade-level and content groups for the remainder of the day.

Evaluations

Overall, the committee member evaluations were positive and showed an appreciation of the work being done throughout the course of the meeting.

In the evaluations, committee members responded to a series of statements using a four-point Likert scale. The majority of committee members strongly agreed that participating in the Content, Bias, and Data Review meeting was professionally rewarding. The materials provided to the committee were appropriate for completing the tasks and members were comfortable using them. Committee members strongly agreed that the work accomplished will help Georgia students. Overall, committee members felt the meeting accomplished its purpose, which was for educators to review all tasks for quality and appropriateness (after considering item statistics and guidelines for test development, fairness, equity, bias, and sensitivity) and to provide recommendations to GaDOE for accepting, revising, or rejecting the tasks. Committee members shared the following statements regarding their overall perceptions of the Content, Bias, and Data Review Meeting:

- "I felt that the entire review process, including the review of potential questions and answers, as well as the data analysis of previous test question results was extremely helpful with creating an effective meaningful testing experience for students."
- "This was my first Content, Bias, and Data Review Meeting. I never knew how much went into creating each item on the GAA assessment. As an educator, this was rewarding and fulfilling to see. These assessments are not just made to test children and go through the motions of testing but truly see what our most challenged students can do in a way that highlights their abilities…"
- "I felt that the whole process was definitely beneficial as there were both people who were familiar with the GAA and people who were content specialist…"
- I thoroughly enjoyed being a part of the GAA 2.0 content and data review team. As a content teacher, it helped me to understand more about the GAA assessment. I always feel when attending these meetings, it helps me as a classroom teacher as I am able to truly collaborate with other educators across the state.
- "The opportunity to provide and hear feedback from educators in the classroom was valuable to me. The various perspectives was meaningful. I gained an understanding of how the percentages are interpreted and the reasoning for various answer choices. I enjoyed the statistics portion and the process of determining if the tasks aligned with the standards."
- "I really enjoyed this process. It is so heart-warming to know that test developers care this much about what is best for our beautiful students with disabilities. I truly felt like the entire process took place because of the best interest of the students! Thank you for allowing me to be a part of it."

Contents

Executive Summary i

Panels i
Workshop i
Evaluations ii
Contents ii
List of Tables & Figures v
Chapter 1. Introduction 1
Purpose and Organization of the Report 1
List of Acronyms 1
Relevant Terminology 2
Chapter 2. Meeting Design and Committee Selection 3
Meeting Design 3
Committee Selection 3
Chapter 3: Workshop Roles 5
Georgia Department of Education Staff 5
NWEA Staff **Error! Bookmark not defined.**
Program Management 6
Chapter 4. Meeting Materials 6
PowerPoint Presentations 7
Content Materials 7
Final Committee Evaluations 9
Chapter 5. Meeting Design Implementation 10
Meeting Agenda 11
Content and Bias Review Workshop Summary 12
Data Review Workshop Summary 14
Evaluations 15
Appendix A. Content and Bias Review Presentation Slides 15
Appendix B. Content and Bias Review Educator Checklist 27
Appendix C. Sample Feedback Sheet 30
Appendix D. Data Review Presentation Slides 31
Appendix E. Data Card Statistic Descriptions 39
Appendix F. Guiding Questions for Data Review 40

List of Tables & Figures

Table 2.1 Number of Committee Members by Grade/Content Area Group, Gender, and Ethnicity..... 3

Table 2.2 Years of Experience by Content Area and Grade Level 4

Table 2.3 Committee Member Highest Degree of Educational Attainment Category Count by Content Area and Grade Level 4

Table 3.1 Roles of Georgia Department of Education Staff..... 5

Table 3.2 Roles of NWEA Staff..... 6

Table 5.1 Meeting Agenda..... 12

Table 5.2 Number of Content and Bias Items Reviewed and Rejected or Revised..... 13

Table 5.3 Number of Data Review Items Reviewed and Rejected..... 14

Table 5.4 Content and Bias Review Meeting Evaluation..... 15

Table 5.5 Data Review Meeting Evaluation..... 16

Chapter 1. Introduction

GaDOE has developed the GAA 2.0 for the assessment of students with the most significant cognitive disabilities. The GAA 2.0 is based on alternate achievement standards aligned to the state's approved general content standards, the GSE.

GaDOE invited educators from throughout the state to engage in a three-day Content, Bias, and Data Review meeting. Participants provided input regarding the newly developed tasks as well as the Spring 2022 field test tasks for ELA, mathematics, science, and social studies. Staff from NWEA (the state's assessment vendor) and GaDOE facilitated the meeting sessions that took place June 14–16, 2022.

Purpose and Organization of the Report

The purpose of this document is to provide a comprehensive report of the processes and procedures used to review the newly developed tasks and the Spring 2022 field test tasks for the GAA 2.0. Accurate and detailed process documentation establishes procedural validity; this contributes to the overall validity argument for the GAA 2.0 assessments and provides evidence in support of the United States Department of Education (ED) Peer Review requirements. Although the ED requires submission of evidence for Peer Review only for assessments in ELA, mathematics, and science, GaDOE recognizes the steps outlined in this report as important evidence in support of the validity argument for *all* content area assessments.

This chapter describes the rationale for the GAA 2.0 Content, Bias, and Data Review meeting. A list of acronyms and relevant definitions used throughout the report is also found in this chapter. Chapter 2 provides an overview of the meeting design, goals, and preparation. Chapter 3 provides information about the meeting facilitators and committee members and their roles and responsibilities. Chapter 4 describes the materials that were used during the meeting. Chapter 5 describes the implementation of the meeting design, including committee feedback and perceptions of the workshop.

List of Acronyms

The following acronyms are found throughout the text of this report. The first time an acronym is used, it will be preceded by the term spelled out in its entirety. Each subsequent reference will include only the acronym. This list provides a quick reference for the reader.

CBR—Content and Bias Review

DR—Data Review

ECS—Extended Content Standards

ED—United States Department of Education

ELA—English language arts

GAA 2.0—Georgia Alternate Assessment 2.0

GaDOE—Georgia Department of Education

GSE–Georgia Standards of Excellence

SB—Student Booklet

TEB—Test Examiner's Booklet

Relevant Terminology

The following terms will appear throughout this report:

Georgia K-12 Mathematics Standards—Georgia's new math standards adopted in August 2021. The newly developed math tasks reviewed for content and bias were developed and aligned to the new standards.

Georgia Standards of Excellence—Georgia's content standards, or the specific knowledge, skills, and abilities students are expected to demonstrate within a content area and grade level or grade range.

Extended Content Standards—Statements of knowledge, skills, and processes that align to grade-level content standards for a general assessment program to ensure that all students who take the test, including students with the most significant cognitive disabilities, are included in all appropriate learning opportunities.

The GAA 2.0 ECS can be found at the link below:

https://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Pages/GAA-2.0-Extended-Content-Standards.aspx

Chapter 2. Meeting design and Committee Selection

Meeting Design

In preparation for the GAA 2.0 Content, Bias, and Data Review meeting, GaDOE and NWEA collaborated to develop a meeting designed for educators to 1) review the newly developed tasks, 2) review the statistics for the field test tasks administered during the Spring 2022 Administration, and 3) provide recommendations to GaDOE on whether the three items within each task should be accepted, revised, or rejected. The meeting design called for a three-day meeting during which GaDOE and NWEA would facilitate collaborative work among GaDOE staff, NWEA content specialists, and groups of educators identified by the GaDOE.

Committee Selection

Per the GAA 2.0 Content, Bias, and Data Review meeting design, GaDOE employed an online application process to solicit a pool of educators from which to select committee members for each grade and content group. Availability and specialization information from nominated educators was then gathered. Selection for committees prioritized educators with content and population expertise, especially expertise in serving students with significant cognitive disabilities.

GaDOE then provided the names of 60 participants to NWEA, which sent invitations on behalf of the department; ultimately 55 participants registered for and attended the meeting. Table 2.1 shows the number of committee members that participated in each content area/grade-span group disaggregated by gender and ethnicity. The committees consisted of 49 female and 6 male participants. Races/ethnicities represented by participating educators included White, African American/black, and multi-racial.

Table 2.1 Number of Committee Members by Grade/Content Area Group, Gender, and Ethnicity

Content Area	Grades	Female	Male	White	African American /black	Multi-Racial	Native American	Declined to State
ELA	3, 4, 5	7	0	5	2	0	0	0
ELA	6, 7	8	0	7	0	1	0	0
ELA	8, HS	6	1	6	1	0	0	0
Math	3, 4, 5	8	0	5	3	0	0	0
Math	6, 7	6	1	5	2	0	0	0
Math	8, HS	4	2	5	1	0	0	0
Science	5, 8, HS	6	0	5	1	0	0	0
Social Studies	8, HS	4	2	5	1	0	0	0
	TOTAL	**49**	**6**	**43**	**11**	**1**	**0**	**0**

Table 2.2–2.3 shows the number of years of experience, current professional roles, and highest degree of educational attainment of the committee members. Most committees were comprised of educators of special education students with 43 of 55 participants currently teaching special education students. Some of the educators served additional populations, such as visually-impaired (VI), deaf, gifted, etc.

Most of the educators listed multiple content areas that they teach; the following tables reflect the content area committee they worked in during the meeting.

Table 2.2 Years of Experience by Content Area and Grade Level

Content Area	Grades	1-5	6-10	11-15	16+	Educators of General Education Students	Educators of Special Education Students	Educators of both Special Education and General Education Students
ELA	3, 4, 5	1	1	2	3	2	2	3
ELA	6, 7	0	2	3	3	1	1	6
ELA	8, HS	0	1	3	3	2	2	3
Math	3, 4, 5	1	1	2	4	1	4	3
Math	6, 7	0	1	4	2	2	2	3
Math	8, HS	0	2	4	0	0	5	1
Science	5, 8, HS	3	1	0	2	1	3	2
Social Studies	8, HS	1	1	0	4	3	2	1
	TOTAL	**6**	**10**	**18**	**21**	**12**	**21**	**22**

Table 2.1 Committee Member Highest Degree of Educational Attainment Category Count by Content Area and Grade Level

Content Area	Grades	Bachelors	Masters	Specialist	Doctorate
ELA	3, 4, 5	1	3	3	0
ELA	6, 7	2	4	2	0
ELA	8, HS	0	2	4	1
Math	3, 4, 5	1	3	3	1
Math	6, 7	1	2	4	0
Math	HS	2	3	1	0
Science	5, 8, HS	2	2	2	0
Social Studies	8, HS	0	3	2	1
	TOTAL	**9**	**22**	**21**	**3**

Chapter 3: Workshop Roles

At any workshop, there are various leadership roles that need to be filled for the workshop to be successful. GaDOE staff, NWEA facilitators, NWEA program management staff, and committee members filled these roles during the meeting.

Georgia Department of Education Staff

Representatives from GaDOE Assessment, Curriculum, and Special Education attended the meeting to provide guidance from the state perspective. Table 3.1 lists GaDOE representatives in attendance.

Table 3.1 Roles of Georgia Department of Education Staff

Name/Division	Title	Role
Allison Timberlake/Assessment and Accountability	Deputy Superintendent of Assessment and Accountability	Supervisory
Jan Blose/Assessment and Accountability	Director of Assessment Development	Supervisory
Sandy Greene/Assessment and Accountability	Director of Assessment Administration	Supervisory
Lisa Hardman/Assessment and Accountability	Assessment Specialist	Advisory
Mary Nesbit-McBride/Assessment and Accountability	Assessment Specialist	Advisory
Tiffani Taylor/Assessment and Accountability	Assessment Specialist	Advisory
Joe Blessing/Assessment and Accountability	Program Manager	Advisory
Missy Shealy/Assessment and Accountability	Assessment Specialist	Advisory
Adrienne Walker/Assessment and Accountability	Data, Quality, and Reporting Program Manager	Advisory
Supriya Mishra/Assessment and Accountability	Assessment Specialist	Advisory
Elena Nightingale/Assessment and Accountability	Lead Psychometrician	Advisory
Anisha Donald/Curriculum and Instruction	Elementary ELA Program Specialist	Advisory
Isa Sanchez/Curriculum and Instruction	Math ESOL Program Specialist	Advisory
Jenise Sexton/Curriculum and Instruction	Mathematics Special Education Program Specialist	Advisory
Mike Wiernicki/Curriculum and Instruction	Math Program Specialist	Advisory
Laura Canepa-Redondo/Curriculum and	Science ESOL Program Specialist	Advisory
Joy Hatcher/Curriculum and Instruction	Social Studies Program Manager	Advisory
Lynn Holland/Special Education	Special Education Program Manager	Advisory
Monique Frazier/Special Education	Special Education Specialist	Advisory
Anne Myers/Special Education	Special Education Specialist	Advisory

NWEA Staff

A staff member from NWEA facilitated each grade level/content area group. The room facilitators guided the groups as they started reviewing tasks and facilitated room-level discussions. The training for facilitators included a walk-through of the Content, Bias, and Data Review presentations, and instruction on interpreting basic item statistics. Multiple training sessions were held during which facilitators could practice sharing their screens, guiding the participants through the review, and working with the notetakers (which would be selected within each room on Day 1) to clearly communicate comments that should be captured in the feedback sheets. The facilitator training also included modeling of how to utilize the training materials and strategies for engaging all participants in the meeting.

Table 3.2 shows the names of the NWEA staff and his or her role in the Content, Bias, and Data Review meeting.

Table 3.2 Roles of NWEA Staff

Name	Meeting Role
Hakan Bergon	State Director
Ricky Foust	Senior Program Manager
Molly Curtis	Program Manager
Kwang-Lee Chu	Data Review General Session Co-Facilitator
Nicole Pettit	General Session Facilitator
Catherine Lyon	ELA Facilitator: G3–G5
Mary Woo	ELA Facilitator: G6–G7
Shoshanah Dietz	ELA Facilitator: G8–HS
Anna Sewell	Math Facilitator: G3–G5
Bradley Madden	Math Facilitator: G6–G7
Shelly Vojdani	Math Facilitator: G8–HS
Beneta Brown	Science Facilitator
John Haglund	Social Studies Facilitator

Program Management

Hakan Bergon, Ricky Foust, and Molly Curtis represented NWEA's program management team.

Lindsey Amoako setup the meeting planning and coordinated between GaDOE's assessment development team and NWEA's assessment development team. She set up the educator and contract facilitator folders on SharePoint and ensured the committee participants received the non-disclosure agreements, reimbursement information, and evaluation forms.

Molly Curtis met with facilitators prior to the meeting to go over the meeting plan and to ensure everything was in place for the meeting to start. During the meeting, she helped educators and GaDOE staff with questions and support as needed.

Ricky Foust managed logistics with the meeting facility as well as made sure that facilitators had the support they needed to be successful.

Chapter 4. Meeting Materials

Training materials for the Content, Bias, and Data Review meeting included PowerPoint slides, Educator Review Checklists, feedback sheets, Data Card Statistic Descriptions, and Data Review Guiding Questions (Appendix A–G). Feedback sheets were utilized during the meeting for facilitators and notetakers to record consensus committee recommendations, and for educators to record their personal recommendations, notes, and/or comments. Educators also had access to content- and grade-specific ECS and vocabulary lists. The educators in the math committee groups had access to two sets of ECS – one set to utilize during the review of the Spring 2022 field test items (which were aligned to the GSE) and one set to utilize during the review of the newly developed tasks (which were aligned to the Georgia K-12 Mathematics Standards).

PowerPoint Presentations

On Day 1, Nicole Pettit from NWEA facilitated the training portion of the Content and Bias Review General Session (Appendix A). The PowerPoint presentation included information about the various elements of item development used when developing and reviewing GAA 2.0 tasks, such as Universal Design for Learning, content and graphic considerations, and bias and sensitivity considerations. The Content and Bias task review process was explained, as were the roles and responsibilities of the facilitators, notetakers, and reviewers.

On Day 2, a Data Review General Session was convened, and Kwang-lee Chu and Nicole Pettit from NWEA facilitated the PowerPoint presentation (Appendix D). Dr. Chu provided training on reading the data cards and interpreting item statistics (P-value, item-total correlation, and DIF). Nicole Pettit explained the Data Review process, as well as the roles and responsibilities of the facilitators, notetakers, and reviewers.

Content Materials

PDFs of the tasks were loaded onto chrome books, which were then used by educators during the meeting. The facilitator also projected the tasks via PDFs within each break-out room. Each committee member had access to the following resources on their respective chrome books:

<u>Content and Bias Review</u>

- PDFs of the grade-and content-specific TEB
- PDFs of the grade-and content-specific SB
- ECS for the applicable grade level(s) and content area. For the math rooms, the ECS were labeled _CBR (Content and Bias Review) to indicate the ECS were aligned to the new Georgia K-12 Mathematics Standards.
- Content and Bias Review Educator Checklist (Appendix B)
- Content-specific vocabulary list
- Feedback Sheet (Appendix C)

<u>Data Review</u>

- Data Review Booklet – a PDF of the tasks administered in Spring 2022, along with the associated Data Cards for each item.
- ECS for the applicable grade level(s) and content area. For the math rooms, the ECS were labeled _DR (Data Review) to indicate the ECS aligned to the GSE.
- Data Card Statistic Descriptions document describing the data card statistics and how to interpret the statistics (Appendix E).
- Data Review Guiding Questions document (Appendix F)
- Feedback Sheet

Final Committee Evaluations

A final participant evaluation was created to gather evidence of the overall effectiveness of the workshop processes. The final evaluation responses contribute to the overall validity of the GAA 2.0 and the validity of the resulting claims and targets. Committee members were asked to indicate their level of agreement with a series of statements about the workshop processes and products. They were also given an opportunity to provide any additional feedback that they wished to share about the workshop. The final evaluation also helped to gather information about committee member demographics and their areas of expertise. The final committee evaluation is included, and the results are discussed, in Chapter 5.

Chapter 5. Meeting Design Implementation

This chapter details the implementation of the three-day GAA 2.0 Content, Bias, and Data Review meeting.

Day 1

The Content and Bias Review portion of the workshop began at approximately 9 am, EST on June 14, 2022. Jan Blose from GaDOE opened the meeting, welcomed GaDOE staff, NWEA staff, and educators, discussed the agenda, meeting norms and security, and COVID protocols. Next, Dr. Blose presented the GAA 2.0 Assessment Overview portion of the PPT, describing the features, task structure, and scaffolding of the tasks.

Nicole Pettit from NWEA then delivered the Content and Bias Review training to the committee members. The committee members were instructed to review the three parts (i.e., three items) within each task — Part A, Part B, and Part C. The committee members where then asked to provide a recommendation to GaDOE for each item: accept as is, accept with revisions, or reject. The training included a discussion of elements that committee members should take into consideration while making their recommendations, as well as guiding questions to ask themselves as they reviewed, including:

- Universal Design for Learning. Do the tasks represent multiple means of representation, action, and engagement? Is the content within the tasks accessible to the widest range of students possible?
- Graphic considerations. Are the graphics as simple as possible for the task? Do captions provide concise and relevant information to students with and without visual impairments?
- Quality. Do the tasks align to the specified extended content standard? Do the tasks represent the appropriate depth and breadth of knowledge? Are the tasks written concisely, with language that is at or below grade level?
- Complexity levels. Do the three parts of the task assess different complexity levels? Is the scaffolding appropriate for the assigned task part?
- Bias and Sensitivity. Are the tasks free from issues that might bias students, or be potentially offensive to students, within both the text and images? Do the tasks respect the diversity of students within the tested population?

Information about the review materials and review process was then shared with the committee members, including how the feedback sheets would be used during the meeting. Notetakers—which would be selected in each room—would capture the consensus committee recommendations on a master feedback sheet and would include rationales for all recommendations made. Individual committee members could use their feedback sheets (accessed on the chrome books) to record their individual comments and notes as they reviewed, and to capture their personal recommendations if their recommendations differed from the consensus opinion.

After the general session, the committee members went into their respective content/grade-span breakout rooms. The committee members reviewed newly developed tasks for the remainder of Day 1.

Day 2

Kwang-Lee Chu provided the training for the data review. Dr. Chu informed the educators that they would be reviewing all three parts (i.e., items) of the tasks field tested in 2022, and that some items had no flags, some items had only one flag, and some items had multiple flags. She provided details on the meaning of each statistic, how items were flagged, and what committee members were expected to do as part of the data review process. The training included explanations of three basic item statistics:

- *P*-value - item difficulty.
- item correlation – the correlation between students' scores on individual items and their total scores.
- DIF – the comparison of two groups' performances, controlling for the students' abilities. For GAA 2.0, DIF compares the performances of male to female, and white and non-white.

These three statistics can be helpful in evaluating the appropriateness and fairness of the GAA 2.0 items because flagged items can indicate if an item may be too easy or too difficult (P-value), if an item may have been mis-keyed or have two keys (item-total correlation), or if an item may be biased based on gender or ethnicity (DIF). Dr. Chu emphasized that statistic flags do not automatically mean that items are flawed or unfair, but that reviewers should consider the flag(s) when evaluating the items. As Dr. Chu proceeded through the training, she showed reviewers where to find the specific item statistics on the data cards and explained the difference between the set of data for initial responses (students' first attempts at answering the questions correctly, without scaffolding) and post-scaffolding responses (students' second attempts at answering the questions correctly).

Next, Nicole Pettit discussed the test development process and explained how the Data Review process fits into the overall test development process. Ms. Pettit instructed the committee members to review the three items within each task (Part A, Part B, and Part C), and to make recommendations to either accept the item as is, or to reject the item. Notetakers would again capture consensus committee recommendations, comments, and rationales on a master feedback sheet, and reviewers would have their individual feedback sheets to use as needed. Ms. Pettit then walked the committee members through the process that reviewers in each break-out room would follow in reviewing the item statistics.

Upon completion of the training session, committee members joined their respective content area breakout rooms to begin reviewing the data with the NWEA facilitators.

The committee members reviewed the 2022 field test tasks until lunchtime. After lunch, some breakout rooms continued to review field test tasks, while other rooms returned to newly developed tasks after completing the Data Review portion of the day. By the end of Day 2, all break-out groups had completed the Content, Bias, and Date Review except ELA Grades 3–5 and Math Grades 3–5 groups.

Day 3

The ELA Grades 3–5 and Math Grades 3–5 groups returned on Day 3 to complete the Content and Bias Review. Reviewers in these two rooms completed reviewing all tasks by lunchtime on Thursday, June 16.

MEETING AGENDA

Table 5.1 shows the original high-level agenda for the Content, Bias, and Data Review meeting.

Table 5.1 Meeting Agenda

Georgia Alternate Assessment 2.0
Data and Content/Bias Review Agenda
June 14–16, 2022

Tuesday, 6/14/22: All Content Areas		
Time	Activity	Location
8:00–9:00	Breakfast	Peachtree Corners
9:00–10:00	Content/Bias General Session	Peachtree Corners
10:00–12:00	Content/Bias Review	Content Breakout rooms
12:00–1:00	Lunch	Peachtree Corners
1:00–4:00	Content/Bias Review	Content Breakout rooms
4:00	Dismissal	

Wednesday, 6/15/22: All Content areas		
Time	Activity	Location
8:00–9:00	Breakfast	Peachtree Corners
9:00–10:00	Data Review General Session	Peachtree Corners
10:00–12:00	Data Review	Content Breakout rooms
12:00–1:00	Lunch	Peachtree Corners
1:00–4:00	Data Review and Content/Bias Review	Content Breakout rooms
4:00	Dismissal	

Thursday, 6/16/22: ELA G3–G5, Math G3–G5, Science		
Time	Activity	Location
8:00–9:00	Breakfast	Peachtree Corners
9:00–12:00	Content/Bias Review	Content Breakout rooms
12:00–1:00	Lunch	Peachtree Corners
1:00–4:00	Content/Bias Review	Content Breakout rooms
4:00	Dismissal	

CONTENT AND BIAS WORKSHOP SUMMARY

Day 1

Within each committee room, facilitators instructed committee members to read the first task while referring to the Educator Review checklists (Appendix B) and ECS as needed. The facilitators then guided

the group through a discussion of the three items (Part A, Part B, and Part C) within the task. The notetaker recorded the group consensus on the master feedback sheets to accept, accept with revisions, or reject each item within the task; reviewers were reminded to use their personal feedback sheets as well. These steps were repeated until all tasks within the grade were reviewed, and the facilitators answered review-relevant questions throughout the process.

Days 2–3

Day 2 began with a general session for the Data Review. Once Data Review had been completed, each break-out room continued their work reviewing the tasks for the Content and Bias Review. Most content groups had completed their reviews by the end of Day 2 except for ELA Grades 3–5 and Math Grades 3–5. The two content groups that returned for Day 3 of the meeting completed the CBR by lunchtime.

Of the 231 items reviewed during the CBR, eight items were rejected by the committees. There were 123 items accepted as is, and 100 items were accepted with revisions.

Table 5.2 shows the number of items reviewed by the educators and recommended for either rejection or revision. Of the eight items rejected by the committees, five were ELA items and three were science items. The main concern for reviewers for the five ELA items was that the questions were not clear enough, or that the information within the passage needed to answer the question was too obscure, i.e., not concrete, or clear enough for students with significant cognitive disabilities. All three of the ELA content groups felt that passages for GAA should be very clear and unambiguous, and that the details within the passages should be presented in sequential order; passages that jumped in time from the past to the present were too confusing for students taking the GAA 2.0.

In Science, the reviewers felt that the three rejected items went beyond the grade-level standard and were therefore not appropriate for the GAA 2.0.

Overall, committee members felt that the strongest items had scenarios with clear details, were very strongly correlated to the tested standard, and had scaffolding that was appropriate and helpful to the student.

Table 5.2. Number of Items Reviewed and Rejected or Revised

Content Area	Grade	Number of Items Reviewed	Number of Items Rejected	Number of Items Revised
ELA	3	12	0	6
	4	12	1	3
	5	12	0	4
	6	12	2	7
	7	12	2	5
	8	12	0	3
	HS	12	0	7
Mathematics	3	12	0	2
	4	12	0	1

	5	12	0	7
	6	12	0	5
	7	12	0	3
	8	12	0	6
	HS	12	0	4
Science	5	12	0	9
	8	12	3	5
	HS	15	0	5
Social Studies	8	12	0	9
	HS	12	0	9
	TOTAL	**231**	**8**	**100**

Data review Workshop Summary

Within each committee room, facilitators instructed participants to read the first task and then review the item statistics for each of the three items (Part A, Part B, Part C) within the task. Facilitators within the three math rooms instructed reviewers to refer to the ECS specifically for Data Review, since the CBR items were aligned to the ECS for the new Georgia K–12 Mathematics Standards, while the DR items were aligned to the ECS for the GSE. Facilitators guided the committee members through evaluating the data, discussing why an item may have flagged for p-value, item-total correlation, and DIF, and exploring the possibilities for flaws within the items. The facilitators and notetakers recorded the consensus recommendations of the committee members to accept or reject the items. The steps were repeated until all tasks within the grade were reviewed. Kwang-lee Chu was available to all committee groups to answer questions as needed.

Because of the inherent interdependence of the three items within one GAA 2.0 task, all 2021 items (both flagged and unflagged) were reviewed during Data Review. Table 5.3 shows the number of items reviewed and recommended for either revision or rejection. Of the total 201 items, ultimately only three items (in ELA grades 3 and 6) were rejected. The ELA Grades 3–5 committee felt that one of the Grade 3 items (which had distractor flags and item-total correlation flags in both the initial and post-scaffolding response) required abstract thinking, and a distractor in the item was too plausible for the question asked. The committee also noted that a passage that relies on understanding of homonyms [for the Grade 3 passage, “flour” and “flower”] is too challenging for this population. Regarding the other rejected Grade 3 item (which had a distractor flag and an item-total correlation flag) in the initial response), the committee felt that distractor Option C (selected by 41% of students in the initial response) was too attractive because it contained wording used twice in the scenario.

The ELA Grades 6–7 recommended that the Grade 6 item (with a distractor flag in the initial response, item total correlation flags in both the initial and post-scaffolding response, and a DIF flag) be rejected because they felt it had multiple keys. The question asked students to “describe,” and the committee felt that all three options were plausibly correct since all three options were descriptions. The committee did not feel that there were any bias or sensitivity issues within the item.

Table 5.3 Number of Data Review Items Reviewed and Rejected

Content Area	Grade	Number of Items Reviewed	Number of Items Rejected
ELA	3	9	2
	4	9	0
	5	9	0
	6	9	1
	7	9	0
	8	9	0
	HS	12	1
Math	3	9	0
	4	6	0
	5	9	0
	6	9	0
	7	9	0
	8	3	0
	HS	15	0
Science	5	9	0
	8	9	0
	HS	24	0
Social Studies	8	9	0
	HS	24	0
	TOTAL	**201**	**4**

Evaluations

Following the completion of the Content and Bias Review meeting, committee members accessed and completed a final evaluation. Their responses show an overall successful meeting that met the stated goals and provided educators an opportunity to guide the evolution of the Georgia Alternate Assessment 2.0. The final evaluation included a variety of statements about the workshop process or products which committee members evaluated by indicating they "Strongly Agree," "Agree," "Disagree," or "Strongly Disagree." Table 5.4 summarizes quantitative results from the workshop evaluations. All committee members completed the final evaluation.

Table 5.4 Content and Bias Review Meeting Evaluation

Quantitative Results	Strongly Agree	Agree	Disagree	Strongly Disagree
I understood how Content & Bias Review fits in the big picture of the assessment.	47	5	0	0
The specific purpose of the Content & Bias Review meeting was clearly explained during the opening session.	44	8	0	0

My role on the Content & Bias Review committee was made clear to me and I understood the goals.	48	4	0	0
GaDOE staff was helpful and provided appropriate leadership.	43	9	0	0
NWEA staff was helpful and provided appropriate leadership.	41	10	1	0
The Content & Bias Review meeting was structured effectively (e.g., allotted time, technology set-up).	42	10	0	0
The facilitator was knowledgeable in the subject area.	39	13	0	0
The facilitator was well prepared and fostered a positive environment.	39	12	1	0
My content knowledge was valued and appreciated in the tasks I performed.	42	7	3	0
The Content & Bias Review meeting accomplished its purpose.	46	6	0	0
Participating in the Content & Bias Review meeting was professionally rewarding.	49	3	0	0
The work accomplished will help Georgia students.	41	9	2	0

Following the completion of the Data Review meeting, committee members accessed and completed a final evaluation. Their responses show an overall successful meeting that met the stated goals and provided educators an opportunity to guide the evolution of the Georgia Alternate Assessment 2.0. The final evaluation included a variety of statements about the workshop process or products which committee members evaluated by indicating they "Strongly Agree," "Agree," "Disagree," or "Strongly Disagree." Table 5.5 summarizes quantitative results from the workshop evaluations. All committee members completed the final evaluation.

Table 5.5 Data Review Meeting Evaluation

Quantitative Results	Strongly Agree	Agree	Disagree	Strongly Disagree
I understood how Data Review fits in the big picture of the assessment.	45	7	0	0
The specific purpose of the Data Review meeting was clearly explained during the opening session.	47	5	0	0
My role on the Data Review committee was made clear to me and I understood the goals.	46	6	0	0

GaDOE staff was helpful and provided appropriate leadership.	43	9	0	0
NWEA staff was helpful and provided appropriate leadership.	41	10	1	0
The Data Review meeting was structured effectively (e.g., allotted time, technology set-up).	43	9	0	0
The facilitator was knowledgeable in the subject area.	39	13	0	0
The facilitator was well prepared and fostered a positive environment.	39	12	1	0
My content knowledge was valued and appreciated in the tasks I performed.	42	7	3	0
The Data Review meeting accomplished its purpose.	48	4	0	0
Participating in the Data Review meeting was professionally rewarding.	45	7	0	0
The work accomplished will help Georgia students.	41	9	2	0

Appendix A. Content and Bias Review Presentation Slides

Georgia Alternate Assessment 2.0

Content and Bias Review
June 14–June 16, 2022

1

Welcome

Georgia Department of Education

Assessment and Accountability

Allison Timberlake, Ph.D., Deputy Superintendent of Assessment and Accountability
Jan Blose, Ed.D., Director of Assessment Development
Sandy Greene, Ed.D., Director of Assessment Administration
Nijia Byrd, Ed.D., Test Development Program Manager
Lisa Hardman, Ed.S., Assessment Specialist, Test Development
Tiffani Taylor, Assessment Specialist, Test Development
Joe Blessing, Assessment Administration Program Manager
Mary Nesbit-McBride, Ph.D., Assessment Specialist, Assessment Administration
Missy Shealy, Assessment Specialist, Assessment Administration
Adrienne Walker, Data, Quality, and Reporting Program Manager
Supriya Mishra, Assessment Specialist, Data, Quality, and Reporting
Elena Nightingale, Ph.D., Lead Psychometrician

2

Georgia Department of Education

Curriculum and Instruction

Breanne Huston, Ph.D., ELA Program Manager
Anisha Donald, Elementary ELA Program Specialist
Sarah Welch, ELA Program Specialist
Asha Jassani, ELA and ESOL Program Specialist
Lya Snell, Ph.D., Math Program Manager
Isa Sanchez, Math ESOL Program Specialist
Mike Wiernicki, Math Program Specialist
Karla Cwetna, Ph.D. Math Program Specialist
Keith Crandall, Science Program Specialist
Renee Shirley-Stevens, Science Content Integration and Special Ed. Specialist
Laura Canepa-Redondo, Science ESOL Program Specialist
Joy Hatcher, Social Studies Program Manager
Ashley Melville, Social Studies Program Specialist
Jennifer Zoumberis, Social Studies Program Specialist

Special Education

Monique Frazier, Special Education Specialist
Anne Myers, Special Education Specialist
Lynn Holland, Special Education Program Manager

3

NWEA

Hakan Bergon, State Director
Lindsey Amoako, Program Manager
Ricky Foust, Program Manager
Kwang-lee Chu, Senior Psychometrician
Nicole Pettit, Alternate Assessment Lead
Beneta Brown, Science Content Lead
Shoshanah Dietz, ELA Content Lead
John Haglund, Social Studies Facilitator
Shelly Vojdani, Mathematics Content Lead
Catherine Lyon, ELA Facilitator
Mary Woo, ELA Facilitator
Bradley Madden, Math Facilitator
Anna Sewell, Math Facilitator

4

Agenda

- Committee Member Forms
- Review of Meeting Schedule
- Meeting Security and Norms
- GAA 2.0 Assessment Overview
- Content and Bias Review (CBR) Overview
- Content Considerations
- Bias and Sensitivity Considerations
- CBR Process

5

Committee Member Forms

- Non-Disclosure Agreement
- Reimbursement Information
- Evaluation Form

Contact Lindsey Amoako if you have questions about the forms. lindsey.amoako@nwea.org

6

Meeting Schedule

Tue, 6/14	ELA			Math			Science	Social Studies
9:00–10:00	Content/Bias Review General Session							
10:00–12:00	Content/Bias Review			Content/Bias Review			Content/Bias Review	Content/Bias Review
	G3–G5	G6–G7	G8–H1	G3–G5	G6–G7	G8–H5	G5, G8, H5	G8–H5
12:00–1:00	Lunch			Lunch			Lunch	Lunch
1:00–4:00	G3–G5	G6–G7	G8–H1	G3–G5	G6–G7	G8–H5	G5, G8, H5	G8–H5

Wed, 6/15	ELA			Math			Science	Social Studies
9:00–10:00	Data Review General Session							
10:00–12:00	Data Review			Data Review			Data Review	Data Review
	G3–G5	G6–G7	G8–H1	G3–G5	G6–G7	G8–H5	G5, G8, H5	G8–H5
12:00–1:00	Lunch			Lunch			Lunch	Lunch
1:00–4:00	G8	G6	G8	G6	G8	G8	Data Review	Data Review
	G3–G5	G6–G7	G8–H1	G3–G5	G6–G7	G8–H5	G5, G8, H5	G8–H5

Thu, 6/16	ELA	Math	Science
9:00–12:00	Content/Bias Review	Content/Bias Review	Content/Bias Review
	G3–G5	G3–G5	G5, G8, H5
12:00–1:00	Lunch	Lunch	Lunch
1:00–4:00	G3–G5	G3–G5	G5, G8, H5

7

Meeting Security & Norms

8

Meeting Security

- Meeting materials and information are the exclusive property of GaDOE and NWEA.
- Discussions involving materials need to be limited to the meeting rooms only.
- Documents shall not be reproduced, discussed, or released or distributed to unauthorized persons in any form.

9

Meeting Security

- E-mailing, copying, printing, posting, or taking screenshots of materials is not permitted.
- Note taking during the meeting is permitted, but notes cannot leave the meeting rooms and will be destroyed after the meeting.
- Only authorized persons are permitted to view the materials.

10

Meeting Norms

- All conversations are confidential.
- Please do not use phones or other devices during the meetings.

11

COVID Protocols

All parties are working together to ensure a safe environment that aligns with protocols recommended by health experts to mitigate the spread of COVID-19.

- Meeting and dining rooms have been arranged to allow as much space as possible.
- Individual supplies (pens/pencils, etc.) will be provided for you to use and keep.
- Masks are optional.

12

COVID Protocols

By registering for the meeting, you agreed to follow an Attendee Code of Conduct:

- Do not attend if you are feeling ill, have COVID or flu-like symptoms, or have been in close contact with someone who tested positive for COVID in the past 14 days.
- Practice appropriate social distancing.
- Email to report an absence or illness to: lindsey.amoako@nwea.org

13

Share the Process

- When you return to your districts and schools, please do talk about the process we undertake, but do not disclose specific content information.
- Share the GAA 2.0 website: testing.gadoe.org

14

GAA 2.0 Assessment Overview

15

Overview

The Georgia Alternate Assessment (GAA) 2.0 is the state's alternate assessment based on alternate academic achievement standards (AA-AAAS) for those students with significant cognitive disabilities who cannot participate in the general statewide assessment program, even with maximum allowable accommodations.

16

Overview

The GAA 2.0 has been developed to ensure that students with the most significant cognitive disabilities are provided access to the state academic content standards and given the opportunity to demonstrate achievement of the knowledge, concepts, and skills inherent in the standards.

17

Design Overview

- The GAA 2.0 includes standardized and scripted tasks with multiple access points.
- Each content area assessment includes 11 tasks.
- Each task includes a scenario and three parts.
- Each part includes built-in supports and additional scaffolding, when applicable.

18

Features Overview

- Scenarios engage student interest and activate background knowledge.
- Tasks use large, simple graphics and an accessible font size.
- Short, simple sentences with directive questions are used in Part A and B items, and more complex sentences may be used in Part C items.
- Common proper nouns (names) are included based on the language of the standards or within the vocabulary list.

19

Task Structure

20

Task Structure

Scenario
- Engages student in concept of standard
- Sets stage for the following activities (Parts A–C)
- Reminder of previous learning
- May provide relevant graphics or figures

Parts A, B and C
- Items move from least complex to most complex.

Additional scaffolding, when applicable
- Scaffolding is scripted to be administered for each part as a second attempt if the student is unable to accurately complete the part on the first attempt.

21

Scaffolding

Part A
- Low complexity content
- Model/Verbal – direct step-by-step model without directly cueing the answer

Part B
- Moderate complexity content
- Gestural – covering one incorrect response option

Part C
- High complexity content
- Verbal – verbally provide additional content information (i.e., definition, re-read passage part or sentence(s) with correct answer in it; describe first step or steps to solve a problem; etc.)

22

Content and Bias Review Overview

23

Purpose

To review newly-developed tasks and make recommendations to:

- Accept (keep without changes)
- **Accept with suggested revisions**
- Reject (task/part discarded with detailed rationale)

24

Test Development Process

25

Roles and Responsibilities

- Your role is essential for verifying the appropriateness of the test items for students throughout the state.
- Each group will have a facilitator who will lead the discussion of each item.
- Using the guidelines provided, groups will discuss each item and provide recommendations to GaDOE.

26

Task Review

Content
- Aligns with Extended Content Standards (ECS)
- Appropriate for students being assessed
- No extraneous information
- Art is appropriate and accessible
- Inclusive of vocabulary needed for students using AT/AAC/ASL
- Improves opportunity to generalize vocabulary from instruction to assessment

Bias
- Fairness and equity
- Bias and sensitivity

27

Universal Design for Learning (UDL)

Multiple Means of Representation:
presenting information in a variety of ways to address the needs of different types of learners.

Multiple Means of Action and Expression:
providing a variety of ways for students to interact with the instructional materials and to demonstrate understanding.

Multiple Means of Engagement:
providing a variety of ways to engage and motivate students to learn.

UDL

28

Universal Design for Assessment

An assessment system that uses UDL includes these main principles:

- There are minimal to no barriers to student learning.
- Content is accessible for the widest range of students possible.
- Questions/content are framed within short sentences, stating the most important ideas first.

29

Universal Design for Assessment

- One idea, fact, or process is introduced at a time.
- Graphics and line drawings are simple and uncluttered.

https://nceo.info/Assessments/universal_design

30

Graphic Considerations

- Graphics (picture images, charts, graphs, etc.) should focus on relevant parts and omit details that are purely decorative or distracting.
- Picture images should be rendered as simply as possible with black line drawings.

31

Art Captions

- Art captions are descriptions of art that appear under the art in the test booklet.
- Captions should provide the same concise, relevant information to both students with and without visual impairments (VI).

Computer Plugged into an Electric Outlet

32

Example

Solar Energy

Solar Panels Storing Energy from the Sun

33

Content Considerations

34

Task Structure

35

Task Structure

Scenario

- Engages student in concept of standard
- Sets stage for the following activities (Parts A–C)
- Reminder of previous learning
- May provide relevant graphics or figures

Parts A, B and C

- Items move from least complex to most complex.

Additional scaffolding, when applicable

- Scaffolding is scripted to be administered for each part as a second attempt if the student is unable to accurately complete the part on the first attempt.

36

Scaffolding

Part A
- Low complexity content
- Model/Verbal – direct step-by-step model without directly cueing the answer

Part B
- Moderate complexity content
- Gestural – covering one incorrect response option

Part C
- High complexity content
- Verbal – verbally provide additional content information (i.e., definition, re-read passage part or sentence(s) with correct answer in it; describe first step or steps to solve a problem; etc.)

37

A quality task:

- Engages the student and is grade- and age-appropriate.
- Strongly aligns to the specified ECS and is written to assess mastery of that standard.
- Clearly communicates student expectations for each part of the task.
- Elicits the full range of responses intended or expected.

38

A quality task:

- Represents the appropriate depth and breadth of knowledge or skill.
- Supports a logical organizing principle.
- Provides a logical order to answer options, when present.
- Provides parallel answer option structure.
- Avoids clueing within a task and across tasks.

39

A quality task:

- Uses language that is at or below grade level.
- Eliminates any unnecessary or nonfunctional words.
- Avoids multiple meaning words, unless being specifically assessed.
- Has scaffolding that is appropriate and accessible based on the standard being assessed.
- Has scaffolding that does not directly cue the answer.

40

Bias and Sensitivity Considerations

41

Bias and Sensitivity Review

- Review and evaluate the tasks and graphics for:
  - fairness and equity.
  - bias and sensitivity.
- Provide recommendations to improve fairness and equity.
- Eliminate bias and sensitivity in the tasks and graphics.

42

Fairness and Equity

Fair and equitable tasks are:

- presented in the most straightforward way possible.
- engaging to students.
- composed of vocabulary for the content area that is at or below grade-level expectations.

43

Fairness and Equity

Fair and equitable tasks DO NOT:

- rely on personal knowledge gained outside the classroom.
- include content that might unfairly advantage or disadvantage any student group.

44

Bias

When considering bias issues, evaluate whether the task:

- Is appropriate in terms of familiarity, interest, age, and grade.
- Avoids stereotyping any group.
- Avoids language and vocabulary that may be considerably more familiar to some groups than others.

45

Sensitivity

When considering sensitivity issues, evaluate whether the task:

- Respects the diversity of the students participating in the assessment.
- Is sensitive to students who are not native English speakers.
- Is sensitive to students' physicality (i.e., weight, disability).

46

Sensitivity

- Avoids content offensive to any group (based on race, gender, sexual orientation, age, religion, ethnicity, socioeconomic status, holidays, or regional origin).
- Avoids language that might be offensive to any group.

47

A quality task:

- Does not provide an unfair advantage or disadvantage for any group.
- Is fair and free of bias.
- Is free of content that may be insensitive or offensive to some groups.

48

Content and Bias Review Process

49

Task Review Process

- The facilitator orients the committee and coordinates the committee members' activities specific to the content and bias review.
- The facilitator leads discussions among the committee members.
- Committee members discuss and review all parts (items) within each task.

50

Task Review Process

- Committee members make recommendations to **Accept, Accept with revisions,** or **Reject** each task part.
- Scribes will document recommendations and comments.
- After the meeting, facilitators will present and reconcile committee recommendations with GaDOE.

51

Review Materials

- GAA 2.0 Extended Content Standards
- Content and Bias Review Test Examiner Booklets (TEBs)
- Content and Bias Review Student Booklets (SBs)
- Feedback Sheets
- Content and Bias Review Educator Checklist
- Vocabulary Lists

52

Extended Content Standards

Grade 5: ELA: Reading Informational (RI)

ELA5RI1.RI.1	Determine the main idea of a text; recount the key details and explain how they support the main idea.		
Least complex ←			→ Most complex
	Part A	**Part B**	**Part C**
Make a differentiated response to provide the main idea of an informational text. Make a differentiated response to provide a key detail of an informational text.	Identify key details from an informational text which could lead to the identification of a main idea.	Identify the main idea of an informational text. Recount key details in an informational text.	Match the main idea and at least one supporting detail within an informational text. Describe the main idea off an informational text using at least one supporting detail.

53

Review Checklist

Content and Bias Review - Educator Checklist

Review of Task Scenario	
During **Content and Bias Review**, committee members determine if the **Scenario**:	
☐	Content is grade-level appropriate and relevant to the task.
☐	Vocabulary and text difficulty are grade-level appropriate.
☐	Contains subject matter appropriate for the population being assessed.
☐	Does not include extraneous information.
☐	Language generally uses simple and compound sentences with short sentence lengths. ✓ Literary text: average sentence length is 10 or fewer words in grades 3-5; 10-12 words in the upper grades. ✓ Informational text: average sentence length is 11 or fewer words in grades 3-5; 11-13 words in the upper grades.
☐	Will not interfere with the student's ability to demonstrate knowledge or understanding.
☐	Scenario meets the specifications for bias and sensitivity. ✓ Avoids stereotyping any group ✓ Is sensitive to students who are not native English speakers ✓ Avoids language that may be considerably more familiar to some groups than others ✓ Avoids language that might be offensive to a particular group ✓ Is sensitive to the way a group has been represented over time ✓ Is sensitive to students' physicality (i.e., weight, disability) ✓ Is accessible to students from different geographical locations ✓ Does not give unfair advantage to one group over another

54

Feedback Sheet

Year	Subject	Grade	DN	Part	Standard	DOK	Key	A/R/R	All Revisions	Comments
2022	Social Studies	HS	GAI.E1E.SS001	Scenario	SS.HS.E.A	1	A			
2022	Social Studies	HS	GAI.E1E.SS001	Part A	SS.HS.E.A	1	A			
2022	Social Studies	HS	GAI.E1E.SS001	Part B	SS.HS.E.A	1	A			
2022	Social Studies	HS	GAI.E1E.SS001	Part C	SS.HS.E.A	1	A			
2022	Social Studies	HS	GAI.E1E.SS002	Scenario	SS.HS.E.TL	1	A			
2022	Social Studies	HS	GAI.E1E.SS002	Part A	SS.HS.E.TL	1	A			
2022	Social Studies	HS	GAI.E1E.SS002	Part B	SS.HS.E.TL	1	A			
2022	Social Studies	HS	GAI.E1E.SS002	Part C	SS.HS.E.TL	3	B			
2022	Social Studies	HS	GAI.E1E.SS003	Scenario	SS.HS.EM	1	A			
2022	Social Studies	HS	GAI.E1E.SS003	Part A	SS.HS.EM	1	A			
2022	Social Studies	HS	GAI.E1E.SS003	Part B	SS.HS.EM	1	B			
2022	Social Studies	HS	GAI.E1E.SS003	Part C	SS.HS.EM	3	C			
2022	Social Studies	HS	GAI.E1E.SS004	Scenario	SS.HS.E.Cc	1	A			
2022	Social Studies	HS	GAI.E1E.SS004	Part A	SS.HS.E.Cc	1	A			
2022	Social Studies	HS	GAI.E1E.SS004	Part B	SS.HS.E.Cc	1	B			
2022	Social Studies	HS	GAI.E1E.SS004	Part C	SS.HS.E.Cc	2	B			

55

Task Review Process

- As a group, review and discuss the tasks.
- Come to a consensus (if possible) as to whether each item will be Accepted, Accepted with Revisions, or Rejected.
- Facilitators will capture edits in the projected PDF, and scribes will record committee suggestions and comments.
- Facilitators will provide comments and suggestions to GaDOE.

56

Expectations of All Reviewers

- Seek clarification and ask questions.
- Listen to and collaborate with other committee members.
- Support high expectations for task quality and of student ability.
- Provide honest and constructive feedback.
- Focus on desired outcomes and provide actionable recommendations

57

Thank You!

~

Questions?

58

Appendix B. Content and Bias Review Educator Checklist

nwea

Content and Bias Review - Educator Checklist

Review of Task Scenario	
During **Content and Bias Review**, committee members determine if the **Scenario**:	
☐	Content is grade-level appropriate and relevant to the task.
☐	Vocabulary and text difficulty are grade-level appropriate.
☐	Contains subject matter appropriate for the population being assessed.
☐	Does not include extraneous information.
☐	Language generally uses simple and compound sentences with short sentence lengths. ✓ Literary text: average sentence length is 10 or fewer words in grades 3-5; 10-12 words in the upper grades. ✓ Informational text: average sentence length is 11 or fewer words in grades 3-5; 11-13 words in the upper grades.
☐	Will not interfere with the student's ability to demonstrate knowledge or understanding.
☐	Scenario meets the specifications for bias and sensitivity. ✓ Avoids stereotyping any group ✓ Is sensitive to students who are not native English speakers ✓ Avoids language that may be considerably more familiar to some groups than others ✓ Avoids language that might be offensive to a particular group ✓ Is sensitive to the way a group has been represented over time ✓ Is sensitive to students' physicality (i.e., weight, disability) ✓ Is accessible to students from different geographical locations ✓ Does not give unfair advantage to one group over another

Review of Task Parts	
During **Content and Bias Review**, committee members determine if the task **Parts**:	
Part A ☐ Part B ☐ Part C ☐	Meet requirements for **alignment to the extended standards**. ✓ Task part aligns to assigned extended standards. ✓ Task part contains subject matter and language that is grade-level appropriate. ✓ Task part maintains the progression of complexity across the three parts. ✓ Task part meets bias and sensitivity guidelines (see above).
Part A ☐ Part B ☐	Reflect best practices for **technical quality**. ✓ The language, including the script, used in the task part is clear and concise. ✓ The key is correct, and no other choices are correct.

Part C ☐	✓ The source of rigor in the task part derives from the knowledge and skills required to answer the item correctly (i.e., not from construct-irrelevant sources). ✓ No answer option is conspicuous or grammatically incorrect. (Unless grammatical correctness is being assessed – ELA) ✓ The task part reflects best practices for universal design. ✓ Scaffolding improves content access without providing the answer.
Part A ☐	Reflect best practices for **art**. ✓ Art is appropriate to the standard and task. ✓ Art has a corresponding caption that is appropriate. ✓ Art is acceptable for students who may have a visual impairment. ✓ Consideration to visual support in Part C.
Part B ☐	
Part C ☐	

Scaffolding Considerations	
During **Content and Bias Review**, committee members ensure that:	
☐	Scaffolding provided for each task does not directly cue the answer
☐	Scaffolding for each task part is appropriate and accessible based on the standard being assessed
☐	Scaffolding provides the most support in Part A, decreasing support through Part B, and the least amount of support in Part C.

Universal Design/Braille Considerations	
During **Content and Bias Review**, committee members ensure that:	
☐	Graphics focus on the relevant parts and omit details that are purely decorative or distracting.
☐	One idea, fact, or process introduced at a time in the task parts.
☐	Keys and legends clearly defined if needed.
☐	No vertical or diagonal text.
☐	No extraneous graphs or pictures that are irrelevant to what is being tested; graphics are related to task/part.
☐	Avoid map items that depend on visually recognizable outlines of continents, countries, or states.
☐	Items that depend on reading of graphic representations also have verbal/textual descriptions.

Appendix c. Sample Feedback Sheet

Year	Subject	Grade	UIN	Part	Standard	DOK	Key	A/AR/R	Art Revisions	Comments and Rationales
2022	Social Studies	8	GA08SS23001	Scenario	SS8CG5b			Accept Accept with Revisions Reject		
2022	Social Studies	8	GA08SS23001	Part A	SS8CG5b	1	B			
2022	Social Studies	8	GA08SS23001	Part B	SS8CG5b	1	A			
2022	Social Studies	8	GA08SS23001	Part C	SS8CG5b	2	A			
2022	Social Studies	8	GA08SS23002	Scenario	SS83Ea					
2022	Social Studies	8	GA08SS23002	Part A	SS83Ea	1	A			
2022	Social Studies	8	GA08SS23002	Part B	SS83Ea	1	C			
2022	Social Studies	8	GA08SS23002	Part C	SS83Ea	2	C			
2022	Social Studies	8	GA08SS23003	Scenario	SS8CG1c					
2022	Social Studies	8	GA08SS23003	Part A	SS8CG1c	1	B			
2022	Social Studies	8	GA08SS23003	Part B	SS8CG1c	1	A			
2022	Social Studies	8	GA08SS23003	Part C	SS8CG1c	2	A			
2022	Social Studies	8	GA08SS23004	Scenario	SS8G1b					
2022	Social Studies	8	GA08SS23004	Part A	SS8G1b	1	A			
2022	Social Studies	8	GA08SS23004	Part B	SS8G1b	2	C			
2022	Social Studies	8	GA08SS23004	Part C	SS8G1b	2	B			

Appendix d. Data Review Presentation Slides

Georgia Alternate Assessment 2.0

Data Review
June 15, 2022

1

Agenda

- Meeting Schedule
- Data Review Overview
- Training on Statistics and Flags
  - Understand and interpret item statistics
  - Apply knowledge of item statistics to evaluate items
- Data Review Process

2

Meeting Schedule

Wed, 6/15	ELA			Math			Science	Social Studies
9:00–10:00	Data Review General Session							
10:00–12:00	Data Review			Data Review			Data Review	Data Review
	G3–G5	G6–G7	G8–H5	G3–G5	G6–G7	G8–H5	G5, G8, H5	G8–H5
12:00–1:00	Lunch			Lunch			Lunch	Lunch
1:00–4:00	DR	DR	DR	DR	CGR	CBR	Data Review	Data Review
	G3–G5	G6–G7	G8–H5	G3–G5	G6–G7	G8–H5	G5, G8, H5	G8–H5

3

GAA 2.0 Data Review Overview

4

Definition and Purpose

Definition: the process of reviewing items for quality and appropriateness based on item statistics

Purpose:

- To review the Spring 2022 field test tasks and their statistics and assign each item into a category.
  - **Accept** (keep without changes)
  - **Reject** (task/part discarded with detailed rationale)
- Flagged items have one or more features that fall outside of established statistical bounds and should be reviewed.

5

Test Development Process

6

Roles and Responsibilities

Your role is essential for:

- Verifying the appropriateness of the test items for students throughout the state.
- Establishing evidence that the items are performing as expected.

Each group will have a facilitator who will lead the discussion of each item based on summary data.

7

Training on Statistics and Flags

8

Training on Statistics and Flags

The following statistics are provided:

- Item difficulty (*p*-value)
- Item-total correlation
- Differential item functioning (DIF)
- Parts are called items for GAA 2.0

All Parts of field-tested tasks will be reviewed.

- Some may not have any flags.
- Some may have one flag.
- Some may have multiple flags.

9

Training on Statistics and Flags

Knowing how to interpret these statistics will allow you to use the relevant data to evaluate:

- Whether an item is of appropriate difficulty,
- Whether the item distinguishes between those who know the content and those who do not,
- Whether an item may be biased.

10

Data Card Example

GAA 2.0 Data Card

Administration Information

Year	Subject	Grade	Form
2013	ELA	3	1

Item Information

Item Identifier	Content Standard	DOK	Task	Part	Key
GAA2.ELA.3.1_v1	ELA.ELA3.W.w4	2	11	C	B

Item Statistics

	N	p-value	Item-Total Correlation	DIF-Male/Female		DIF-Ethnicity	
				Category	Favored	Category	Favored
Value	541	0.51	0.57	A		A	
Flag							

Score Points	Percent, %
0	10
1	30
2	40
No Response	0

Distractor Analysis

Student Responses	Initial Response				Post-Scaffolding Response			
	Percent, %	Percent Flag	Item-Total Correlation	Item-Total Flag	Percent, %	Percent Flag	Item-Total Correlation	Item-Total Flag
A*	51		0.61		50		0.55	
B	22		-0.11		22	Y	-0.01	
C	22		-0.14		28		-0.19	
No Response	4		-0.25		0		.043	
Total N	506				506			

11

P-Values on the Data Card

The item *p*-value is shown in the item statistics table.

Item Statistics

	N	p-value	Item-Total Correlation	DIF-Male/Female		DIF-Ethnicity	
				Category	Favored	Category	Favored
Value	541	0.51	0.57	A		A	
Flag							

12

P-Value

- The *p*-value indicates the item difficulty.
- The *p*-value is the average item score earned by all students divided by the item's total possible points.

13

Visualizing *P*-Values

- *p-value* = 0.0 means all students answered this item incorrectly. This is a hard item.
- *p*-value=1.0 means all students answered this item correctly. This is an easy item.

14

P-Value Flag

	P-Value Flagging Criteria	Indicates
p-value	• Part A: has two options If *p*-value of correct response < 0.40 • Parts B & C: have three options If *p*-value of correct response < 0.30	Difficult Item
	All parts: If *p*-value of correct response > 0.95	Easy Item

15

P-Value Question for You

Item 1, *p*-value = 0.40
Item 2, *p*-value = 0.60
Item 3, *p*-value = 0.80

Which item is the hardest?
Which item is the easiest?

16

Item-Total Correlations on the Data Card

The item-total correlation is shown in the item statistics table.

Item Statistics

	N	p-value	Item-Total Correlation	DIF-Male/Female		DIF-Ethnicity	
				Category	Favored	Category	Favored
Value	541	0.51	0.57	A		A	
Flag							

17

Item-Total Correlation

- The item-total correlation is the correlation between students' scores on an item and their total scores.
- Ranges from -1 to 1
- It indicates how well an item distinguishes between low- and high-performing students.
  - An item with a high correlation indicates that:
    - Students who do well on the test tend to answer the item correctly.
    - Students who do poorly on the test tend to answer the item incorrectly.

18

Visualizing Item-Total Correlation

When item-total correlation is **positive**, it means that the high-performing students tended to select the correct answer choice.

19

Visualizing Item-Total Correlation

When item-total correlation is **negative**, it means that the high-performing students tended to select the incorrect answer choice.

20

Item-Total Correlation Flag

	Item-Total Correlation Flagging Criteria	Indicates
Item-Total Correlation	If item-total correlation of correct response < 0.20	Poorly discriminating item

21

Item-Total Correlation

What does the item-total correlation tell us?

- correlation = 0.35
  - The item differentiates between low- and high-performing students.
- correlation = -0.15
  - Is the item mis-keyed?
- correlation = 0
  - Does the item measure content differently from the rest of the items?

22

Distractor Analysis on the Data Card

The percentage of selection options are shown in the distractor analysis table.

Distractor Analysis: Part C Item

Student Responses	Initial Response				Post-Scaffolding Response			
	Percent, %	Percent Flag	Item-Total Correlation	Item-Total Flag	Percent, %	Percent Flag	Item-Total Correlation	
A*	41		0.45		36		0.55	
B	25		-0.11		32		-0.01	
C	29		-0.14		26	Y	-0.15	
No Response	4		-0.73		6		-0.82	
Total N	531				306			

23

Item-Total Correlations on the Data Card

Their item-total correlations are shown in the distractor analysis table.

Distractor Analysis: Part C Item

Student Responses	Initial Response				Post-Scaffolding Response			
	Percent, %	Percent Flag	Item-Total Correlation	Item-Total Flag	Percent, %	Percent Flag	Item-Total Correlation	Item-Total Flag
A*	41		0.45		36		0.55	
B	25		-0.11		32		-0.01	
C	29		-0.14		26	Y	-0.15	
No Response	4		-0.73		6		-0.82	
Total N	531				306			

24

Distractor Flags

	Distractor Flagging Criteria	Indicates
Distractor Analysis	If % of correct response < % of distractor	Possible mis-key
	Initial Response • Part A: If % of distractor > 40 • Parts B & C: If % of distractor > 30 Post-Scaffolding • Parts A & B: If % of distractor > 40 • Part C: If % of distractor > 30	Possible second correct option
	If item-total correlation of a distractor is > 0.05	Possible second correct option

25

DIF on the Data Card

The DIF is shown in the item statistics table.

Item Statistics

	N	p-value	Item-Total Correlation	DIF-Male/Female		DIF-Ethnicity	
				Category	Favored	Category	Favored
Value	1,638	0.50	0.40	B	F	A	
Flag							

26

Differential Item Functioning (DIF)

DIF compares two groups' performances controlling for the students' ability.

- Reference and focal groups are defined.
- Significant DIF indicates the item may measure achievement differently across groups matched by student ability.

27

Differential Item Functioning (DIF)

DIF	Focal	Reference
Male/Female	Female (F)	Male (M)
Ethnicity (White/Non-white)	Non-white (Non-W)	White (W)

Note: DIF is computed when each group has about 200 students.

28

Differential Item Functioning (DIF)

- DIF is classified into three categories:
  - A = no DIF (difference is negligible)
  - B = moderate DIF
  - C = severe DIF
- DIF categories B and C does not mean that the item is biased. It does mean the item needs to be examined closely at this meeting to determine whether it is biased.

29

DIF on the Data Card

The DIF is shown in the item statistics table.

Item Statistics

	N	p-value	Item-Total Correlation	DIF-Male/Female		DIF-Ethnicity	
				Category	Favored	Category	Favored
Value	1,638	0.50	0.40	B	F	A	
Flag							

This example shows female students had higher performances than their counterparts and the difference is moderate (B). The DIF between ethnicity subgroups is negligible, so no Favored group is shown in the table.

30

Differential Item Functioning (DIF)

A question to ask in the presence of DIF:

Is there anything in the task/part that may cause the two groups to perform differently on the item?

31

DIF on the Data Card

What do the following examples mean?

DIF-Male/Female		DIF-Ethnicity	
Category	Favored	Category	Favored
B	M	A	

DIF-Male/Female		DIF-Ethnicity	
Category	Favored	Category	Favored
A		C	Non-W

32

Flagging Criteria Review

	Flagging Criteria	Indicates
p-value	• Part A: If *p*-value of correct response < 0.40 • Parts B & C: If *p*-value of correct response < 0.30	Difficult item
	If *p*-value of correct response > 0.95	Easy item
Item-Total Correlation	If item-total correlation of correct response < 0.20	Poorly discriminating item

33

Flagging Criteria Review

	Flagging Criteria	Indicates
Distractor Analysis	If % of correct response < % of distractor	Possible mis-key
	Initial Response • Part A: If % of distractor > 40 • Parts B & C: If % of distractor > 30 Post-Scaffolding • Parts A & B: If % of distractor > 40 • Part C: If % of distractor > 30	Possible second correct option
	If item-total correlation of a distractor is > 0.05	Possible second correct option

34

Flagging Criteria Review

	Flagging Criteria	Indicates
DIF	If ETS DIF classification is B or C	Possible bias in item

35

Data Review Process

36

Data Review Process

- The facilitator will present and explain the materials that will be used during Data Review.
- The facilitator will coordinate the committee members' activities specific to the data review.
- The facilitator will lead the discussion among the committee members.

37

Data Review Process

- Committee members discuss and review all task parts (three related items) within each task.
- Committee members make recommendations to **Accept** or **Reject** each task part.
- Scribes will document recommendations and comments.
- After the meeting, facilitators will present and reconcile committee recommendations with GaDOE.

38

Data Review Materials

- Data Review Booklet
- Data Review Feedback Sheet
- GAA 2.0 Extended Content Standards
- Data Card Statistic Descriptions
- Guiding Questions for Data Review

39

Extended Content Standards

Grade 1: ELA: Reading Informational (RI)

ELGRI.1.RI.2	Determine the main idea of a text; recount the key details and explain how they support the main idea.		
Least complex ←		→ Most complex	
	Part A	**Part B**	**Part C**
Make a differentiated response to provide the main idea of an informational text. Make a differentiated response to provide a key detail of an informational text.	Identify key details from an informational text which could lead to the identification of a main idea.	Identify the main idea of an informational text. Recount key details in an informational text.	Match the main idea and at least one supporting detail within an informational text. Describe the main idea of an informational text using at least one supporting detail.

40

Data Card Statistic Descriptions

Georgia Alternate Assessment 2.0
Data Card Statistic Descriptions

The intent of these criteria is to flag items that should be reviewed, but these flags are not necessarily indicative of an issue with the item. There are many reasons the experts may want to keep an item in spite of statistical flags.

Terms	Flagging Criteria	Explanation
p-value	**Too difficult** Part A < 0.40 Parts B and C < 0.30 **Too easy** Any Parts > 0.95	The p-value is the average item score divided by maximum possible points. Consequently, the p-value of an item is always between 0 and 1. Higher values indicate more students got higher scores and the item is easy. The flagging criterion is different because the number of options is different for Part A versus Parts B and C.
Item-Total Correlation	**Low correlation** Any parts < 0.20	Item-total correlation is the connection between the scores of an item and overall test scores. It is based on the assumption that students with higher abilities have better chances of answering an item correctly and earning higher raw scores. Low item-total correlation values indicate that a correct answer does not relate to how well students did on the test.

41

Guiding Questions for Data Review

- Which extended content standard is being measured by this task/part?
- Is the task/part of an appropriate difficulty for the grade that is being assessed?
- Does the task/part have any technical flaws?

42

Guiding Questions for Data Review

- Is there one and only one correct answer?
- Does the task/part differentiate based on ability?
- Is there anything in the task/part that may cause the focal group and the reference group to perform differently on the item?

43

Data Review Process

- As a group, review and discuss the tasks.
- Come to a consensus (if possible) as to whether each item will be Accepted or Rejected.
- Facilitators will record consensus committee recommendations.
- Facilitators will provide committee comments and recommendations to GaDOE.

44

Expectations of All Reviewers

- Seek clarification and ask questions.
- Listen to and collaborate with other committee members.
- Support high expectations for task quality and of student ability.
- Provide honest and constructive feedback.
- Focus on desired outcomes and provide actionable recommendations

45

Final Reminders

- Use the item statistics to aid your professional judgment, not to automatically approve or reject an item.
- Use the content and context of the item along with the item statistics to guide your decision.

46

Thank You!

~

Questions?

47

Appendix E. Data Card Statistic Descriptions

Georgia Alternate Assessment 2.0
Data Card Statistic Descriptions

The intent of these criteria is to flag items that should be reviewed, but these flags are not necessarily indicative of an issue with the item. There are many reasons the experts may want to keep an item in spite of statistical flags.

Terms	Flagging Criteria	Explanation
p-value	**Too difficult** Part A < 0.40 Parts B and C < 0.30 **Too easy** Any Parts > 0.95	The *p*-value is the average item score divided by maximum possible points. Consequently, the *p*-value of an item is always between 0 and 1. Higher values indicate more students got higher scores and the item is easy. The flagging criteria are different because the number of options are different for Part A versus Parts B and C.
Item-Total Correlation	**Low correlation** Any parts < 0.20	Item-total correlation is the connection between the achievement on an item and achievement on the overall test. It is based on the assumption that students with higher abilities have better chances of answering an item correctly and earning higher total scores. Low item-total correlation values indicate that a correct answer does not relate to how well students did on the test.
Distractor Analysis	1. Any distractors have a higher response rate (%) than the correct answer, or 2. Initial Response • Part A > 40 • Parts B and C > 30, or 3. Post-scaffolding Response • Parts A and B > 40 • Part C > 30. High Item-Total Correlation • Any parts > 0.05	The distractor analysis examines the statistics of the incorrect response options. High response rates and high item-total correlations indicate item distractors should be reviewed to assess whether a) the answer key is assigned correctly, or b) the item has multiple correct answers. Note that other item-level statistics should be considered when reviewing distractor analysis.
DIF Gender and Ethnicity	A = no DIF B = moderate DIF C = severe DIF	DIF is used to investigate whether an item measures achievement differently between two subgroups of students (e.g., male vs. female). Different flags are used to indicate the degree of the presence of DIF from no DIF (A) to severe DIF (C).

Appendix F. Guiding Questions for Data Review

Guiding Questions for Data Review

Why am I here?

- Before each task is field-tested, other educator committees review it for content and bias/sensitivity.
- After each task is field-tested, the same tasks are reviewed again with their data.

What is my task?

- Review all parts within the tasks.
- Use the item statistics to aid your professional judgement, rather than to automatically approve or reject an item.
- Use the context of the item <u>and</u> the item statistics to guide your decision.
- Limit content judgments to item alignment to the extended content standards.
- Try to avoid instructional validity judgments. (e.g., Should we test this?)

Questions you should consider for Data Review:

- Which extended content standard is being measured by this task/part?
- Is the task/part of an appropriate difficulty for the grade that is being assessed?
- Does the task/part have any technical flaws?
- Is there one and only one correct answer?
- Does the task/part differentiate based on ability? (Remember – the greater the item total correlation, the more it differentiates.)
- Is there anything in the task/part that may cause students in the focal group and students in the reference group to perform differently on the item?

A Report on the Standards Validation for the Georgia High School Alternate Assessment 2.0

November 21, 2022

Submitted by:
Chad Buckendahl, Ph.D.
402.770.0085
cbuckendahl@acsventures.com

Contents

Chapter 1: Introduction and Purpose *3*

Chapter 2: Panelists *4*

Chapter 3: Process *5*

Methodology 5

Conducting the Workshop 6

Chapter 4: Results *8*

Round 1 Ratings 8

Round 2 11

Vertical Articulation 14

Chapter 5: Evaluating the Standard Setting Study *15*

Chapter 6: Recommendations *17*

References *18*

Appendices *19*

Appendix A – Panelists Demographics 19

Appendix B – Workshop Agenda 22

Appendix C – Orientation and Training Slides 23

Appendix D – Readiness Survey Form 39

Appendix E – Sample Validations Rating Form 40

Appendix F – Evaluation Survey Results 41

Chapter 1: Introduction and Purpose

Under the Every Student Succeeds Act (ESSA) of 2015 and the Individuals with Disabilities Education Act (IDEA) of 2004, all states are required to implement high-quality alternate assessments for students with significant cognitive disabilities. These alternate assessments must be aligned to the academic content standards designed for all students in English language arts (ELA) and mathematics in grades 3 through 8, once in High School, and in science in at least one grade in each of the 3 – 5, 6 – 8, and 9 – 12 grade ranges. The Georgia Department of Education (GaDOE) contracted with Questar Assessment (now NWEA) to conduct a standards validation study of the High School assessments within the Georgia Alternate Assessment 2.0 (GAA 2.0) system. The goal of a standards validation is to evaluate existing cut scores associated with achievement levels that describe student performance while remaining anchored in the content of what students at each level know and can do. Questar (NWEA) contracted with ACS Ventures (ACS) to design and implement a standards validation setting study to evaluate the existing threshold performance levels for achievement on the High School GAA 2.0.

Important assumptions of a standards validation study are that the policy-level definitions of student performance remain consistent, the target population of students remains stable, and the intended construct being measured is comparable to the assessment on which the original achievement standards were established. For the High School GAA 2.0, the policy Achievement Level Descriptors (ALDs) and the target population of students remain unchanged; however, the range ALDs have changed to reflect a shift in content. The intended constructs of English language arts (ELA), mathematics, science, and social studies measured by each High School assessment, respectively remain constant, but beginning with the 2022 administration of the GAA 2.0, the content representation changed with some course content being removed.

Because the current achievement levels were established in 2019, conducting an exploratory standard setting study does not promote consistency in the system. However, because the assessments have undergone changes to the content representation, evaluation of the current locations of the cut scores that distinguish achievement levels is warranted. A standards validation study that systematically evaluates the cut scores is used to make recommendations about whether the existing cut scores should continue to be used or revised. The *Standards for Educational and Psychological Testing* states that "…when proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly" (AERA, APA, & NCME, 2014, Standard 5.21).

This report provides an overview of the standard setting process along with the recommendations that resulted from the standards validation study.

Chapter 2: Panelists

A total of 35 educators from Georgia participated in the standards validation study. The panelists were recruited in spring 2022 through efforts by NWEA and input from GaDOE with the intent to be representative of the diversity of educators and districts across the state. Additionally, special attention was placed on the background of the panelists to ensure the panel would consist of special educators, content-area teachers, program administrators, school and district level administration, and post-secondary educators. Several of the panelists also took part in the July 2019 standard setting process.

Panelist demographic information was collected in advance of the study and as part of an onsite survey during the workshop. This survey collected information about panelist professional experience, educational experience, as well as the certifications and current positions held. As shown in Table 1, the panel had representation from a range of degrees earned and teaching experience. Summary demographic information is in Appendix A.

Table 1. Key Panelist Demographic Information

Panelist Demographic Information		
Highest Degree Received	**# of panelists**	**%**
Bachelors	5	14.3%
Specialist	8	22.9%
Masters	17	48.6%
Doctorate	5	14.3%
Years of Teaching Experience*	**# of panelists**	**%**
Less than 5 years	1	2.9%
5 to 10 years	10	29.4%
11 to 20 years	10	29.4%
20 or more years	13	38.2%
Gender*	**# of panelists**	**%**
Female	26	76.5%
Male	8	23.5%
Race/Ethnicity*	**# of panelists**	**%**
American Indian or Alaskan Native	1	2.9%
Black or African American	9	26.5%
White	24	70.6%
Region	**# of panelists**	**%**
CE	2	6.7%
CW	3	10.0%
Metro	9	30.0%
NE	5	16.7%
NW	3	10.0%
SC	1	3.3%
SE	2	6.7%
SW	5	16.7%
Not applicable**	5	16.7%

*One panelist declined to state **Region not applicable for higher education and transition specialists

Chapter 3: Process

Methodology

The standards validation study utilized an Ordered Item Booklet (OIB) that is associated with item-mapping methods within standard setting literature. In the OIB, each score point was presented to panelists in order from the empirically easiest to the most difficult for students who took the assessment in spring 2022. Following discussion of the threshold characteristics of the range achievement level descriptors, panelists reviewed the existing cut scores and determined whether any changes were needed to the cut scores that were established in July 2019. Those standards were determined using the Item-Descriptor Matching (ID Matching) methodology.

The primary reason for conducting the standards validation was to determine if the updates made to the GAA 2.0 blueprints had any effect on the performance standards that were established during the 2019 standard setting. The blueprint updates were associated with changes to the range Achievement Level Descriptors (ALDs) and with modifications to the content assessed on the GAA 2.0. Updates to the High School ALDs for ELA, mathematics, science, and social studies were finalized in August 2021 and reflected on the 2022 administration of the GAA 2.0.

In advance of the study, NWEA completed a psychometric review of the scaling impacts on the scoring of the assessments and the density of the items to construct an Ordered Item Booklet (OIB) based on the empirical data collected from the 2022 administration. The primary purpose of this review was to evaluate whether the empirical evidence supports the existing cut scores and that the results represent student achievement in the four content areas: ELA, mathematics, science, and social studies. The results helped determine critical elements of the OIB. For example, the information to put on the OIB, formats that ease the use of OIB. When a cut score was absent on the OIB, the closest higher score would be used for computing impact data and contextual information (scale score of the absent cut, observed items in standard error range) was highlighted on the item map and in training to support the judgement task.

Once the statistical review was completed by NWEA, the focus of the validation workshop was for panelists to review the content and location of achievement levels for the current assessments. Each panel evaluated whether the content associated with the threshold expectations for each achievement level is still appropriate to support interpretation and use of the cut score when also considering item performance and impact data. The work was completed with participants being divided into four panels based on content-area expertise, which resulted in four panels (one for each content area) consisting of 8-9 panelists each.

In contrast to the initial cut score determination, which considered the whole range of performance present in the ALDs, this standards validation process focused on the level of determination between Level 1 and Level 2, between Level 2 and Level 3, and between Level 3 and Level 4. To do so, panelists were presented with the range ALDs specific to the content area and asked to identify the descriptors that represent a student performing at the *threshold* of each achievement level. Panelists were then instructed to use these threshold ALDs, the OIB, and the empirical item performance data to evaluate whether the location of the current cut scores reasonably represented the expectations present in the updated blueprints. To facilitate the meeting, panelists were provided with the following information:

- A summary of how the original cut scores were determined
- Copies of the range and threshold ALDs for each content area
- Copies of scoring guides/rubrics to describe score points in the current assessments
- Locations of existing interim cut scores for each assessment content area

The study occurred over two days and included a vertical articulation process that was completed on the last day of the workshop. The agenda for the workshop is provided in Appendix B.

Conducting the Workshop

General Orientation and Panelist Training

The workshop began on Monday, July 18th with a general orientation session for all 35 panelists. A welcome and introductions were offered, and GaDOE provided an overview of the timeline of GAA 2.0, the structure of the GAA 2.0, and changes to the High School assessments. Following this, Dr. Buckendahl, the lead facilitator, provided an overview of the purpose of the study:

- Develop GAA 2.0 thresholds for ALDs at the Level 2, Level 3, and Level 4 performance levels
- Recommend cut scores for those threshold ALDs
- Communicate recommendations to policymakers who will make the final decision

After reviewing the purpose of the study, the lead facilitator provided participants an explanation of the role of cut scores, and why the GaDOE was completing a standards validation study. There are two primary reasons for completing this study:

1. Changes have been made in the assessment blueprints
2. Changes have been made to the ALDs that were used to develop the existing cut scores

Additional explanation was provided to participants on how standard setting fits within the normal test development model, prior standard setting methodology that was used, and how standards *validation* differs from standard *setting*. It was also noted to participants there is no evidence indicating that the cut scores are problematic. However, because changes were made to the assessments, the GaDOE wants to ensure that the interpretation of the cut scores and achievement levels continue to remain meaningful.

Dr. Allison Timberlake with GaDOE began the workshop by sharing the purpose of the study and providing panelists with an overview of the GAA 2.0 assessment components and scoring process. This overview also described how the assessment was developed based on alternate achievement standards that were meant to provide students with significant cognitive disabilities the opportunity to demonstrate their knowledge and skills. Special attention was paid to how the tasks are scored.

After the general orientation, Dr. Buckendahl trained panelists on the validation methodology and provided an overview of how panelists would be expected to complete the process outlined above. This training reviewed how the standard validation activities of the workshop fit into the overall standard setting process, use of the Ordered Item Booklet (OIB) and Item Map, use of the rating sheets, and the purpose of validating the cut score at all three achievement level thresholds. The training also included a review of understanding item difficulty, understanding cognitive complexity, the difference between each, and how they impact the panelists' evaluation of the current cut score. The training also showed how impact data would be used to evaluate the group's recommended cut scores using data from the 2022 administration to estimate the percentage of students that would be classified into each achievement level. Finally, the training included a discussion on the importance of meeting security protocols and participants' confidentiality agreements. The full orientation and training presentation can be found in Appendix C.

In addition to the large-group training, content area representatives from NWEA provided the individual panels with an overview of the specific changes made to the assessment blueprints and updated ALDs. This overview covered which domains where affected, the percent to which they were altered, and what content replaced areas that were eliminated. Content areas shifted at differential rates with ELA having changes of 10%, Mathematics, 30%, Science, 50%, and Social Studies, 50%. This means that the potential for influence on panelists' perceptions of the cut scores may have been differentially observed.

Threshold ALDs

Following the orientation and overview of the assessment, each panel followed a systematic process for the creation of the threshold ALDs. First, the panelists reviewed the range ALDs to develop an understanding of the content and achievement expectations of the assessment. After this review, panelists were tasked with identifying descriptors within the range ALDs that could be used to demonstrate a student is performing at the threshold of each performance level.

The facilitator led a discussion that focused on the threshold level expectations identified by panelists. Facilitators recorded when the panelists reached a consensus on the descriptors that best represented threshold student performance for each achievement level. Threshold ALDs were created for Level 2, Level 3, and Level 4 (no threshold was needed for the lowest level since there is no cut score that needs to be evaluated). This threshold ALD document was shared with the panelists to be used when evaluating the existing cut scores.

Cut Score Evaluations

Panelists were asked to independently review at minimum several items in the OIB that appeared before and after the locations of the current cut scores and to determine whether those items reflected a reasonable representation of the range of performance present at the transition between each achievement level. The range of items around each cut score was determined using the conditional standard error of measurement (CSEM) at each score point. The range was determined by identifying items within 1 CSEM of each current cut score. In addition to the OIB, panelists were also provided item maps that included all items in the OIB along with an estimate of their item difficulty. The item maps, in combination with the OIBs and the current cut scores, allowed panelists to directly observe how each item could be linked via its difficulty level to a specific level. Panelists used this information in combination with the range ALDs and the threshold descriptors for each level to determine if items would still be considered appropriate representations of the knowledge and skills identified in the threshold descriptions. For each cut score, panelists provided a rating indicating if they believed the current recommended cut score is appropriate for the updated ALDs; if they denoted it did NOT, they were asked to indicate where they believe the cut score should be placed. Panelists were told that if they provided ratings that were within the CSEM range identified, the ratings would effectively validate the current locations. An example of the Validation Ratings Form used to collect these ratings can be found in Appendix E.

After an initial round of ratings, panelists were provided summary data from the round (individual recommendations, group mean and median OIB page number, and the distribution of recommendations), as well as impact data demonstrating the estimated percent of students that would be categorized at each performance level based on the panel's initial recommendation (median OIB page number). The impact data was estimated using data collected from the 2022 administration of the assessment. After reviewing these sources of evidence, the panelists discussed the impact of their first-round ratings and the rationale behind each judgment. Following this discussion the panelists completed a second round of ratings, which allowed them the opportunity to independently revise any of their previous judgments. These final ratings were then reviewed again as a panel.

With the completion of these activities, each panel had reviewed the items associated near each cut point and reached a judgment on whether the location of the cut scores is consistent with threshold expectations or whether they recommended any modifications to the current cut scores. The results of those reviews and recommendations are included in the following section of this report.

Chapter 4: Results

Round 1 Ratings

After panelists completed making their Round 1 judgments for the page numbers that represented each achievement level, ACS compiled the data into an Excel worksheet for each content area. Each score point (i.e., 1 or 2) on the assessment was presented to panelists over two pages of the OIB. This means that the OIB would have 120 pages to represent 60 score points. In other words, a page number recommendation of 30 represents the point value of 15 on the assessment. These page numbers represented the cut scores that panelists recommended for the threshold performance levels and the rating forms were given back to the panelists individually for reference in the second round of judgments. Facilitators presented the minimum, maximum, range, median and mean of the data (OIB page numbers), the distribution of individual panelist recommendations, as well as the impact data based on the overall panel's recommendations (median). These data were also reviewed in comparison to current interim cut scores. Table 2, Table 3, and Table 4 display the mean and median values of the Round 1 results for each assessment content area, along with the current interim cut scores for each assessment. The distribution of recommended cut scores used as feedback can be found in Figures 1 – 4 below. A sample Validation Ratings Form can be found in Appendix E.

Table 2. Round 1 Recommended Page Numbers of Cut Scores for Level 2

Level 2					
Test	Current Cut Score	Round 1 Recommendations			
		Min	Max	Mean	Median
ELA	30	30	32	30.3	30
MATH	26	22	34	26.9	26
SCI	36	22	36	33.5	36
SOC	30	26	36	29.1	28

Table 3. Round 1 Recommended Page Numbers of Cut Scores for Level 3

Level 3					
Test	Current Cut Score	Round 1 Recommendations			
		Min	Max	Mean	Median
ELA	90	90	90	90.0	90
MATH	84	68	106	82.9	82
SCI	82	42	84	72.9	82
SOC	72	66	76	70.0	72

Table 4. Round 1 Recommended Page Numbers of Cut Scores for Level 4

Level 4					
Test	Current Cut Score	Round 1 Recommendations			
		Min	Max	Mean	Median
ELA	120	118	120	119.3	120
MATH	> 120	94	120	112.2	116
SCI	> 120	60	120	103.5	118
SOC	106	90	106	100.4	106

Figure 1. English Language Arts Round 1 Distribution of the Page Numbers of Cut Score Recommendations from Panelists

Figure 2. Mathematics Round 1 Distribution of the Page Numbers of Cut Score Recommendations from Panelists

Figure 3. Science Round 1 Distribution of the Page Numbers of Cut Score Recommendations from Panelists

Figure 4. Social Studies Round 1 Distribution of the Page Numbers of Cut Score Recommendations from Panelists

After data were presented to panelists, the facilitators led the panelists in a discussion of the results. Panelists discussed whether the results were consistent with their expectations, the rationale behind the ratings they provided, the characteristics of items that they viewed as critical to their ratings and how they mapped the items to the threshold ALD descriptions. Note that because there were fewer shifts for some content areas (e.g., ELA), this may help to explain the earlier consensus in the panelists' recommendations; while content areas with larger changes (e.g., Mathematics, Science) had greater variation in the initial round of judgments.

Round 2

Following the Round 1 recommendations and group discussions, each panelist reviewed the OIB, impact data, and their previous recommendations to provide their Round 2 recommendations. Round 2 data was again compiled by ACS and presented to the panelists at the end of the workshop as part of a vertical articulation activity. Table 5, Table 6, and Table 7 display the mean and median values for the page numbers in the OIB as they were presented to panelists.

Table 5. Round 2 Recommended Page Numbers of Cut Scores for Level 2

Level 2					
	Current	Round 2 Recommendations			
Test	Cut Score	Min	Max	Mean	Median
ELA	30	30	32	30.4	30
MATH	26	22	34	26.0	26
SCI	36	22	36	34.0	36
SOC	30	26	30	28.0	28

Table 6. Round 2 Recommended Page Numbers of Cut Scores for Level 3

Level 3					
	Current	Round 2 Recommendations			
Test	Cut Score	Min	Max	Mean	Median
ELA	90	90	90	90.0	90
MATH	84	74	84	79.8	80
SCI	82	82	82	82.0	82
SOC	72	66	72	69.0	68

Table 7. Round 2 Recommended Page Numbers of Cut Scores for Level 4

Level 4					
	Current	Round 2 Recommendations			
Test	Cut Score	Min	Max	Mean	Median
ELA	120	118	120	119.3	120
MATH	> 120	106	116	109.6	108
SCI	> 120	118	118	118.0	118
SOC	106	102	106	105.0	106

Figures 5 – 8 below show the estimated distribution of students in each performance category for each of the assessments based upon their Round 2 ratings.

Figure 5. English Language Arts Estimated Percent of Students in each Performance Category

Figure 6. Mathematics Estimated Percent of Students in each Performance Category

Figure 7. Science Estimated Percent of Students in each Performance Category

Figure 8. Social Studies Estimated Percent of Students in each Performance Category

Vertical Articulation

After the Round 2 ratings were reviewed, the lead facilitator in collaboration with GaDOE and NWEA staff led the panelists through the vertical articulation process. In standard setting, vertical articulation is usually completed to help evaluate the consistency and trends of impact across grade levels within a given content area. In this part of the workshop, the focus of the review was not solely on cross-grade articulation of achievement levels and included coherence across content areas. Panelists were asked to review the consistency of the impact data across the four content areas that were being reviewed during the workshop. While this does represent a slight modification, much of the policies and practices could be followed, with the panelists looking for the relationship across content areas rather than grade level. During the workshop, panelists were presented with the estimated impact data based upon the Round 2 recommendations for all four assessments. Panelists were able to provide recommendations for changes to cut scores and immediately observe how the change would impact the impact data for the given assessment. During this review, panelists were given some key policies that needed to be followed during the process, including:

- Panelists were told that there should be a compelling reason to change a cut score if the recommendation was within the CSEM range around the existing cut scores
- If a change was recommended, the group was asked to provide a written rationale for the change
- Each panel reviewed their impact data along with the cut score recommendation and then presented their recommendations for changes (or no changes) with any rationale for the change

The vertical articulation panel did consider revisions to the cut score; however, all the potential changes that were considered fell within the CSEM error bands identified for the cut scores. In addition, an additional factor that the vertical articulation panel did consider was that for most Round 2 recommendations, the subject matter expert panel reached their recommendations with a strong degree of consensus. For two of the cut scores across subject areas, the subject matter committee recommended modifications that were discussed with the cross-grade groups (see Table 8) and for the others, the range of recommendations was narrow (see Table 10 below). As a result, the vertical articulation panels final recommendations were to maintain all existing cut scores because even the two recommended modifications were minor and within the CSEM.

The vertical articulation panel continued to anchor their judgments in the content of the achievement level descriptors while also considering the content changes that occurred for the assessments with comments about the nature of those content changes and the expected effect on student achievement. It is also important to note that for some assessments, the exact cut scores were not always represented in the OIB. For example, one of the two instances where a panel recommended a change, the recommended cut score that emerged out of Vertical Articulation and going into Policy Review was as close as could be selected to the existing cut score without going beyond the existing cut score. Table 8 shows the existing cut score, the recommended cut score from Round 2 of the standards validation process, and then the recommended scale score based on the discussion and consensus of the Vertical Articulation panel.

Table 8. Final Cut Score Recommendations by Page Number and Scale Score

GAA 2.0 Recommended Cut Scores										
Test		**Current Cut Score**			**Round 2 Recommended Cut Score**			**Vertical Articulation**		
		Level 2	**Level 3**	**Level 4**	**Level 2**	**Level 3**	**Level 4**	**Level 2**	**Level 3**	**Level 4**
ELA	*Cut*	30	90	120	30	90	120	--	--	--
	Scaled	1350	1400	1439	1351	1402	1441	1351	1402	1441
MATH	*Cut*	26	84	>120	26	80	108	--	--	--
	Scaled	1350	1400	1442	1357	1399	1433	1351	1401	1435
SCI	*Cut*	36	82	>120	36	82	118	--	--	--
	Scaled	1350	1400	1478	1350	1404	1453	1351	1402	1479
SOC	*Cut*	30	72	106	28	68	106	--	--	--
	Scaled	1350	1400	1434	1347	1389	1437	1351	1398	1437

Chapter 5: Evaluating the Standard Setting Study

Throughout the study, Kane's (2001) framework for evaluating standard setting studies was applied. Kane (1994) suggested three sources of evidence should be considered in the validation process: external, procedural, and internal. Evidence for each of these sources is discussed in this section.

External Evidence

As external evidence, the evaluation of the assessment and scale properties in advance was conducted to determine whether the assessment had the capacity to support the existing cut scores and potential changes that might be recommended during the standards validation study. This statistical review of performance was completed by NWEA in advance of constructing the OIBs for the study. This review evaluated the stability of item and test form performance since the time the existing cut scores were determined and supported the consistency of scores on the test over this period of time.

Procedural Evidence

Procedural evidence can be observed through the selection of qualified panelists, selection and application of appropriate methods and procedures, and panelists' feedback provided throughout the standard setting workshop. One important aspect of the procedural validity evidence is the completion of readiness surveys by panelists. During the workshops, all panelists completed a readiness survey prior to beginning their Round 1 ratings to signify their understanding and readiness to proceed with the study. The facilitators reviewed the readiness survey results to ensure all panelists indicated individual readiness and understanding prior to beginning the Round 1 rating process. Across content areas, all panelists indicated that they understood the task and were prepared to complete their ratings prior to beginning the rating process.

Panelists also completed surveys designed to evaluate how well-prepared they felt to provide their ratings and whether they felt the ratings were appropriate. Panelists completed evaluation surveys at the conclusion of each day of the study. A summary of the panelists' responses for one question are provided in Table 9 below: *How confident were you about your [day] recommendations for achievement levels for students?* The table displays the distribution of responses for the consecutive evaluations. Based upon the survey feedback, it appears most panelists felt the training on the standard setting process was complete, and 100% were at least somewhat

confident with the overall cut score recommendations. Some variation in confidence on the first day could be attributed to the differential changes to the content for the assessments with content areas that had fewer changes resulting in initial higher levels of confidence. The results from all evaluation surveys can be found in Appendix F.

Table 9. Survey of Panelist Confidence

Confidence in Recommendations								
	Not at all Confident		Not Very Confident		Somewhat Confident		Very Confident	
Evaluation Survey	N	%	N	%	N	%	N	%
Day 1 (N=33)	0	0%	2	0.1%	13	39.4%	18	54.5%
Day 2 (N = 31)	0	0%	0	0%	6	19.4%	25	80.6%

Internal Evidence

The primary source of internal validity evidence can be observed when looking at the variability of the cut point recommendations. For each recommended cut score, the standard error was calculated and displayed along with the associated cut score in Table 10. The variability of the standard error did decline as panelists moved from the first to the second. This general reduction in standard error between rounds and across assessments is indicative of an increased degree of agreement across panelists and reinforces the consensus of the respective groups.

Table 10. Standard Error

GAA 2.0 Standard Error						
	Level 2		Level 3		Level 4	
Test	Median Cut Score	SE_{Mean}	Median Cut Score	SE_{Mean}	Median Cut Score	SE_{Mean}
English Language Arts						
Round 1	30.0	18.6	90.0	9.3	120.0	1.0
Round 2	30.0	0.9	90.0	0.0	120.0	1.0
Mathematics						
Round 1	26.0	3.9	82.0	10.5	116.0	8.3
Round 2	26.0	3.5	80.0	2.5	108.0	3.7
Science						
Round 1	36.0	4.9	82.0	16.4	118.0	23.5
Round 2	36.0	4.9	82.0	0.0	118.0	0.0
Social Studies						
Round 1	28.0	3.0	72.0	3.6	106.0	7.5
Round 2	28.0	1.5	68.0	2.1	106.0	1.9

Chapter 6: Recommendations

At the conclusion of the study, the cut score recommendations for each assessment were provided to GaDOE for a policy review with district leaders, educators, and other stakeholders and population experts. Based upon the evidence collected and the review of the performance of panelists, the study resulted in recommendations for the cut scores supported by educators, higher education and transition specialists, and subject matter experts. These recommendations provide exact validation of most existing cut scores and secondary validity evidence of the two remaining cut scores wherein recommendations represent a slight change to the current cut score. The results of the standards validation study were carried forward to GaDOE leadership and the policy review committee and were approved as validation of the current cut scores.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004). https://www.congress.gov/108/plaws/publ446/PLAW-108publ446.pdf

Appendices

Appendix A – Panelists Demographics

Name	Education	Certification field(s)	Current position	Year of Experience	Grade level(s) taught	Specialization
Panelist 1	Masters		Asst Director of Transition Services	12		Transition to Employment SWPS
Panelist 2	Masters	Social Science (6-12), Special Education (K-12)	Social Science Teacher	19	9th-12th	
Panelist 3	Specialist	ECE	Special Ed self-contained	28	PreK, 2nd, 9th-12th, Self-contained	Ed Leadership, Co-teaching, Gen Curr.
Panelist 4	Doctorate	Special Education	Higher Education of Special Education	25	K-12, Inclusive PSE	Inclusive Post-Secondary Education, Intellectual Disabilities, Autism
Panelist 5	Masters	SPED Adaptive, SPED Academic, SPED General Curriculum	M/D Special Education Teacher	9	9th-11th	
Panelist 6	Masters	History, GenEd/SPED, Adaptive	Teacher/SPED Leadership	10	9th-12th	
Panelist 7	Bachelors	SPED PK-12	SPED teacher High School adapted curriculum	25	9th-12th+	Reading Endorsement
Panelist 8	Bachelors	SPED K-12 Math, ELA, SS, Science, Adapted Curriculum	Self-contained	8	9th-12th	
Panelist 9	Masters	Early childhood/Special education	Director of Assessments	10	5th-12th	
Panelist 10	Masters	Special Education, Adapted Curriculum	Alternate Access Support Teacher	13	K-5, 9-12	
Panelist 11	Specialist	Adaptive Special Education K-5	Special Education HS SID/PID	7	HS 9th-12th	

Panelist 12	Masters	Special Education Content Areas (ELA, Math, Sci, SS), SPED General & Adapted Curriculum	In-School Coordinator	15	PK-12th	
Panelist 13	Bachelors	General and Adapted SPED Math, Science, Social Studies, ELA	Resource Adapted Special Education	7	9th-12th	
Panelist 14	Doctorate	English High School, Gifted, AP Literature, AP Language	Teacher	8	11th-12th	
Panelist 15	Bachelors	ELA & Social Studies	Data and Technology Specialist	9	5th & 9th-12th	Reading Endorsement
Panelist 16	Masters	Early childhood	MD/MI/AU 9th-12th	20	9th-12th	MO/MI/AU
Panelist 17	Specialist	Adapted Curriculum K-8, Ed Leadership, IRR SPED, Multiple Severe Cognitive Disabilities	Orthopedic Impairment Lead Teacher	25	6th-8th	All SPED except vision/hearing
Panelist 18	Specialist	Special Ed/Alt. Assessed	Special Ed/Alt Assessment	17	9th-12th	Special Ed
Panelist 19	Doctorate	Tier II Leadership	Coordinator/Reading Intervention/Instructional Coordinator	20	K-12 & Adults	Special Education
Panelist 20	Masters	SPED Academic Content Concentrations, English 6-12, SPED Adapted Curriculum	HS SPED teacher (Moderate Intellectual Disabilities)	12	K-2, 6-8, 9-12	Autism Spectrum Disorder, Moderate/Severe Disabilities
Panelist 21	Masters	SPED/all science 6-12/AP+IB Biology/Ed.S. instructional technology	Interrelated teacher	10	6th-12th science - physical, biology, earth systems	
Panelist 22	Bachelors	Middle ELA/Sci/Adaptive/SPEC Gen. Curriculum	High School SPED Teacher/Lead	22	9th-12th	
Panelist23	Masters	Science 6-12; ESOL	Teacher	29	9th-12th	Mortuary Science

Panelist 24	Masters	Special Education/General Curriculum PK-12	AVID	10	9th-12th	BD
Panelist 25	Specialist	Math 6-12, Science 6-8, Social Studies 6-8	Assessment, Accountability, & Strategy Coordinator	19	6th-12th	
Panelist 26	Specialist	Interrelated & Adaptive	Special Education Coordinator	18	K-12	GAA - Access
Panelist 27	Masters	SPED Adapted, SPED General Curriculum	Self-contained 6th-8th	15	PreK-12	Autism
Panelist 28	Doctorate	ELA 6th-12th	Faculty	33	Graduate, Undergraduate , 9th-12th	Ed Leadership, Reading, ESOL
Panelist 29	Specialist	Speech Language Pathology, PK-12 Adaptive Curricula, PK-5 Gen. Ed, Tier I Leadership	Support specialist for SPED	25	PK-12th	
Panelist 30	Masters	Secondary Math Education	Instructor	38	Middle School (6 yrs), HS (2 yrs), College (30 years)	
Panelist 31	Masters	Adaptive & General SPED & Elem Ed (K-5)	High School SPED	4	K, 9th-12th+	
Panelist 32	Doctorate	Special Ed Adaptive Curricula K-8	Director of Higher Education Program	20	9th-12th, 1-3 (ESY)	
Panelist 33	Specialist	Adaptive SPED & Gen. SPED	SPED Adaptive Teacher	15	9th-12th	MI/MO
Panelist 34	Masters	Certified Rehabilitation Counselor	Transition Coordinator			
Panelist 35	Masters	Early childhood (SPED)	SPED Teacher	23	K-12	

Appendix B – Workshop Agenda

Georgia Department of Education
July 18 – 19, 2022

GAA 2.0 Standards Validation Agenda

Agenda	Activities
Day 1: July 18th	
Day 1 Check In 7:30 – 9:00	Breakfast and Check-in
Welcome and Overview 9:00 – 10:15	Confidentiality and Nondisclosure Form Orientation to Standards Validation
Break 10:00 – 10:15	Morning break
Split into Content Area Groups 10:30 – Noon	Review blueprint and changes to standards Review and Discuss Achievement Level Descriptors (ALDs) Review sample task as a group
Lunch Noon – 1:00	Lunch break
Evaluating Existing Cut Scores 1:00 – 3:45	Review Content in Ordered Item Booklet (OIB) Orientation to OIB Item Map Begin Making Initial Recommendations Complete Initial Recommendations First Evaluation
Day 1 Wrap-up 3:45 – 4:00	First Evaluation and Collection of Materials
Day 2: July 19th	
Day 2 Check In 7:30 – 9:00	Breakfast and Check-in
Feedback and Discussion of Initial Recommendations 9:00 – 11:45	Group Level Results Impact Data Consensus Recommendations Individual Final Recommendations
Day 2 Wrap-up 11:45 – 12:00	Second Evaluation and Collection of Materials
Lunch 12:00 – 1:00	Lunch break
Vertical Articulation 1:00 – 3:00	Discussion

Georgia Alternate Assessment 2.0
(GAA 2.0)
High School
Standards Validation Meeting
July 18-19, 2022
Richard Woods, Georgia's School Superintendent | Georgia Department of Education | Educating Georgia's Future
Opening Session
Allison Timberlake
Deputy Superintendent, Assessment & Accountability
Georgia Department of Education
Educating Georgia's Future by graduating students who are ready to learn, ready to live, and ready to lead.
Participants
• Georgia Educators
• High school
• Special and general education teachers
• District special education and assessment directors
• University System of Georgia and Technical College System of Georgia
• Competitive Integrated Employment
Richard Woods, Georgia's School Superintendent | Georgia Department of Education | Educating Georgia's Future
Support Staff
Georgia Department of Education
Assessment and Accountability
Curriculum and Instruction
Special Education
Governor's Office of Student Achievement
NWEA
ACS Ventures
Richard Woods, Georgia's School Superintendent | Georgia Department of Education | Educating Georgia's Future

What is Standard Setting?

- An iterative process by which panels of content-matter experts recommend cut scores, by grade and content area, for each of the achievement levels on the GAA 2.0 assessments.
- Achievement levels give meaning and context to scale scores.
  - What knowledge and skills must a student demonstrate to perform at each level?
  - What scale score must a student earn to "meet" the alternate achievement standards?
- The State School Superintendent and the State Board of Education review the recommended results from the meeting and make the final decision on the cut scores.

Standard setting is a policy-setting exercise that requires the input of content experts. As such, you are *recommending* important policy for our state based on your content expertise, but you are *not* the policy makers – the Superintendent and State Board set policy.

Richard Woods, Georg

Standards Validation

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Achievement Levels

- **Level 1**: Students at this level demonstrate a **limited** understanding of the knowledge and skills specified in Georgia's Extended Content Standards and **may need substantial academic support** as they transition to the next grade/course, inclusive post-secondary education, or competitive integrated employment.
- **Level 2**: Students at this level demonstrate a **partial** understanding of the knowledge and skills specified in Georgia's Extended Content Standards and **may need frequent academic support** as they transition to the next grade/course, inclusive post-secondary education, or competitive integrated employment.
- **Level 3**: Students at this level demonstrate an **adequate** understanding of the knowledge and skills specified in Georgia's Extended Content Standards and **may need occasional academic support** as they transition to the next grade/course, inclusive post-secondary education, or competitive integrated employment.
- **Level 4**: Students at this level demonstrate a **thorough** understanding of the knowledge and skills specified in Georgia's Extended Content Standards and **may need little to no academic support** as they transition to the next grade/course, inclusive post-secondary education, or competitive integrated employment.

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Achievement Level Descriptors

- Range ALDs are narrative descriptions of the knowledge and skills expected at each of the four achievement levels that have been established for the GAA 2.0.
- Revised range ALDs were developed for each content area by committees of Georgia educators in October and November 2020.
- Range ALDs are based on the state-adopted content standards and extended content standards.
- Range ALDs lay the groundwork for setting or validating standards.

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Types of ALDs

- **Policy ALDs** – High-level descriptors that are the same for all content areas and grade levels
- **Range ALDs** – Grade-level, content-specific descriptors which link back to the standards
- **Threshold ALDs** – Specifically define what it takes for a student to attain each achievement level
- **Reporting ALDs** – Summary of Range ALDs in public friendly language for reports provided to parents and the public

Overview

The Georgia Alternate Assessment (GAA) 2.0 is the state's alternate assessment based on alternate academic achievement standards (AA-AAAS) for those students with significant cognitive disabilities who cannot participate in the general statewide assessment program, even with maximum allowable accommodations.

GAA 2.0 Test Design

- The GAA 2.0 includes standardized and scripted tasks with multiple access points.
- Each content area assessment includes 11 tasks.
- Each task includes a scenario and three parts.
- Each part includes built-in supports and additional scaffolding, when applicable.

GAA 2.0 Test Design

- The scenario – passages, graphs, images, etc. – contains content on which the three parts will be based.
- The levels of complexity are aligned with levels 2, 3, and 4 of the extended content standard.
- If a student answers incorrectly, the student is provided with additional scaffolding and another opportunity to respond.

GAA 2.0 Test Design

Part A

- Low complexity content
- **Model/Verbal** – direct step-by-step model without directly cueing the answer

Part B

- Moderate complexity content
- **Gestural** – covering one incorrect response option

Part C

- High complexity content
- **Verbal** – verbally provide additional content information (i.e., definition, re-read passage part or sentence(s) with correct answer in it; describe first step or steps to solve a problem; etc.)

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

GAA 2.0 Test Design

- Administration
  - Administered by teachers in a one-on-one setting.
  - Teachers have flexibility to customize materials, within guidelines, based on individual student needs (e.g., cutting up responses, using manipulatives).
  - Students use their established communication method(s) to respond (e.g., written, verbal, eye gaze)
  - The entire six-week state window can be used for administration.
  - Students can take multiple breaks, as needed.

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Revised HS GAA 2.0 Development Process

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Updated HS GAA 2.0 Design English Language Arts

Blueprint prior to 2021

Claim/Target	Standards Assessed	Weight
Students comprehend text in increasingly complex ways.		**50%**
Target 1: Analyze and draw inferences from the text, citing textual evidence as support.	ELAGSE 11-12.RL.1	
	ELAGSE11-12.RI.1	
Target 2: Determine at least two themes or central ideas of a text, analyze their development over the course of the text, and summarize the text.	ELAGSE11-12.RL.2	
	ELAGSE11-12.RI.2	
Target 3: Analyze the impact of the author's choices regarding how to develop and relate literary elements.	ELAGSE11-12.RL.3	
Students produce writing for a range of purposes and audiences.		**20%**
Target 1: Write arguments to support claims.	ELAGSE9-10.W.1	
Target 2: Write informative/explanatory texts.	ELAGSE9-10.W.2	
Students demonstrate an understanding of the function and meaning of language.		**30%**
Target 1: Use various types of phrases and clauses to convey meaning.	ELAGSE9-10.L.1b	
Target 2: Use context to determine the meaning of words and phrases.	ELAGSE9-10.L.4a	
Target 3: Use reference materials to determine a word's pronunciation, meaning, and/or part of speech.	ELAGSE9-10.L.4c	

Current Blueprint w/ High School Redesign

Claim/Target	Standards Assessed	Weight
Students comprehend text in increasingly complex ways.		**50%**
Target 1: Analyze and draw inferences from the text, citing textual evidence as support.	ELAGSE11-12.RL.1	10%
	ELAGSE11-12.RI.1	10%
Target 2: Determine at least two themes or central ideas of a text, analyze their development over the course of the text, and summarize the text.	ELAGSE11-12.RL.2	10%
	ELAGSE11-12.RI.2	10%
Target 3: Analyze the impact of the author's choices regarding how to develop and relate literary elements.	ELAGSE11-12.RL.3	10%
Students produce writing for a range of purposes and audiences.		**20%**
Target 1: Write arguments to support claims.	ELAGSE11-12.W.1	10%
Target 2: Write informative/explanatory texts.	ELAGSE11-12.W.2	10%
Students demonstrate an understanding of the function and meaning of language.		**30%**
Target 1: Produce work that shows accurate spelling and correct use of the conventions of punctuation and capitalization.	ELAGSE11-12.L.2c	10%
Target 2: Use context to determine the meaning of words and phrases.	ELAGSE11-12.L.4a	10%
Target 3: Use reference materials to determine a word's pronunciation, meaning, and/or part of speech.	ELAGSE11-12.L.4c	10%

Updated HS GAA 2.0 Design Mathematics

Blueprint prior to 2021

Claim/Target	Standards Assessed	Weight
Students demonstrate increasingly complex understanding of number sense.		**20%**
Target 1: (Algebra I) Rewrite expressions and solve equations involving square roots.	MGSE.N.RN.2	
	MGSE.A.REI.4b	
Students demonstrate increasingly complex spatial reasoning and understanding of geometric principles.		**30%**
Target 1: (Geometry) Understand and use geometric properties of two- and three-dimensional shapes.	MGSE.G.MG.1	
Target 2: (Geometry) Solve problems involving area, perimeter, and volume.	MGSE.G.GMD.3	
Target 3: (Geometry) Understand congruence and similarity as related to transformations.	MGSE.G.SRT.2	
Students demonstrate increasingly complex understanding of measurement, data, and analytic procedures.		**20%**
Target 1: (Algebra I) Represent and interpret data displays.	MGSE.S.ID.6a	
	MGSE.S.ID.6c	
Students solve increasingly complex mathematical problems using algebraic thinking.		**30%**
Target 1: (Algebra I) Use operations to simplify expressions and solve equations using algebraic properties and models.	MGSE.A.CED.1	
Target 2: (Algebra I) Interpret graphs of functions.	MGSE.A.REI.6	
	MGSE.F.IF.4	

Current Blueprint w/ High School Redesign

Claim/Target	Standards Assessed	Weight
Students demonstrate increasingly complex understanding of number sense.		**20%**
Target 1: Rewrite expressions and solve equations involving square roots.	MGSE.N.RN.2	10%
	MGSE.A.REI.4b	10%
Students demonstrate increasingly complex understanding of measurement, data, and analytic procedures.		**30%**
Target 1: Represent and interpret data displays.	MGSE.S.ID.1	10%
	MGSE.S.ID.6a	10%
	MGSE.S.ID.6c	10%
Students solve increasingly complex mathematical problems using algebraic thinking.		**50%**
Target 1: Use operations to simplify expressions and solve equations using algebraic properties and models.	MGSE.A.CED.1	20%
Target 2: Interpret graphs of functions.	MGSE.A.REI.6	10%
	MGSE.F.IF.4	10%
Target 3: Understand the relationship between two quantities.	MGSE.F.BF.1a	10%

Updated HS GAA 2.0 Design Science

Blueprint prior to 2021

Claim/Target	Standards Assessed	Weight
Students explain the properties of solutions and the factors that affect reactions involving solutions.		**30%**
Target 1: (Physical Science) Develop and use models to describe properties of solutions and investigate how temperature, surface area, and agitation affect rates that solutes dissolve in a solvent.	SPS6a	
	SPS6b	
Target 2: (Physical Science) Plan and conduct experiments to determine pH of household products.	SPS6e	
Students use evidence to explain motion of objects in support of Newton's laws.		**20%**
Target 1: (Physical Science) Use evidence to describe Newton's laws of motion.	SPS8b	
Target 2: (Physical Science) Explain the benefits of simple machines.	SPS8d	
Students explain or use models to demonstrate the role of cellular structures in maintaining homeostasis and the relationship between structure and function within a cell.		**20%**
Target 1: (Biology) Recognize that cell structures interact as a system to maintain homeostasis and determine the role of cellular transport in maintaining homeostasis.	SB1a	
Target 2: (Biology) Investigate the role of cellular transport in maintaining homeostasis. (e.g., active, passive, and osmosis).	SB1d	
Students describe the interrelationships between organisms and the environment within an ecosystem.		**30%**
Target 1: (Biology) Develop and use a model to describe cycling of matter and flow of energy in an ecosystem.	SB5b	
Target 2: (Biology) Predict an organism's ability to survive when an ecosystem changes and design a basic solution to reduce the impact of human activity on the environment.	SB5d	
	SB5e	

Current Blueprint w/ High School Redesign

Claim/Target	Standards Assessed	Weight
Students explain or use models to demonstrate the role of cellular structures in maintaining homeostasis, the relationship between structure and function within a cell, and the cycling of matter through photosynthesis and respiration.		**30%**
Target 1: Recognize that cell structures interact as a system to maintain homeostasis and determine the role of cellular transport in maintaining homeostasis.	SB1a	10%
Target 2: Investigate the role of cellular transport in maintaining homeostasis (e.g., active, passive, and osmosis).	SB1d	10%
Target 3: Use a model to show the cycling of matter and flow of energy through photosynthesis and cellular respiration within a single-celled organism.	SB1e	10%
Students use models to explain common ancestry and differences across groups of organisms.		**30%**
Target 1: Use models to draw conclusions about common ancestry among major groups of organisms.	SB4b	10%
Target 2: Use evidence to explain the similarities and differences of viruses and other organisms.	SB4c	20%
Students describe the interrelationships between organisms and the environment within an ecosystem.		**40%**
Target 1: Develop and use a model to describe cycling of matter and flow of energy in an ecosystem.	SB5b	20%
Target 2: Predict an organism's ability to survive when an ecosystem changes and design a basic solution to reduce the impact of human activity on the environment.	SB5d	10%
	SB5e	10%

Updated HS GAA 2.0 Design Social Studies

Blueprint prior to 2021

Claim/Target	Standards Assessed	Weight
Students describe key events and documents that contributed to the development of the United States.		**20%**
Target 1: Describe English settlement and colonization.	SSUSH1a	
Target 2: Explain key features of the United States Constitution.	SSUSH5c	
Students evaluate conflicts that influenced the development of the United States.		**30%**
Target 1: Evaluate the growing economic disparity between the North and the South.	SSUSH9a	
Target 2: Explain the importance of the Reconstruction Amendments.	USH10c	
Target 3: Analyze the origins of U.S. involvement in World War II.	SSUSH19a	
Students explain the economics of personal finance.		**50%**
Target 1: Create a personal budget.	SSEPF1b	
Target 2: Understand, compare, and explain the concepts of banks, credit, and insurance.	SSEPF2a	
	SSEPF4a	
	SSEPF5a	
Target 3: Identify skills necessary for success in the workplace.	SSEPF6a	

Current Blueprint w/ High School Redesign

Claim/Target	Standards Assessed	Weight
Students describe key events that contributed to the development of the United States.		**40%**
Target 1: Describe English settlement and colonization.	SSUSH1a	10%
Target 2: Examine the influence of key inventions on U.S. infrastructure including, but not limited to, the telegraph, telephone, and electric lightbulb.	SSUSH11c	10%
Target 3: Describe the causes, including overproduction, underconsumption, and stock market speculation, that led to the stock market crash of 1929 and the Great Depression.	SSUSH17a	10%
Target 4: Explain the social and political impact of widespread unemployment that resulted in developments such as Hoovervilles.	SSUSH17c	10%
Students explain key documents that contributed to the development of the United States.		**30%**
Target 1: Explain key features of the United States Constitution.	SSUSH5c	10%
Target 2: Explain how objections to the ratification of the Constitution were addressed in the Bill of Rights.	SSUSH5e	10%
Target 3: Explain the importance of the Reconstruction Amendments.	SSUSH10c	10%
Students evaluate conflicts that influenced the development of the United States.		**30%**
Target 1: Evaluate the growing economic disparity between the North and the South before the Civil War.	SSUSH9a	10%
Target 2: Analyze the origins of U.S. involvement in World War II.	SSUSH19a	10%
Target 3: Investigate the domestic impact of World War II including war mobilization, as indicated by rationing, wartime conversion, and the role of women and African Americans or Blacks.	SSUSH19d	10%

Successful Transition

- Alternate academic achievement standards must be aligned to ensure that a student who meets the alternate academic achievement standards is on track to pursue postsecondary education or competitive integrated employment, consistent with ESEA section 1111(b)(1)(E)(i)(V) and 34 CFR § 200.2(b)(3)(ii)(B)(2).
- For GAA 2.0, the **Level 3** ALD represents a student who **meets** the alternate academic achievement standards.

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Achievement Levels

- **Level 1**: Students at this level demonstrate a **limited** understanding of the knowledge and skills specified in Georgia's Extended Content Standards and **may need substantial academic support** as they transition to the next grade/course, inclusive post-secondary education, or competitive integrated employment.
- **Level 2**: Students at this level demonstrate a **partial** understanding of the knowledge and skills specified in Georgia's Extended Content Standards and **may need frequent academic support** as they transition to the next grade/course, inclusive post-secondary education, or competitive integrated employment.
- **Level 3**: Students at this level demonstrate an **adequate** understanding of the knowledge and skills specified in Georgia's Extended Content Standards and **may need occasional academic support** as they transition to the next grade/course, inclusive post-secondary education, or competitive integrated employment.

MEET ALTERNATE ACHIEVEMENT STANDARDS

- **Level 4**: Students at this level demonstrate a **thorough** understanding of the knowledge and skills specified in Georgia's Extended Content Standards and **may need little to no academic support** as they transition to the next grade/course, inclusive post-secondary education, or competitive integrated employment.

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Postsecondary Education

- More than 308 colleges/universities support students with intellectual disabilities throughout the United States.
- There are currently eight inclusive college education programs in Georgia, with one college preparing to admit students in Fall 2022.
  - 48% of 2021 graduates of inclusive college programs are employed.

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Competitive Integrated Employment

- To satisfy the definition of "competitive integrated employment," the employment must satisfy the requirements for all three components:
  - Competitive earnings;
  - Integrated location; and
  - Opportunities for advancement.

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Your Role

- For the next two days, you will review the high school assessment and achievement level descriptors for the four GAA 2.0 achievement levels *and* make a series of recommendations for cut scores.
- You are here as a representative of Georgia educators.
  - You are the experts and we need your input.
  - Please keep the end goal in mind – successful transition to the next grade/course, inclusive postsecondary education and/or competitive integrated employment – this is about ensuring Georgia's students with significant cognitive disabilities are prepared for their future.

Next Steps

- Standards validation
- Vertical articulation/policy review meeting
- State Board of Education adoption of achievement standards
- Score reporting for high school begins in September

Meeting Norms

- All conversations and data are confidential.
  - What happens here stays here.
- Do not use phones, tablets, or computers in the meeting rooms.
- When you return to your districts and schools, please do talk about the process, but do not disclose the specifics (test items, recommended cut scores, etc.).
- Remember, your recommendations will be shared with the Superintendent and State Board of Education – they will make the final decisions.
- Achievement level descriptors, cut scores, and scores will be released publicly after State Board adoption.

COVID Protocols

All parties are working together to maximize a safe environment that aligns with protocols recommended by health experts to mitigate the spread of COVID-19.

- Meeting and dining rooms have been arranged to allow as much space as possible.
- Individual supplies (pens/pencils, etc.) will be provided for you to use and keep.
- Masks are optional.

COVID Protocols

By registering for the meeting, you agreed to follow an Attendee Code of Conduct:

- Do not attend if you are feeling ill, have COVID or flu-like symptoms, or have been in close contact with someone who tested positive for COVID in the past 14 days.
- Practice appropriate social distancing.
- Email to report an absence or illness to: ricky.foust@nwea.org

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Preparing students

for life.

www.gadoe.org

@georgiadeptofed

youtube.com/georgiadeptofed

Georgia Alternate Assessment 2.0

(GAA 2.0)

Standards Validation Training

July 18-19, 2022

Why evaluate existing achievement standards for GAA 2.0?

- Allows GaDOE to obtain evaluation **by Georgia content experts** – teachers, administrators, curriculum specialists, and workforce.
- Opportunity for **educator review of existing cut scores** used to define achievement levels.
- To make **recommendations to the Georgia State Board of Education** regarding existing cut scores for the Georgia Alternate Assessment (GAA 2.0).
- To ensure cut scores are **consistent with expectations stated in the Extended Content Standards**.

GAA 2.0 Blueprints

Georgia Alternate Assessment 2.0
Mathematics High School
Assessment Blueprint

Claim/Target	Standards Assessed	Weight
Students demonstrate increasingly complex understanding of number sense.		20%
Target 1: Rewrite expressions and solve equations involving square roots.	MGSE.N.RN.2	10%
	MGSE.A.REI.4b	10%
Students demonstrate increasingly complex understanding of measurement, data, and analytic procedures.		30%
Target 1: Represent and interpret data displays.	MGSE.S.ID.1	10%
	MGSE.S.ID.6a	10%
	MGSE.S.ID.6c	10%
Students solve increasingly complex mathematical problems using algebraic thinking.		50%
Target 1: Use operations to simplify expressions and solve equations using algebraic properties and models.	MGSE.A.CED.1	20%
Target 2: Interpret graphs of functions.	MGSE.A.REI.6	10%
	MGSE.F.IF.4	10%
Target 3: Understand the relationship between two quantities.	MGSE.F.BF.1a	10%

GAA 2.0 Extended Standards

High School : Mathematics: Quantities

MGSE9.12.N.Q.3	Choose a level of accuracy appropriate to limitations on measurement when reporting quantities. *For example, money situations are generally reported to the nearest cent (hundredth). Also, an answers' precision is limited to the precision of the data given.*		
Least complex ⟷ Most complex			
Respond differentially when presented with real-world materials about accuracy and appropriateness of measurements. Communicate a response to identify more accurate or more appropriate measurement in real-world situations.	Identify the most appropriate unit of measure to be used in a real-world situation (e.g., measure the length of a grape in cm/m/km, paying with coins/bills).	Identify the most accurate measurement to be used in a real-world situation (e.g., nearest cup, ½ cup or ¼ cup related to cooking).	Identify which degree of accuracy and unit of measurement is accurate and appropriate in a real-world problem which may include estimation and rounding. (e.g., A quotient of 3.5 people would be 4 people. A quotient of $1.666667 would be $1.67).

Task Part A | Task Part B | Task Part C

Achievement Levels

- On the GAA 2.0, students will earn one of four achievement levels, based on their scale score.
  - Achievement Level Descriptors (ALDs) describe the content-based expectations for students in each achievement level.

ESEA Requirement

- Alternate academic achievement standards must be aligned to ensure that a student who meets the alternate academic achievement standards is on track to pursue postsecondary education or competitive integrated employment, consistent with ESEA section 1111(b)(1)(E)(i)(V) and 34 CFR § 200.2(b)(3)(ii)(B)(2).
- **NOTE:** For GAA 2.0, the **Level 3** ALD represents a student who **meets** the alternate academic achievement standards.

Achievement level descriptors (ALDs)

- ALDs describe the specific knowledge and skills that a student at a given achievement level should be able to demonstrate with any contextual factors.
- Represent the full range of students across an entire level (e.g., Level 1, Level 2, Level 3, Level 4).
- Updated to represent revised content

Types of ALDs

Policy ALDs	• High-level descriptors that are the same for all subjects and grade levels
Range ALDs	• Grade-level, content specific descriptors which link back to the standards
Threshold ALDs	• Specifically defines what it takes for a student to attain each achievement level
Reporting ALDs	• Puts Range ALDs in public friendly language for reports provided to parents and the public

What are we doing here?

What is a Cut Score?

A cut score is the minimum test score a student must earn to be considered at a specific achievement level.
Three cut scores result in four levels of achievement.

Cut Score 1 | Cut Score 2 | Cut Score 3

Level 1 | Level 2 | Level 3 | Level 4

Georgia Alternate Assessment 2.0 Score Distribution

Overview of Method Used to Recommend Existing Cut Scores

- For the ID Matching method, panelists reviewed each score point of a part in an Ordered Item Booklet (OIB).
- Panelists made one of two judgments:
  - Matched the expectations of the score point with an achievement level.

    OR
  - Determined that the score point was in the threshold area between achievement levels.

Standards Validation Process

1. Review the relevant knowledge and skills to be taught and assessed in each content area to support the goals of the state
2. Review the expectations associated with each Achievement Level
3. Convene a committee of educators to review and confirm existing cut scores given changes in content
4. Policy makers and other stakeholders review the impact associated with the recommended cut scores
5. The Superintendent reviews the results and provides recommended cut scores to the State Board for approval

Ordered Item Booklet (OIB)

- Each page represents a score point within a part of a task
  - For the GAA 2.0, a score point represents an "item" in the OIB
- Each part is worth up to two points and will appear once for *each score point assigned*
  - *A single part, worth 2 points, will appear on two separate locations, based upon the likelihood of getting 1 point and 2 points*

Panelists' Evaluation Procedures

1. Panelists review revised range ALDs with a focus on the **threshold** to develop an understanding of the content and achievement expectations.
2. Panelists review the OIB to evaluate **whether existing cut scores continue to be a reasonable representation** of the threshold between each achievement level.
3. Panelists discuss their evaluation and **determine whether they would recommend any modifications** to the current cut scores.
4. Cross subject and grade level articulation discussion.

Ordered Item Booklet (OIB)

Panelist Recommendations

- Panelists will complete an initial round of judgments and discuss to group consensus.
  - Evaluate existing cut score locations relative to threshold expectations of ALDs.
  - Focus on area around the existing cut scores to determine whether they continue to be reasonable representations of the ALDs.
  - Group level discussion about initial judgments.
  - Consensus group recommendation.

Understanding Item Difficulty

- The expected knowledge, skills, and abilities required to answer the question correctly
- Are students responding to the item accurately and independently?
- Are students responding to the item accurately following scaffolding?
  - What types of scaffolding are being provided for the student?

Item Difficulty – cont.

- Item difficulty is based on performance of students on each. **Cognitive complexity** is another consideration for evaluating items.
- Cognitive complexity refers to the cognitive processes associated with an item and the expected knowledge, skills, and abilities required to answer the question correctly.
  - Answering a question about a topic vs. Identifying something embedded within a larger body of information
  - Requiring a straightforward answer vs. Engaging in multiple steps to solve a problem

Item Difficulty vs. Cognitive Complexity

- It is possible for an item to be **high difficulty** but **low cognitive complexity**
- Examples:
  - What is heteroscedasticity?
  - Whose autobiography is "A Dream About Lightning Bugs?
- Items are ordered in booklets by difficulty level, but cognitive complexity should be kept in mind when making judgments, particularly because of the partial credit that students can achieve through scaffolding.

OIB Item Map

OIB Part Unit #	OIB Page #	Task	Part	Scaffolding	Content Standard	Scale Location	ALD	Note Cut Scores: Level 2 = 1350, Level 3 = 1400, level 4 = 1439
. . .								
7	14	11	B	Y	ELAGSE11-12.L.4c	1334	1	
8	16	11	A	Y	ELAGSE11-12.L.4c	1337	1	
9	18	2	A	Y	ELAGSE11-12.RI.1	1339	1	
10	20	4	A	Y	ELAGSE11-12.W.1	1343	1	
11	22	7	A	Y	ELAGSE11-12.W.2a	1343	1	
12	24	1	A	N	ELAGSE11-12.RL.1	1344	1	
13	26	4	B	Y	ELAGSE11-12.W.1	1345	1	
14	28	7	B	Y	ELAGSE11-12.W.2b	1349	1	
15	**30**	**1**	**B**	**Y**	**ELAGSE11-12.RL.1**	**1350**	**2**	
16	32	8	A	N	ELAGSE11-12.RL.3	1356	2	
17	34	6	A	N	ELAGSE11-12.RI.2	1357	2	
18	36	1	C	Y	ELAGSE11-12.RL.1	1359	2	
19	38	5	B	Y	ELAGSE11-12.RL.2	1361	2	
20	40	2	A	N	ELAGSE11-12.RI.1	1362	2	
21	42	9	B	Y	ELAGSE11-12.L.2c	1363	2	
. . .								
56	112	5	C	N	ELAGSE11-12.RL.2	1425	3	
57	114	9	B	N	ELAGSE11-12.L.2c	1427	3	
58	116	4	C	N	ELAGSE11-12.W.1	1434	3	
59	118	7	C	N	ELAGSE11-12.W.2d	1438	3	
60	**120**	**2**	**C**	**N**	**ELAGSE11-12.RI.1**	**1441**	**4**	

What are impact data?

- Using these initial cut scores, data from the 2022 administration can be evaluated to ***estimate the percentage of students*** that would be classified ***into each achievement level.***

Record your ratings

- Record your initial recommendations on your rating sheet
- The OIB Sequence number should be recorded on your rating sheet
- Sample rating sheet:

Rating: Level 2 Cut Score on Page: XX

Rating: Level 3 Cut Score on Page: XX

Rating: Level 4 Cut Score on Page: XX

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Discussion

Review	Review the initial recommendations, distribution, and impact data.
Engage	Engage in table level discussions to evaluate the initial recommendations.
Identify	Identify the knowledge and skills necessary to respond to a few of the items above and below the recommendations.
Discuss	Discuss what knowledge and skills distinguish the items for the achievement levels.
Evaluate	Evaluate the ratings you initially provided and determine if changes are necessary.
Consensus	Come to consensus with your subject area assessment.

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Next steps

- Conduct vertical articulation for all content areas.
- Complete process evaluation surveys.

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Meeting Security & Norms

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Meeting Security

- Meeting materials and information are the exclusive property of GaDOE and NWEA.
- Discussions involving materials need to be limited to the meeting rooms only.
- Documents shall not be reproduced, discussed, or released or distributed to unauthorized persons in any form.

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Meeting Security

- E-mailing, copying, printing, posting, or taking screenshots of materials is not permitted.
- Note taking during the meeting is permitted, but notes cannot leave the meeting rooms and will be destroyed after the meeting.
- Only authorized persons are permitted to view the materials.

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Meeting Security

- Exam materials may not be shared, discussed, or disseminated.
  - See Confidentiality and Nondisclosure agreement
- Use of personal devices (e.g., phones, tablets, laptops) permitted ONLY <u>outside</u> the meeting room.
- Sharing information about the standards validation process is encouraged, but not the items or recommended cut scores.

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Discussion Guidelines

- *Be present*
  - Everyone participates
  - No side conversations
- *Be open and ask questions*
  - Courteous discourse
  - New information to contribute
- *Be respectful*
  - All opinions are equally valued
- *Observers may not participate*

Richard Woods, Georgia's School Superintendent | Georgia Department of Education | *Educating Georgia's Future*

Preparing students
for life.
www.gadoe.org
@georgiadeptofed
youtube.com/georgiadeptofed
GaDOE
Georgia Department of Education

Georgia Department of Education
GAA 2.0 Standards Validation Workshop
July 18-19, 2022
Readiness Survey

NAME: _____

SUBJECT: _____

GRADE LEVEL: _____

PLEASE RESPOND TO THE FOLLOWING QUESTIONS BELOW BEFORE BEGINNING YOUR ORDERED ITEM BOOKLET REVIEW. YOU CAN RESPOND TO BOTH ITEMS BY CIRCLING YOUR RESPONSE. AFTER ANSWERING BOTH ITEMS, PLEASE RETURN THIS FORM TO YOUR FACILITATOR.

1. I understand how to make my standards evaluation judgments for each of the three cut scores.

   YES NO

2. I am prepared to begin my review and evaluation.

   YES NO

Appendix E – Sample Validations Rating Form

Name:_____
Subject:_____
Grade:_____

Standards Validation Recommendations

Initial Rating

Level 2 Threshold Cut Score on Page: _____

Level 3 Threshold Cut Score on Page: _____

Level 4 Threshold Cut Score on Page: _____

Final Rating

Level 2 Threshold Cut Score on Page: _____

Level 3 Threshold Cut Score on Page: _____

Level 4 Threshold Cut Score on Page: _____

Appendix F – Evaluation Survey Results

Process Evaluations					
Day 1 Evaluation					
Question		***Very Unsuccessful***	***Unsuccessful***	***Successful***	***Very Successful***
Rate the success of each workshop orientation component: Overview of the workshop	Frequency	0	1	5	29
	Rate	0.0%	2.9%	14.3%	82.9%
Rate the success of each workshop orientation component: Content review of the OIB	Frequency	0	0	5	30
	Rate	0.0%	0%	14.3%	85.7%
Rate the success of each workshop orientation component: Discussion of the ALDs	Frequency	0	1	2	32
	Rate	0.0%	2.9%	5.7%	91.4%
Rate the success of each workshop orientation component: Training on the evaluation process	Frequency	1	1	6	27
	Rate	2.9%	2.9%	17.1%	77.1%
		Too little time was allocated to the orientation	***The right amount of time was allocated to the orientation***	***Too much time was allocated to the orientation***	
How would you rate the amount of time allocated to the orientation?	Frequency	0	18	17	
	Rate	0.0%	51.4%	48.6%	

Please provide any comments about the orientation that would help in planning future workshops.	Perhaps a pre-survey of expertise surrounding vocabulary terms related to standards & assessment. Grouping panelists by area of knowledge (content, assessment population) and then dividing strategically into balanced groups Because I was involved in previous development sessions, it was easy to quickly come up to speed with current work. All elements were beneficial including training, review, discussion It was great as is While I respect that there are committee members who haven't participated in this process before, it is important for the overview, the 2nd pard of the orientation (powerpoint) was repeated in our breakout session. I didn't see the need to go through the same slides a second time. The 1st could have been more generalized w/specifics in break-out group or repeat just a few highlights during break-out. n-a no improvements needed Evaluator training should have a few guidelines and be more concrete. I was many times where it was a look and then ask questions. One sample can be done together first Show videos or demonstrations for what was expected during the entire process Provide a model/example of expected task. Indicate group vs. individual task Great workshop - our duties were very well explained Comment related to use of smart watches? I appreciate the opportunity to be involved in the validation process I'm grateful for the opportunity to be involved in the process Although a bit overwhelming for a non-educator, I felt it was thorough. Have one person present the information instead of two people presenting almost the exact same information. Cut down the information and leave room for more of it during the breakout sessions. Need more of the details in the small group sessions. I thought it was great and not too much if you were new to this but also just enough to understand. Simplified. Kelley was great and so helpful, as well as the orientation speakers. Morning speech is a bit long. Perhaps give 5 minute break within the morning session. A lot of people talk to each other (icebreaker at the table?) to break up the morning.

		Very Confident	***Somewhat Confident***	***Not Very Confident***	***Not at all Confident***
How confident were you about your Day 1 recommendations for achievement levels for students?	Frequency	18	13	2	0
	Rate	54.5%	39.4%	6.1%	0
		More than enough times was available	***Sufficient time was available***	***Barely enough time was available***	***There was not enough time available***
How did you feel about the time allocated for making these recommendations?	Frequency	11	22	0	0
	Rate	33.3%	66.7%	0.0%	0

Please provide any comments about the Day 1 evaluation activities that would	Please see 2nd part of Q3 on 1st page. Facilitator was key to our success - Good humor; kept group moving forward in positive manner A great day I would have liked a 2nd OIB sheet to show where questions fill in terms of scaffolding v. no scaffolding of same question. Kind of a cross reference

helpful in planning future workshops.	great day! facilitator was clear, organized, and helpful Loved collaborating in small groups Day 1 evaluation time was good. I would like more guidance to structure it better Presenting the ALDs group range by level may be helpful. It can aid in also seeing them represented across the different standards grouped by range level. Provide examples of what is expected Allow extra time to model/sample task Great facilitator in my science!

Day 2 Evaluation					
Question		***Very Confident***	***Somewhat Confident***	***Not Very Confident***	***Not at all Confident***
How confident were you about your Day 2 recommendations for achievement levels for students?	Frequency	25	6	0	0
	Rate	80.6%	19.4%	0.0%	0.0%
		More than enough times was available	***Sufficient time was available***	***Barely enough time was available***	***There was not enough time available***
How did you feel about the time allocated for making these recommendations?	Frequency	18	11	1	1
	Rate	58.1%	35.5%	3.2%	3.2%
		Very Confident	***Somewhat Confident***	***Not Very Confident***	***Not at all Confident***
How confident were you about your groups' vertical articulation discussion?	Frequency	24	7	0	0
	Rate	77.4%	22.6%	0.0%	0.0%
	Very Confident	***Somewhat Confident***	***Not Very Confident***	***Not at all Confident***	***Very Confident***
How confident were you about your groups' vertical articulation recommendations?	Frequency	24	7	0	0
	Rate	77.4%	22.6%	0.0%	0.0%
		More than enough times was available	***Sufficient time was available***	***Barely enough time was available***	***There was not enough time available***
How did you feel about the time allocated for this discussion?	Frequency	16	14	0	1
	Rate	51.6%	45.2%	0.0%	3.2%
		Very Successful	***Successful***	***Unsuccessful***	***Very Unsuccessful***
Overall, how would you rate the success of the standards validation workshop?	Frequency	20	9	1	0
	Rate	66.7%	30.0%	3.3%	0.0%
		Very Organized	***Organized***	***Unorganized***	***Very Unorganized***
How would you rate the organization of the standards validation workshop?	Frequency	20	6	2	2
	Rate	66.7%	20.0%	6.7%	6.7%
Please provide any comments about the standards validation	I feel for new members to the process. Some more information could be provided prior to explain procedures and rationale for the work being completed.				

activities that would be helpful in planning future workshops.	Given the parameters of the data provided, there was more than enough time to make recommendations. Increased Covid protocols (basics!). Overall - more focus on students as a cognitively and socially diverse group. Less paper! More online options for the work. Pre-test or survey to assess gaps in participant knowledge. 1) Create balanced committees for each content area. An equal number of content, assessment, and "population" experts. 2) Facilitators with actual special ed/modified curriculum background. 3) Inclusion of data about student demographics, placements, areas of eligibility, and cognitive scores. 4) Richer discussion surrounding the actual needs of students as they relate to both standards-based content, life skill instruction, and assessment. 5) More respect for participants' time and more accessibility options. 6) GaDOE and partner orgs observe instruction and assessment in pilot classrooms to understand the population range. Be respectful of time; use ours wisely. Larger selection of OIB questions to consider. Standards validation was hard to understand. There should have been more of everyone at the central activities and making a sound decision as a group. It was very different to understand other group's decisions and follow along at times. Also, if we are referring to cognitive student abilities who is being measured would adjust the standard. I enjoyed the time spent talking to my co-workers about the standards and testing required for our students. The open-ended questions were very helpful with Russ. The standards validation activities are multi-faceted. While there is a range in each category, there is also a range in student abilities that will need to be considered. More attention to the weights of the curriculum areas as well as a consideration of the many confounding variables can inform the rationale. This is a valuable process and I was honored to be able to be a part of the conversation. Loved the collaboration and input from colleagues. Such a great experience! Great workshop. Very informative Collaboration was awesome! Thank you! Day 2 solidified my understanding of the process. Initial training including modeling, sample reviews would be helpful with making the process smoother. None Bound pages in booklet format so turning pages the top page is not upside down. Clearer information on ability to keep some folder items and what needs to be collected. Clarification on smart watches as devices. Truly enriching. Would love to attend more sessions! The 2nd day could have been condensed into day 1!