Maine Through Year Assessment
Spring 2024 Technical Report Appendices

nwea

Table of Contents

The Maine Through Year Assessment

Assessment Coordinator Training

Spring 2024

Welcome

- Krista Averill, Maine DOE Assessment Coordinator
- Fred Valenzuela, NWEA Sr. Program Manager
- Mindy Stobbe, NWEA Program Manager
- Hailey Westphal, NWEA Sr. Solution Delivery Consultant
- Ricky Foust, NWEA Director of Program Management

nwea State Solutions

Sections Covered

nwea State Solutions

Maine Department of Education

nwea State Solutions

Maine Through Year Assessment Overview

Subjects, Grades, and Delivery

- Content Areas
  - Mathematics (3 – 8 and 2nd year of High School)
  - Reading (3 – 8 and 2nd year of High School)
- Administration Windows
  - Fall: October 2 – November 3, 2023
  - Winter (Optional): January 2 – February 16, 2024
  - **Spring: April 22 – May 31, 2024**
- Modes of Delivery
  - Online
  - Paper-Based Accommodated Forms
    - Standard (Print on Demand)
    - Braille & Large Print (Order)
    - **All paper-based forms require Maine DOE approval**
- Scores
  - Fall/Winter: RIT
  - **Spring: Maine-Specific Scale Score and RIT**

nwea State Solutions

Maine Department of Education

Testing Time & Scheduling Recommendations – Spring 2024

Grade level	Content area	Number of assessment questions*	Recommended scheduled assessment-taking time
3 – 5	Mathematics	50	2 hours total, 2 - 3 sessions
3 – 5	Reading	46	2.5 hours total, 3 - 4 sessions
6 – 8	Mathematics	50	2 hours total, 1 - 2 sessions
6 – 8	Reading	46	2 hours total, 1 - 2 sessions
2nd Year of High School	Mathematics	54	1.5 hours, 1 session
2nd Year of High School	Reading	49	1.5 hours, 1 session

*All students in the same grade, given the same assessment, will receive the same number of assessment items.

- SAUs / Schools have flexibility in scheduling the assessment
  - Student needs should be prioritized when developing the assessment schedule.
  - Assessments can be worked on over multiple days.
  - The assessments are untimed. Proctors should not pace students.

nwea State Solutions

Maine Department of Education

Day of Assessment

- Estimated assessment time does not include:
  - System check test
  - Test ticket distribution
  - Launching the secure browser
  - Student log in
- Students' assessment can be paused by logging out.
- Students will automatically be logged out of the assessment after 15 minutes of inactivity.

Note: *No proctor action required for the student to resume the assessment, students must log back in using the information on the test ticket*

nwea State Solutions

Maine Department of Education

Student Tutorial

- An interactive video for the Maine Through Year Assessment is available for students to learn how to use the online assessment platform. During this tutorial, the student will be shown the following:
  - How to use the online tools
  - How to navigate through the assessment
  - How to respond to different items types
  - Tips for taking the assessment

Resource and Link: Maine Through Year Student Tutorial

nwea State Solutions

Maine Department of Education

Item Type Sampler

- An item type sampler or *practice assessment* will provide students an opportunity to practice each item type and gain familiarity with the platform.
- Includes all item types and tools for each grade and subject.
- Accessible by the Maine Connections Page, the Maine DOE webpage, or a link in the secure browser.
- Paper item type samplers are also provided as PDFs for schools to download and print (including answer keys).
- **Utilizing the item type sampler in the secure browser is also a great way to ensure that devices meet all the system requirements before the actual day of the assessment.**

Resource and Link: Maine Online Item Type Sampler

nwea State Solutions

Maine Department of Education

Item Type Sampler vs Through Year Assessment

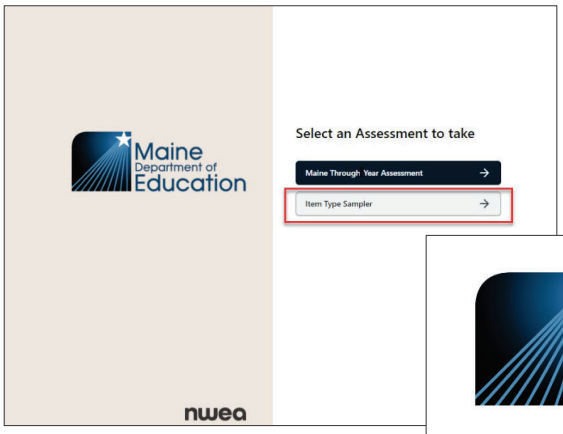
- The Item Type Sampler *(practice assessment)* contains 15 questions in Math and 17 questions in Reading.
- A test ticket is not needed to take the Item Type Sampler.
- Rarely, a student may mistakenly open the Item Type Sampler within the secure lockdown browser rather than the actual Through Year Assessment.
  - When in the Item Type Sampler, "Test Student Name" appears where the student's name should appear.
- Always refer to Manage Online Testing to confirm the student's test status and their response progress.

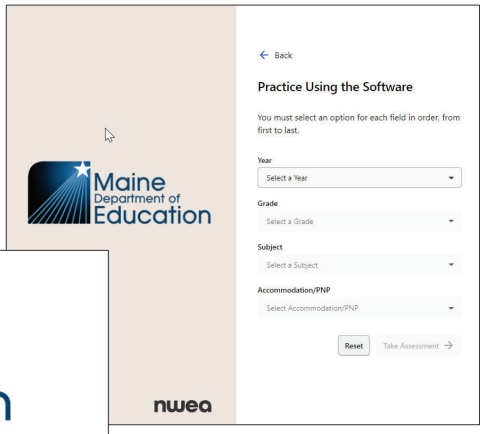
Resource and Link: Maine Online Item Type Sampler

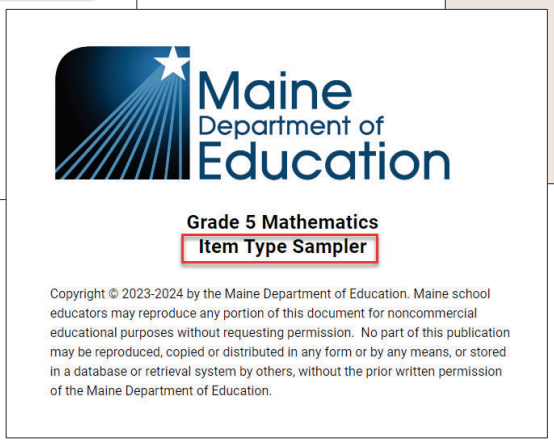
nwea State Solutions

Maine Department of Education

Item Type Sampler







Resource and Link: [Maine Online Item Type Sampler]

nwea State Solutions

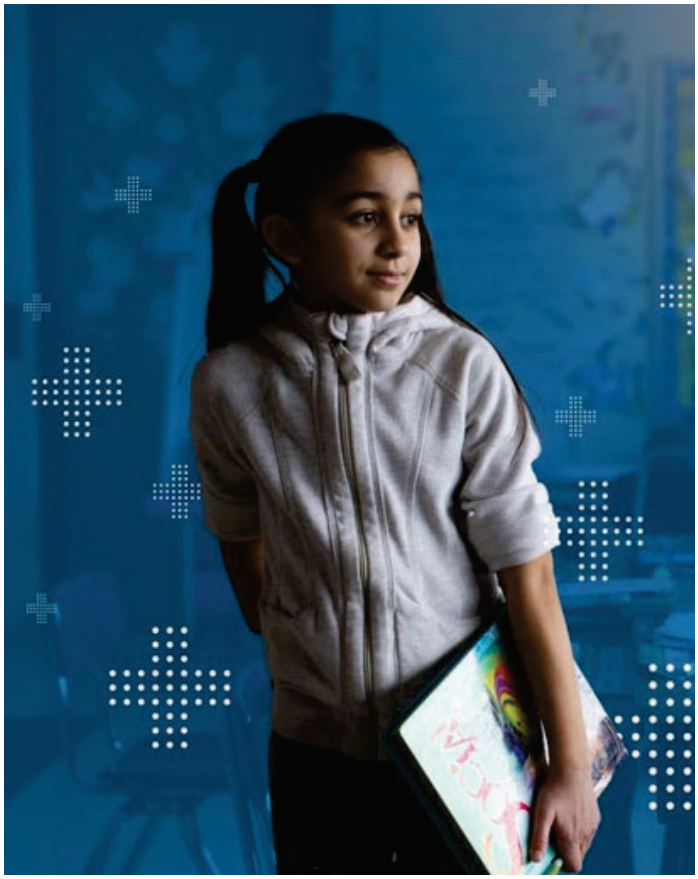
Maine Department of Education

Questions from the Maine Through Year Overview section?

FAQ document and slide deck will be shared out after training session.

nwea State Solutions

Technology Readiness



nwea State Solutions

State Solutions Secure Testing Browser

- System Requirements
  - Check to make sure the Operating System is currently supported before any updates are made.
  - NWEA State Solutions Secure Browser – REQUIRED for all devices.

    https://securebrowser.state.nwea.org

Note: *This is a different Secure Testing Browser than what was used for MAP Growth.*

For those using Chromebooks, devices that are not managed will be unable to download the secure browser.

Device and Application
Mac Secure Testing Browser
Windows Secure Testing Browser
Chromebook App
iPad App

State Solutions Secure Browser

MAP Growth Secure Browser

Resource and Link: NWEA State Solutions System and Technology Guide

nwea State Solutions

Maine Department of Education

Tips for Installing the State Secure Browser

- Partner Code: ME
- Multiple Device Management (MDM) Installation available
- Secure Browser can be downloaded from the Management System or via NWEA provided link.

Reminders:

- Be sure to turn off auto updates on student devices during the administration window.
- Previous versions of the State Solution Secure Browser must be uninstalled before the new version is installed.
- *MAP Growth Secure Browser is a different application and does not need to be uninstalled/reinstalled.*

nwea State Solutions

Maine Department of Education

Updated Version for macOS

- New secure browser for macOS for the Spring 2024 Admin available here.
- Additional macOS installation steps and bulletin are available on the Maine Connections page.
  - macOS NWEA State Solutions Secure Browser Bulletin March 2024
  - macOS Installation Steps
- We highly recommend updating to version 4.3.1 for a smooth testing experience.

Reminders:

- Be sure to turn off auto updates on student devices during the administration window.
- The previous macOS versions of the State Solution Secure Browser must be uninstalled before the new version is installed.
- *MAP Growth Secure Browser is a different application and does not need to be uninstalled/reinstalled.*

nwea State Solutions

Maine Department of Education

Supported Devices

+ NWEA State Solutions Secure Browser

Device	Supported OS Versions
Windows PC	Windows 10 and Windows 11
macOS	macOS 12, macOS 13, and macOS 14
Chromebook	Release Channel Only – version 109 or later
iPads	iOS 15 and iOS 16

Resources and Links:
NWEA State Solutions System and Technology Guide and
NWEA State Solutions System Requirements Guide

nwea State Solutions

Maine Department of Education

Minimum System Requirements for Acacia

+ The Acacia Management and Reporting System is supported on the latest versions of the following browsers:
  - Google Chrome
  - Mozilla Firefox and Firefox LTS
  - Microsoft Edge
  - Safari
  - Safari on iPad

Note: Internet Explorer is no longer supported

nwea State Solutions

Maine Department of Education

Technology Readiness

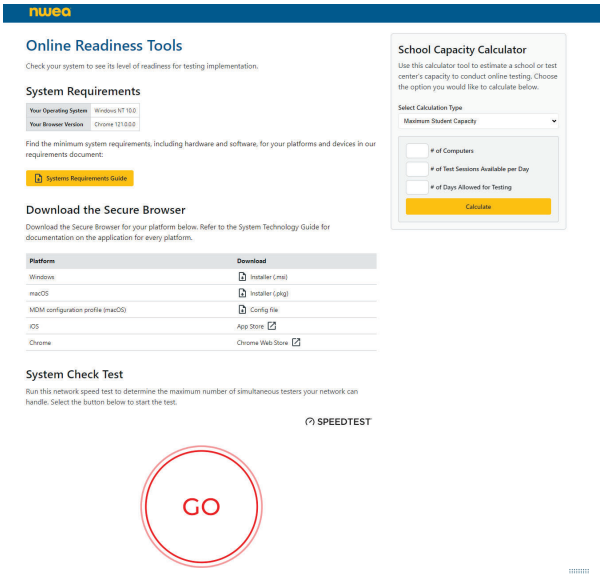
- NWEA State Solutions System and Technology Guide
  - IT Readiness
  - Network and System Requirements
  - State Solutions Secure Browser Installations
  - Allowed Lists
- Online Readiness Tools
  - Upon launching the Secure Browser
  - Additional site available for checks

nwea State Solutions

Maine Department of Education

Online Readiness Tools

- Online Readiness Tools
  - Upon launching the Secure Browser
  - Additional site available for checks
- The System Check Test should be performed prior to students taking the assessment to confirm there are no issues with connectivity and your network is ready for simultaneous testers.
- If there are reports of connectivity issues, please be ready to provide NWEA Partner Support with a screenshot of the results from the System Check Test.



nwea State Solutions

Maine Department of Education

System Maintenance & Releases

- Periodically, systems are unavailable due to platform and software maintenance.
- There is a software release scheduled for the weekend of April 13 but currently no other planned software releases or hardware maintenance during the Spring 2024 Testing Window starting on April 22nd and ending on May 31st.
- A reminder window will also pop up upon logging into MARC, which is used for the single sign on to Acacia.
  - **Reminder:** Turn off auto updates on student devices during the Assessment Administration Window.

Resource and Link: Platform and Software Maintenance Windows

nwea State Solutions

Maine Department of Education

Questions from the Technology Readiness section?

FAQ document and slide deck will be shared out after training session.

nwea State Solutions

nwea State Solutions

Assessment Management in Acacia™

NWEA Platforms & Terminology

- Acacia is the assessment platform used to deliver and manage the Maine Through Year Assessment.
- MAP Growth (teach.mapnwea.org) is the platform used for Single Sign On (SSO) to access Acacia.
  - The MAP Growth platform is also where users are managed and where MAP Growth Reports with RIT score data from the Through Year Assessment can be accessed.
    - **Reminder:** Students must also be rostered by the SAU in MAP Growth *for each administration, by the last day of the assessment window,* to have MAP Growth Reports with Through Year Assessment RIT score data.
- Assessments are delivered to students within the state solutions secure browser, unless they are taking a paper-based assessment.
  - The Item Type Sampler is also available within the secure browser.

Acacia™ Components

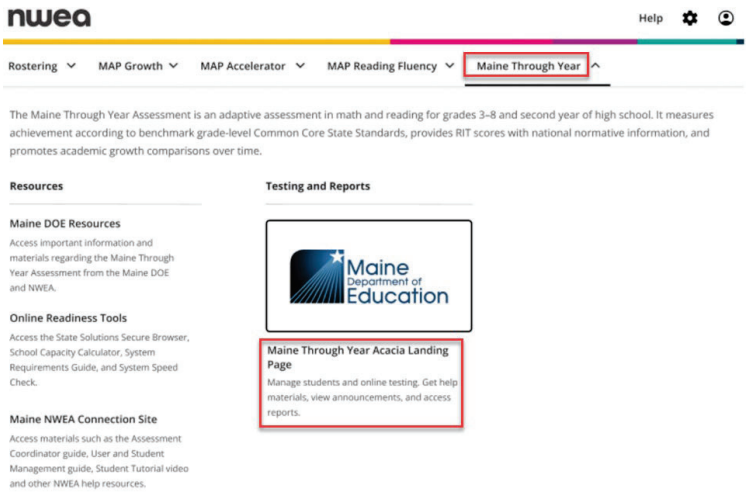
- Acacia™ Manage
  - The management system allows administrators and teachers to smoothly manage the entire assessment process including managing students, online test assignments, monitor test status, analyze data reports, and much more – all in one place!
- Acacia™ Assess
  - The online test delivery platform that delivers assessments to students, more commonly known as the state solutions secure browser.
- Acacia™ Reports
  - The online reporting suite (ORS) provides a dynamic, *real-time, easy-to-use reporting for assessments.
  - ** Real-time reports available within 24-72 hours after test has been submitted.*

nwea State Solutions

Maine Department of Education

Accessing Acacia - Single Sign On

- Single Sign On (SSO) connects your access from MAP Growth (aka MARC) to Acacia Manage.
  - One less username and password to remember.
  - User roles will be managed through MARC.
  - Same user roles in MARC and Acacia, permissions may vary slightly.
  - *Having a missing or incorrect **School State Code** may prohibit you from accessing Acacia.*
  - After logging into MARC, users will see the 'Maine Through Year' tab and the 'Maine Through Year Acacia Landing Page'.



nwea State Solutions

Maine Department of Education

Acacia Home Screen

Note: What you will see on this screen is determined by your role.

+ Help ?
+ Profile
+ Logout

nwea State Solutions

Maine Department of Education

Roles for Acacia Setup

	Manage Users (in MAP Growth)	Roster Students (in MAP Growth)	View Students (in MAP Growth)	View Student Registrations (in Acacia)	View students and Accommodations (in Acacia)	Manage Online Testing (in Acacia)
District Assessment Coordinator	X	X	X	X	X	X
Data Administrator	X	X	X	X	X	
Proctor				X		X
School Assessment Coordinator			X	X	X	X

nwea State Solutions

Maine Department of Education

Roles for Testing Students

	District Assessment Coordinator	School Assessment Coordinator	Proctor	Maine DOE
Assign Accommodations	X	X		X
Assign Not Tested Codes **(Maine DOE Only)**				X
Create and Manage Student Groups	X	X		X
View Manage Online Testing Dashboard	X	X	X	X
Print Test Tickets	X	X	X	X
Proctor Assessments			X	

nwea State Solutions

Maine Department of Education

Acacia Rostering

- Maine DOE will be responsible for rostering students in Acacia prior to the assessment window.
- Students will be rostered to their "Reporting School," which is the school they attend and at which they receive their instruction.
- NEO is the source of truth for which students are rostered. NEO rosters are based on student information entered by the SAU into Synergy.
  - SAUs should ensure that student demographics are correct in Synergy.
- Each weekday morning during the assessment administration window, Maine DOE will upload daily delta/roster file for any changes made in Synergy the prior weekday.

nwea State Solutions

Maine Department of Education

Registration

- Test registrations are created automatically when students are rostered.
- Any needed edits to registrations will be done by SAUs.
  - Upload into Acacia via the registration report or update within a student's profile.
- Edits to registrations include adding supports and accommodations.
- Students will have a line for each subject in the registration report.
- Registration Report template and Registration Report Upload and Report Format are available in the Acacia Help Resources.

nwea State Solutions

Maine Department of Education

Maine Through Year Assessment: SAU Tasks

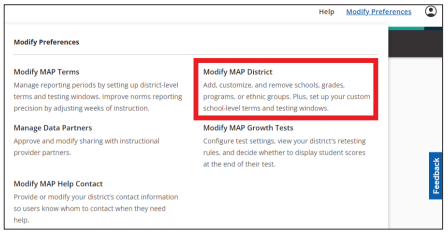
- Confirm School State Codes (SchoolOrgID) in MAP Growth are correct.
- Add/confirm supports and accommodations to student registrations, as needed.
- Print test tickets.
- Monitor student progress.
- *Optional action to receive MAP Growth reports*.*
  - Import of Student Roster will need to be done by the SAU in MAP Growth.
- Data clean up.
- Access reports via Acacia and MAP Growth*

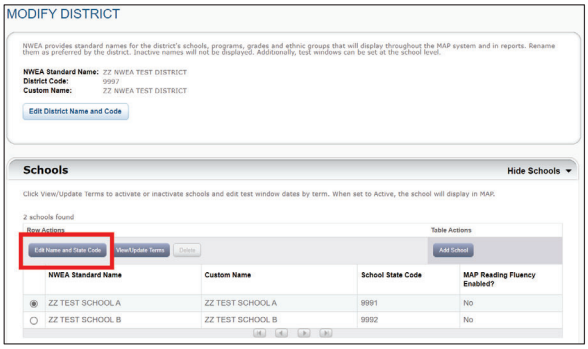
nwea State Solutions

Maine Department of Education

MAP Growth School State Codes

- School State Codes are an important part of the SSO connection between MAP Growth and Acacia.
- Having a missing or incorrect code may prohibit you from accessing Acacia.
- School State Codes need to align with the **School Org IDs** in the Infrastructure Data for 2023/2024 SY located on the Maine DOE website.
  - Maine School State Codes 2023/2024 SY
- Leading zeroes should not be included in the School State Code (School Org ID).
- The District Code and the School State Code are different codes, please be sure the same code is not being used in both fields.
- School State Codes should be reviewed before and during each assessment window.
  - Users that can make these changes will have a role of System Administrator or District Assessment Coordinator (DAC).
  - This can be done in MARC. Select Modify Preferences > Modify MAP District > Select the bubble next to your school in the Schools table > Edit Name and State Code.





nwea State Solutions

Maine Department of Education

Rostering with Clever

- Rostering with Clever
  - If Clever is being used, confirm that Clever is sharing the **State_ID** field with NWEA; this can be located under the school you will be sharing.
  - The **State_ID** field maps to the School State Code in NWEA
  - In Infinite Campus the field shared with Clever is **sch_number.**

Note: *If Clever is used for rostering and the school state code is only updated in MARC, the nightly Clever sync will override those updates.*

nwea State Solutions

Maine Department of Education

Rostering for MAP Growth Reports

- MAP Growth reports are available in MARC **for RIT scores** from the Maine Through Year Assessment.
- For these reports to be available, student rostering will need to be done in both MAP Growth and Acacia.
  - Maine DOE will roster for the Maine Through Year Assessment.
  - SAUs will need to roster in MAP Growth *by the last day of the assessment window*, to have RIT data from the Maine Through Year Assessment in MAP Growth reports.
  - Student State ID must be the same in both platforms – this is the connector / unique identifier for MAP Growth reporting.
    - We recommend having this unique ID populated in both the Student ID and the Student State ID field in MAP Growth. The Student State ID field is needed for Acacia. If you would like to use your local ID in the Student ID field, you may do so in the MAP Growth Roster File template.

Resource and Link: MAP Growth Rostering Quick Guide

nwea State Solutions

Maine Department of Education

Student Groups – Online Testing and Reporting

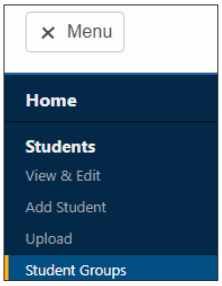
- Online Testing groups are optional, but Reporting groups **are required for each administration** for instructors to be able to see their student results.
- Students can be **grouped by grade** by their **teacher (with a group name) or grouped by assigned test administered (with a group name).**
- **Online Testing Groups**
  - Allow proctors to view smaller groups of students in Manage Online Testing in Acacia.
  - Test tickets can be printed by grade by these assigned groups.
- **Reporting Groups**
  - Provide educator access to students' score reports.
- A student does **not** need to be in the same **testing** and **reporting** group.
  - Students can be assigned to multiple groups.
- Groups can be added for previous administrations **within** the same school year.

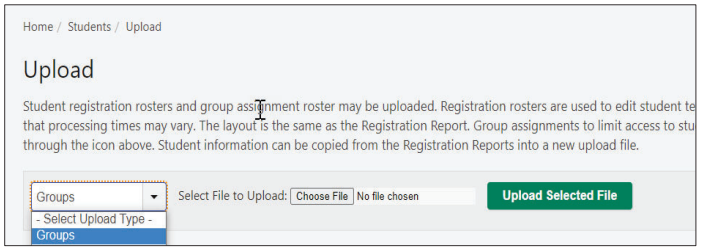
nwea State Solutions

Maine Department of Education

Creating Student Groups

- Located under Students section in the Menu.
  - Visibility based on user role permissions.
- Select Student Groups to create, view & edit to create a Student Group manually.
- Select Upload in the Menu to create and upload Student Groups in bulk.



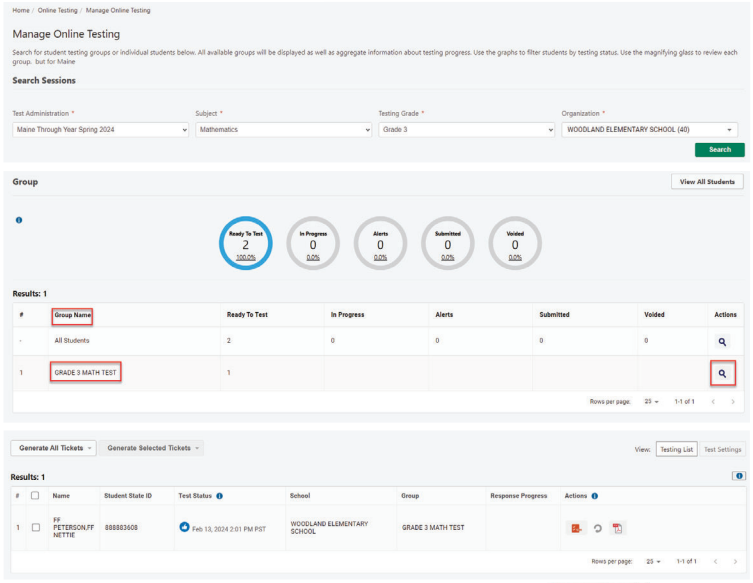


nwea State Solutions

Maine Department of Education

Viewing Student Groups

- You can view Student Groups in Manage Online Testing.
- Find the group you are looking for and select the magnifying glass.
- You can then view students within that group for info on students test status, response progress, etc.



nwea State Solutions

Maine Department of Education

Questions from the Assessment Management in Acacia section?

FAQ document and slide deck will be shared out after training session.

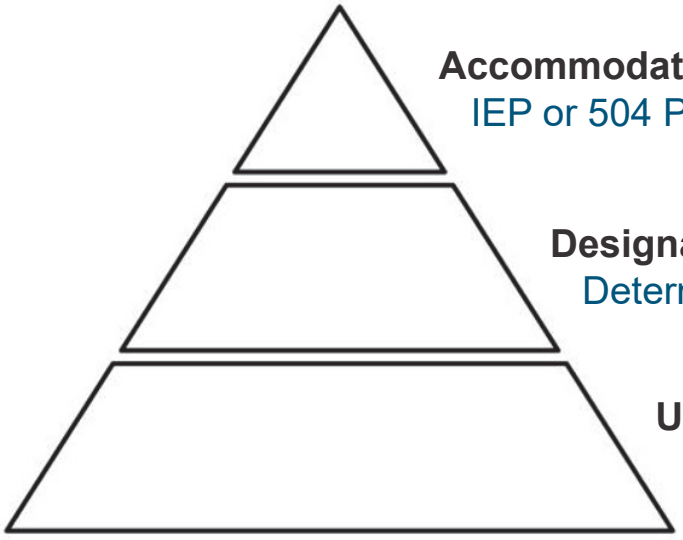
nwea State Solutions

nwea State Solutions

Accessibility

Universal Tools, Designated Supports, and Accommodations

Accessibility Features



Accommodations
IEP or 504 Plan ONLY

Designated Supports
Determined on an Individual Basis

Universal Tools
All Students

nwea State Solutions

Types of Accessibility Features

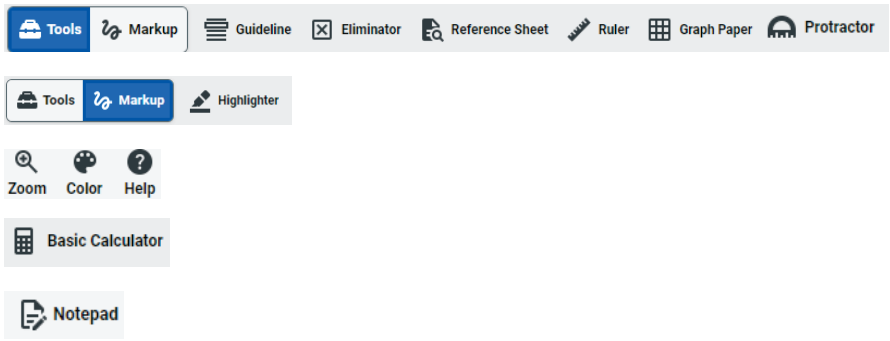
+ **Non-embedded**: Features provided locally that do not change the assessment within the platform.

+ **Embedded**: Impacts delivery of the assessment within the platform.

nwea State Solutions

Universal Tools

- Non-embedded: Scratch Paper
- Embedded Universal Tools:
  - Calculator (Math only and with specific items)
  - Color Contrast
  - Graph Paper (Math only)
  - Guideline
  - Help Videos
  - Highlighter
  - Keyboard Navigation
  - Notepad
  - Protractor (Math only and with specific items)
  - Reference Sheet (Math only)
  - Ruler (Math only and with specific items)
  - Zoom (Zoom icon available on devices except iPads, iPads have a responsive zoom using the touch screen to zoom in and out)



Maine Department of Education

Designated Supports

- Increase accessibility without altering the construct of any assessment item.
- Determined on an individual basis by an educational team.
- An educational team is two or more education professionals with knowledge of a student's performance.
- Designated supports must be consistent with the student's normal routine during classroom instruction.

Non-Embedded Designated Supports

Non-embedded designated supports can be viewed and edited via both the registration file and student's profile.

+ Individual / Small Group Setting
+ Bilingual Word Glossary for Multilingual Learners
+ Mathematical Supports (for Math Assessment Only)

Examples of supports that can be provided to students and do **not** need to be indicated in the assessment platform include:

- Translated versions of the Mathematics Reference Sheet, which can be found on the Maine Through Year Assessment webpage
- Assistive technology
- Medical devices
- Visual aids
- Auditory devices
- Student reads assessment aloud to self in individual setting
- Directions clarification

nwea State Solutions

Maine Department of Education

Embedded Designated Support: Text to Speech (TTS)

+ Available in English.
+ Guidance for Text to Speech is in the Accessibility Guide.
+ Need for this designated support will be indicated on the student's profile.
+ All text will be read aloud in Math.
+ Passages will not be read in Reading.

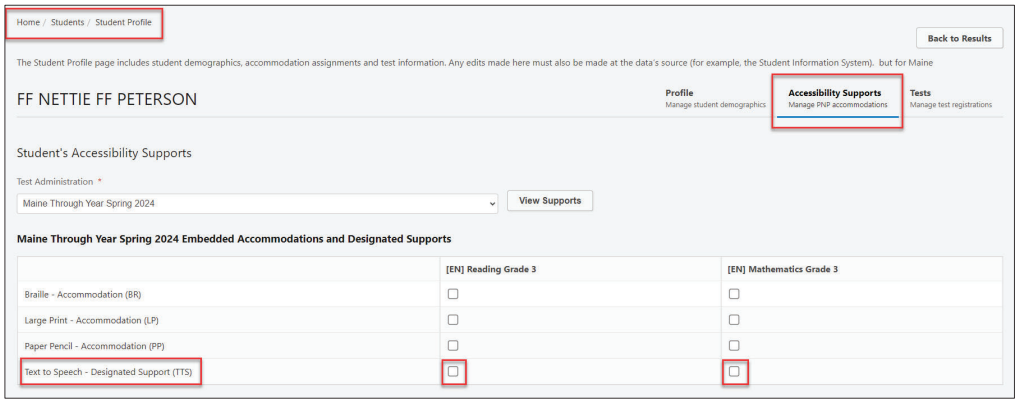
Resource and Link: Maine Through Year Accessibility Guide

nwea State Solutions

Maine Department of Education

Embedded Designated Support: Text to Speech (TTS)

- Assigning Text to Speech Manually.
- Under Student's profile, select Accessibility Supports and the subject for TTS, be sure to Save changes at the bottom.



nwea State Solutions

Maine Department of Education

Adding TTS as a Designated Support after Starting the Assessment

- What if a student has already started their assessment and needs to have TTS added?
  - TTS can be added by the SAU either before or during the assessment.
  - The Proctor should ask the student to log out of their assessment while they contact the DAC or SAC to go into the student's profile and add TTS. Once TTS has been added, the student can log back in and they will have TTS for the remainder of the assessment.

nwea State Solutions

Maine Department of Education

Accommodations

- Accommodations are changes in procedures or materials that are used to increase equitable access during the assessment for students with documentation of the need on an Individualized Education Plan (IEP) or 504 Plan.

nwea State Solutions

Maine Department of Education

Non-Embedded Accommodations

Non-embedded accommodations can be viewed and edited via both the Registration File and the student's profile.

- Human Reader (Paper-Based Tests ONLY).
- Scribe
  - There are no constructed response questions on the Maine Through Year Assessment.
- American Sign Language.
- Calculator (for entire Math assessment).
- Human Reader for Reading Passages (Students in grades 6+ with a documented print disability, Reading Assessment only).

nwea State Solutions

Maine Department of Education

Embedded Accommodations: Paper-Based Forms

Embedded accommodations can be viewed and edited in both the Registration File and the student's profile.

- Standard Print
- Large Print
- Braille

Note: *Paper-based forms are not adaptive*

nwea State Solutions

Maine Department of Education

Paper-Based Forms: Overview

- For standard and large-print forms, the student's IEP or 504 Plan requires assessments to be paper-based and not administered online.
- Standard print (size 12 font) assessments are print-on-demand. Large print (size 18 font) and braille assessment forms are shipped to schools.
- **The use of paper-based forms must be approved by the Maine DOE.**
- To request a paper-based form, schools must **complete the request form**. The form opens for Spring '24 on April 1.
- After paper-based forms are complete, the proctor (or scribe) must transcribe the responses into the online assessment delivery system exactly as student has responded.
- Transcribing must be completed by the last day of the window.
- All paper-based materials must be destroyed on-site by the last day of the administration window.

Note: *Paper-based forms are not adaptive.*

nwea State Solutions

Maine Department of Education

nwea State Solutions

Not Tested Codes

Not Tested Codes (NTCs)

- Not Tested Codes (NTCs) are used solely by the Maine DOE to track special circumstances in which students' assessment data will not be included in an SAU's or school's aggregated data.
- Only Maine DOE will enter NTCs into the Acacia platform. SAUs should not enter NTCs, and any NTCs entered by SAUs will be removed.

Questions from the Accessibility and Not Tested Codes sections?

FAQ document and slide deck will be shared out after training session.

nwea State Solutions

Preparing for and Monitoring the Assessment

nwea State Solutions

Print Student Test Tickets

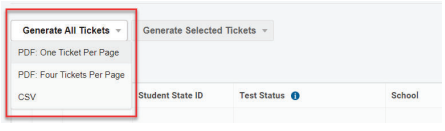
- Available in these formats:
  - PDF Format (one per page).
  - PDF Format (four per page).
  - CSV Export (for bulk printing, can export 100 students).
- Ability to print in two ways.
  - Manage Online Testing page.
  - Individual Student Profile > Test Registrations tab > View Test Session under Actions > PDF icon under Actions.
- Proctors can print on demand!
- Students don't have to be in an online testing group to take their tests, they just need their test tickets.

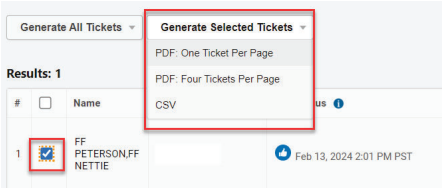
nwea State Solutions

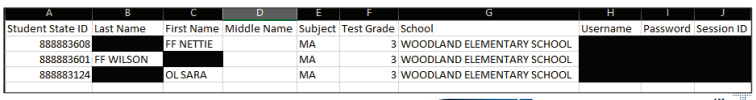
Maine Department of Education

Student Test Tickets

- Online Testing > Manage
- Multiple student test tickets can be printed at once or printed individually.







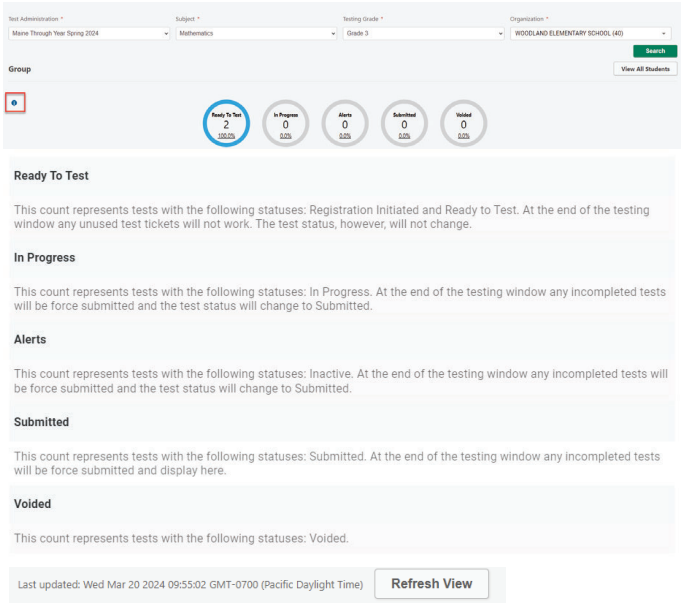
Student State ID	Last Name	First Name	Middle Name	Subject	Test Grade	School	Username	Password	Session ID
888883608	[redacted]	FF NETTIE		MA	3	WOODLAND ELEMENTARY SCHOOL			
888883601	FF WILSON	[redacted]		MA	3	WOODLAND ELEMENTARY SCHOOL			
888883124	[redacted]	OL SARA		MA	3	WOODLAND ELEMENTARY SCHOOL			

nwea State Solutions

Maine Department of Education

Testing Progress

- Testing progress can be viewed at the group, SAU, or school level.
  - Ease of use to allow proctors to more efficiently monitor students.
  - Select the info icon to see what each test status indicates
- The page will retain your filter selections when you refresh the page.
- Testing Status Report in Operational Reports can also help understand where your students are in the assessment as a file export.



nwea State Solutions

Maine Department of Education

Test Resets and Maine DOE Policy

Any assessment that needs to be reset must be approved by the Maine DOE.

- *A reset is when a student will receive a new test ticket and upon logging into the assessment will start at the beginning. All previous answers and results are deleted.*

Below are some situations in which a student's assessment may be reset:

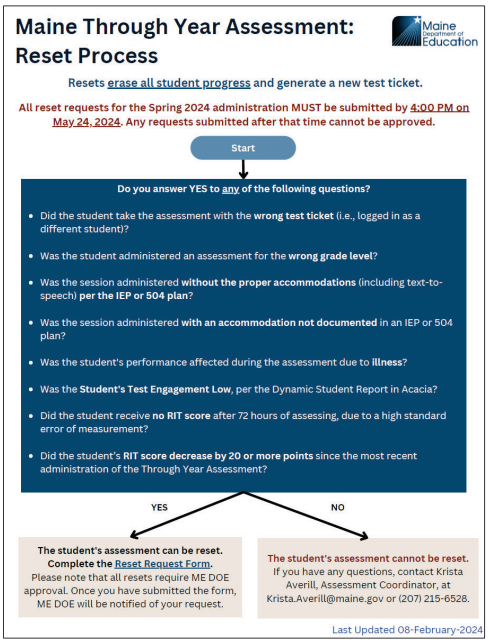
- The student took the assessment with the **wrong test ticke**t (i.e., logged in as a different student).
- The student began or completed assessment for the **wrong grade level**.
- The student began or completed assessment **without the proper accommodations** per the IEP or 504 plan.
- The student began or completed assessment **with an accommodation not documented** in IEP or 504 plan.
- The student's performance was affected during the assessment due to **illness**.
- After completion of the assessment, the Engagement Metric in the Dynamic Student Report shows **low engagement** for the assessment.
- The student received **no RIT score** after 72 hours of assessing, due to a high standard error of measurement.
- The student's **RIT score decreased by 20 or more points** since the most recent administration of the Through Year Assessment.

nwea State Solutions

Maine Department of Education

Reset Process Flow Chart

- Refer to the Process Flow Chart to see if a student can have their assessment reset.
- If the student can have a reset done, you will submit the following form to kick off the approval process.
  - Maine Reset Requests
- **Resets for Spring 2024 must be submitted via the form by 4:00pm EST on May 24, 2024.**
  - *Requests submitted after that time cannot be approved.*



Maine Through Year Assessment: Reset Process

Maine Department of Education

Resets erase all student progress and generate a new test ticket.

All reset requests for the Spring 2024 administration MUST be submitted by 4:00 PM on May 24, 2024. Any requests submitted after that time cannot be approved.

Start

Do you answer YES to any of the following questions?

- Did the student take the assessment with the **wrong test ticket** (i.e., logged in as a different student)?
- Was the student administered an assessment for the **wrong grade level**?
- Was the session administered **without the proper accommodations** (including text-to-speech) **per the IEP or 504 plan**?
- Was the session administered **with an accommodation not documented** in an IEP or 504 plan?
- Was the student's performance affected during the assessment due to **illness**?
- Was the **Student's Test Engagement Low**, per the Dynamic Student Report in Acacia?
- Did the student receive **no RIT score** after 72 hours of assessing, due to a high standard error of measurement?
- Did the student's **RIT score decrease by 20 or more points** since the most recent administration of the Through Year Assessment?

YES

The student's assessment can be reset. Complete the Reset Request Form. Please note that all resets require ME DOE approval. Once you have submitted the form, ME DOE will be notified of your request.

NO

The student's assessment cannot be reset. If you have any questions, contact Krista Averill, Assessment Coordinator, at Krista.Averill@maine.gov or (207) 215-6528.

Last Updated 08-February-2024

nwea State Solutions

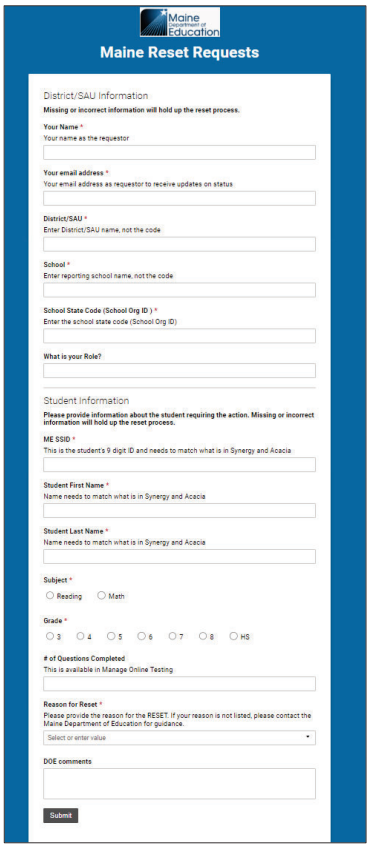
Maine Department of Education

Form for Resets

- Assessments that need a reset must be done by submitting a request via the form found on the right-side pane of the Maine Connections Page.

Maine: Reset requests

- Form to request reset
- Reset process flow



Maine Department of Education

Maine Reset Requests

District/SAU Information

Missing or incorrect information will hold up the reset process.

Your Name *
Your name as the requestor

Your email address *
Your email address as requestor to receive updates on status

District/SAU *
Enter District/SAU name, not the code

School *
Enter reporting school name, not the code

School State Code (School Org ID) *
Enter the school state code (School Org ID)

What is your Role?

Student Information

Please provide information about the student requiring the action. Missing or incorrect information will hold up the reset process.

ME SSID *
This is the student's 9 digit ID and needs to match what is in Synergy and Acacia

Student First Name *
Name needs to match what is in Synergy and Acacia

Student Last Name *
Name needs to match what is in Synergy and Acacia

Subject *
Reading Math

Grade *
3 4 5 6 7 8 HS

# of Questions Completed
This is available in Manage Online Testing

Reason for Reset *
Please provide the reason for the RESET. If your reason is not listed, please contact the Maine Department of Education for guidance.
Select or enter value

DOE comments

Submit

nwea State Solutions

Maine Department of Education

Reset Form Fields

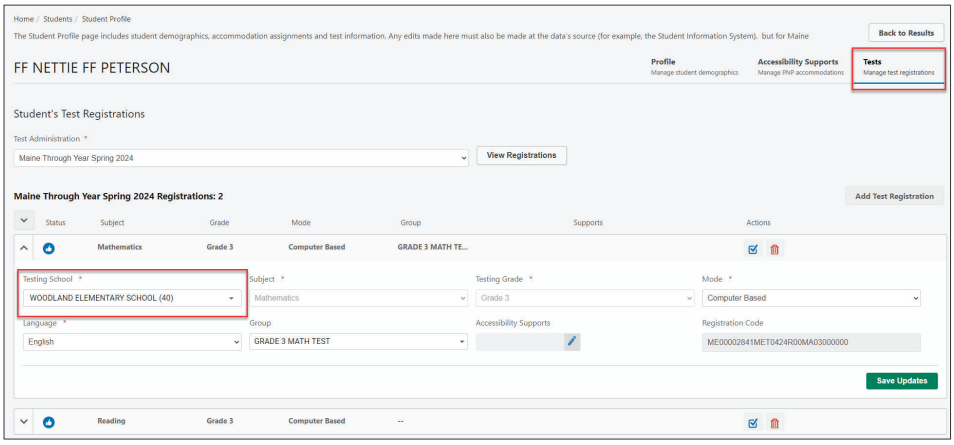
- Please be sure you are providing the correct information in the fields.
  - **Your Name:** The person completing the reset form.
  - **Your email address:** Email address where you can receive updates on status.
  - **District/SAU:** Name of district/SAU - this is not a code.
  - **School:** Name of school – this is not a code.
  - **School State Code (School Org ID):** This is the **school** code.
  - **Student's First & Last Name:** Names need to be what is reflected in Synergy and Acacia.
  - **ME SSID:** This is the student's nine-digit ID
- Incorrect or missing information will hold up the reset process.

nwea State Solutions

Maine Department of Education

Testing School

- The Reporting School field will populate the Testing School.
  - **Reminder:** Reporting School is the school the student attends and at which they receive instruction.
- The Testing School can be changed should the student be taking the assessment at a location other than the Reporting School.
  - Student reports will go to the Reporting School.
- Testing School is located by going to the Student Profile > Tests tab > Testing School



nwea State Solutions

Maine Department of Education

Rostering for Students at Regional and OOS Programs

- Students will be rostered to their attending school. This is known as the “Reporting School”.
- All student reports will be provided to the attending school.
- This will allow educators and staff at the program location to administer the assessment and have access to student assessment results to inform instruction.
- The attending school will share student testing status and/or performance information with the responsible SAU at the request of the responsible SAU.

nwea State Solutions

Maine Department of Education

SAU Transfer Process in Acacia

- Student Mobility
  - Students who move into a school must be enrolled in Synergy by the new attending school immediately after being exited from the previous school.
  - Maine DOE will upload a daily roster/delta file directly to Acacia.
  - It is the responsibility of the new school to ensure that students have the opportunity to finish incomplete portions of the assessment.
    - The new school will need to contact the Maine DOE Assessment Team or the student's old school to get the student's test ticket information if a test is already in progress. This information will be provided to the new school in a secure manner.

nwea State Solutions

Maine Department of Education

SAU Transfer Process in MAP Growth

- Student Mobility
  - The transfer will also need to be done in MAP Growth.
    - The old SAU would remove **current term** from Students profile in MAP Growth.
    - The new SAU would roster the student as normal within MAP Growth.
    - If the same Student State ID is in two different SAUs within MAP Growth, student data from Acacia will not be updated until the conflict is resolved.
- **The steps above are crucial to ensure that the Student State IDs in MAP Growth and Acacia match.**
- For information on removing a testing term from a Student Profile, see the NWEA Connections Article linked here:
  - Can a testing term be removed from a student profile? (nwea.org)

nwea State Solutions

Maine Department of Education

SAU Transfers and MAP Growth Errors

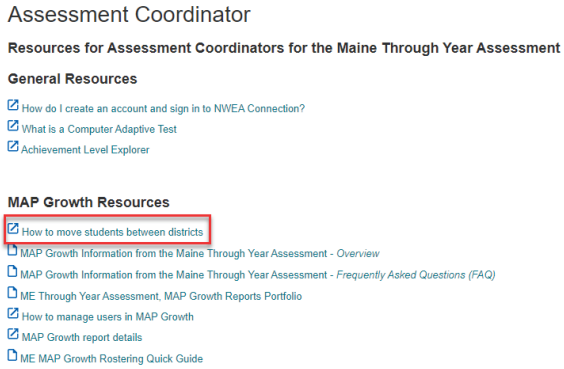
- Users with a System Admin, Data Admin or Assessment Coordinator (DAC) in Acacia will have access to 'Student Import Errors' in the Operational Reports section.
  - Starting in Spring 24, this report will be available throughout the admin windows through the last day of the SAU Cleanup Window.
- If a student has transferred from one SAU to another, it's crucial that both Synergy and MAP Growth reflect the current roster details or the MAP Growth Error of 'This student is duplicated in the upload file' will occur.
  - Error occurs when the student is rostered to both their prior SAU/school and their current SAU/school.
  - To resolve this error, the student must be rostered to the correct reporting SAU/school and term in MAP Growth.
- If one SAU/school needs to know which other SAU/school also has the student on their roster, NWEA Partner Support can provide the school and school state code (i.e. SchoolOrgID) so that the partner can coordinate the correction needed.
- **It is the responsibility of the prior SAU/school to remove the student from the MAP Growth roster for the current term.**

nwea State Solutions

Maine Department of Education

SAU Transfer Process & Historical Data in MAP Growth

- Student Mobility
  - To maintain historical data, SAUs can refer to this process and submit form:
    - How to move students between districts
  - Clever users, if you stop sharing the student as part of your regular Clever sync for that term, this will automatically unenroll them and you won't need to manually do it within the system.



nwea State Solutions

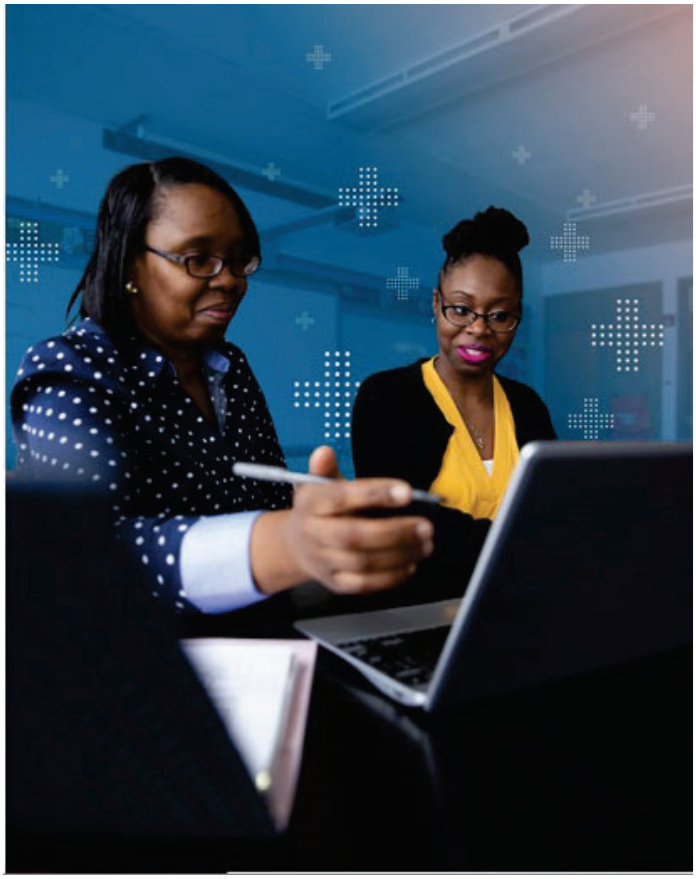
Maine Department of Education

Questions from the Preparing for and Monitoring the Assessment section?

FAQ document and slide deck will be shared out after training session.

nwea State Solutions

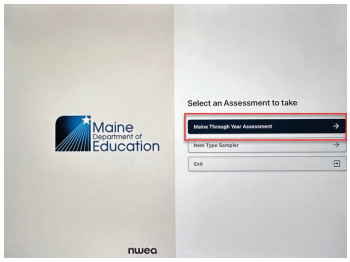
Proctor and Student Experience

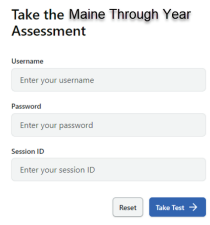


nwea State Solutions

Student Experience - Login

- **Step 1:** Student launches Secure Browser.
- **Step 2:** Communicate that the student should be selecting the **Maine Through Year Assessment** option and not the Item Type Sampler option.
- **Step 3:** From Test Ticket, student enters username, password, and Session ID.
- **Step 4:** Student verifies text on screen is accurate while Proctor monitors.
- **Step 5:** Proctor gives verbal approval to begin assessment.

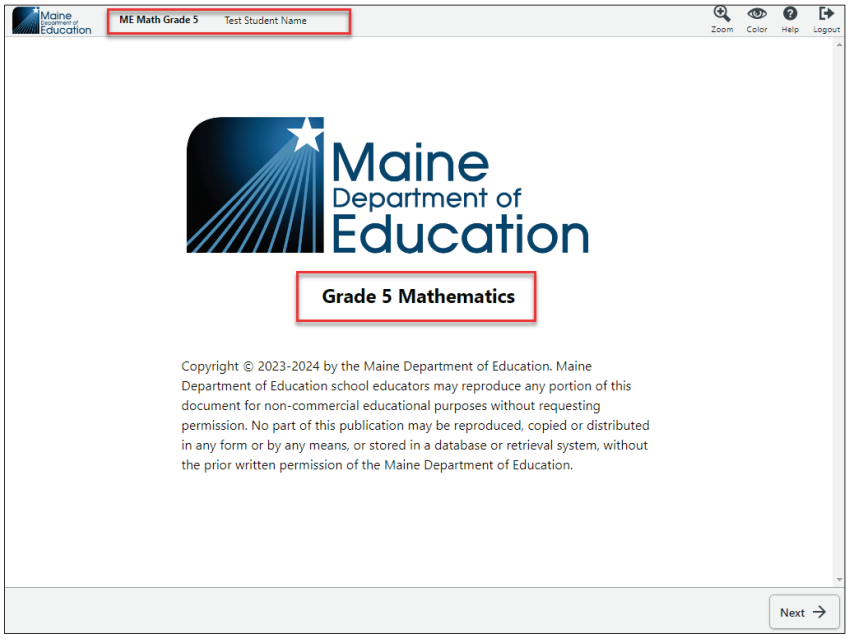




Maine Department of Education

nwea State Solutions

Student Experience – Summary Screen

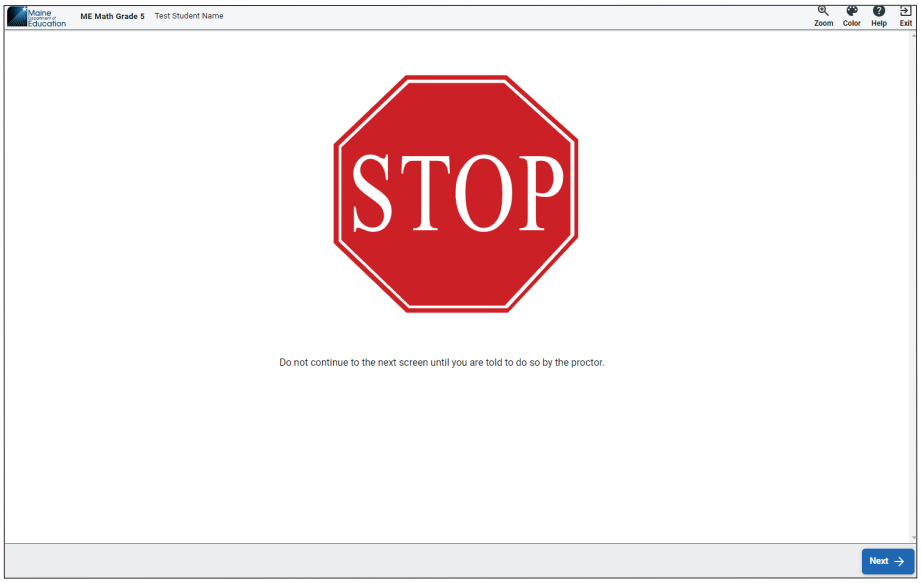


- Student to verify name, grade, and subject on this screen is correct.

nwea State Solutions

Maine Department of Education

Student Experience – Proctor Screen



- Student to wait on this screen until Proctor gives approval to select next to begin assessment.

nwea State Solutions

Maine Department of Education

Student Experience - Logout

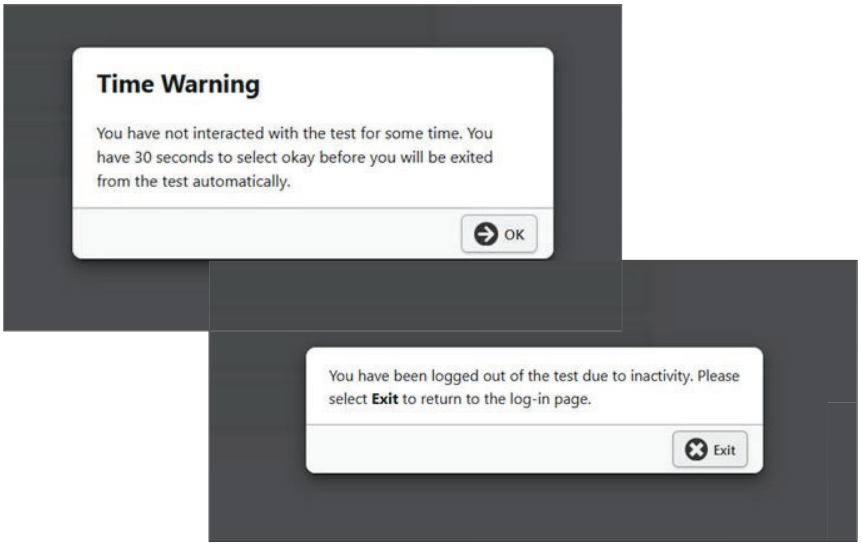
- If a student needs to step away, they can exit/logout of the assessment.
- Once they log back in, they will pick up where they left off and all questions previously answered will be saved.

nwea State Solutions

Maine Department of Education

Student Experience - Inactivity

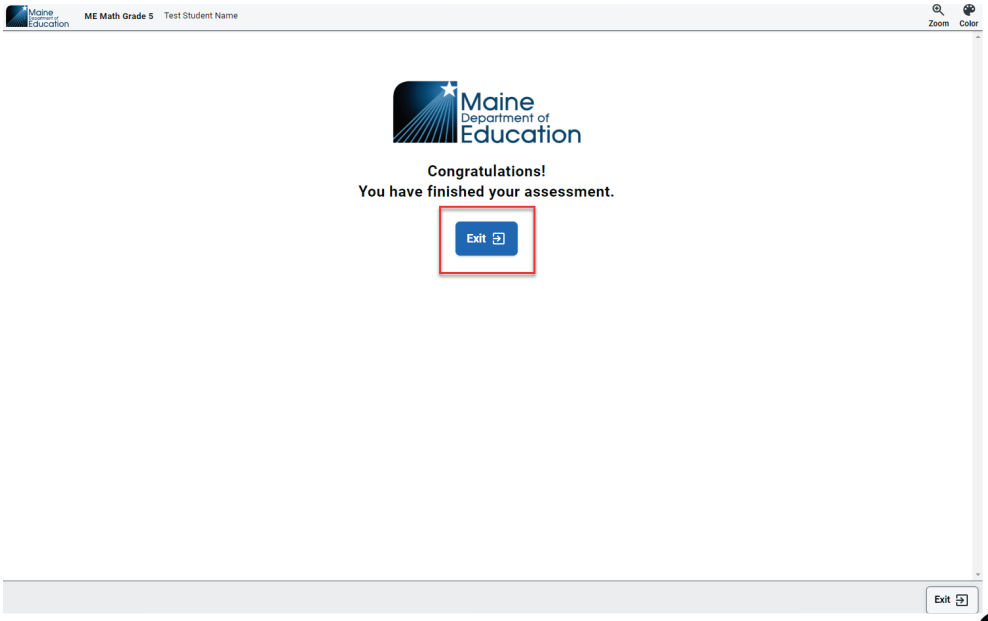
- Message appears when student has been idle for 14.5 minutes.
- If student doesn't click within the screen, then they will get the time out message.
- Once they receive this message, clicking exit is their only option.



nwea State Solutions

Maine Department of Education

Student Experience – End of Assessment



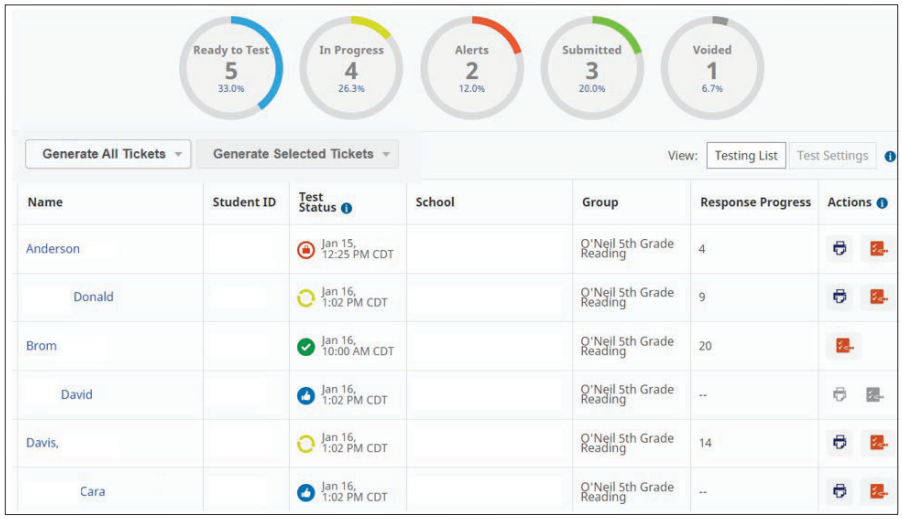
- Starting in the Spring 24 admin, there will be a blue exit button in the middle of the 'Congratulations' page to ensure the test is being fully submitted by the student.

nwea State Solutions

Maine Department of Education

Proctor Experience - Testing Progress

- Ease of use to allow proctors to more efficiently monitor students from the test group, school, or by looking up students individually.
- Testing Status Report can help understand where your students are in testing.

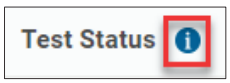


nwea State Solutions

Maine Department of Education

Proctor Experience - Testing Progress

- Icons and Descriptions for monitoring testing progress.
- Select the info icon next to Test Status to bring up the Test Status Icon Key.



Test Status Icon Key

In Progress	Test session is currently active.
Deactivated	Registration has been deactivated and replaced.
Inactive	Test session has become inactive.
Saved	Score is saved and ready to view.
Finished	Test has been submitted by student.
Expired	Test session was submitted at end of testing window.
Processing Reset	Test session is being reset.
Registration Initiated	Registration with testing platform has been initiated.
Ready To Test	Test session is ready for student.

nwea State Solutions

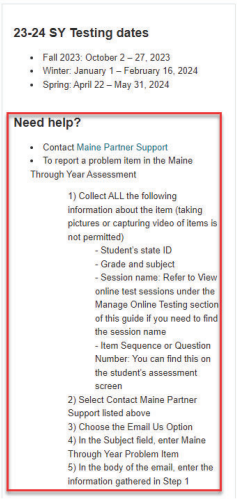
Maine Department of Education

Reporting Issues

- Problem item reports
  - Should students experience an item that is potentially problematic, a problem item report can be submitted via the Maine Connections Page.
  - Click Contact Maine Partner Support under Need Help? Section.
  - Choose the Email Us option.
  - The following information will be needed:
    - Subject Name: *Maine Through Year Problem Item*
    - State Student ID
    - Grade and Subject
    - Session Name
    - Item Sequence or Question Number

Note: *Do not take photos or provide details around the content of the item.*

Maine Department of Education



23-24 SY Testing dates

- Fall 2023: October 2 – 27, 2023
- Winter: January 1 – February 16, 2024
- Spring: April 22 – May 31, 2024

Need help?

- Contact Maine Partner Support
- To report a problem item in the Maine Through Year Assessment

1) Collect ALL the following information about the item (taking pictures or capturing video of items is not permitted)
   - Student's state ID
   - Grade and subject
   - Session name: Refer to View online test sessions under the Manage Online Testing section of this guide if you need to find the session name
   - Item Sequence or Question Number: You can find this on the student's assessment screen
2) Select Contact Maine Partner Support listed above
3) Choose the Email Us Option
4) In the Subject field, enter Maine Through Year Problem Item
5) In the body of the email, enter the information gathered in Step 1

nwea State Solutions

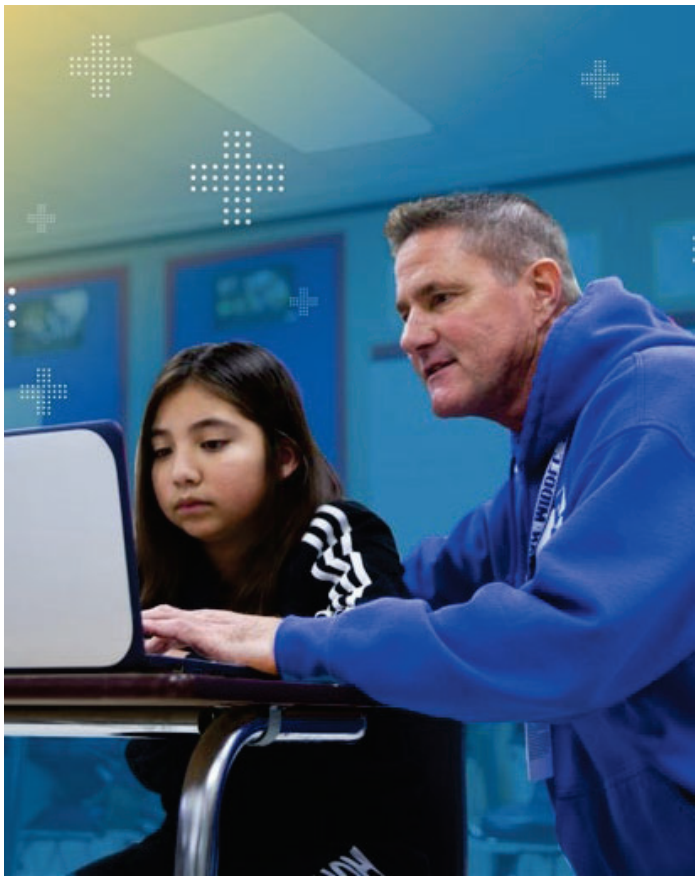
Maine Department of Education

Questions from the Proctor and Student Experience section?

FAQ document and slide deck will be shared out after training session.

nwea State Solutions

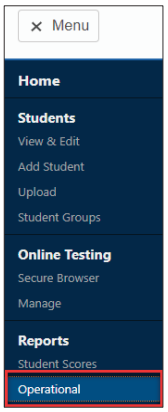
Operational Reports



nwea State Solutions

Operational Reports

- Operational Reports are designed to help DACs and SACs monitor the testing status and the status of materials.
- To access Operational Reports:
  1. In the main menu, select **Reports > Operational.**
  2. Select **Organization** and **Report Type** from the drop-down lists.
  3. Select Find.
  4. Information about the report will appear below. Select the icon in the Download column to download the report.



nwea State Solutions

Maine Department of Education

Operational Reports

Operational Report	Description
Registration Report	Report details the students that were rostered to the administration.
NTC Usage Report	Report details student assessments that have NTCs assigned.
Summary Test Status Report	Report is a summarization of testing statuses.
Testing Status Report	Report details the status of each student's assessment.
Student Mobility Report	Report details students that have been transferred from one school and/or district to another.
Material Orders Report	This report summarizes the quantity of assessments by school that were assigned a paper, large print, or braille accommodation.
Organization Report	This report details the organizational hierarchy data in the system; source of data is the state org file.
Student Score Data File	Student Score Data file will contain all valid test events for assessments completed within the administration by grade and content area.
MAP Growth Roster Errors	This report will contain all the MAP Growth Roster errors that did not sync from MAP Growth to Acacia. This will be at the organization level.

nwea State Solutions

Maine Department of Education

nwea State Solutions

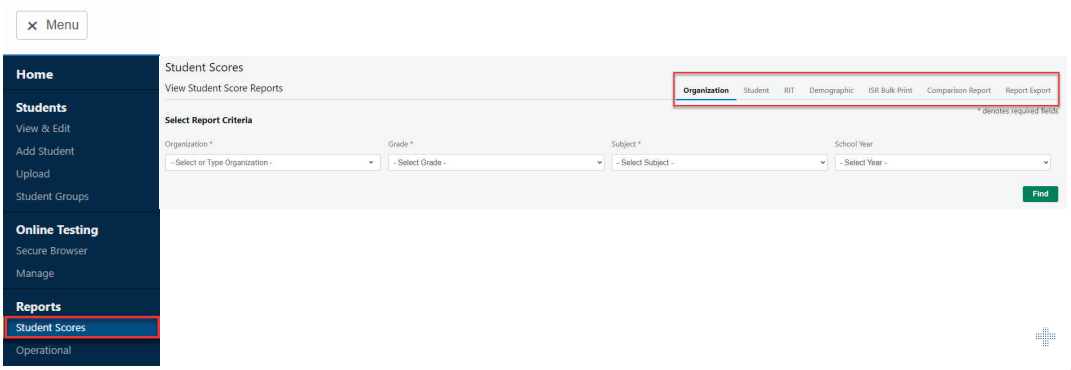
Data and Reporting

New Acacia Reporting Features for Spring 2024

- Report download icons have been moved to the top of the page.
- Report view icons within a report have been moved to the left side of the page in a tab format.
- Demographic filters have been moved to the top of the page.
- New pencil icon available on some reports to adjust filters more quickly.
- Organization data export now available in csv format.

Accessing Reports

- To access reports, go the Student Scores under the Reports section in the menu.
- The top right tabs will show you the categories you can select.



nwea State Solutions

Maine Department of Education

Data and Reporting – What is Available

Report / File	Access	Description
Student Score Data File (SSDF)	State and SAU Level	• Will contain all valid test events for assessments completed within the administration by grade and subject • Will include Maine scale score, Maine scale score SEM, overall RIT for Math and Reading, RIT SEM, RIT Achievement Percentile for Math and Reading, Instructional Area sub scores
Organization Report – By District **SPRING ONLY**	DACs and Admins	• Demographic filters • Averages for the SAU • List view and histogram view
Organization Report – By School **SPRING ONLY**	SACs and Admins	• Average for the school • Graphic views of student performance • List view and histogram view
Organization Report – By Group **SPRING ONLY**	Instructors and above	• Will have averages for the group • Graphic views of student performance • Users will be able to create groups • List view and histogram view
Dynamic Student Report	Instructors and above	• Available on a rolling basis • Student performance data in an easy printable format focused on each content area separately • Item level information by standard, item type, and difficulty

nwea State Solutions

Maine Department of Education

Organization Reports – SAU & School Level
Spring ONLY

What this report offers

+ Summative data by achievement level by group, school and district.
+ Includes number of students tested and percentages by achievement levels.
+ Individual student achievement data for students in a specific group.

Questions it helps answer

+ How are our students doing overall?
+ How are we performing compared to Maine benchmarks?
+ Which is our lowest reporting category? Our highest?

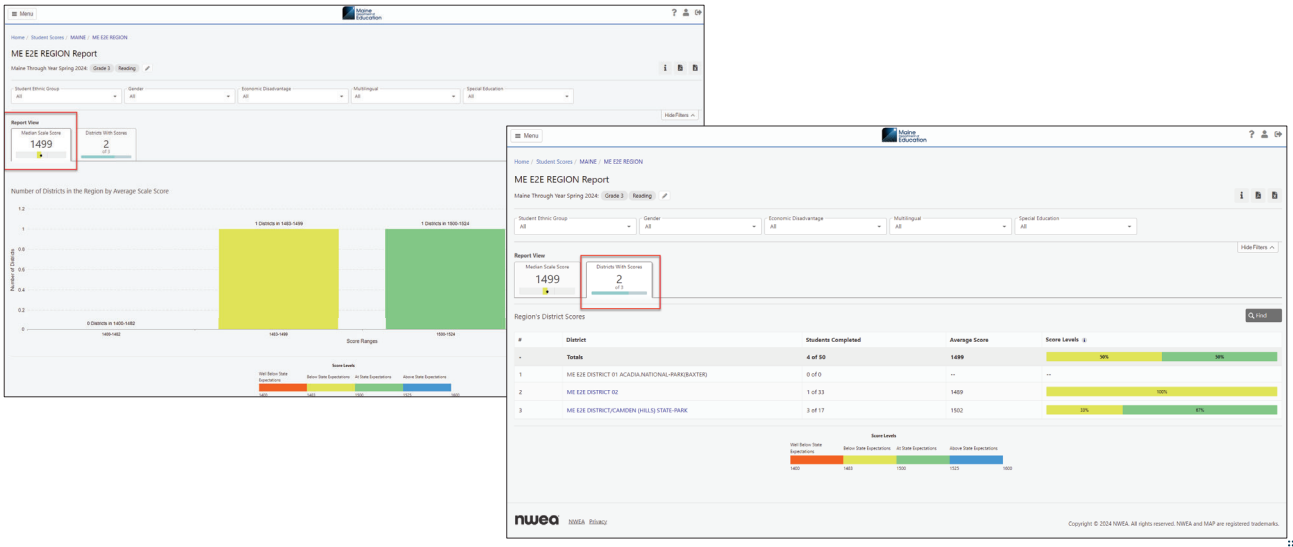
When to use & what to consider

+ After testing, to see results for Spring admins.
+ As part of instructional decision-making process.
+ When you want to use data to inform student grouping.
+ Displays data from a single session.
+ Can be downloaded as pdf or csv file.
+ Columns can be sorted.

nwea State Solutions

Maine Department of Education

Organization Reports – SAU & School Level
Spring ONLY



nwea State Solutions

Dynamic Student Reports – Key Information

What this report offers

+ Student-level data to support each student's progress.
+ Identifies which standards students were able to successfully answer questions relating to.
+ Test details around student engagement, test duration, and tools used.

Questions it helps answer

+ Is this student on track?
+ What are this student's relative strengths and suggested areas of focus?
+ How can I leverage those relative strengths and suggested areas of focus to help this student?

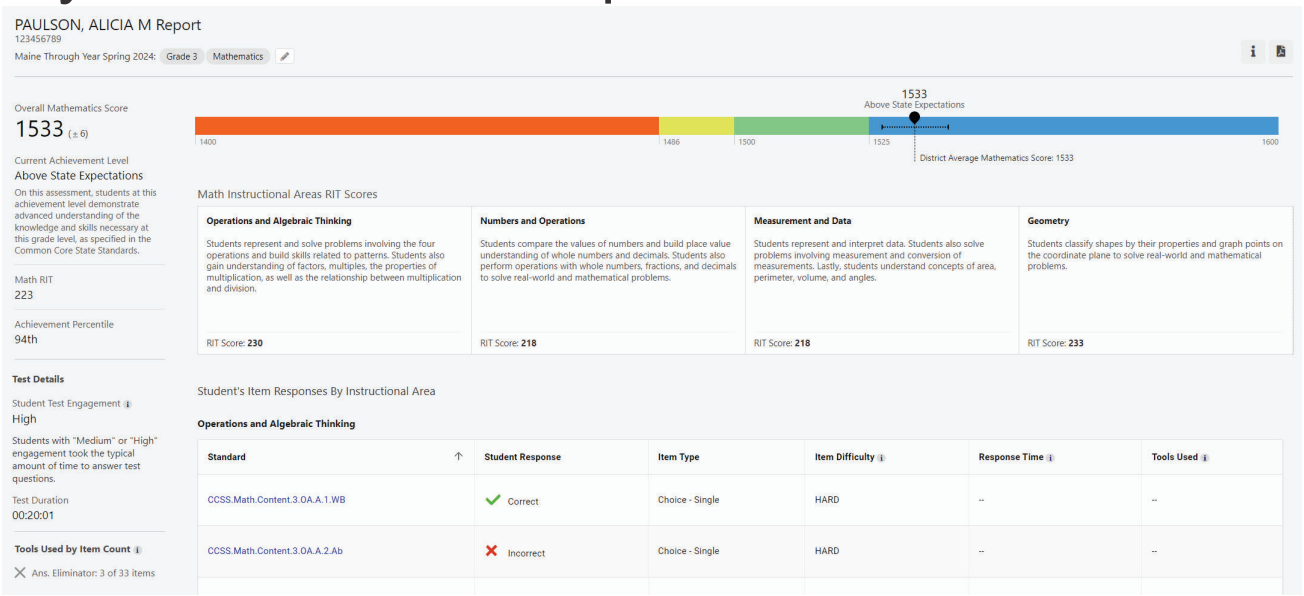
When to use & what to consider

+ Can be downloaded as a pdf file and printed.

nwea State Solutions

Maine Department of Education

Dynamic Student Report



Note: Maine-Specific Scale Score is only available in Spring.

nwea State Solutions

Maine Department of Education

Data and Reporting – What is Available

Report / File	Access	Description
RIT Report	Instructors and above	• Will contain RIT scores for students in an organization (SAU and school) organized by student group • Will include Overall RIT for Math and Reading, along with the Achievement Percentile and RIT score for each reporting category
Demographic Report **SPRING ONLY**	Instructors and above	• Will contain the Average Scale Score in Math and Reading for students in various demographic or targeted groups • Helps educators identify achievement trends for specific genders, ethnicities, or other groups such as Limited English or Economically Disadvantaged
Individual Student Report **SPRING ONLY**	Instructors and above Parents and families	• Designed to show a student's achievement on the Maine Through Year Assessment in Math and Reading to parents and families • Educators will be able to print these reports in batches, making it easier to distribute after Spring testing is complete

nwea State Solutions

Maine Department of Education

RIT Reports

What this report offers

+ RIT information for all students matching the search criteria.
+ Includes RIT score achievement percentile and reporting category RIT.

Questions it helps answer

+ How is the reporting group doing overall?
+ How does the achievement percentile compare for this student?
+ What is the lowest and highest instructional area?

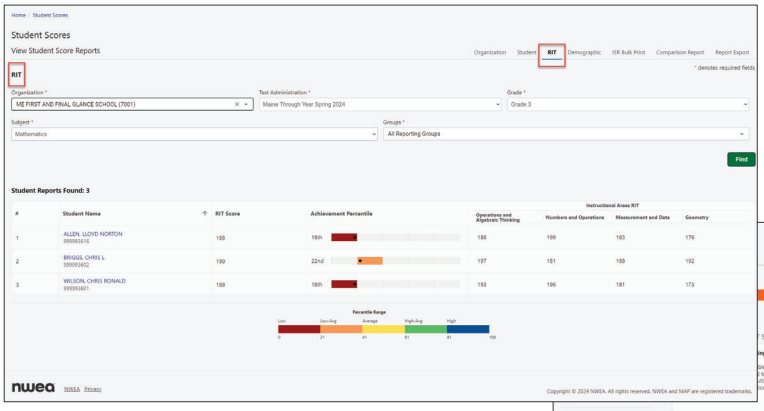
When to use & what to consider

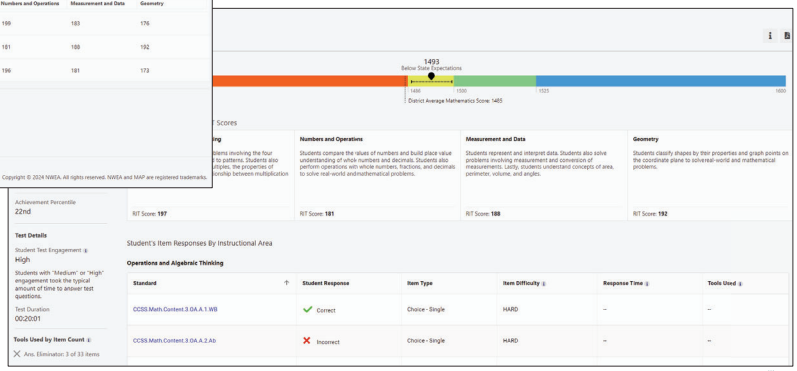
+ After testing, to see achievement data.
+ As part of the instructional decision-making process.
+ When you want to use data to inform student grouping.
+ All columns can be sorted for flexibility in looking at data.

nwea State Solutions

Maine Department of Education

RIT Reports





nwea State Solutions

Maine Department of Education

Demographic Report – **Spring ONLY**

What this report offers

+ Average scale scores.
+ Average reporting category scores.
+ Distribution of scale scores for demographic groups such as gender, ethnicity/race, and targeted groups.

Questions it helps answer

+ Are there any trends or differences among genders, ethnicities, or other groups such as Multilingual Learners or Economically Disadvantaged?

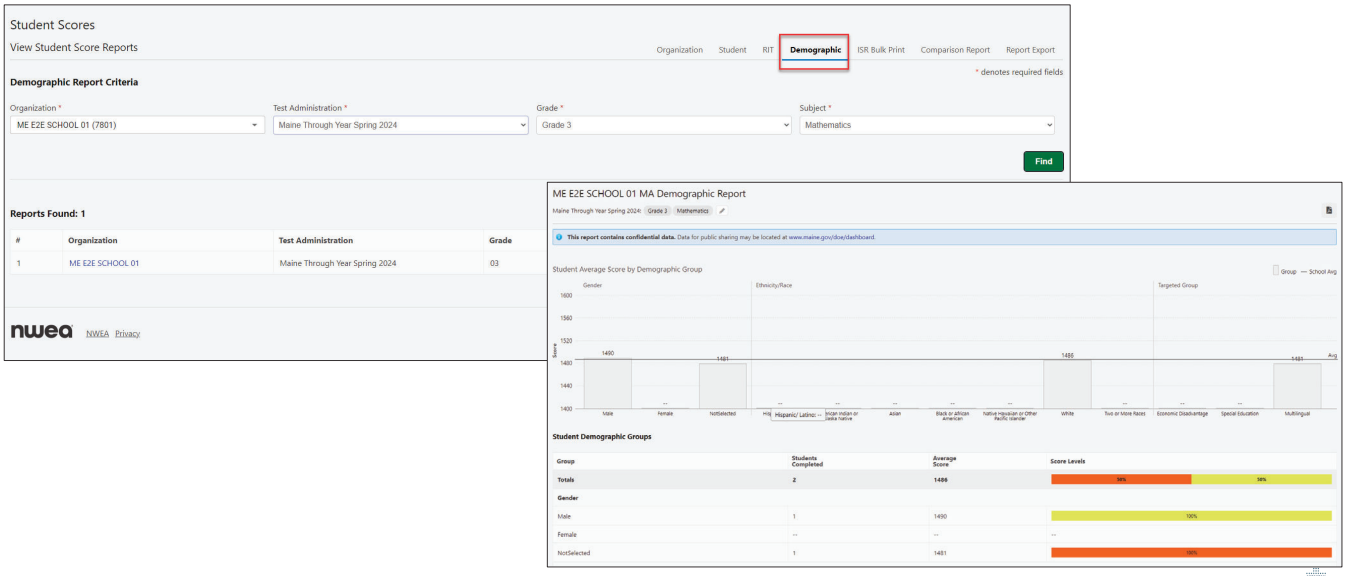
When to use & what to consider

+ The Demographic Report will only be available with your Spring results and will not be available in the Fall and Winter.

nwea State Solutions

Maine Department of Education

Demographic Reports – **Spring ONLY**



nwea State Solutions

Maine Department of Education

Individual Student Reports (ISRs) - **Spring ONLY**

What this report offers

+ Student-level data to support each student's progress.

Questions it helps answer

+ How is the student performing relative to grade-level expectations in reading and math?
+ What are this student's relative strengths and suggested areas of focus?

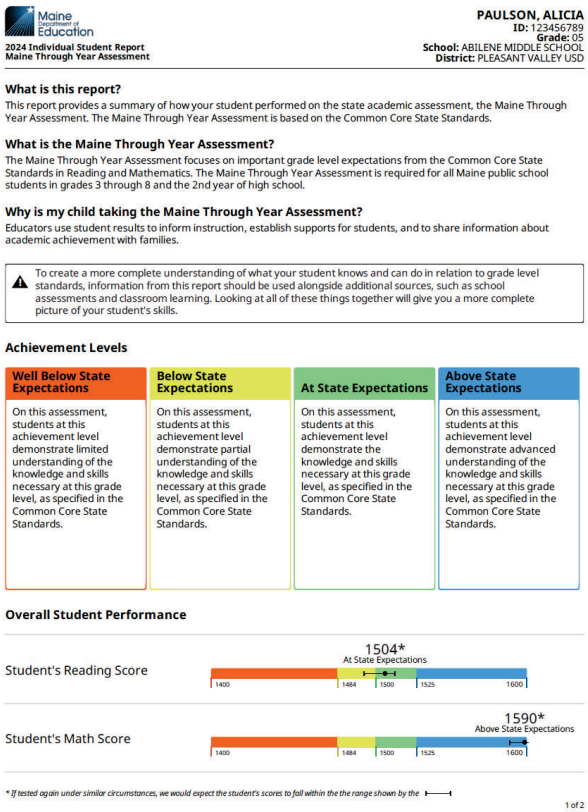
When to use & what to consider

+ ISR's will only be available for the Spring admins.
+ Printed and distributed by SAUs/Schools after administration.
+ To provide to parents and families to provide student performance.

nwea State Solutions

Maine Department of Education

ISR
Available in Spring Only



Maine Department of Education
2024 Individual Student Report
Maine Through Year Assessment

PAULSON, ALICIA
ID: 123456789
Grade: 05
School: ABILENE MIDDLE SCHOOL
District: PLEASANT VALLEY USD

What is this report?
This report provides a summary of how your student performed on the state academic assessment, the Maine Through Year Assessment. The Maine Through Year Assessment is based on the Common Core State Standards.

What is the Maine Through Year Assessment?
The Maine Through Year Assessment focuses on important grade level expectations from the Common Core State Standards in Reading and Mathematics. The Maine Through Year Assessment is required for all Maine public school students in grades 3 through 8 and the 2nd year of high school.

Why is my child taking the Maine Through Year Assessment?
Educators use student results to inform instruction, establish supports for students, and to share information about academic achievement with families.

⚠ To create a more complete understanding of what your student knows and can do in relation to grade level standards, information from this report should be used alongside additional sources, such as school assessments and classroom learning. Looking at all of these things together will give you a more complete picture of your student's skills.

Achievement Levels

Well Below State Expectations	Below State Expectations	At State Expectations	Above State Expectations
On this assessment, students at this achievement level demonstrate limited understanding of the knowledge and skills necessary at this grade level, as specified in the Common Core State Standards.	On this assessment, students at this achievement level demonstrate partial understanding of the knowledge and skills necessary at this grade level, as specified in the Common Core State Standards.	On this assessment, students at this achievement level demonstrate the knowledge and skills necessary at this grade level, as specified in the Common Core State Standards.	On this assessment, students at this achievement level demonstrate advanced understanding of the knowledge and skills necessary at this grade level, as specified in the Common Core State Standards.

Overall Student Performance

Student's Reading Score — 1504* At State Expectations (1400 | 1484 | 1500 | 1525 | 1600)

Student's Math Score — 1590* Above State Expectations (1400 | 1484 | 1500 | 1525 | 1600)

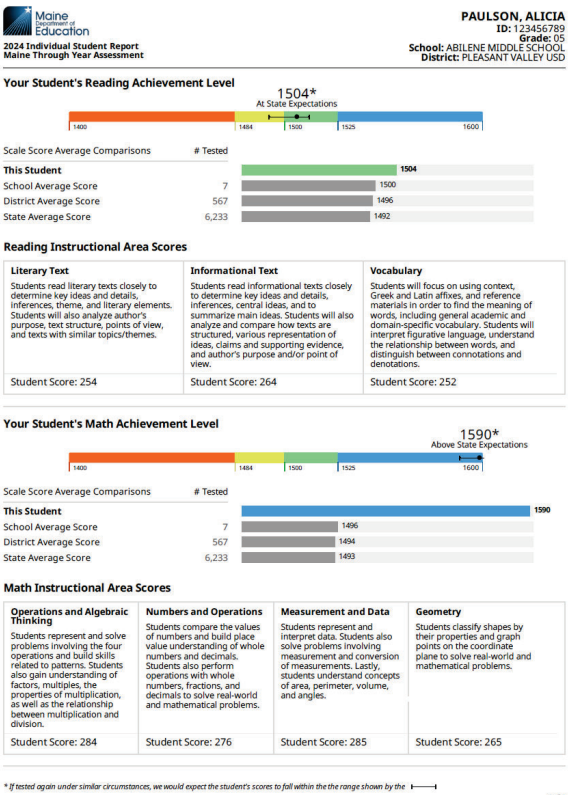
** If tested again under similar circumstances, we would expect the student's scores to fall within the the range shown by the ⊢⊣*

1 of 2

nwea State Solutions

Maine Department of Education

ISR
Available in Spring Only



Maine Department of Education
2024 Individual Student Report
Maine Through Year Assessment

PAULSON, ALICIA
ID: 123456789
Grade: 05
School: ABILENE MIDDLE SCHOOL
District: PLEASANT VALLEY USD

Your Student's Reading Achievement Level

1504* At State Expectations (1400 | 1484 | 1500 | 1525 | 1600)

Scale Score Average Comparisons	# Tested	
This Student		1504
School Average Score	7	1500
District Average Score	567	1496
State Average Score	6,233	1492

Reading Instructional Area Scores

Literary Text	Informational Text	Vocabulary
Students read literary texts closely to determine key ideas and details, inferences, theme, and literary elements. Students will also analyze author's purpose, text structure, points of view, and texts with similar topics/themes.	Students read informational texts closely to determine key ideas and details, inferences, central ideas, and to summarize main ideas. Students will also analyze and compare how texts are structured, various representation of ideas, claims and supporting evidence, and author's purpose and/or point of view.	Students will focus on using context, Greek and Latin affixes, and reference materials in order to find the meaning of words, including general academic and domain-specific vocabulary. Students will interpret figurative language, understand the relationship between words, and distinguish between connotations and denotations.
Student Score: 254	Student Score: 264	Student Score: 252

Your Student's Math Achievement Level

1590* Above State Expectations (1400 | 1484 | 1500 | 1525 | 1600)

Scale Score Average Comparisons	# Tested	
This Student		1590
School Average Score	7	1496
District Average Score	567	1494
State Average Score	6,233	1493

Math Instructional Area Scores

Operations and Algebraic Thinking	Numbers and Operations	Measurement and Data	Geometry
Students represent and solve problems involving the four operations and build skills related to patterns. Students also gain understanding of factors, multiples, the properties of multiplication, as well as the relationship between multiplication and division.	Students compare the values of numbers and build place value understanding of whole numbers and decimals. Students also perform operations with whole numbers, fractions, and decimals to solve real-world and mathematical problems.	Students represent and interpret data. Students also solve problems involving measurement and conversion of measurements. Lastly, students understand concepts of area, perimeter, volume, and angles.	Students classify shapes by their properties and graph points on the coordinate plane to solve real-world and mathematical problems.
Student Score: 284	Student Score: 276	Student Score: 285	Student Score: 265

** If tested again under similar circumstances, we would expect the student's scores to fall within the the range shown by the ⊢⊣*

2 of 2

nwea State Solutions

Maine Department of Education

Data and Reporting – **New Reports!**

Report / File	Access	Description
Comparison Summary Report **SPRING ONLY**	State, SAU, and School Level	• View summary of student performance at one or more organizations • Aggregated State, District, and School comparison reports based on overall scale score • Ability to select organizations, school years, test administrations, grade(s), and subject(s) • Once report is created, ability to drill down to student demographics: gender, student ethnic group, economic disadvantage, multilingual, and special education • Ability to save reports for easy access to frequent queries
Student Results File	State and SAU Level	• Reportable student-level assessment results for an organization • All grades and subjects for a district or state in one file • Includes: student data (enrollment, demographics, etc.), test event data (including Student Engagement Metrics), overall scale score and reporting category data, RIT score and instructional area data, and accommodations

nwea State Solutions

Maine Department of Education

Comparison Summary Report – **Spring ONLY**

What this report offers

+ Report creator form to create reports that compare aggregate student performance at one or more orgs based on the Maine scale score.
+ Compares aggregate student performance by org.
+ Ability to view by multi-grade and subject.
+ Ability to filter results by student demographics.

Questions it helps answer

+ Number of students tested within the org by grade and subject.
+ Average score of students within the org by grade and subject.
+ Visual of score level percentages within the org by grade and subject.

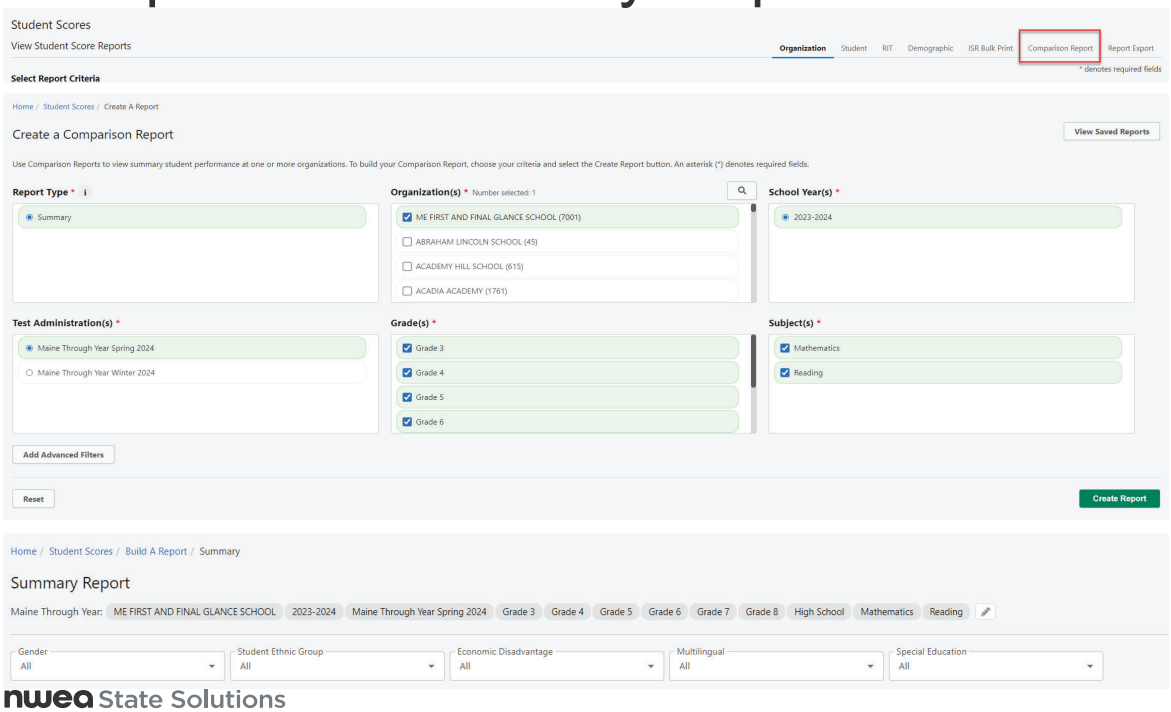
When to use & what to consider

+ Ability to rerun saved report queries by utilizing the bookmark feature.
+ Report will be available starting in the Spring 24 admin and available for the Spring admins only (including Spring 23).

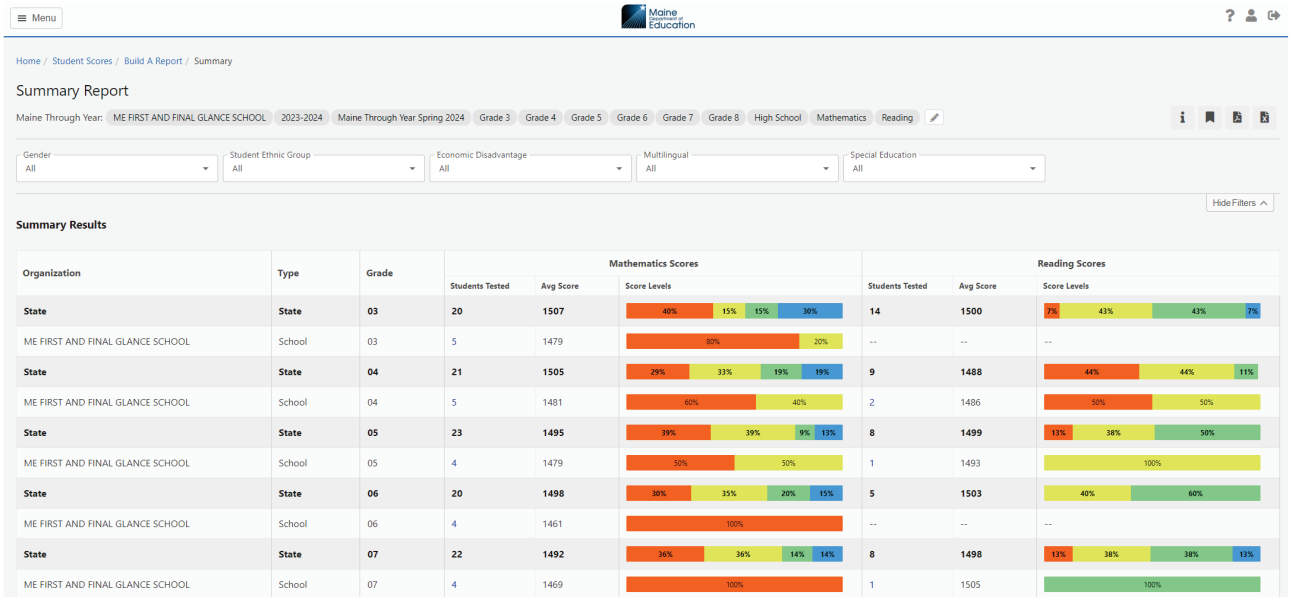
nwea State Solutions

Maine Department of Education

Comparison Summary Report Builder

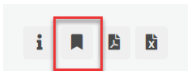


Comparison Summary Report

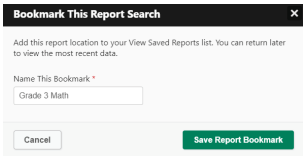


Comparison Summary Report Bookmark

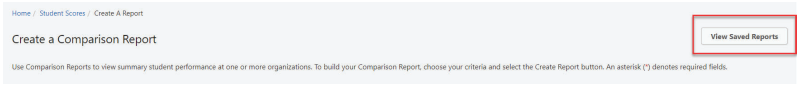
+ Once a Summary report has been run, a bookmark option will be available.



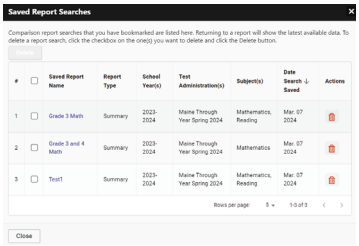
+ Name your bookmark and save.



+ Access your saved reports from the Report Builder screen.



+ View your list of saved reports.



nwea State Solutions

Student Results File

What this report offers

+ Report creator form for downloadable reports.
+ Student results file can be regenerated several times throughout the day during the admin window.

Questions it helps answer

+ Report will include student results, student engagement data, instructional areas, demographics, etc.

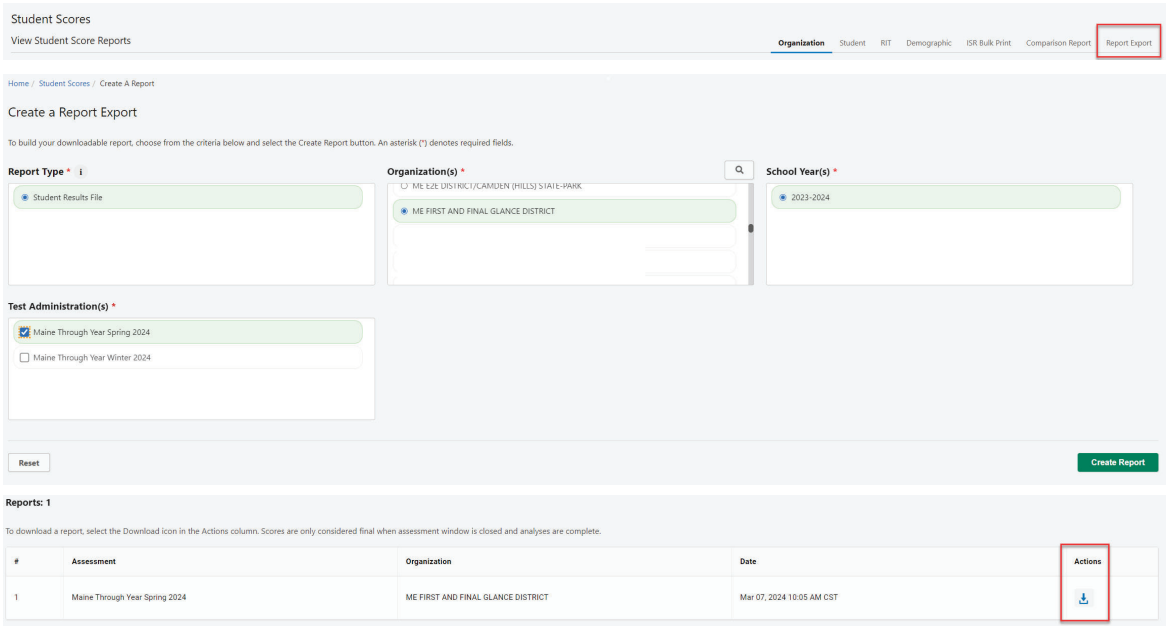
When to use & what to consider

+ Can be downloaded as a csv file.
+ Real time report available throughout the admin window.
+ Report will be available starting in the Spring 24 admin.
+ Report also available for Fall 23 and Winter 24, however the scale score columns will be blank.

nwea State Solutions

Maine Department of Education

Student Results File Creator



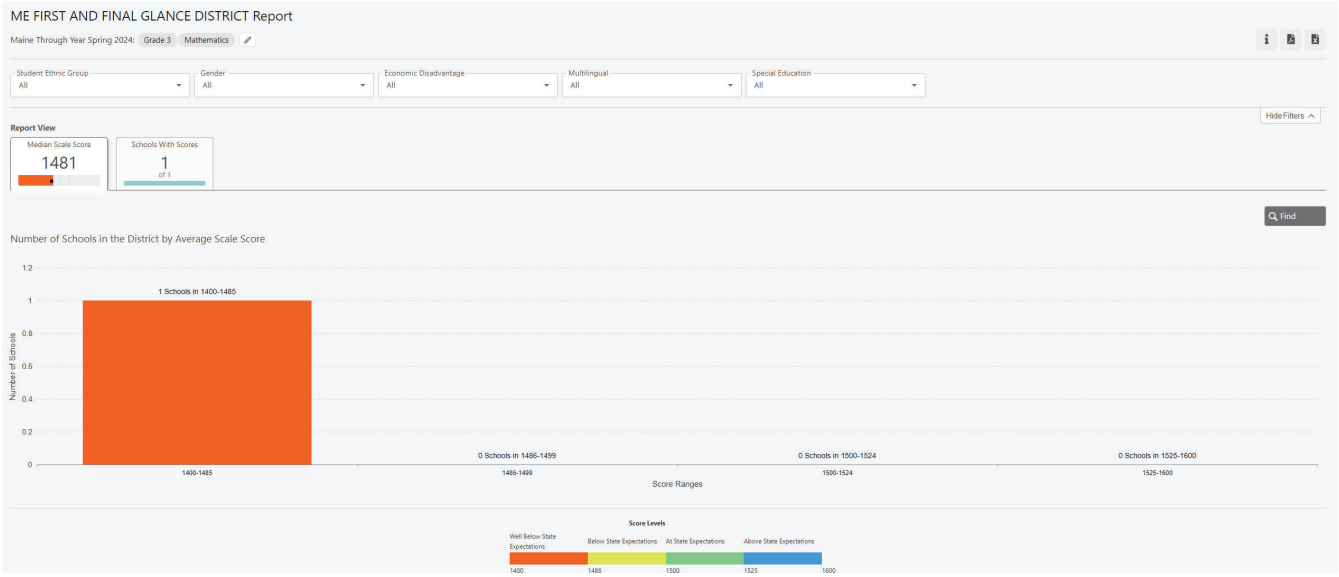
nwea State Solutions

Student Results File

Test Admi	School Ye	Testing Di	Testing Di	Testing Sc	Testing Sc	Reporting	Reporting	Reporting	Reporting	Student Sta	Student Li	Student Fi	Student M	Enrolled C	Test Type	Test Name	Subject	Test Grade	Test Comp	Test Elaps	Valid Atte	Invalid Att	Not Teste	Student Ei	Scale Scor	Scale Scor	Scale Scor	RIT Scale S	RIT SEM	RIT Achiev	Norms Re	Estimated	Instructio	Instructio	Instructio	Instru
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999911603	BURGES	BRITTANY	DANIELLA	11		Maine Thr	RE	HS	########	45:44.2	0	NEI										Literary Text		Informational Te		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999911605	VASQUEZ	JERRY	GABRIEL	11		Maine Thr	MATH	HS	########	10:08.0	1				1488	9	Well Belo	State Expectations				Operations and Alge		The Real and Cor		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999911605	VASQUEZ	JERRY	GABRIEL	11		Maine Thr	RE	HS	3/1/2024	02:04.7	1				1483	6	Well Belo	202	3.41	11	2020	32	Literary Te	210	Informatic		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999911614	BOLES	LIGIA		11		Maine Thr	MATH	HS	3/1/2024	14:56.8	1				1475	9	Well Belo	202	4.28	7	2020	32	Operation	215	The Real a		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999911619	AREVALO	JENELLE		11		Maine Thr	MATH	HS	########	18:03.0	1				1493	7	Below Sta	219	3.33	24	2020	32	Operation	220	The Real a		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999912604	BERRY	PERRY		12		Maine Thr	MATH	HS	########	56:12.0	1				1488	9	Well Belo	229	3.78	42	2020	32	Operation	213	The Real a		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999993601	WILSON	CHRIS	RONALD	3		Maine Thr	MATH	3	########	12:59.9	1				1481	8	Well Belo	188	3.94	18	2020	32	Operation	193	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999993602	BRIGGS	CHRIS	L	3		Maine Thr	MATH	3	########	41:50.6	1				1493	6	Below Sta	190	3.85	22	2020	32	Operation	197	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999993602	BRIGGS	CHRIS	L	3		Maine Thr	RE	3	2/1/2024	36:17.0	0	NEI										Literary Text		Informational Te		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999993616	ALLEN	LLOYD	NORTON	3		Maine Thr	MATH	3	########	08:51.0	1				1480	6	Well Belo	188	3.1	18	2020	32	Operation	188	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999993618	BALLARD	ANGELICA	WAGNER	3		Maine Thr	MATH	3	3/1/2024	06:00.6	1				1464	6	Well Belo	164	3.73	1	2020	32	Operation	151	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999993623	COOPER	RYDER		3		Maine Thr	MATH	3	3/4/2024	32:33.0	1				1476	5	Well Belo	175	3.14	3	2020	32	Operation	183	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999994602	DAVIS	TAYLOR	RODRIGUE	4		Maine Thr	MATH	4	########	29:43.8	1				1494	5	Below Sta	199	3.82	23	2020	32	Operation	204	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999994604	WILSON	PAMELA		4		Maine Thr	RE	4	########	10:24.8	1				1491	6	Below State Expectations						Literary Text		Informational Te		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999994605	BRIGGS	PRILLA	FRANCESA	4		Maine Thr	MATH	4	########	37:33.3	1				1489	5	Below Sta	194	4.15	14	2020	32	Operation	190	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999994605	BRIGGS	PRILLA	FRANCESA	4		Maine Thr	RE	4	########	14:21.9	1				1481	8	Well Below State Expectations						Literary Text		Informational Te		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999994617	COOK	CLINT		4		Maine Thr	MATH	4	3/1/2024	06:58.0	1				1467	4	Well Belo	170	3.16	1	2020	32	Operation	183	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999994618	BECK	MYRA	NORENE	4		Maine Thr	MATH	4	########	42:25.1	1				1474	4	Well Belo	172	3.4	1	2020	32	Operation	169	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999994650	CASTANEI	AANANDINI		4		Maine Thr	MATH	4	3/4/2024	47:33.8	1				1482	4	Well Belo	192	2.91	12	2020	32	Operation	198	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999995605	DAVIS	CLINT	GEOFFREY	5		Maine Thr	MATH	5	########	54:41.1	1				1485	5	Below Sta	196	4.38	9	2020	32	Operation	187	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999995605	DAVIS	CLINT	GEOFFREY	5		Maine Thr	RE	5	########	03:29.4	1				1493	7	Below Sta	193	4.86	13	2020	32	Literary Te	203	Informatic		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999995617	CRAIG	NATHANII	G	5		Maine Thr	MATH	5	3/1/2024	04:39.2	1				1468	4	Well Below State Expectations						Operation	171	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999995623	BARNES	CAROLYN		5		Maine Thr	MATH	5	########	20:21.8	1				1478	4	Well Belo	194	2.9	7	2020	32	Operation	197	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999995645	DIXON	CHRIS		5		Maine Thr	MATH	5	3/4/2024	36:05.0	1				1485	4	Below Sta	198	2.93	11	2020	32	Operation	202	Numbers		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999996601	HALL	JESSICA		6		Maine Thr	MATH	6	########	23:19.7	1				1469	8	Well Belo	200	4.45	10	2020	32	Operation	208	The Real a		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999996609	HERRERA	MATHEW		6		Maine Thr	RE	6	########	04:58.5	0	NEI										Operations and Alge		The Real and Cor		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999996616	CORTEZ	STEPHEN		6		Maine Thr	MATH	6	3/4/2024	33:47.5	1				1458	5	Well Belo	182	3.17	1	2020	32	Operation	186	The Real a		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999996621	CASEY	LOUIS		6		Maine Thr	MATH	6	########	02:11.1	1				1460	5	Well Belo	183	3.15	1	2020	32	Operation	206	The Real a		
Maine Thr	2024	7000	ME FIRST /	7001	ME FIRST /	7000	ME FIRST /	7001	ME FIRST /	999996623	CHAVEZ	KAREN	MELINDA	6		Maine Thr	MATH	6	3/1/2024	12:49.2	1				1456	5	Well Belo	178	3.3	1	2020	32	Operation	180	The Real a		

nwea State Solutions

Reports at SAU Level



nwea State Solutions

Maine Department of Education

Reports at School Level

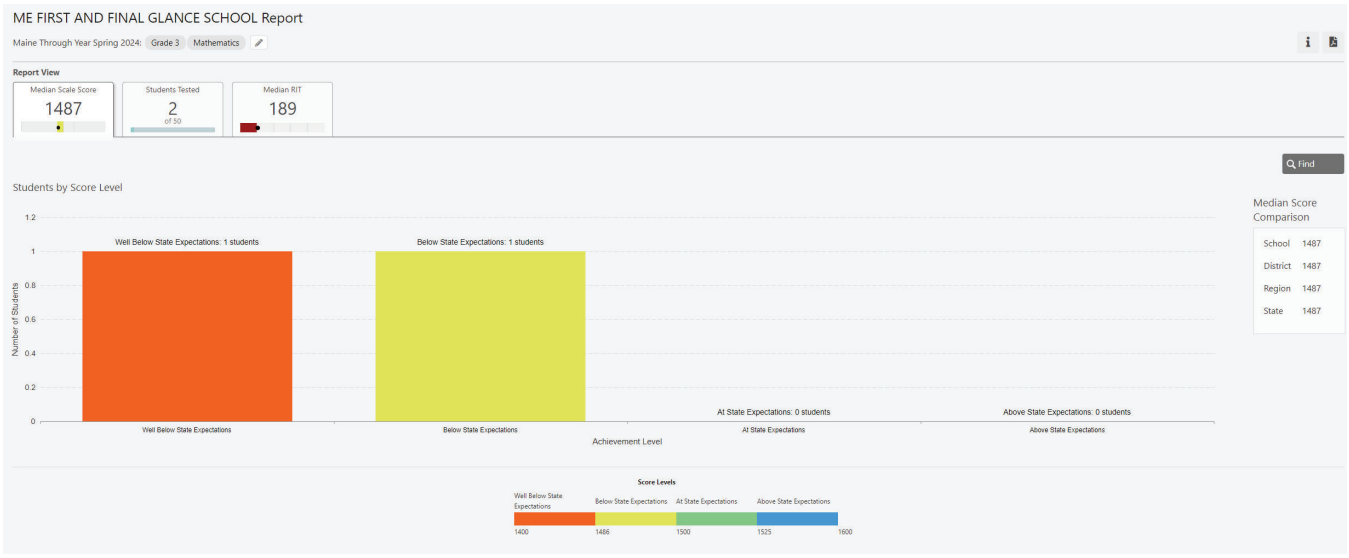
Results: 20 Schools

#	School	Students Completed	Average Score	Score Levels
1	Anderson Middle	24 of 24	2477	18% 30% 32% 20%
2	Thomasville Middle	202 of 202	2478	16% 30% 34% 20%
3	Bethany East Middle	54 of 54	2480	20% 32% 30% 18%
4	Williams Middle	30 of 30	2480	18% 30% 32% 20%
5	Coli Lake Middle	14 of 14	2486	16% 30% 34% 20%
6	Davton Middle	41 of 41	2490	20% 32% 30% 18%
7	Doristi Middle	244 of 244	2496	18% 30% 32% 20%
8	Eastview Middle	182 of 182	2496	16% 30% 34% 20%
9	Everton Middle	168 of 168	2500	20% 32% 30% 18%
10	Harris Middle	24 of 24	2508	18% 30% 32% 20%

nwea State Solutions

Maine Department of Education

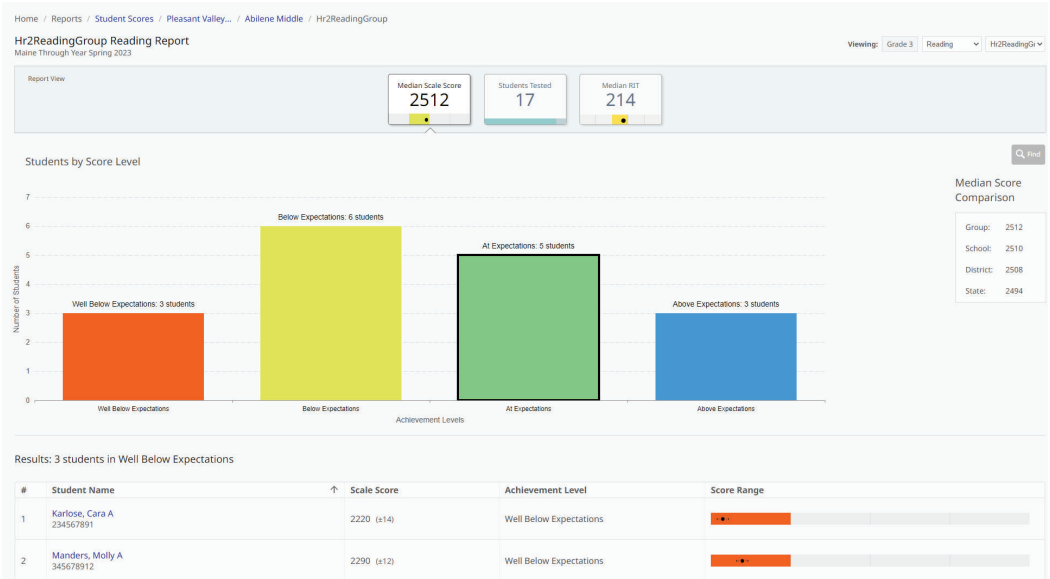
Reports at School Level



nwea State Solutions

Maine Department of Education

Reports at School / Group Level



nwea State Solutions

Maine Department of Education

When Reports Become Available

- **Operational Reports** are available throughout the assessment window.
  - Exceptions:
    - Student Score Data File will be available on 7/15/24, after the cleanup windows and all test events have been reconciled. SSDF's from prior admins will continue to be available.
- Data and Reporting in Acacia will be available within 24-72 hours from when an assessment is completed.
  - ISRs for Spring 24
    - ISR's will be available on 7/22/24.
- Starting in Spring 24, MAP Growth Reports with RIT from the Through Year Assessment will be available 24-72 hours from when an assessment is completed **if** the student has been rostered in MAP Growth.

nwea State Solutions

Maine Department of Education

Questions from the Operational Reports and Data & Reporting sections?

FAQ document and slide deck will be shared out after training session.

nwea State Solutions

nwea State Solutions

Preparation, Resources, and Tips

Preparation

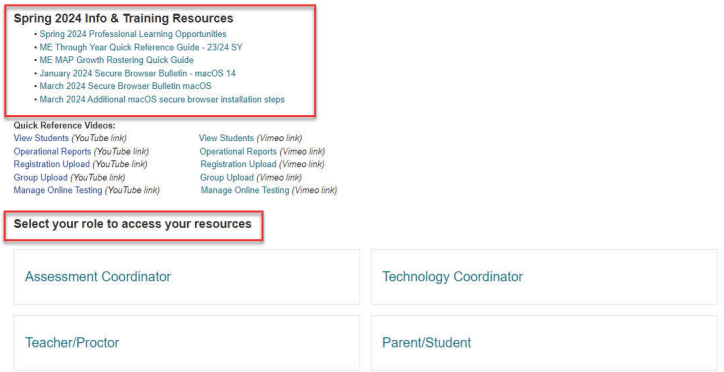
- Review technical requirements for the Maine Through Year Assessment.
  - Run the System Check Test to test your network leading up to a testing day and on testing days.
- Confirm you are on the latest version of NWEA State Solution Secure Browser.

  Note: *Previous versions of the State Solution Secure Browser **must be uninstalled** before the new version is installed.* Reminder, that the MAP Growth Secure Browser is a different application and does not need to be uninstalled/reinstalled.
- Review Maine DOE guidelines for accessibility and identify students in need of specific accommodations / supports.
- Review Scheduling Guidance from Maine DOE.
- Review Maine DOE Assessment Security Handbook.

Resources

Resources for Spring 2024 should be updated by no later than 4/8/24 and available on the Maine Connections Page.

- **For Assessment Coordinators:**
  - ME Through Year Assessment Checklist
  - ME Through Year Assessment Coordinator Guide
  - ME Through Year User and Student Management Guide
  - ME Through Year Accessibility Guide
- **For Educators:**
  - Item Type Samplers (Online and Paper Form)
  - Online Student Tutorial Video
  - ME Reports Interpretive Guide (Acacia Reports)
- **For Proctors:**
  - ME Through Year Assessment Administration Guide
  - ME Through Year Manage Online Testing



nwea State Solutions

Maine Department of Education

Technology & Security Resources

- NWEA State Solutions System and Technology Guide
  - NWEA State Solutions System and Technology Guide
- NWEA Online Readiness Tools
  - System Requirements Guide
  - Downloads for the Secure Browser
  - System Check Test
  - School Capacity Calculator
- Maine Assessment Security Handbook
  - Assessment Security Webpage

nwea State Solutions

Maine Department of Education

Suggestions for a Smooth Assessment Experience

- Enable audio on devices used for TTS and provide headphones.
- Ensure all students have appropriate accessibility features assigned, as needed.
- Validate School Proctor roles have been assigned and Instructor roles are active in MARC.
- Use the Manage Online Testing Dashboard to monitor testing progress through the assessment window.
  - **Reminder:** Simply refresh the dashboard to see updated information.

nwea State Solutions

Maine Department of Education

Troubleshooting Tips

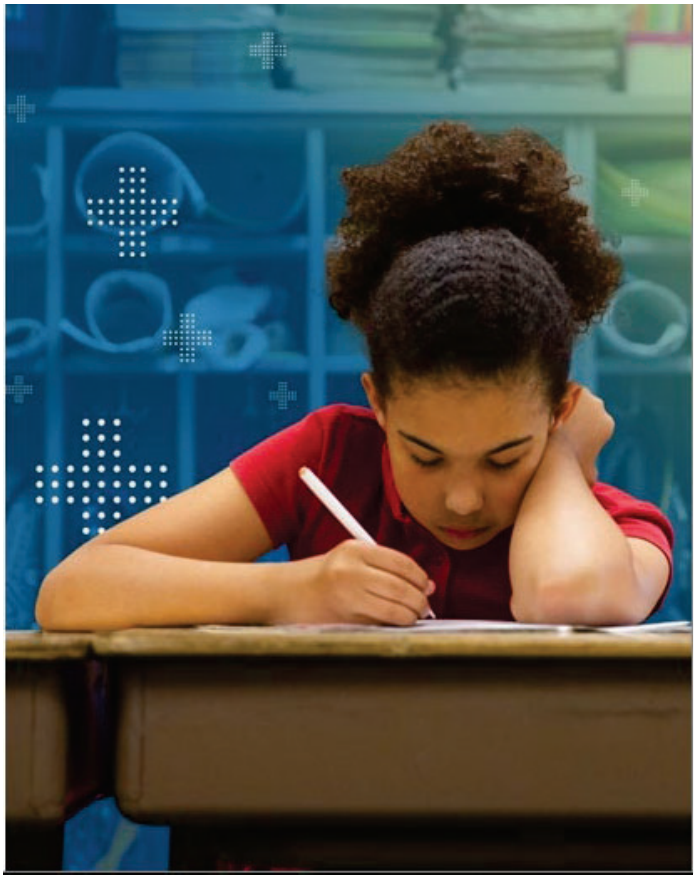
- In Acacia, the student's assessment is saved after every answer.
- If a student runs into issues, the first step would be to log out, close app, and log back in.
- Second step would be a full device reboot.

  Note: *Proctor action is not needed to log students back in.*
- If the first two steps do not resolve issue, contact Partner Support at (855) 430-1777.

nwea State Solutions

Maine Department of Education

Communication and Support



nwea State Solutions

NWEA Maine Partner Support

Phone: (855) 430-1777

Days & Hours: Monday – Friday, 7:00am – 8:00pm EST

Maine Connections Page

- Technical issues with the Acacia platform.
- Technical issues with the State Solutions Secure Browser.
- Technical issues with the MARC (MAP Growth) platform, including rostering and accessing MAP Growth reports.
- Support with pre-administration activities in the Acacia platform (e.g., creating groups or assigning accommodations).
- Support administering the online assessment.
- Information regarding procedures for students taking accommodated paper-based forms.
- SAU data clean-up tasks.

nwea State Solutions

Maine Department of Education

Maine DOE MEDMS Helpdesk

MEDMS.Helpdesk@maine.gov

(207) 624-6896

+ Fixing a student who incorrectly appears or does not appear in your assessment roster in NEO.
+ For help determining if a student is eligible for the Maine Through Year Assessment, based on information entered by the SAU into Synergy.

nwea State Solutions

Maine Department of Education

Maine DOE Assessment Team

Krista Averill

krista.averill@maine.gov

(207) 215-6528

+ Questions related to assessment content, accessibility, scoring, and reporting.
+ Policy-related questions.
+ Fixing a student who appears on your assessment roster in NEO but does not appear in Acacia.
  – *Please wait at least 36 hours after making the update in Synergy.*
+ Any problems that NWEA Maine Partner Support or MEDMS Helpdesk are not able to resolve.

nwea State Solutions

Maine Department of Education

Spring 2024 Important Dates

- **April 1st:** Acacia Management system opens.
  - SAUs can begin to roster students in MAP Growth.
  - Begin management activities in Acacia.
  - Paper-based accommodated forms (standard, large print, and braille) requests can begin.
- **April 22nd – May 31st:** Spring 2024 Assessment Window.
- **May 31st:** Last day to roster students in MAP Growth for Through Year RIT scores in MAP Growth reports.
- **June 7th:** Last day to update supports/accommodations and fix MAP Growth rostering errors.

nwea State Solutions

Maine Department of Education

Spring 2024 SAU Cleanup Window & Tasks

- **June 3rd – June 7th:** June 7th will be the last day to update supports/accommodations and fix MAP Growth rostering errors.
- Tasks to complete during SAU cleanup window.
  - Confirm that the school state code in MARC is correct.
  - Update any student demographic information that is missing or incorrect with the MAP Growth Roster Errors report.
    - Student ID is the connection between MAP Growth and Acacia and must match. This could indicate that a student needs to be unenrolled from a MAP Growth Term. Note: This action is taken in the MAP Growth (MARC) platform, not Synergy.
  - **Refer to the SAU Spring 2024 Cleanup Window Checklist** available on the Maine Connections page under Spring 2024 Info & Training Resources section on the home page.

nwea State Solutions

Maine Department of Education

Questions from the Preparation, Resources, & Tips and Communication & Support sections?

FAQ document and slide deck will be shared out after training session.

nwea State Solutions

Thank you for your time!

nwea State Solutions

Appendix B: Content Standard Coverage

Table B.1. Content Standard Coverage—Reading

Grade	Standard	Item Count	Student Count	Total Student	Percent
3	CCSS.ELA-Literacy.RI.3.1	35	10759	11669	92.2
	CCSS.ELA-Literacy.RI.3.3	16	8927	11669	76.5
	CCSS.ELA-Literacy.RI.3.2	13	8529	11669	73.09
	CCSS.ELA-Literacy.RI.3.8	12	8333	11669	71.41
	CCSS.ELA-Literacy.RI.3.6	13	7918	11669	67.86
	CCSS.ELA-Literacy.RI.3.5	14	6254	11669	53.59
	CCSS.ELA-Literacy.RI.3.9	4	4962	11669	42.52
	CCSS.ELA-Literacy.RI.3.7	5	3925	11669	33.64
	CCSS.ELA-Literacy.RI.4.1	6	1507	11669	12.91
	CCSS.ELA-Literacy.RI.4.2	7	1443	11669	12.37
	CCSS.ELA-Literacy.RI.4.5	4	1174	11669	10.06
	CCSS.ELA-Literacy.RI.4.7	4	358	11669	3.07
	CCSS.ELA-Literacy.RI.4.3	3	43	11669	0.37
	CCSS.ELA-Literacy.RI.4.8	2	3	11669	0.03
	CCSS.ELA-Literacy.RL.3.3	27	11666	11669	99.97
	CCSS.ELA-Literacy.RL.3.2	21	11574	11669	99.19
	CCSS.ELA-Literacy.RL.3.6	15	11422	11669	97.88
	CCSS.ELA-Literacy.RL.3.5	10	10846	11669	92.95
	CCSS.ELA-Literacy.RL.3.1	29	9789	11669	83.89
	CCSS.ELA-Literacy.RL.3.9	7	6691	11669	57.34
	CCSS.ELA-Literacy.RL.4.3	11	1346	11669	11.53
	CCSS.ELA-Literacy.RL.4.1	19	1302	11669	11.16
	CCSS.ELA-Literacy.RL.4.2	14	1067	11669	9.14
	CCSS.ELA-Literacy.RL.4.5	3	911	11669	7.81
	CCSS.ELA-Literacy.RL.4.6	3	258	11669	2.21
	CCSS.ELA-Literacy.RL.3.7	3	182	11669	1.56
	CCSS.ELA-Literacy.RI.3.4	9	11392	11669	97.63
	CCSS.ELA-Literacy.L.3.4.a	22	11003	11669	94.29
	CCSS.ELA-Literacy.L.3.6	7	7620	11669	65.3
	CCSS.ELA-Literacy.L.3.4.b	7	6118	11669	52.43
	CCSS.ELA-Literacy.L.3.4.d	5	4763	11669	40.82
	CCSS.ELA-Literacy.L.3.4.c	2	1953	11669	16.74
	CCSS.ELA-Literacy.L.4.5.c	7	1525	11669	13.07
	CCSS.ELA-Literacy.L.4.5.a	7	548	11669	4.7
	CCSS.ELA-Literacy.L.4.4.a	10	321	11669	2.75
	CCSS.ELA-Literacy.L.4.6	3	177	11669	1.52
	CCSS.ELA-Literacy.L.4.5.b	2	23	11669	0.2
	CCSS.ELA-Literacy.L.4.4.c	1	4	11669	0.03
	CCSS.ELA-Literacy.L.4.4.b	1	1	11669	0.01
4	CCSS.ELA-Literacy.RI.4.9	9	12230	12236	99.95
	CCSS.ELA-Literacy.RI.4.2	23	12099	12236	98.88

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.ELA-Literacy.RI.4.7	11	10799	12236	88.26
	CCSS.ELA-Literacy.RI.4.1	16	10342	12236	84.52
	CCSS.ELA-Literacy.RI.4.5	10	6420	12236	52.47
	CCSS.ELA-Literacy.RI.4.3	12	6409	12236	52.38
	CCSS.ELA-Literacy.RI.4.8	5	5450	12236	44.54
	CCSS.ELA-Literacy.RI.3.1	20	2465	12236	20.15
	CCSS.ELA-Literacy.RI.3.5	12	2193	12236	17.92
	CCSS.ELA-Literacy.RI.3.8	9	2168	12236	17.72
	CCSS.ELA-Literacy.RI.3.3	12	1344	12236	10.98
	CCSS.ELA-Literacy.RI.3.7	5	642	12236	5.25
	CCSS.ELA-Literacy.RI.5.1	7	606	12236	4.95
	CCSS.ELA-Literacy.RI.5.3	8	520	12236	4.25
	CCSS.ELA-Literacy.RI.5.2	3	505	12236	4.13
	CCSS.ELA-Literacy.RI.3.2	6	288	12236	2.35
	CCSS.ELA-Literacy.RI.3.6	6	107	12236	0.87
	CCSS.ELA-Literacy.RI.5.5	2	102	12236	0.83
	CCSS.ELA-Literacy.RI.5.9	1	21	12236	0.17
	CCSS.ELA-Literacy.RI.3.9	4	9	12236	0.07
	CCSS.ELA-Literacy.RL.4.3	29	12140	12236	99.22
	CCSS.ELA-Literacy.RL.4.1	36	11979	12236	97.9
	CCSS.ELA-Literacy.RL.4.2	22	11650	12236	95.21
	CCSS.ELA-Literacy.RL.4.6	3	6396	12236	52.27
	CCSS.ELA-Literacy.RL.3.2	12	4766	12236	38.95
	CCSS.ELA-Literacy.RL.4.5	5	4731	12236	38.66
	CCSS.ELA-Literacy.RL.3.3	10	3702	12236	30.25
	CCSS.ELA-Literacy.RL.3.1	14	2573	12236	21.03
	CCSS.ELA-Literacy.RL.5.1	11	2035	12236	16.63
	CCSS.ELA-Literacy.RL.5.6	4	1494	12236	12.21
	CCSS.ELA-Literacy.RL.3.6	7	1412	12236	11.54
	CCSS.ELA-Literacy.RL.5.2	12	1404	12236	11.47
	CCSS.ELA-Literacy.RL.5.3	7	1394	12236	11.39
	CCSS.ELA-Literacy.RL.3.5	4	1347	12236	11.01
	CCSS.ELA-Literacy.RL.4.7	1	763	12236	6.24
	CCSS.ELA-Literacy.RL.5.5	4	187	12236	1.53
	CCSS.ELA-Literacy.RL.5.7	3	175	12236	1.43
	CCSS.ELA-Literacy.RL.3.7	2	112	12236	0.92
	CCSS.ELA-Literacy.L.4.4.a	15	11936	12236	97.55
	CCSS.ELA-Literacy.RI.4.4	4	9771	12236	79.85
	CCSS.ELA-Literacy.L.4.5.a	12	8255	12236	67.46
	CCSS.ELA-Literacy.L.4.5.c	13	7298	12236	59.64
	CCSS.ELA-Literacy.RL.4.4	8	6494	12236	53.07
	CCSS.ELA-Literacy.L.4.6	10	4440	12236	36.29
	CCSS.ELA-Literacy.L.3.4.b	4	2506	12236	20.48
	CCSS.ELA-Literacy.L.3.4.a	11	2198	12236	17.96

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.ELA-Literacy.RL.3.4	11	2140	12236	17.49
	CCSS.ELA-Literacy.L.4.4.b	6	1963	12236	16.04
	CCSS.ELA-Literacy.L.5.5.a	1	1622	12236	13.26
	CCSS.ELA-Literacy.L.3.6	3	1453	12236	11.87
	CCSS.ELA-Literacy.RL.5.4	6	1176	12236	9.61
	CCSS.ELA-Literacy.L.3.4.c	2	946	12236	7.73
	CCSS.ELA-Literacy.L.4.4.c	3	736	12236	6.02
	CCSS.ELA-Literacy.L.5.4.a	2	515	12236	4.21
	CCSS.ELA-Literacy.L.5.5.c	3	462	12236	3.78
	CCSS.ELA-Literacy.RI.3.4	3	331	12236	2.71
	CCSS.ELA-Literacy.RI.5.4	2	270	12236	2.21
	CCSS.ELA-Literacy.L.4.5.b	1	204	12236	1.67
	CCSS.ELA-Literacy.L.3.4.d	1	144	12236	1.18
	CCSS.ELA-Literacy.L.5.5.b	2	144	12236	1.18
	CCSS.ELA-Literacy.L.5.4.c	1	6	12236	0.05
	CCSS.ELA-Literacy.L.5.6	2	2	12236	0.02
5	CCSS.ELA-Literacy.RI.5.1	28	11772	12213	96.39
	CCSS.ELA-Literacy.RI.5.2	19	11725	12213	96
	CCSS.ELA-Literacy.RI.5.3	21	11482	12213	94.01
	CCSS.ELA-Literacy.RI.5.7	8	8189	12213	67.05
	CCSS.ELA-Literacy.RI.5.8	7	6567	12213	53.77
	CCSS.ELA-Literacy.RI.4.1	10	4986	12213	40.83
	CCSS.ELA-Literacy.RI.6.3	9	4579	12213	37.49
	CCSS.ELA-Literacy.RI.4.2	12	4282	12213	35.06
	CCSS.ELA-Literacy.RI.6.1	14	3825	12213	31.32
	CCSS.ELA-Literacy.RI.6.8	2	2943	12213	24.1
	CCSS.ELA-Literacy.RI.4.5	4	2388	12213	19.55
	CCSS.ELA-Literacy.RI.6.5	3	2199	12213	18.01
	CCSS.ELA-Literacy.RI.4.8	3	2146	12213	17.57
	CCSS.ELA-Literacy.RI.5.9	3	2113	12213	17.3
	CCSS.ELA-Literacy.RI.6.6	4	1909	12213	15.63
	CCSS.ELA-Literacy.RI.4.7	4	1893	12213	15.5
	CCSS.ELA-Literacy.RI.5.5	7	1690	12213	13.84
	CCSS.ELA-Literacy.RI.4.3	5	1674	12213	13.71
	CCSS.ELA-Literacy.RI.6.2	4	1016	12213	8.32
	CCSS.ELA-Literacy.RI.6.7	7	577	12213	4.72
	CCSS.ELA-Literacy.RI.4.9	6	363	12213	2.97
	CCSS.ELA-Literacy.RI.6.9	1	3	12213	0.02
	CCSS.ELA-Literacy.RL.5.2	26	11615	12213	95.1
	CCSS.ELA-Literacy.RL.5.1	24	10397	12213	85.13
	CCSS.ELA-Literacy.RL.5.3	13	9518	12213	77.93
	CCSS.ELA-Literacy.RL.5.7	6	8551	12213	70.02
	CCSS.ELA-Literacy.RL.5.9	7	8134	12213	66.6
	CCSS.ELA-Literacy.RL.5.6	7	6763	12213	55.38

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.ELA-Literacy.RL.5.5	5	6185	12213	50.64
	CCSS.ELA-Literacy.RL.6.3	15	3928	12213	32.16
	CCSS.ELA-Literacy.RL.6.1	13	3814	12213	31.23
	CCSS.ELA-Literacy.RL.6.2	7	1905	12213	15.6
	CCSS.ELA-Literacy.RL.4.3	10	1861	12213	15.24
	CCSS.ELA-Literacy.RL.4.2	12	1612	12213	13.2
	CCSS.ELA-Literacy.RL.4.1	20	1522	12213	12.46
	CCSS.ELA-Literacy.RL.6.5	6	1380	12213	11.3
	CCSS.ELA-Literacy.RL.4.6	1	992	12213	8.12
	CCSS.ELA-Literacy.RL.4.5	2	341	12213	2.79
	CCSS.ELA-Literacy.RL.6.6	2	15	12213	0.12
	CCSS.ELA-Literacy.RL.6.9	1	2	12213	0.02
	CCSS.ELA-Literacy.RL.5.4	19	11336	12213	92.82
	CCSS.ELA-Literacy.L.5.5.c	6	9305	12213	76.19
	CCSS.ELA-Literacy.L.5.4.a	23	7572	12213	62
	CCSS.ELA-Literacy.L.5.5.a	10	7427	12213	60.81
	CCSS.ELA-Literacy.RI.5.4	7	6518	12213	53.37
	CCSS.ELA-Literacy.L.6.4.a	7	3739	12213	30.61
	CCSS.ELA-Literacy.L.4.5.c	9	2531	12213	20.72
	CCSS.ELA-Literacy.L.5.6	5	2241	12213	18.35
	CCSS.ELA-Literacy.RL.6.4	4	2229	12213	18.25
	CCSS.ELA-Literacy.L.5.4.c	2	1577	12213	12.91
	CCSS.ELA-Literacy.L.5.4.b	4	1565	12213	12.81
	CCSS.ELA-Literacy.L.6.5.a	3	1349	12213	11.05
	CCSS.ELA-Literacy.L.5.5.b	5	1332	12213	10.91
	CCSS.ELA-Literacy.L.4.4.a	7	1196	12213	9.79
	CCSS.ELA-Literacy.L.6.5.c	2	900	12213	7.37
	CCSS.ELA-Literacy.L.4.4.c	1	596	12213	4.88
	CCSS.ELA-Literacy.L.4.6	4	412	12213	3.37
	CCSS.ELA-Literacy.RI.6.4	4	318	12213	2.6
	CCSS.ELA-Literacy.RI.4.4	1	207	12213	1.69
	CCSS.ELA-Literacy.L.4.4.b	3	139	12213	1.14
	CCSS.ELA-Literacy.L.4.5.a	3	118	12213	0.97
	CCSS.ELA-Literacy.L.6.4.b	2	88	12213	0.72
	CCSS.ELA-Literacy.RL.4.4	2	69	12213	0.56
	CCSS.ELA-Literacy.L.6.5.b	2	5	12213	0.04
	CCSS.ELA-Literacy.L.6.4.c	1	3	12213	0.02
	CCSS.ELA-Literacy.L.6.6	1	2	12213	0.02
	CCSS.ELA-Literacy.L.4.5.b	1	1	12213	0.01
	CCSS.ELA-Literacy.L.6.4.d	1	1	12213	0.01
6	CCSS.ELA-Literacy.RI.6.3	18	11948	11958	99.92
	CCSS.ELA-Literacy.RI.6.8	6	11918	11958	99.67
	CCSS.ELA-Literacy.RI.6.5	9	9819	11958	82.11
	CCSS.ELA-Literacy.RI.6.1	36	9526	11958	79.66

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.ELA-Literacy.RI.6.2	14	7240	11958	60.55
	CCSS.ELA-Literacy.RI.7.1	17	3312	11958	27.7
	CCSS.ELA-Literacy.RI.7.5	13	3195	11958	26.72
	CCSS.ELA-Literacy.RI.6.6	11	2791	11958	23.34
	CCSS.ELA-Literacy.RI.6.7	17	2729	11958	22.82
	CCSS.ELA-Literacy.RI.7.6	5	2666	11958	22.29
	CCSS.ELA-Literacy.RI.5.3	6	692	11958	5.79
	CCSS.ELA-Literacy.RI.7.2	14	585	11958	4.89
	CCSS.ELA-Literacy.RI.5.1	7	516	11958	4.32
	CCSS.ELA-Literacy.RI.5.2	4	468	11958	3.91
	CCSS.ELA-Literacy.RI.7.3	9	329	11958	2.75
	CCSS.ELA-Literacy.RI.5.7	2	234	11958	1.96
	CCSS.ELA-Literacy.RI.7.8	2	24	11958	0.2
	CCSS.ELA-Literacy.RL.6.3	28	11937	11958	99.82
	CCSS.ELA-Literacy.RL.6.1	30	11915	11958	99.64
	CCSS.ELA-Literacy.RL.6.5	6	9649	11958	80.69
	CCSS.ELA-Literacy.RL.6.2	8	7805	11958	65.27
	CCSS.ELA-Literacy.RL.5.2	11	4808	11958	40.21
	CCSS.ELA-Literacy.RL.5.5	3	2707	11958	22.64
	CCSS.ELA-Literacy.RL.5.3	7	2418	11958	20.22
	CCSS.ELA-Literacy.RL.6.6	6	2035	11958	17.02
	CCSS.ELA-Literacy.RL.7.3	16	1667	11958	13.94
	CCSS.ELA-Literacy.RL.5.1	8	1245	11958	10.41
	CCSS.ELA-Literacy.RL.7.6	6	1006	11958	8.41
	CCSS.ELA-Literacy.RL.7.2	6	854	11958	7.14
	CCSS.ELA-Literacy.RL.5.6	4	825	11958	6.9
	CCSS.ELA-Literacy.RL.7.1	5	361	11958	3.02
	CCSS.ELA-Literacy.RL.5.7	3	17	11958	0.14
	CCSS.ELA-Literacy.RL.7.9	2	2	11958	0.02
	CCSS.ELA-Literacy.L.6.6	3	11941	11958	99.86
	CCSS.ELA-Literacy.RI.6.4	7	7374	11958	61.67
	CCSS.ELA-Literacy.L.6.4.a	16	7171	11958	59.97
	CCSS.ELA-Literacy.L.6.5.a	5	7136	11958	59.68
	CCSS.ELA-Literacy.L.6.5.c	3	4234	11958	35.41
	CCSS.ELA-Literacy.L.6.4.b	12	3559	11958	29.76
	CCSS.ELA-Literacy.RI.7.4	14	3254	11958	27.21
	CCSS.ELA-Literacy.L.5.5.c	1	3201	11958	26.77
	CCSS.ELA-Literacy.RL.6.4	6	2135	11958	17.85
	CCSS.ELA-Literacy.L.6.4.d	4	1832	11958	15.32
	CCSS.ELA-Literacy.RL.5.4	9	1759	11958	14.71
	CCSS.ELA-Literacy.RL.7.4	7	1629	11958	13.62
	CCSS.ELA-Literacy.L.7.5.a	6	1094	11958	9.15
	CCSS.ELA-Literacy.L.5.5.a	2	705	11958	5.9
	CCSS.ELA-Literacy.L.5.4.b	1	464	11958	3.88

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.ELA-Literacy.L.6.4.c	1	333	11958	2.78
	CCSS.ELA-Literacy.L.5.4.a	3	248	11958	2.07
	CCSS.ELA-Literacy.L.7.4.c	1	234	11958	1.96
	CCSS.ELA-Literacy.L.7.4.b	4	57	11958	0.48
	CCSS.ELA-Literacy.L.6.5.b	2	23	11958	0.19
	CCSS.ELA-Literacy.L.5.6	1	8	11958	0.07
	CCSS.ELA-Literacy.RI.5.4	1	2	11958	0.02
	CCSS.ELA-Literacy.L.5.5.b	1	1	11958	0.01
	CCSS.ELA-Literacy.L.7.6	1	1	11958	0.01
7	CCSS.ELA-Literacy.RI.7.3	18	12175	12212	99.7
	CCSS.ELA-Literacy.RI.7.6	17	12129	12212	99.32
	CCSS.ELA-Literacy.RI.7.5	23	11712	12212	95.91
	CCSS.ELA-Literacy.RI.7.1	18	10757	12212	88.09
	CCSS.ELA-Literacy.RI.7.2	21	8876	12212	72.68
	CCSS.ELA-Literacy.RI.8.6	10	4260	12212	34.88
	CCSS.ELA-Literacy.RI.8.5	19	3671	12212	30.06
	CCSS.ELA-Literacy.RI.8.1	16	3084	12212	25.25
	CCSS.ELA-Literacy.RI.8.3	10	2837	12212	23.23
	CCSS.ELA-Literacy.RI.8.2	9	2445	12212	20.02
	CCSS.ELA-Literacy.RI.7.7	1	2386	12212	19.54
	CCSS.ELA-Literacy.RI.6.1	24	1173	12212	9.61
	CCSS.ELA-Literacy.RI.7.8	1	953	12212	7.8
	CCSS.ELA-Literacy.RI.6.6	9	936	12212	7.66
	CCSS.ELA-Literacy.RI.6.7	10	318	12212	2.6
	CCSS.ELA-Literacy.RI.6.5	5	198	12212	1.62
	CCSS.ELA-Literacy.RI.7.9	3	155	12212	1.27
	CCSS.ELA-Literacy.RI.6.8	2	122	12212	1
	CCSS.ELA-Literacy.RI.6.2	9	106	12212	0.87
	CCSS.ELA-Literacy.RI.6.3	6	74	12212	0.61
	CCSS.ELA-Literacy.RI.8.7	2	46	12212	0.38
	CCSS.ELA-Literacy.RL.7.9	8	12167	12212	99.63
	CCSS.ELA-Literacy.RL.7.1	16	11612	12212	95.09
	CCSS.ELA-Literacy.RL.7.6	13	11352	12212	92.96
	CCSS.ELA-Literacy.RL.7.3	31	11149	12212	91.3
	CCSS.ELA-Literacy.RL.7.2	15	10935	12212	89.54
	CCSS.ELA-Literacy.RL.6.3	11	1363	12212	11.16
	CCSS.ELA-Literacy.RL.6.1	11	1349	12212	11.05
	CCSS.ELA-Literacy.RL.7.5	2	1135	12212	9.29
	CCSS.ELA-Literacy.RL.6.5	4	351	12212	2.87
	CCSS.ELA-Literacy.RL.8.1	8	217	12212	1.78
	CCSS.ELA-Literacy.RL.8.6	1	164	12212	1.34
	CCSS.ELA-Literacy.RL.8.3	7	79	12212	0.65
	CCSS.ELA-Literacy.RL.6.6	1	20	12212	0.16
	CCSS.ELA-Literacy.RL.6.2	2	12	12212	0.1

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.ELA-Literacy.RL.8.2	2	5	12212	0.04
	CCSS.ELA-Literacy.RL.7.4	18	12041	12212	98.6
	CCSS.ELA-Literacy.L.7.6	3	8728	12212	71.47
	CCSS.ELA-Literacy.RI.7.4	19	7965	12212	65.22
	CCSS.ELA-Literacy.L.7.5.a	15	7346	12212	60.15
	CCSS.ELA-Literacy.RI.8.4	17	5809	12212	47.57
	CCSS.ELA-Literacy.L.7.4.b	9	3032	12212	24.83
	CCSS.ELA-Literacy.L.7.4.a	2	2093	12212	17.14
	CCSS.ELA-Literacy.L.8.4.a	1	1832	12212	15
	CCSS.ELA-Literacy.RL.6.4	3	1323	12212	10.83
	CCSS.ELA-Literacy.L.8.4.b	8	1204	12212	9.86
	CCSS.ELA-Literacy.L.7.4.c	2	486	12212	3.98
	CCSS.ELA-Literacy.L.6.4.a	8	463	12212	3.79
	CCSS.ELA-Literacy.L.7.5.b	1	336	12212	2.75
	CCSS.ELA-Literacy.L.8.5.a	3	317	12212	2.6
	CCSS.ELA-Literacy.L.6.4.d	1	138	12212	1.13
	CCSS.ELA-Literacy.L.6.5.b	2	138	12212	1.13
	CCSS.ELA-Literacy.L.6.4.b	5	101	12212	0.83
	CCSS.ELA-Literacy.RI.6.4	1	63	12212	0.52
	CCSS.ELA-Literacy.RL.8.4	11	57	12212	0.47
	CCSS.ELA-Literacy.L.6.5.a	2	10	12212	0.08
	CCSS.ELA-Literacy.L.6.5.c	2	5	12212	0.04
	CCSS.ELA-Literacy.L.6.4.c	1	1	12212	0.01
8	CCSS.ELA-Literacy.RI.8.5	44	12145	12295	98.78
	CCSS.ELA-Literacy.RI.8.6	20	11724	12295	95.36
	CCSS.ELA-Literacy.RI.8.1	36	10914	12295	88.77
	CCSS.ELA-Literacy.RI.8.3	28	10570	12295	85.97
	CCSS.ELA-Literacy.RI.8.2	18	7617	12295	61.95
	CCSS.ELA-Literacy.RI.8.8	5	5362	12295	43.61
	CCSS.ELA-Literacy.RI.7.5	13	4368	12295	35.53
	CCSS.ELA-Literacy.RI.7.1	11	3934	12295	32
	CCSS.ELA-Literacy.RI.8.7	2	3636	12295	29.57
	CCSS.ELA-Literacy.RI.7.6	9	2564	12295	20.85
	CCSS.ELA-Literacy.RI.7.3	12	1138	12295	9.26
	CCSS.ELA-Literacy.RI.7.2	14	739	12295	6.01
	CCSS.ELA-Literacy.RI.7.8	2	27	12295	0.22
	CCSS.ELA-Literacy.RI.7.7	1	4	12295	0.03
	CCSS.ELA-Literacy.RL.8.3	34	12202	12295	99.24
	CCSS.ELA-Literacy.RL.8.1	24	12145	12295	98.78
	CCSS.ELA-Literacy.RL.8.2	13	10413	12295	84.69
	CCSS.ELA-Literacy.RL.8.6	11	7926	12295	64.47
	CCSS.ELA-Literacy.RL.7.3	16	3910	12295	31.8
	CCSS.ELA-Literacy.RL.7.2	7	2151	12295	17.49
	CCSS.ELA-Literacy.RL.8.9	1	1915	12295	15.58

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.ELA-Literacy.RL.7.6	4	1560	12295	12.69
	CCSS.ELA-Literacy.RL.7.1	4	63	12295	0.51
	CCSS.ELA-Literacy.RL.7.9	2	2	12295	0.02
	CCSS.ELA-Literacy.RI.8.4	25	10560	12295	85.89
	CCSS.ELA-Literacy.RL.8.4	42	9730	12295	79.14
	CCSS.ELA-Literacy.L.8.6	6	8404	12295	68.35
	CCSS.ELA-Literacy.L.8.4.b	15	6622	12295	53.86
	CCSS.ELA-Literacy.L.8.5.a	5	5886	12295	47.87
	CCSS.ELA-Literacy.RI.7.4	9	4016	12295	32.66
	CCSS.ELA-Literacy.RL.7.4	10	3768	12295	30.65
	CCSS.ELA-Literacy.L.8.4.a	6	3230	12295	26.27
	CCSS.ELA-Literacy.L.8.5.b	4	1882	12295	15.31
	CCSS.ELA-Literacy.L.7.5.a	6	916	12295	7.45
	CCSS.ELA-Literacy.L.7.4.c	1	457	12295	3.72
	CCSS.ELA-Literacy.L.8.4.d	1	152	12295	1.24
	CCSS.ELA-Literacy.L.7.4.b	3	22	12295	0.18
	CCSS.ELA-Literacy.RI.9-10.1	5	12474	12513	99.69
	CCSS.ELA-Literacy.RI.9-10.5	5	12469	12513	99.65
	CCSS.ELA-Literacy.RI.9-10.2	7	12447	12513	99.47
	CCSS.ELA-Literacy.RI.9-10.3	8	12120	12513	96.86
	CCSS.ELA-Literacy.RI.9-10.7	2	6899	12513	55.13
	CCSS.ELA-Literacy.RI.9-10.8	2	3706	12513	29.62
	CCSS.ELA-Literacy.RI.9-10.6	3	2663	12513	21.28
HS	CCSS.ELA-Literacy.RL.9-10.2	2	12513	12513	100
	CCSS.ELA-Literacy.RL.9-10.3	3	12509	12513	99.97
	CCSS.ELA-Literacy.RL.9-10.1	2	12505	12513	99.94
	CCSS.ELA-Literacy.RL.9-10.5	1	12504	12513	99.93
	CCSS.ELA-Literacy.L.9-10.4.a	4	12504	12513	99.93
	CCSS.ELA-Literacy.RL.9-10.4	1	12504	12513	99.93
	CCSS.ELA-Literacy.RI.9-10.4	3	12473	12513	99.68
	CCSS.ELA-Literacy.L.9-10.6	1	12447	12513	99.47

Table B.2. Content Standard Coverage—Mathematics

Grade	Standard	Item Count	Student Count	Total Student	Percent
3	CCSS.Math.Content.3.G.A.2	13	10977	11723	93.64
	CCSS.Math.Content.4.G.A.1	10	10683	11723	91.13
	CCSS.Math.Content.3.G.A.1	8	8406	11723	71.71
	CCSS.Math.Content.4.G.A.2	5	177	11723	1.51
	CCSS.Math.Content.4.G.A.3	4	120	11723	1.02
	CCSS.Math.Content.3.MD.C.6	10	10794	11723	92.08
	CCSS.Math.Content.3.MD.B.4	28	9050	11723	77.2
	CCSS.Math.Content.3.MD.B.3	37	9038	11723	77.1
	CCSS.Math.Content.3.MD.A.2	18	7663	11723	65.37
	CCSS.Math.Content.3.MD.A.1	23	7651	11723	65.26
	CCSS.Math.Content.3.MD.D.8	16	4035	11723	34.42
	CCSS.Math.Content.3.MD.C.5	10	3013	11723	25.7
	CCSS.Math.Content.3.MD.C.7.a	8	2523	11723	21.52
	CCSS.Math.Content.3.MD.C.7.b	7	2499	11723	21.32
	CCSS.Math.Content.3.MD.C.7.d	4	2274	11723	19.4
	CCSS.Math.Content.3.MD.C.7.c	3	2030	11723	17.32
	CCSS.Math.Content.4.MD.B.4	3	114	11723	0.97
	CCSS.Math.Content.4.MD.C.6	4	82	11723	0.7
	CCSS.Math.Content.4.MD.A.1	1	1	11723	0.01
	CCSS.Math.Content.3.NF.A.3.c	14	10000	11723	85.3
	CCSS.Math.Content.3.NBT.A.1	10	7990	11723	68.16
	CCSS.Math.Content.3.NF.A.3.d	12	7445	11723	63.51
	CCSS.Math.Content.3.NF.A.3.b	21	7101	11723	60.57
	CCSS.Math.Content.3.NF.A.1	18	7020	11723	59.88
	CCSS.Math.Content.3.NF.A.2.a	15	6783	11723	57.86
	CCSS.Math.Content.3.NBT.A.3	7	5818	11723	49.63
	CCSS.Math.Content.3.NF.A.2.b	15	5446	11723	46.46
	CCSS.Math.Content.3.NBT.A.2	6	4107	11723	35.03
	CCSS.Math.Content.3.NF.A.3.a	7	3580	11723	30.54
	CCSS.Math.Content.4.NBT.B.5	2	1205	11723	10.28
	CCSS.Math.Content.4.NF.A.1	2	1056	11723	9.01
	CCSS.Math.Content.4.NF.A.2	3	818	11723	6.98
	CCSS.Math.Content.4.NF.B.4.b	2	84	11723	0.72
	CCSS.Math.Content.4.NBT.A.2	3	82	11723	0.7
	CCSS.Math.Content.4.NF.C.6	4	72	11723	0.61
	CCSS.Math.Content.4.NF.C.7	2	12	11723	0.1
	CCSS.Math.Content.4.NBT.A.3	1	10	11723	0.09
	CCSS.Math.Content.4.NBT.B.4	1	7	11723	0.06
	CCSS.Math.Content.4.NF.C.5	1	2	11723	0.02
	CCSS.Math.Content.4.NBT.A.1	1	1	11723	0.01
	CCSS.Math.Content.4.NBT.B.6	1	1	11723	0.01
	CCSS.Math.Content.3.OA.D.8	24	7911	11723	67.48

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.Math.Content.3.OA.C.7	12	7675	11723	65.47
	CCSS.Math.Content.3.OA.D.9	15	7315	11723	62.4
	CCSS.Math.Content.3.OA.A.1	26	7139	11723	60.9
	CCSS.Math.Content.3.OA.B.5	14	5500	11723	46.92
	CCSS.Math.Content.3.OA.A.4	9	4187	11723	35.72
	CCSS.Math.Content.3.OA.B.6	8	3978	11723	33.93
	CCSS.Math.Content.3.OA.A.3	13	3782	11723	32.26
	CCSS.Math.Content.3.OA.A.2	7	2182	11723	18.61
	CCSS.Math.Content.4.OA.B.4	7	523	11723	4.46
	CCSS.Math.Content.4.OA.C.5	1	290	11723	2.47
	CCSS.Math.Content.4.OA.A.3	2	90	11723	0.77
4	CCSS.Math.Content.4.G.A.1	23	11136	12291	90.6
	CCSS.Math.Content.4.G.A.2	10	6895	12291	56.1
	CCSS.Math.Content.5.G.A.1	3	6025	12291	49.02
	CCSS.Math.Content.4.G.A.3	7	4094	12291	33.31
	CCSS.Math.Content.5.G.A.2	1	1507	12291	12.26
	CCSS.Math.Content.5.G.B.4	2	904	12291	7.35
	CCSS.Math.Content.3.G.A.2	2	110	12291	0.89
	CCSS.Math.Content.5.G.B.3	2	52	12291	0.42
	CCSS.Math.Content.4.MD.B.4	28	10497	12291	85.4
	CCSS.Math.Content.4.MD.C.6	14	10147	12291	82.56
	CCSS.Math.Content.4.MD.A.1	12	6999	12291	56.94
	CCSS.Math.Content.3.MD.B.3	11	3484	12291	28.35
	CCSS.Math.Content.4.MD.C.5	5	3355	12291	27.3
	CCSS.Math.Content.4.MD.C.7	5	3185	12291	25.91
	CCSS.Math.Content.4.MD.A.2	6	2276	12291	18.52
	CCSS.Math.Content.3.MD.A.1	6	2216	12291	18.03
	CCSS.Math.Content.4.MD.A.3	3	1729	12291	14.07
	CCSS.Math.Content.3.MD.C.6	3	1271	12291	10.34
	CCSS.Math.Content.5.MD.B.2	2	705	12291	5.74
	CCSS.Math.Content.5.MD.A.1	3	676	12291	5.5
	CCSS.Math.Content.5.MD.C.4	2	374	12291	3.04
	CCSS.Math.Content.3.MD.A.2	2	153	12291	1.24
	CCSS.Math.Content.5.MD.C.5.b	1	59	12291	0.48
	CCSS.Math.Content.3.MD.D.8	1	44	12291	0.36
	CCSS.Math.Content.5.MD.C.5.c	1	1	12291	0.01
	CCSS.Math.Content.4.NBT.B.5	27	10465	12291	85.14
	CCSS.Math.Content.4.NBT.A.3	19	8939	12291	72.73
	CCSS.Math.Content.4.NF.C.7	17	7994	12291	65.04
	CCSS.Math.Content.4.NF.A.2	15	7297	12291	59.37
	CCSS.Math.Content.4.NBT.A.2	26	7152	12291	58.19
	CCSS.Math.Content.4.NF.B.3.d	12	7118	12291	57.91
	CCSS.Math.Content.4.NF.A.1	16	6182	12291	50.3

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.Math.Content.4.NBT.B.4	6	5709	12291	46.45
	CCSS.Math.Content.4.NF.C.6	8	5187	12291	42.2
	CCSS.Math.Content.4.NF.B.4.b	8	4714	12291	38.35
	CCSS.Math.Content.4.NF.B.3.b	7	4226	12291	34.38
	CCSS.Math.Content.4.NF.C.5	11	3958	12291	32.2
	CCSS.Math.Content.4.NBT.B.6	10	3886	12291	31.62
	CCSS.Math.Content.3.NF.A.3.c	3	3674	12291	29.89
	CCSS.Math.Content.4.NF.B.3.c	6	3446	12291	28.04
	CCSS.Math.Content.4.NBT.A.1	9	3398	12291	27.65
	CCSS.Math.Content.5.NBT.B.7	6	1781	12291	14.49
	CCSS.Math.Content.5.NF.B.5	1	1564	12291	12.72
	CCSS.Math.Content.4.NF.B.4.c	5	1367	12291	11.12
	CCSS.Math.Content.4.NF.B.3.a	5	1282	12291	10.43
	CCSS.Math.Content.4.NF.B.4.a	4	632	12291	5.14
	CCSS.Math.Content.5.NF.B.7.b	3	519	12291	4.22
	CCSS.Math.Content.3.NF.A.1	3	468	12291	3.81
	CCSS.Math.Content.5.NF.A.1	5	364	12291	2.96
	CCSS.Math.Content.5.NF.A.2	4	299	12291	2.43
	CCSS.Math.Content.3.NF.A.3.b	1	203	12291	1.65
	CCSS.Math.Content.5.NF.B.4.a	4	178	12291	1.45
	CCSS.Math.Content.5.NBT.B.5	2	141	12291	1.15
	CCSS.Math.Content.5.NBT.A.2	2	80	12291	0.65
	CCSS.Math.Content.3.NF.A.3.a	1	29	12291	0.24
	CCSS.Math.Content.5.NBT.A.3.b	1	29	12291	0.24
	CCSS.Math.Content.3.NBT.A.3	1	26	12291	0.21
	CCSS.Math.Content.5.NBT.A.1	1	9	12291	0.07
	CCSS.Math.Content.3.NF.A.2.b	1	6	12291	0.05
	CCSS.Math.Content.5.NBT.A.3.a	2	3	12291	0.02
	CCSS.Math.Content.5.NBT.B.6	1	3	12291	0.02
	CCSS.Math.Content.3.NBT.A.2	1	1	12291	0.01
	CCSS.Math.Content.3.NF.A.2.a	1	1	12291	0.01
	CCSS.Math.Content.4.OA.B.4	20	8926	12291	72.62
	CCSS.Math.Content.4.OA.C.5	12	7781	12291	63.31
	CCSS.Math.Content.4.OA.A.1	12	6807	12291	55.38
	CCSS.Math.Content.4.OA.A.3	10	5284	12291	42.99
	CCSS.Math.Content.3.OA.C.7	4	4936	12291	40.16
	CCSS.Math.Content.4.OA.A.2	7	4914	12291	39.98
	CCSS.Math.Content.3.OA.D.9	2	740	12291	6.02
	CCSS.Math.Content.5.OA.B.3	1	397	12291	3.23
	CCSS.Math.Content.5.OA.A.1	5	260	12291	2.12
	CCSS.Math.Content.3.OA.D.8	2	176	12291	1.43
	CCSS.Math.Content.3.OA.A.1	4	169	12291	1.37
	CCSS.Math.Content.3.OA.A.3	1	13	12291	0.11

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.Math.Content.3.OA.A.2	1	1	12291	0.01
5	CCSS.Math.Content.5.G.A.1	25	9722	12261	79.29
	CCSS.Math.Content.5.G.B.4	8	7941	12261	64.77
	CCSS.Math.Content.5.G.B.3	8	5559	12261	45.34
	CCSS.Math.Content.5.G.A.2	4	3572	12261	29.13
	CCSS.Math.Content.4.G.A.2	5	999	12261	8.15
	CCSS.Math.Content.6.G.A.4	8	776	12261	6.33
	CCSS.Math.Content.6.G.A.2	4	728	12261	5.94
	CCSS.Math.Content.6.G.A.1	4	572	12261	4.67
	CCSS.Math.Content.4.G.A.1	4	445	12261	3.63
	CCSS.Math.Content.4.G.A.3	4	405	12261	3.3
	CCSS.Math.Content.6.G.A.3	4	46	12261	0.38
	CCSS.Math.Content.5.MD.C.4	11	8572	12261	69.91
	CCSS.Math.Content.5.MD.B.2	25	6831	12261	55.71
	CCSS.Math.Content.5.MD.C.5.c	9	6205	12261	50.61
	CCSS.Math.Content.5.MD.C.5.a	5	4557	12261	37.17
	CCSS.Math.Content.5.MD.A.1	9	4302	12261	35.09
	CCSS.Math.Content.5.MD.C.3	5	4154	12261	33.88
	CCSS.Math.Content.5.MD.C.5.b	6	3740	12261	30.5
	CCSS.Math.Content.4.MD.A.1	4	294	12261	2.4
	CCSS.Math.Content.4.MD.C.6	1	164	12261	1.34
	CCSS.Math.Content.6.SP.A.3	3	49	12261	0.4
	CCSS.Math.Content.6.SP.B.5	3	29	12261	0.24
	CCSS.Math.Content.4.MD.B.4	3	10	12261	0.08
	CCSS.Math.Content.5.NBT.B.7	41	12066	12261	98.41
	CCSS.Math.Content.5.NBT.A.2	15	8674	12261	70.74
	CCSS.Math.Content.5.NBT.A.3.b	14	7563	12261	61.68
	CCSS.Math.Content.5.NBT.B.5	10	7531	12261	61.42
	CCSS.Math.Content.5.NF.A.2	25	7384	12261	60.22
	CCSS.Math.Content.5.NF.B.5	9	6539	12261	53.33
	CCSS.Math.Content.5.NF.A.1	16	6334	12261	51.66
	CCSS.Math.Content.5.NBT.A.3.a	13	6100	12261	49.75
	CCSS.Math.Content.5.NF.B.4.a	11	6062	12261	49.44
	CCSS.Math.Content.5.NBT.B.6	11	5417	12261	44.18
	CCSS.Math.Content.5.NBT.A.4	11	4984	12261	40.65
	CCSS.Math.Content.5.NF.B.7.a	6	3127	12261	25.5
	CCSS.Math.Content.5.NBT.A.1	6	2221	12261	18.11
	CCSS.Math.Content.5.NF.B.4.b	7	1857	12261	15.15
	CCSS.Math.Content.5.NF.B.7.c	4	1856	12261	15.14
	CCSS.Math.Content.5.NF.B.7.b	5	1479	12261	12.06
	CCSS.Math.Content.4.NBT.B.5	4	1404	12261	11.45
	CCSS.Math.Content.5.NF.B.3	5	1219	12261	9.94
	CCSS.Math.Content.4.NBT.A.2	15	1072	12261	8.74

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.Math.Content.5.NF.B.6	4	893	12261	7.28
	CCSS.Math.Content.4.NBT.A.3	3	663	12261	5.41
	CCSS.Math.Content.6.RP.A.3.c	7	536	12261	4.37
	CCSS.Math.Content.4.NF.C.7	3	477	12261	3.89
	CCSS.Math.Content.4.NF.C.6	2	224	12261	1.83
	CCSS.Math.Content.4.NF.A.1	4	217	12261	1.77
	CCSS.Math.Content.4.NF.B.4.a	1	189	12261	1.54
	CCSS.Math.Content.6.NS.A.1	3	152	12261	1.24
	CCSS.Math.Content.4.NF.C.5	1	124	12261	1.01
	CCSS.Math.Content.6.RP.A.3.b	2	88	12261	0.72
	CCSS.Math.Content.6.NS.C.5	1	68	12261	0.55
	CCSS.Math.Content.6.NS.C.6.c	2	63	12261	0.51
	CCSS.Math.Content.6.RP.A.3.a	1	60	12261	0.49
	CCSS.Math.Content.6.NS.C.7	1	48	12261	0.39
	CCSS.Math.Content.4.NF.A.2	2	28	12261	0.23
	CCSS.Math.Content.6.NS.C.6.a	1	13	12261	0.11
	CCSS.Math.Content.6.NS.C.7.d	1	3	12261	0.02
	CCSS.Math.Content.4.NF.B.3.c	1	1	12261	0.01
	CCSS.Math.Content.6.NS.C.6.b	1	1	12261	0.01
	CCSS.Math.Content.5.OA.A.1	20	9841	12261	80.26
	CCSS.Math.Content.5.OA.B.3	21	9060	12261	73.89
	CCSS.Math.Content.4.OA.A.2	2	7124	12261	58.1
	CCSS.Math.Content.5.OA.A.2	10	6961	12261	56.77
	CCSS.Math.Content.4.OA.A.3	1	780	12261	6.36
	CCSS.Math.Content.6.EE.A.1	2	427	12261	3.48
	CCSS.Math.Content.6.EE.C.9	1	287	12261	2.34
	CCSS.Math.Content.6.EE.B.6	2	84	12261	0.69
	CCSS.Math.Content.6.EE.B.7	2	18	12261	0.15
	CCSS.Math.Content.6.EE.A.2.c	3	17	12261	0.14
	CCSS.Math.Content.4.OA.C.5	1	15	12261	0.12
	CCSS.Math.Content.6.EE.A.3	1	3	12261	0.02
	CCSS.Math.Content.6.EE.B.8	1	1	12261	0.01
6	CCSS.Math.Content.6.G.A.4	14	10919	11996	91.02
	CCSS.Math.Content.6.G.A.3	13	10111	11996	84.29
	CCSS.Math.Content.5.MD.C.5.b	1	5727	11996	47.74
	CCSS.Math.Content.5.MD.C.4	2	3184	11996	26.54
	CCSS.Math.Content.5.G.B.3	1	1752	11996	14.6
	CCSS.Math.Content.6.G.A.2	11	1611	11996	13.43
	CCSS.Math.Content.6.G.A.1	8	1075	11996	8.96
	CCSS.Math.Content.5.G.B.4	1	805	11996	6.71
	CCSS.Math.Content.7.G.B.6	3	278	11996	2.32
	CCSS.Math.Content.7.G.B.4	1	9	11996	0.08
	CCSS.Math.Content.6.EE.A.2.c	20	9129	11996	76.1

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.Math.Content.6.EE.B.6	14	7396	11996	61.65
	CCSS.Math.Content.6.EE.A.1	17	6886	11996	57.4
	CCSS.Math.Content.6.EE.B.7	21	6704	11996	55.89
	CCSS.Math.Content.6.EE.B.5	9	5195	11996	43.31
	CCSS.Math.Content.6.EE.B.8	13	3742	11996	31.19
	CCSS.Math.Content.6.EE.A.3	7	2213	11996	18.45
	CCSS.Math.Content.6.EE.A.2.a	4	2098	11996	17.49
	CCSS.Math.Content.6.EE.A.2.b	4	1439	11996	12
	CCSS.Math.Content.5.OA.B.3	1	1159	11996	9.66
	CCSS.Math.Content.6.EE.C.9	3	628	11996	5.24
	CCSS.Math.Content.7.EE.A.1	2	287	11996	2.39
	CCSS.Math.Content.5.OA.A.1	3	201	11996	1.68
	CCSS.Math.Content.7.EE.A.2	2	184	11996	1.53
	CCSS.Math.Content.7.EE.B.3	1	143	11996	1.19
	CCSS.Math.Content.7.EE.B.4.b	3	49	11996	0.41
	CCSS.Math.Content.7.EE.B.4.a	3	47	11996	0.39
	CCSS.Math.Content.6.SP.A.3	16	8801	11996	73.37
	CCSS.Math.Content.6.SP.B.5	10	8526	11996	71.07
	CCSS.Math.Content.6.SP.B.4	7	4996	11996	41.65
	CCSS.Math.Content.7.SP.C.5	1	4461	11996	37.19
	CCSS.Math.Content.6.SP.A.2	9	3669	11996	30.59
	CCSS.Math.Content.5.MD.B.2	2	3458	11996	28.83
	CCSS.Math.Content.7.SP.C.7	8	1423	11996	11.86
	CCSS.Math.Content.7.SP.B	3	1103	11996	9.19
	CCSS.Math.Content.7.SP.C.8	3	167	11996	1.39
	CCSS.Math.Content.7.SP.C.6	3	10	11996	0.08
	CCSS.Math.Content.6.NS.B.2	11	8573	11996	71.47
	CCSS.Math.Content.6.NS.B.3	8	7186	11996	59.9
	CCSS.Math.Content.6.RP.A.3.c	21	6572	11996	54.78
	CCSS.Math.Content.6.RP.A.3.b	12	6205	11996	51.73
	CCSS.Math.Content.6.NS.C.6.b	6	6156	11996	51.32
	CCSS.Math.Content.6.NS.C.6.a	6	5771	11996	48.11
	CCSS.Math.Content.6.NS.A.1	28	5654	11996	47.13
	CCSS.Math.Content.6.NS.C.6.c	12	5526	11996	46.07
	CCSS.Math.Content.6.RP.A.3.a	7	5242	11996	43.7
	CCSS.Math.Content.6.NS.C.7.a	9	5212	11996	43.45
	CCSS.Math.Content.6.NS.C.7	1	4797	11996	39.99
	CCSS.Math.Content.6.RP.A.2	6	4386	11996	36.56
	CCSS.Math.Content.5.G.A.1	8	4199	11996	35
	CCSS.Math.Content.6.NS.C.7.d	5	3842	11996	32.03
	CCSS.Math.Content.6.NS.C.7.b	7	3725	11996	31.05
	CCSS.Math.Content.5.NBT.B.7	3	3713	11996	30.95
	CCSS.Math.Content.6.NS.C.8	5	3183	11996	26.53

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.Math.Content.6.NS.C.7.c	4	2932	11996	24.44
	CCSS.Math.Content.6.NS.B.4	3	2261	11996	18.85
	CCSS.Math.Content.6.NS.C.5	6	2186	11996	18.22
	CCSS.Math.Content.6.RP.A.3.d	4	1492	11996	12.44
	CCSS.Math.Content.6.RP.A.1	4	1448	11996	12.07
	CCSS.Math.Content.5.NF.B.5	1	1111	11996	9.26
	CCSS.Math.Content.5.NBT.A.2	1	961	11996	8.01
	CCSS.Math.Content.5.NF.A.1	3	829	11996	6.91
	CCSS.Math.Content.5.MD.A.1	4	828	11996	6.9
	CCSS.Math.Content.5.NF.A.2	2	67	11996	0.56
	CCSS.Math.Content.7.RP.A.2.a	1	21	11996	0.18
	CCSS.Math.Content.7.NS.A.1.d	1	7	11996	0.06
	CCSS.Math.Content.5.NBT.B.5	1	3	11996	0.03
	CCSS.Math.Content.5.NF.B.4.a	1	1	11996	0.01
7	CCSS.Math.Content.7.G.B.6	22	11690	12241	95.5
	CCSS.Math.Content.7.G.A.1	15	10892	12241	88.98
	CCSS.Math.Content.7.G.B.4	11	9362	12241	76.48
	CCSS.Math.Content.7.G.A.3	6	5128	12241	41.89
	CCSS.Math.Content.7.G.B.5	10	3858	12241	31.52
	CCSS.Math.Content.6.G.A.4	4	2628	12241	21.47
	CCSS.Math.Content.6.G.A.3	3	615	12241	5.02
	CCSS.Math.Content.8.G.A.3	5	360	12241	2.94
	CCSS.Math.Content.8.G.A.4	1	304	12241	2.48
	CCSS.Math.Content.8.G.B.8	3	50	12241	0.41
	CCSS.Math.Content.8.G.A.5	4	24	12241	0.2
	CCSS.Math.Content.7.EE.B.4.a	30	10693	12241	87.35
	CCSS.Math.Content.7.EE.B.3	6	6683	12241	54.6
	CCSS.Math.Content.7.EE.B.4.b	29	6683	12241	54.6
	CCSS.Math.Content.8.EE.C.7.b	3	5285	12241	43.17
	CCSS.Math.Content.6.EE.B.7	6	3870	12241	31.62
	CCSS.Math.Content.7.EE.A.1	7	1741	12241	14.22
	CCSS.Math.Content.7.EE.A.2	4	1342	12241	10.96
	CCSS.Math.Content.6.EE.A.1	1	662	12241	5.41
	CCSS.Math.Content.6.EE.C.9	1	498	12241	4.07
	CCSS.Math.Content.8.EE.A.2	1	262	12241	2.14
	CCSS.Math.Content.8.F.A.1	1	227	12241	1.85
	CCSS.Math.Content.6.EE.B.6	2	61	12241	0.5
	CCSS.Math.Content.8.EE.C.7.a	1	25	12241	0.2
	CCSS.Math.Content.6.EE.A.2.c	1	11	12241	0.09
	CCSS.Math.Content.6.EE.B.8	1	11	12241	0.09
	CCSS.Math.Content.8.F.B.4	1	6	12241	0.05
	CCSS.Math.Content.8.EE.A.1	1	1	12241	0.01
	CCSS.Math.Content.7.SP.C.7	31	10749	12241	87.81

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.Math.Content.7.SP.C.6	12	7869	12241	64.28
	CCSS.Math.Content.7.SP.A	17	7331	12241	59.89
	CCSS.Math.Content.7.SP.C.5	7	5117	12241	41.8
	CCSS.Math.Content.7.SP.B	10	4606	12241	37.63
	CCSS.Math.Content.7.SP.C.8	20	3041	12241	24.84
	CCSS.Math.Content.6.SP.B.5	2	303	12241	2.48
	CCSS.Math.Content.6.SP.A.3	1	46	12241	0.38
	CCSS.Math.Content.8.SP.A.3	3	33	12241	0.27
	CCSS.Math.Content.8.SP.A.2	1	11	12241	0.09
	CCSS.Math.Content.7.RP.A.2.d	12	9686	12241	79.13
	CCSS.Math.Content.7.NS.A.1.d	11	8403	12241	68.65
	CCSS.Math.Content.7.RP.A.2.a	9	7928	12241	64.77
	CCSS.Math.Content.7.NS.A.1.c	7	7425	12241	60.66
	CCSS.Math.Content.7.NS.A.1.b	7	7349	12241	60.04
	CCSS.Math.Content.7.NS.A.2.a	8	6891	12241	56.29
	CCSS.Math.Content.7.NS.A.3	9	6777	12241	55.36
	CCSS.Math.Content.7.NS.A.2.d	7	6372	12241	52.05
	CCSS.Math.Content.7.NS.A.2.b	6	5331	12241	43.55
	CCSS.Math.Content.7.RP.A.2.c	10	5292	12241	43.23
	CCSS.Math.Content.7.NS.A.1.a	4	5046	12241	41.22
	CCSS.Math.Content.7.NS.A.2.c	4	4843	12241	39.56
	CCSS.Math.Content.7.RP.A.2.b	3	4129	12241	33.73
	CCSS.Math.Content.7.RP.A.3	9	3812	12241	31.14
	CCSS.Math.Content.7.RP.A.1	6	3544	12241	28.95
	CCSS.Math.Content.6.NS.C.7	1	2713	12241	22.16
	CCSS.Math.Content.6.NS.C.6.a	1	1662	12241	13.58
	CCSS.Math.Content.6.RP.A.3.c	9	1531	12241	12.51
	CCSS.Math.Content.6.NS.B.2	1	1370	12241	11.19
	CCSS.Math.Content.6.NS.C.6.c	4	903	12241	7.38
	CCSS.Math.Content.6.NS.C.7.a	4	691	12241	5.64
	CCSS.Math.Content.6.NS.B.4	1	618	12241	5.05
	CCSS.Math.Content.6.RP.A.3.a	1	370	12241	3.02
	CCSS.Math.Content.8.NS.A.1	3	369	12241	3.01
	CCSS.Math.Content.6.NS.A.1	3	292	12241	2.39
	CCSS.Math.Content.6.NS.C.5	1	252	12241	2.06
	CCSS.Math.Content.8.NS.A.2	3	183	12241	1.49
	CCSS.Math.Content.6.RP.A.3.b	2	128	12241	1.05
	CCSS.Math.Content.6.NS.C.7.b	1	121	12241	0.99
	CCSS.Math.Content.6.NS.B.3	1	63	12241	0.51
	CCSS.Math.Content.6.NS.C.6.b	1	1	12241	0.01
8	CCSS.Math.Content.8.G.A.5	33	10718	12337	86.88
	CCSS.Math.Content.8.G.A.3	21	9191	12337	74.5
	CCSS.Math.Content.8.G.B.7	21	5454	12337	44.21

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.Math.Content.8.G.A.2	10	4259	12337	34.52
	CCSS.Math.Content.8.G.C.9	13	4242	12337	34.38
	CCSS.Math.Content.8.G.B.8	17	3486	12337	28.26
	CCSS.Math.Content.8.G.A.1.a	4	3247	12337	26.32
	CCSS.Math.Content.8.G.A.4	9	2193	12337	17.78
	CCSS.Math.Content.7.G.A.1	2	336	12337	2.72
	CCSS.Math.Content.8.G.A.1.c	2	248	12337	2.01
	CCSS.Math.Content.8.G.A.1.b	2	81	12337	0.66
	CCSS.Math.Content.7.G.B.6	2	5	12337	0.04
	CCSS.Math.Content.7.G.B.5	1	4	12337	0.03
	CCSS.Math.Content.7.G.A.3	1	2	12337	0.02
	CCSS.Math.Content.8.EE.C.7.b	25	12027	12337	97.49
	CCSS.Math.Content.8.EE.A.4	21	11452	12337	92.83
	CCSS.Math.Content.8.F.B.4	22	9998	12337	81.04
	CCSS.Math.Content.8.EE.B.5	21	9918	12337	80.39
	CCSS.Math.Content.8.EE.A.3	15	6932	12337	56.19
	CCSS.Math.Content.8.EE.C.8	10	6471	12337	52.45
	CCSS.Math.Content.8.EE.A.2	14	5711	12337	46.29
	CCSS.Math.Content.8.F.A.3	5	5474	12337	44.37
	CCSS.Math.Content.8.EE.C.7.a	14	5173	12337	41.93
	CCSS.Math.Content.8.F.A.1	10	4761	12337	38.59
	CCSS.Math.Content.8.EE.A.1	9	3969	12337	32.17
	CCSS.Math.Content.8.F.B.5	7	3848	12337	31.19
	CCSS.Math.Content.8.EE.B.6	6	3669	12337	29.74
	CCSS.Math.Content.7.EE.B.3	3	1694	12337	13.73
	CCSS.Math.Content.7.EE.B.4.a	4	1248	12337	10.12
	CCSS.Math.Content.8.F.A.2	5	1118	12337	9.06
	CCSS.Math.Content.7.EE.B.4.b	3	63	12337	0.51
	CCSS.Math.Content.8.SP.A.1	22	10164	12337	82.39
	CCSS.Math.Content.8.SP.A.3	22	8307	12337	67.33
	CCSS.Math.Content.8.SP.A.4	7	5229	12337	42.38
	CCSS.Math.Content.8.SP.A.2	10	4040	12337	32.75
	CCSS.Math.Content.7.SP.C.6	2	2849	12337	23.09
	CCSS.Math.Content.7.SP.C.5	1	2598	12337	21.06
	CCSS.Math.Content.7.SP.B	4	2392	12337	19.39
	CCSS.Math.Content.7.SP.C.7	4	962	12337	7.8
	CCSS.Math.Content.7.SP.C.8	5	743	12337	6.02
	CCSS.Math.Content.8.NS.A.1	11	12096	12337	98.05
	CCSS.Math.Content.8.NS.A.2	9	11714	12337	94.95
	CCSS.Math.Content.7.NS.A.1.c	2	3323	12337	26.94
	CCSS.Math.Content.7.RP.A.2.a	2	2438	12337	19.76
	CCSS.Math.Content.7.NS.A.1.b	2	1977	12337	16.02
	CCSS.Math.Content.7.RP.A.2.c	3	1502	12337	12.17

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.Math.Content.7.RP.A.3	3	836	12337	6.78
	CCSS.Math.Content.7.NS.A.1.d	2	253	12337	2.05
	CCSS.Math.Content.7.NS.A.2.c	1	61	12337	0.49
	CCSS.Math.Content.7.NS.A.3	1	12	12337	0.1
	CCSS.Math.Content.7.NS.A.2.b	2	3	12337	0.02
	CCSS.Math.Content.7.RP.A.2.d	1	1	12337	0.01
HS	CCSS.Math.Content.HSG-SRT.A.1.b	2	12111	12555	96.46
	CCSS.Math.Content.HSG-SRT.B.5	2	11149	12555	88.8
	CCSS.Math.Content.HSG-SRT.C.7	1	10670	12555	84.99
	CCSS.Math.Content.HSG-SRT.C.8	1	10466	12555	83.36
	CCSS.Math.Content.HSG-CO.C.10	1	10267	12555	81.78
	CCSS.Math.Content.HSG-CO.A.1	1	10033	12555	79.91
	CCSS.Math.Content.HSG-CO.A.3	2	9869	12555	78.61
	CCSS.Math.Content.HSG-C.A.2	1	8041	12555	64.05
	CCSS.Math.Content.HSG-CO.A.5	1	2156	12555	17.17
	CCSS.Math.Content.HSG-CO.C.9	1	1865	12555	14.85
	CCSS.Math.Content.HSG-CO.B.7	1	1863	12555	14.84
	CCSS.Math.Content.HSG-SRT.A.3	1	1845	12555	14.7
	CCSS.Math.Content.HSG-GPE.B.7	1	1771	12555	14.11
	CCSS.Math.Content.HSG-GMD.A.3	1	1597	12555	12.72
	CCSS.Math.Content.HSG-CO.A.2	1	1362	12555	10.85
	CCSS.Math.Content.HSG-GMD.B.4	1	1283	12555	10.22
	CCSS.Math.Content.HSG-GPE.A.1	1	911	12555	7.26
	CCSS.Math.Content.HSG-GPE.B.5	1	615	12555	4.9
	CCSS.Math.Content.HSA-APR.A.1	3	12515	12555	99.68
	CCSS.Math.Content.HSA-SSE.B.3.a	2	12283	12555	97.83
	CCSS.Math.Content.HSA-REI.A.1	1	11569	12555	92.15
	CCSS.Math.Content.HSF-LE.A.3	1	11255	12555	89.65
	CCSS.Math.Content.HSF-IF.A.1	1	11239	12555	89.52

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.Math.Content.HSF-BF.A.1.a	1	10920	12555	86.98
	CCSS.Math.Content.HSA-CED.A.1	1	10676	12555	85.03
	CCSS.Math.Content.HSA-SSE.A.2	1	10655	12555	84.87
	CCSS.Math.Content.HSF-BF.A.2	1	10549	12555	84.02
	CCSS.Math.Content.HSF-IF.B.5	1	10414	12555	82.95
	CCSS.Math.Content.HSA-SSE.A.1	1	10403	12555	82.86
	CCSS.Math.Content.HSF-LE.A.1.a	1	9927	12555	79.07
	CCSS.Math.Content.HSF-LE.A.1.b	1	9449	12555	75.26
	CCSS.Math.Content.HSF-BF.A.1.b	1	3880	12555	30.9
	CCSS.Math.Content.HSF-LE.B.5	1	2850	12555	22.7
	CCSS.Math.Content.HSA-REI.B.4.b	1	2415	12555	19.24
	CCSS.Math.Content.HSF-BF.B.3	1	2136	12555	17.01
	CCSS.Math.Content.HSF-IF.A.2	1	1606	12555	12.79
	CCSS.Math.Content.HSF-LE.A.2	1	1602	12555	12.76
	CCSS.Math.Content.HSA-CED.A.2	1	1449	12555	11.54
	CCSS.Math.Content.HSF-IF.B.6	1	1449	12555	11.54
	CCSS.Math.Content.HSA-CED.A.3	1	1403	12555	11.17
	CCSS.Math.Content.HSF-IF.C.8.a	1	1331	12555	10.6
	CCSS.Math.Content.HSA-REI.D.10	1	1188	12555	9.46
	CCSS.Math.Content.HSF-IF.C.7.e	1	1044	12555	8.32
	CCSS.Math.Content.HSF-IF.B.4	1	548	12555	4.36
	CCSS.Math.Content.HSS-ID.A.3	1	11027	12555	87.83
	CCSS.Math.Content.HSS-ID.C.7	1	10961	12555	87.3
	CCSS.Math.Content.HSS-CP.A.4	1	10502	12555	83.65
	CCSS.Math.Content.HSS-ID.B.6.c	1	10487	12555	83.53
	CCSS.Math.Content.HSS-CP.A.1	1	2193	12555	17.47
	CCSS.Math.Content.HSS-ID.A.1	1	1945	12555	15.49
	CCSS.Math.Content.HSS-ID.B.5	1	1817	12555	14.47
	CCSS.Math.Content.HSS-CP.A.2	1	604	12555	4.81
	CCSS.Math.Content.HSS-ID.A.2	1	535	12555	4.26
	CCSS.Math.Content.HSN-Q.A.1	1	12524	12555	99.75
	CCSS.Math.Content.HSN-Q.A.3	1	12524	12555	99.75

Grade	Standard	Item Count	Student Count	Total Student	Percent
	CCSS.Math.Content.HSN-Q.A.2	1	12481	12555	99.41
	CCSS.Math.Content.HSN-RN.B.3	1	12463	12555	99.27

Maine Spring 2024 Adaptive Test Simulation Study Report

Table of Contents

List of Tables

List of Figures

1. Introduction

The Maine Department of Education contracted with NWEA to administer a Maine Through Year Assessment (TYA) for the Maine Comprehensive Assessment System (MECAS) from Spring 2023 to Spring 2024. The Maine Through Year Assessment Program has three administrations per year: fall, winter, and spring. Each administration assesses grades 3−8 and high school (HS) assessments in reading and mathematics.

This report presents the results of the adaptive test study for the Spring 2024 administration. This adaptive test study simulates assessment requirements to evaluate the capacity of the item pool to support those requirements through the NWEA constraint-based engine (CBE). This report focuses on five areas:

- An overview of the spring test design
- An overview of the NWEA constraint-based engine
- An examination of the item pool (summative and MAP Growth)
- The accuracy of item selections against the blueprint
- The relative accuracy of student-proficiency estimation

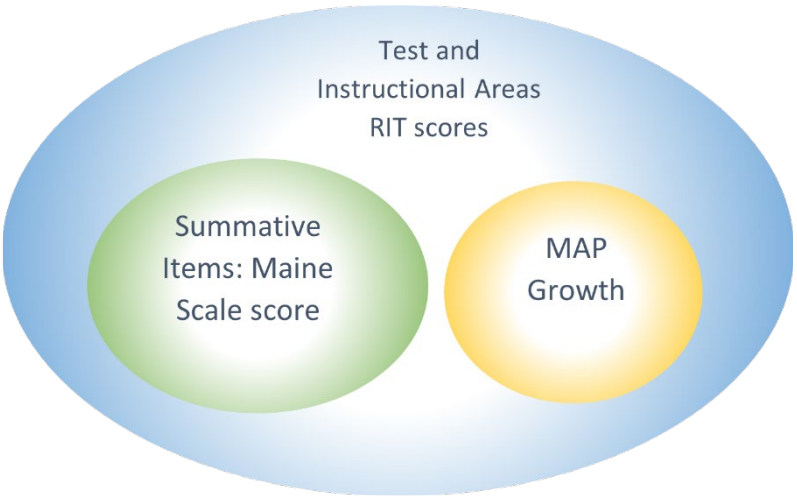
1.1. Maine Spring Test Design Overview

The Maine Through Year Assessment serves different purposes in different administrations. Fall and winter administrations produce growth and national norm scores in addition to the RIT and instructional area scores through the NWEA MAP Growth product. The spring administration serves two purposes: to determine end-of-grade proficiency and to produce growth and national norm scores. To fulfill these two purposes, the spring test consists of two components: the summative and diagnostic portions. Content standards for both portions are aligned to the Common Core State Standards (CCSS). The descriptions of two portions of the spring test are as follows:

- Summative Portion
    - Addresses the breadth and depth of the Maine's grade-level content standards by selecting items from the NWEA state item pool that align to the Common Core State Standards
    - Produces a summative test score and determines proficiency
    - Satisfies the summative test blueprint for a balance of content representation
- Diagnostic Portion
    - Combines with the summative portion to yield scores for instructional area reporting categories as well as the RIT score
    - Accesses the MAP Growth item pool to utilize its RIT scores to generate growth and norm-referenced scores
    - Allows item selection from other grades: grades 3 to 5 can assess items from K to 8, and grades 6 and up allow items from 3 to 12

Figure 1.1 illustrates how these two components work together to produce both a single summative test score and a suite of diagnostic scores, including instructional areas.

Figure 1.1. Items Contributing to Summative, MAP Growth RIT, and Instructional Area Scores



1.1.1. Reading and Mathematics Score Types

Figure 1.2 presents reading score types, and Figure 1.3 and Figure 1.4 present mathematics score types. As shown in the figures, the Maine scale score is used to determine the student achievement classification. Sub scores are reported for each instructional area.

Figure 1.2. Reading Grades 3−8 and High School Score Types

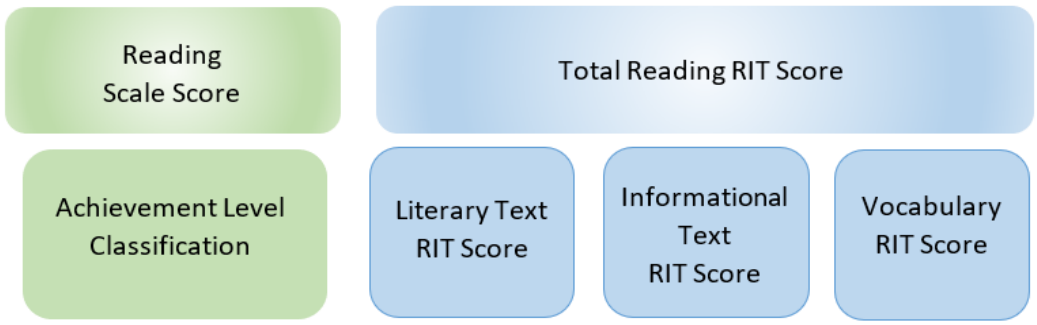


Figure 1.3. Mathematics Grades 3−5 Score Types

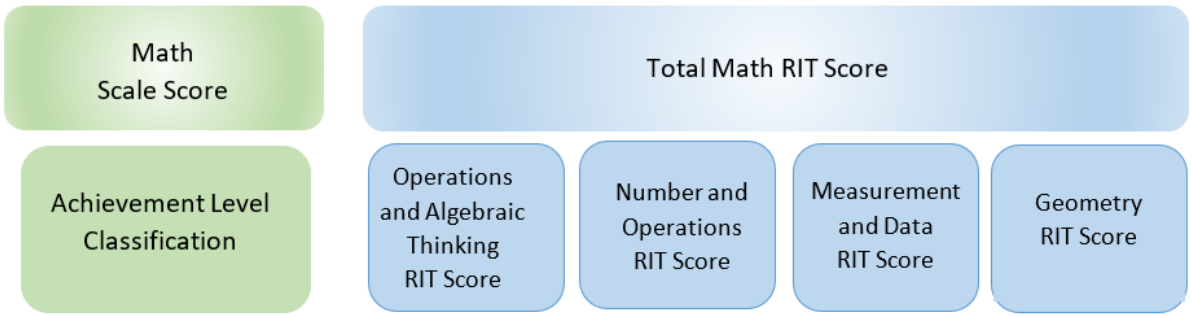
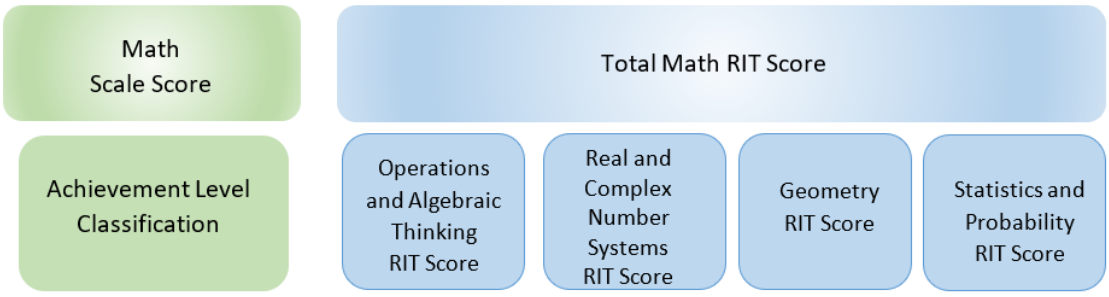


Figure 1.4. Mathematics Grades 6−8 and High School Score Types



1.2. Simulation Purposes

The Maine Through Year Assessment is an item-level computer adaptive test (CAT). One main reason for using a CAT method is to provide a test customized to each student's ability level, which increases the reliability of the student's ability estimate. The proprietary constraint-based engine (CBE) from NWEA is used for the adaptive test. Pattern scoring is used to derive student scores before selecting subsequent items to facilitate the item-selection process. Before operational administration, a simulation study is necessary to investigate whether the item pool can sustain the requirements specified for the MECAS assessment and produce reliable student scores.

The technical purposes of the simulation study are to provide evidence (along with post-administration analyses) supporting test-score interpretation and to support arguments regarding student proficiency relative to the state standards. The simulation is intended to demonstrate that students receive comparable representations of content with sufficient technical adequacy such that the state can infer that test scores have the same meaning across students' individualized test events.

1.3. Constraint-Based Engine (CBE) Overview

The CBE is an adaptive item-selection engine NWEA has developed for computerized adaptive testing. It combines established psychometric approaches with NWEA-specific innovations to deliver tests that strictly adhere to test blueprints while providing flexibility in item-pool construction and item ordering.

The CBE is designed to address the challenges that test designers face during test construction. It is not locked into any specific IRT model, measurement scale, item-selection criteria, or item-ordering configuration. Rather, a test designer specifies what is needed to run the desired test, and the CBE selects items that meet that test design, without adding its own modifications or restrictions. This lets the test designer measure what they desire—including metrics other than student ability—without being encumbered by arbitrary decisions made during administrating the test.

The CBE leverages existing psychometric concepts to achieve its goals but modifies those approaches and blends them with trade-secret innovations. Specifically, it allows the use of any IRT model, such as Rasch, three-parameter logistic (3PL), partial-credit model (PCM), and generalized partial-credit model (GPCM). The CBE also implements a blend of a modified shadow test approach and a modified weighted penalty model. Building on these approaches, it also allows test designers set their own item attributes to measure against, separates item

ordering from item selection, and consumes both constraints and “guidelines”—criteria the designer would prefer to meet but does not require the test to meet.

1.3.1. Item-Selection Approaches

The CBE uses a blended approach for item selection, combining and building on elements of the weighted penalty model (WPM) and the shadow test approach (STA). The critical difference between these two approaches is that the WPM focuses on maximizing information on an item-by-item basis, while the STA focuses on maximizing information for the whole test simultaneously. Combining these approaches lets the CBE both select the best item to present next and select the best items to use in the construction of the test as a whole.

Each item-selection method has its own advantages and disadvantages that the test designer must consider when choosing the best options for a test. The CBE capitalizes on this variety by blending and building on the approaches above to maximize the strengths of each.

The key innovations implanted in the CBE are:

- Blended item-selection approach (shadow test approach and weighted penalty model)
- Separation of item-selection and item-ordering procedures
- Implementation of designer-defined content and item attributes
- Implementation of “guidelines” set by the test designer that they would prefer (but do not require) the test to meet
- Shared stimulus selection by item rather than by stimulus

Both the STA and the WPM are powerful item-selection models in their own right. However, despite these strengths, each has drawbacks that the test designer must consider when choosing between them for implementation on a particular test.

By calculating tests item-by-item, the WPM provides a greater degree of item pool and selection flexibility compared with the STA. Under the WPM, increasingly heavy penalties are applied to items that do not meet the blueprint, while penalties are removed for items that increase information. The item with the best overall “score” (lowest penalty) is then chosen. This increased flexibility, however, means that it is possible to deliver a test that does not meet the blueprint. As the system has no foresight into possible paths that will cause the test to deviate from the blueprint, small deviations can compound over time.

In contrast, the STA guarantees adherence to the test blueprint by calculating the entire test at each selection step. This lets the STA take future selection steps into account and cut off paths that would lead to the test deviating from the blueprint. However, this strict adherence to the test blueprint adds a large degree of rigidity to the item-selection process and requires a large, carefully tailored item pool to be successful. As the best items are positioned early in the test, the likelihood of calculating a feasible test decreases (Robin et al., 2005). This can require the use of a spare item pool to replenish the primary item pool so that feasible tests can continue to be calculated. Maintaining a sufficiently large and manicured item pool, or multiple item pools, is both costly and inefficient.

The CBE combines the item-pool flexibility of the WPM with the foresight of the STA. At each selection step, the CBE calculates all tests that meet all constraints and applies penalties to each of these full tests based on a combination of the degree to which they meet guidelines as well as the extent to which they maximize information. Penalties are applied quadratically rather

than linearly to strongly differentiate the tests' ability to meet guidelines and maximize information. It then selects the test that maximizes the function of information minus guideline penalties. The result is that any test delivered to the student is guaranteed to meet the test-design specification, but the chances that the item pool will eventually run down (such that a spare is required) is eliminated.

After the CBE calculates the optimal items to be included in the student-specific plan, it proceeds to order those items based on business rules and the requirements indicated by the test design. Since each item in the student-specific plan has been confirmed to meet all constraints, and since the plan as a whole maximizes both information and adherence to guidelines, ordering can be done in a more lightweight fashion that conserves system resources and improves performance compared with solving for both item selection and order simultaneously. Additionally, this gives the test designer greater control in measuring domain understanding through the choices indicated by the test design around balancing and item-ordering preferences.

1.3.2. Maximum Likelihood Estimation with Fencing (MLEF)

In the early stages of the test, each student has a perfect response pattern: all correct or all incorrect. Definitionally, this is always the case until at least two items have been answered. To address this, the CBE uses Maximum Likelihood Estimate with Fences (MLEF) (Han, 2016). Under MLEF, imaginary "fence" items are generated with fixed responses in order to provide a log likelihood function that can be used as a starting point for adaptive selection. The log likelihood function estimated in a dichotomous MLEF item is:

$$L = lnP_{LF} + ln(1 - P_{UF}) + \sum_{j=1}^{n} [\mu_j lnP_j + (1 - \mu_j) ln(1 - P_j)]$$

where P_{LF} is the item response function of the lower fence, P_{UF} is the item response function of the upper fence, and μ is a response to a string of j items for polytomous items; the upper and lower fences depend on whether the response pattern is all correct or all incorrect.

For all correct:

$$L = lnP_{LF} + \sum_{j=1}^{n} [\mu_j lnP_j + (1 - \mu_j) ln(1 - P_j)]$$

For all incorrect:

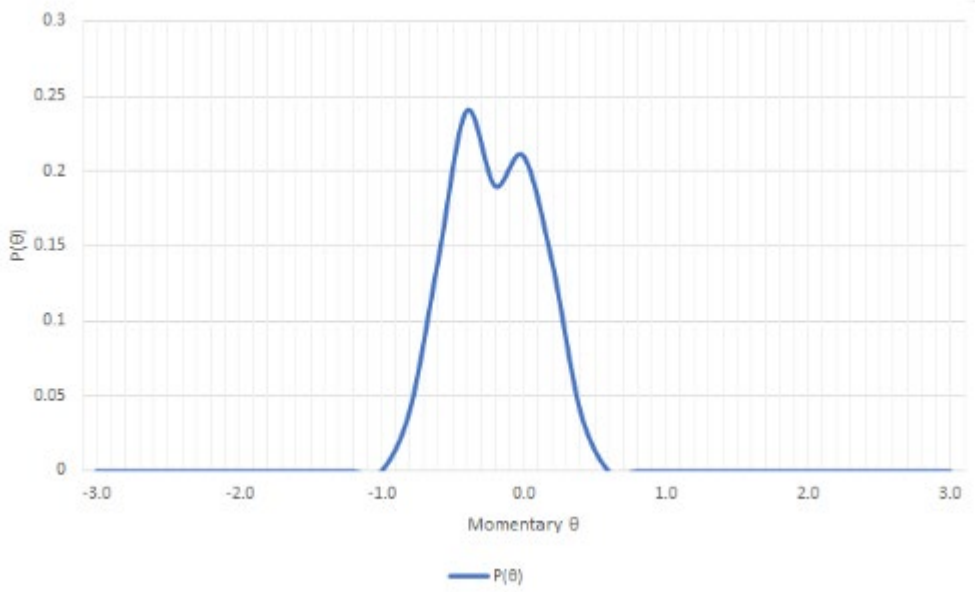
$$L = ln(1 - P_{UF}) + \sum_{j=1}^{n} [\mu_j lnP_j + (1 - \mu_j) ln(1 - P_j)]$$

If fencing items are required, the test designer can indicate appropriate scale properties on the test design and apply constraints or guidelines to utilize them. In the event of a perfect response pattern, the CBE uses these scale properties to select the appropriate item to be used as a fence. Additionally, the CBE determines the minimum or maximum difficulty of items that have been administered to the student.

- If the student response pattern is all correct: the CBE adds the delta to the maximum difficulty of the administered items and inserts this value into the item-difficulty parameter. It adds a virtual response that is incorrect for that item during the process of calculating the student ability estimate using the MLEF method.
- If the student response pattern is all incorrect: the CBE subtracts the delta from the minimum difficulty of the administered items and inserts this value into the item-difficulty parameter. It adds a virtual response that is correct for that item.

In this way, the CBE guarantees there will be a maximum in the MLE process. MLE is calculated by computing the probability of student ability, $p(\theta)$, at every bin defined in the scale and then choosing the theta estimate that has the highest probability. For example, for a scale defined as -3.0 to +3.0 θ, broken into ranges of size 0.1, the CBE calculates the probability of student theta at each bin (-3, -2.9, -2.8, and so on) and then chooses the bin with the highest probability.

Figure 1.5. Sample Student Theta Probability Subject to MLEF



The example illustrated in Figure 1.5 shows that two theta estimates are competing for most likely at -0.4 and 0 (i.e., the two peaks). Although they are similar in probability, -0.4 is identified as most likely; therefore, the CBE proceeds with its test information calculation under the assumption that θ = -0.4.

2. Study Design

2.1. Sample

This simulation study sampled 1,000 students from Maine who took the Spring 2023 Maine Through Year Assessment to obtain a representative sample of the general student population for Maine. This sample represents the ability range and distribution of Maine students' "true ability" and is used in the simulation study. The following tables show the demographic information and student ability estimates of the population of the Maine students and the sampled students.

Table 2.1. Student Demographic Information for Spring 2023—Mathematics

Grade		Total	Gender		Ethnicity							
			Female	Male	American Indian	Asian	African American	Hispanic	Native Hawaiian or Pacific Islander	Caucasian	Two or More Races	Not Specified
3	N	12108	5927	6180	422	76	155	542	17	10425	471	0
	%	100.00	48.95	51.04	3.49	0.63	1.28	4.48	0.14	86.10	3.89	0.00
4	N	12124	5863	6259	386	109	164	574	18	10437	435	1
	%	100.00	48.36	51.62	3.18	0.90	1.35	4.73	0.15	86.09	3.59	0.01
5	N	11891	5822	6065	373	94	149	527	15	10291	442	0
	%	100.00	48.96	51.00	3.14	0.79	1.25	4.43	0.13	86.54	3.72	0.00
6	N	12037	5944	6090	346	88	165	568	5	10437	427	1
	%	100.00	49.38	50.59	2.87	0.73	1.37	4.72	0.04	86.71	3.55	0.01
7	N	12167	5820	6343	378	91	150	571	15	10552	410	0
	%	100.00	47.83	52.13	3.11	0.75	1.23	4.69	0.12	86.73	3.37	0.00
8	N	12556	6074	6477	383	108	188	570	19	10857	430	1
	%	100.00	48.38	51.58	3.05	0.86	1.50	4.54	0.15	86.47	3.42	0.01
HS	N	11879	5812	6063	401	78	225	511	14	10294	356	0
	%	100.00	48.93	51.04	3.38	0.66	1.89	4.30	0.12	86.66	3.00	0.00

Table 2.2. Student Demographic Information for Spring 2023—Reading

			Gender		Ethnicity							
Grade		Total	Female	Male	American Indian	Asian	African American	Hispanic	Native Hawaiian or Pacific Islander	Caucasian	Two or More Races	Not Specified
3	N	12034	5886	6147	414	76	154	500	17	10405	468	0
	%	100.00	48.91	51.08	3.44	0.63	1.28	4.15	0.14	86.46	3.89	0.00
4	N	12026	5827	6197	374	109	160	527	18	10407	431	0
	%	100.00	48.45	51.53	3.11	0.91	1.33	4.38	0.15	86.54	3.58	0.00
5	N	11820	5798	6018	367	93	145	477	15	10280	443	0
	%	100.00	49.05	50.91	3.10	0.79	1.23	4.04	0.13	86.97	3.75	0.00
6	N	11977	5919	6055	341	88	165	534	5	10417	427	0
	%	100.00	49.42	50.56	2.85	0.73	1.38	4.46	0.04	86.98	3.57	0.00
7	N	12074	5781	6289	371	88	147	531	16	10513	408	0

			Gender		Ethnicity							
Grade		Total	Female	Male	American Indian	Asian	African American	Hispanic	Native Hawaiian or Pacific Islander	Caucasian	Two or More Races	Not Specified
	%	100.00	47.88	52.09	3.07	0.73	1.22	4.40	0.13	87.07	3.38	0.00
8	N	12498	6059	6434	378	108	185	527	19	10852	429	0
	%	100.00	48.48	51.48	3.02	0.86	1.48	4.22	0.15	86.83	3.43	0.00
HS	N	11941	5830	6107	397	81	225	484	15	10380	359	0
	%	100.00	48.82	51.14	3.32	0.68	1.88	4.05	0.13	86.93	3.01	0.00

Table 2.3. Summary of Student Ability for Spring 2023 Maine Through Year Assessment

Grade	Reading RIT Theta				Math RIT Theta			
	Mean	SD	Min	Max	Mean	SD	Min	Max
3	-0.46	1.39	-3.86	4.02	-0.03	1.58	-5.19	5.40
4	0.21	1.39	-3.66	4.81	0.70	1.59	-4.85	6.66
5	0.85	1.36	-3.46	5.07	1.31	1.72	-4.18	7.64
6	1.24	1.35	-2.87	5.77	1.66	1.68	-3.75	8.12
7	1.57	1.42	-2.95	6.26	2.10	1.69	-3.17	9.03
8	2.02	1.44	-2.46	7.24	2.74	1.76	-3.03	9.42
HS	1.67	1.21	-2.04	6.40	2.82	1.24	-1.00	9.40

Table 2.4. Student Sample Demographic Information for Simulation Study—Reading

			Gender		Ethnicity							
Grade		Total	Female	Male	American Indian	Asian	African American	Hispanic	Native Hawaiian or Pacific Islander	Caucasian	Two or More Races	Not Specified
3	N	1000	495	505	46	10	10	24	3	869	38	0
	%	100.00	49.50	50.50	4.60	1.00	1.00	2.40	0.30	86.90	3.80	0.00
4	N	1000	495	505	46	10	10	24	3	869	38	0
	%	100.00	49.50	50.50	4.60	1.00	1.00	2.40	0.30	86.90	3.80	0.00
5	N	1000	494	505	29	10	7	39	0	878	37	0

			Gender		Ethnicity							
Grade		Total	Female	Male	American Indian	Asian	African American	Hispanic	Native Hawaiian or Pacific Islander	Caucasian	Two or More Races	Not Specified
	%	100.00	49.40	50.50	2.90	1.00	0.70	3.90	0.00	87.80	3.70	0.00
6	N	1000	459	541	41	10	11	33	1	860	44	0
	%	100.00	45.90	54.10	4.10	1.00	1.10	3.30	0.10	86.00	4.40	0.00
7	N	1000	483	517	28	10	14	34	0	873	41	0
	%	100.00	48.30	51.70	2.80	1.00	1.40	3.40	0.00	87.30	4.10	0.00
8	N	1000	457	543	30	12	12	35	1	881	29	0
	%	100.00	45.70	54.30	3.00	1.20	1.20	3.50	0.10	88.10	2.90	0.00
HS	N	1000	491	509	30	8	16	38	0	869	39	0
	%	100.00	49.10	50.90	3.00	0.80	1.60	3.80	0.00	86.90	3.90	0.00

Table 2.5. Student Sample Demographic Information for Simulation Study—Mathematics

			Gender		Ethnicity							
Grade		Total	Female	Male	American Indian	Asian	African American	Hispanic	Native Hawaiian or Pacific Islander	Caucasian	Two or More Races	Not Specified
3	N	1000	499	501	28	13	7	26	0	882	44	0
	%	100.00	49.90	50.10	2.80	1.30	0.70	2.60	0.00	88.20	4.40	0.00
4	N	1000	499	501	28	13	7	26	0	882	44	0
	%	100.00	49.90	50.10	2.80	1.30	0.70	2.60	0.00	88.20	4.40	0.00
5	N	1000	490	510	29	10	8	42	2	873	36	0
	%	100.00	49.00	51.00	2.90	1.00	0.80	4.20	0.20	87.30	3.60	0.00
6	N	1000	502	497	36	12	11	33	0	852	56	0
	%	100.00	50.20	49.70	3.60	1.20	1.10	3.30	0.00	85.20	5.60	0.00
7	N	1000	493	507	35	10	13	27	0	892	23	0
	%	100.00	49.30	50.70	3.50	1.00	1.30	2.70	0.00	89.20	2.30	0.00
8	N	1000	453	547	33	7	9	29	2	883	37	0
	%	100.00	45.30	54.70	3.30	0.70	0.90	2.90	0.20	88.30	3.70	0.00

			Gender		Ethnicity							
Grade		**Total**	**Female**	**Male**	**American Indian**	**Asian**	**African American**	**Hispanic**	**Native Hawaiian or Pacific Islander**	**Caucasian**	**Two or More Races**	**Not Specified**
HS	N	1000	507	493	30	7	21	47	1	860	34	0
	%	100.00	50.70	49.30	3.00	0.70	2.10	4.70	0.10	86.00	3.40	0.00

Table 2.6. Summary of Student Ability for Simulation Sample

Grade	Reading MG RIT Theta				Math MG RIT Theta			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
3	-0.37	1.56	-5.0	3.5	0.07	1.28	-3.1	3.9
4	0.41	1.48	-4.9	4.7	1.01	1.28	-3.8	5.0
5	0.95	1.50	-4.8	4.5	1.55	1.48	-3.8	5.7
6	1.29	1.45	-3.9	5.7	2.03	1.46	-2.4	8.1
7	1.60	1.48	-3.5	5.5	2.44	1.62	-2.1	8.7
8	2.10	1.49	-4.2	9.0	2.95	1.76	-2.5	9.1
HS	2.71	1.51	-3.1	6.8	3.32	1.81	-2.7	8.9

2.2. Item Pool Characteristics

The Maine Through Year Assessment has two content areas, reading and mathematics, in grades 3–8 and high school. The following tables present the numbers of items in the item pool, including both through-year summative items and MAP Growth diagnostic items, by instructional area. Note that the summative test allows items from +/- one-off (adjacent) grades, and MAP Growth allows items from kindergarten to grade 8 for grades 3 to 5 tests and grades 3 to high school for tests of grade 6 and up. Only operational items that have item statistics for the adaptive process are included in the tables below. Math grades 3 to 5 and 6 to HS have different instructional areas. Thus, zero counts appear in different instructional areas across math grades.

Table 2.7. Item Counts by Item Source and Instructional Area—Reading

Source	Content Category	Grade							
		2	3	4	5	6	7	8	HS
Summative									
	Literary Text		140	129	121	118	129	112	8
	Informational Text		158	116	125	140	143	180	32
	Vocabulary		92	122	103	84	90	122	9
Sub Total			390	367	349	342	362	414	49
Diagnostic									
	Literary Text	140	178	194	157	178	94	88	64
	Informational Text	162	161	175	132	208	96	173	135
	Vocabulary	236	200	295	203	249	152	185	79
Sub Total		538	539	664	492	635	342	446	508
Total		538	929	1031	841	977	704	860	557

Table 2.8. Item Counts by Item Source and Instructional Area—Mathematics

Source	Content Category	Grade							
		2	3	4	5	6	7	8	HS
Summative									
	Operations and Algebraic Thinking		131	61	51	116	76	186	29
	Numbers and Operations		126	220	221				
	Measurement and Data		171	74	70				
	Geometry		21	40	45	46	64	132	21
	The Real and Complex Number Systems					169	114	20	4
	Statistics and Probability					42	97	61	9
Sub Total			449	395	387	373	351	399	63
Diagnostic									
	Operations and Algebraic Thinking	29	27	22	25	24	22	46	40
	Numbers and Operations	37	43	40	38				

Source	Content Category	Grade							
		2	3	4	5	6	7	8	HS
	Measurement and Data	27	20	25	26				
	Geometry	17	17	20	21	27	14	17	43
	The Real and Complex Number Systems					38	39	31	22
	Statistics and Probability					20	23	25	31
Sub Total		110	107	107	110	109	98	119	136
Total		110	556	502	497	482	449	518	199

The figures and tables below present the distribution of item difficulty for summative and MAP Growth item pools.

Figure 2.1. Reading Item Difficulty Distribution

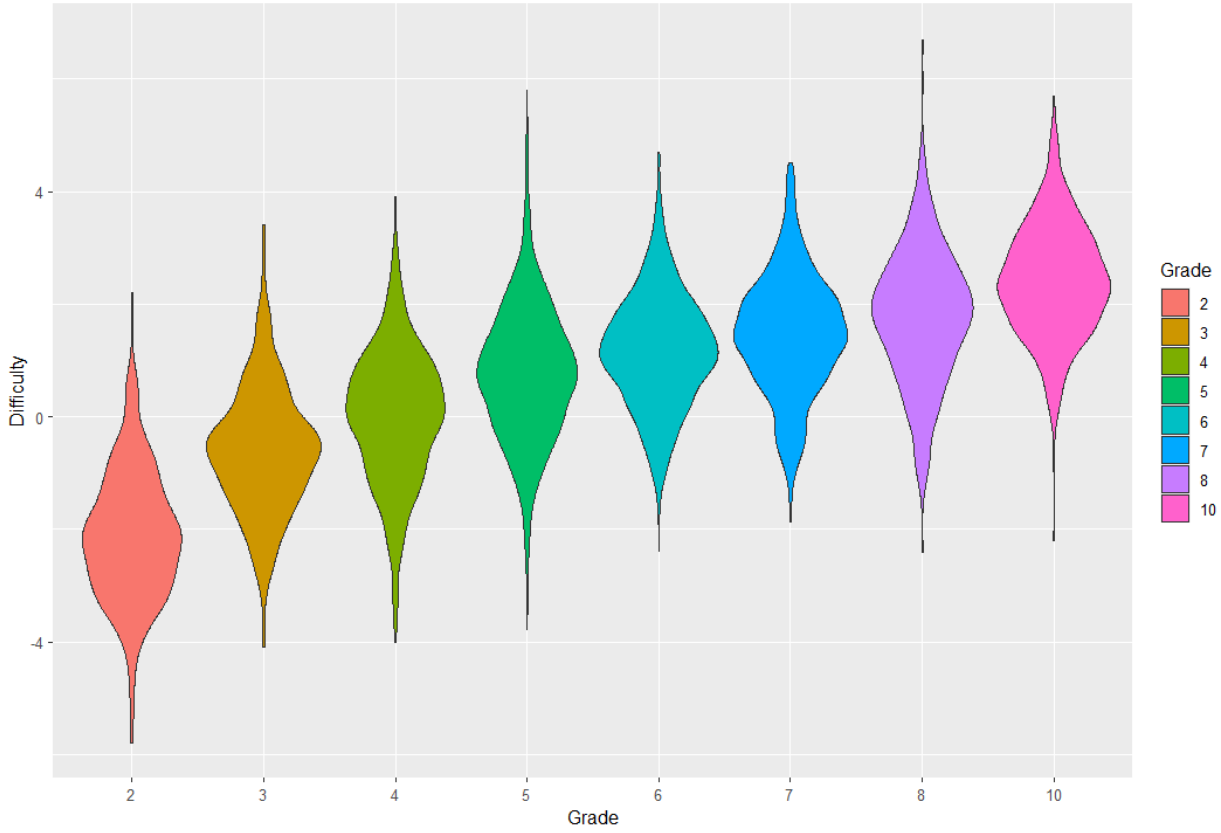


Figure 2.2. Mathematics Item Difficulty Distribution

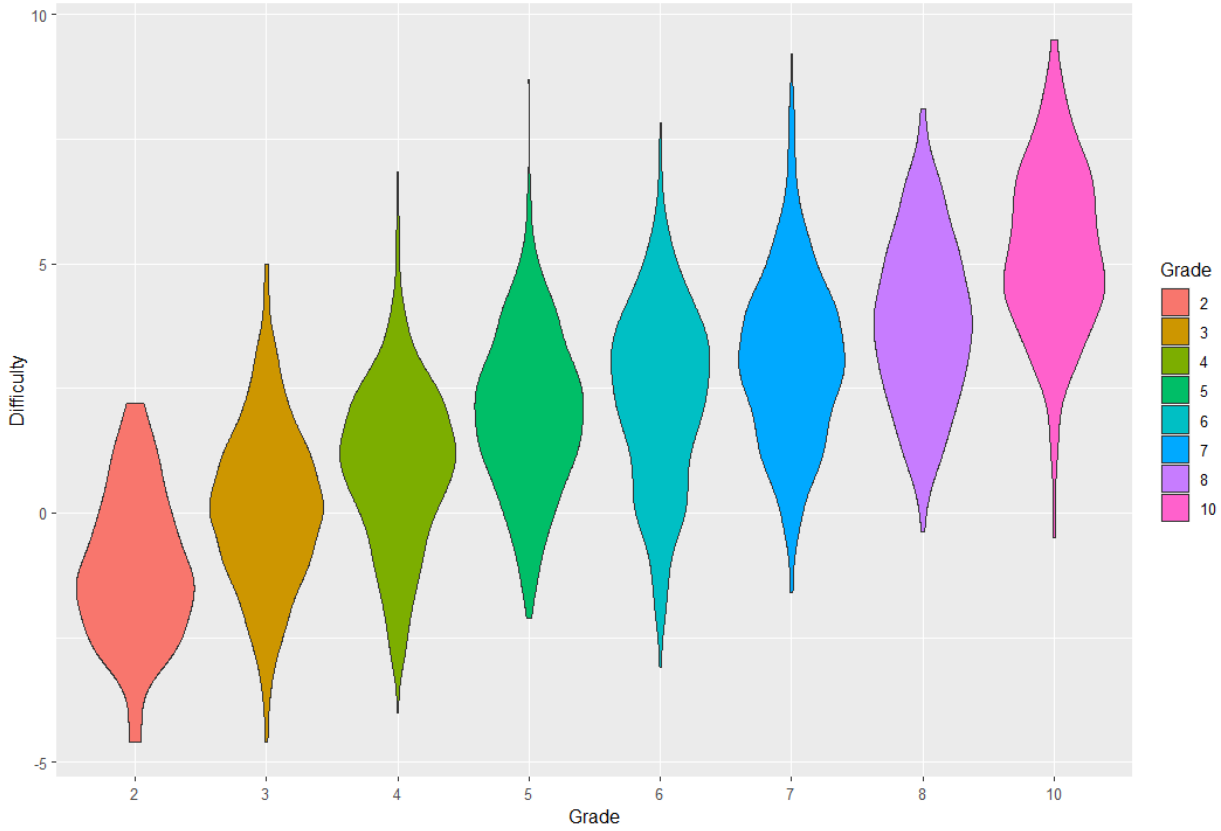


Table 2.9. Summary of Reading and Mathematics Item Difficulties

Subject	Grade	Summative					Total Test				
		N Items	Mean	SD	Min	Max	N Items	Mean	SD	Min	Max
Reading											
	2						538	-2.13	1.17	-5.80	2.20
	3	390	-0.43	1.03	-2.91	3.42	929	-0.60	1.13	-4.10	3.42
	4	367	0.13	1.06	-2.14	3.89	1031	0.01	1.21	-4.00	3.90
	5	349	0.84	1.11	-2.09	4.54	841	0.75	1.21	-3.80	5.80
	6	342	1.14	1.05	-1.39	4.68	977	1.17	1.02	-2.40	4.70
	7	362	1.34	1.09	-1.87	4.36	704	1.45	1.07	-1.87	4.50
	8	414	1.38	1.31	-2.30	6.60	860	1.77	1.22	-2.40	6.70
	HS	49	1.43	0.88	-0.81	3.56	557	2.40	1.01	-2.20	5.70
Math											
	2						110	-1.12	1.49	-4.60	2.20
	3	449	0.13	1.53	-4.60	4.56	556	0.12	1.63	-4.60	5.00
	4	395	0.86	1.59	-3.12	6.84	502	0.92	1.65	-4.00	6.84
	5	387	1.87	1.52	-1.79	6.41	497	2.00	1.61	-2.10	8.70
	6	373	2.31	1.74	-2.32	7.83	482	2.33	1.83	-3.10	7.83
	7	351	2.97	1.57	-1.55	7.84	449	3.08	1.72	-1.60	9.20
	8	399	3.72	1.67	-0.38	7.97	518	3.83	1.68	-0.38	8.10

Subject	Grade	Summative					Total Test				
		N Items	Mean	SD	Min	Max	N Items	Mean	SD	Min	Max
	HS	63	4.52	1.22	1.59	7.09	199	5.23	1.73	-0.50	9.50

2.3. Evaluation Criteria

The simulation results based on the CBE are evaluated and organized focusing on the following questions:

1. Do the selected items meet the requirements of the test blueprints?
2. Is the student's estimated ability matching the difficulty level of the selected items?
3. Do items have adequate exposure rates?
4. Do student ability estimates have adequate precision?
5. Are summative, MAP Growth, and field-test items given in the designated sequence?

The subsections are describing how those questions are answered in the evaluation process.

2.3.1. Meeting Test Blueprint Requirements

The nature of a CAT is for differing students to have different items during test administration. Not all students receive the same items. However, the items administered to students must meet the blueprint requirements outlined in the test specifications (e.g., a minimum number of items per reporting category). This section provides results that show whether the constraint-based engine administered the test based on the blueprints.

Meeting blueprint requirements is evaluated using the matching rate. The matching rate is calculated as the percentage of items that meet the blueprint requirements. For example, if 10 items are required for instructional area 1 and 10 items were administered, the matching rate is 100%.

2.3.2. Ability Estimate Accuracy

The precision of ability estimates assesses how accurately the CBE recovers students' true abilities based on the item pool. The relation between student ability estimate for selection and item difficulty statistics are investigated using correlation, mean squared error (MSE), bias, and standard error (SE). The following indices were used to evaluate the precision of score estimation within the CBE:

- The correlation between true theta and estimated theta is a direct indication of matching between these two data.
- MSE, bias, and SE are a set of error estimates between predicted and observed values. In this study, the two values are replaced by true theta and estimated theta.
- MSE is the mean of differences between a student's theta estimate and the true theta. The smaller the MSE, the closer the fit between a student's theta and the items selected for the student.
- MSE can be partitioned into two parts: Bias and SE.
- Bias is a systematic error that indicates how well the predicted value is at estimating the true value. SE is a random error that indicates how many uncontrolled errors are in the sample of items.

The relation between MSE, bias, and SE is:

$$MSE = Bias^2 + SE^2$$
$$E[(b_i - \theta)^2] = (E[b_i] - \theta)^2 + E(E[b_i] - b_i)^2$$

where θ is student ability, and bi is item difficulty. MSE is $E[(b_i - \theta)^2]$, bias is $(E[b_i] - \theta)$, and SE is the square root of $E(E[b_i] - b_i)^2$.

Additionally, the standard error of measure (SEM) quantile is used as an additional way to examine the interaction of the item-selection rules with the item pool. A higher SEM is an indication of a shallower pool for students within these abilities.

2.3.3. Item Exposure Rates

The exposure rate for each item was calculated as the percentage of students who received that item. Because different students receive different items based on blueprint constraints and their momentary ability estimates during a CAT administration, one indication of a deep pool and effective item selection is a low exposure rate. Once the item parameters are calibrated with appropriate sizes of samples, usually in the first year of an assessment program, a lower exposure rate is beneficial from a test security perspective.

2.3.4. Score Precision and Test Reliability

Score precision is estimated through multiple indicators: standard deviation (SD) of estimated thetas across students, mean SEM, and test reliability. The standard deviation of estimated thetas provides the basic information of the theta distribution produced by the CBE. As for the score precision, traditional reliability coefficients from classical test theory (CTT) are designed under the condition that students take the same test form, whereas in a CAT, students receive different items. Thus, CTT reliability is not available for a CAT. Instead, NWEA uses the marginal reliability coefficient (Samejima, 1994), which uses the item response theory's standard error of measurement (σ) and variance of estimated theta [(var(θ)] to estimate the reliability of student scores:

$$Marginal\ Reliability = \frac{var(\hat{\theta}) - \sigma^2}{var(\hat{\theta})}$$

Another method is analyzing the classification accuracy of the Maine scores. Classification accuracy helps to understand how effective the item pool is at differentiating students at the boundaries defined in standard setting. Higher accuracy implies that the item pool is rich enough to support the classification decisions. This is not absolute classification accuracy but another tool to use when evaluating the test design and item-pool interaction effects.

2.3.5. Item Sequence

Item sequence is defined as having each test start with summative items and then gradually add MAP Growth items. Field-test items are embedded in the test by avoiding the first and last few slots. The exception to this is items that are part of a set with a common reading passage or paired passages; the engine ensures these items are delivered as a group and not broken up.

3. Simulation Results

3.1. Test Blueprint Matching Rate

The tables below present the blueprint constraint accuracy rate for the summative blueprint by content area. The Maine Through Year Assessment reports instructional area scores using a combination of summative and MAP Growth diagnostic items. All summative items will be calibrated to the MAP Growth RIT scale so that summative items can be used for instructional area RIT score reporting along with the MAP Growth items.

Instructional areas are configured as constraints to enforce item selection according to state test blueprint and MAP Growth requirements. When determining the number of items for each instructional area, the state-approved blueprints are used to determine the summative portion of the test for each grade and content area, and MAP Growth requirements are used to determine the total test items. The MAP Growth rules are to have at least 4 MAP Growth items and at least 9 items (sum of summative and MAP Growth items) per instructional area. The following tables show the range of the target number of items configured to each instructional area and the range of the actual numbers of items administered across students. All tests have a 100% match to the test blueprint. Note that the blueprint matching rate uses only operational items.

Table 3.1. Summative Blueprint Constraint Accuracy Rate—Reading

Grade	Instructional Area	Summative (State Blueprint)					Total Test (MAP Growth RIT Scores)				
		Target		Actual		%	Target		Actual		%
		Min	Max	Min	Max	Match	Min	Max	Min	Max	Match
3	Literary Text	12	14	12	14	100	18	21	18	20	100
	Informational Text	8	9	8	9	100	12	14	12	13	100
	Vocabulary	5	7	5	7	100	8	11	9	11	100
4	Literary Text	11	12	11	12	100	17	18	17	18	100
	Informational Text	9	11	9	11	100	14	17	14	16	100
	Vocabulary	5	7	5	7	100	8	11	8	10	100
5	Literary Text	9	11	9	11	100	14	17	14	17	100
	Informational Text	9	11	9	11	100	14	17	14	16	100
	Vocabulary	5	7	5	7	100	8	11	9	11	100
6	Literary Text	9	11	9	11	100	14	17	14	16	100
	Informational Text	11	12	11	12	100	17	18	17	18	100
	Vocabulary	5	7	5	7	100	8	11	8	10	100
7	Literary Text	8	9	8	9	100	12	14	12	14	100
	Informational Text	12	14	12	14	100	18	21	18	20	100
	Vocabulary	5	7	5	7	100	8	11	8	11	100
8	Literary Text	8	9	8	9	100	12	14	12	14	100
	Informational Text	12	14	12	14	100	18	21	18	20	100
	Vocabulary	5	7	5	7	100	8	11	8	11	100
HS	Literary Text	8	9	8	8	100	12	14	12	13	100
	Informational Text	12	14	14	14	100	18	21	18	18	100
	Vocabulary	5	8	8	8	100	8	11	11	12	100

Table 3.2. Summative Blueprint Constraint Accuracy Rate—Mathematics

Grade	Instructional Area	Summative (State Blueprint)					Total Test (MAP Growth RIT Scores)				
		Target		Actual		%	Target		Actual		%
		Min	Max	Min	Max	Match	Min	Max	Min	Max	Match
3	Operations and Algebraic Thinking	6	6	6	6	100	10	10	10	10	100
	Numbers and Operations	9	9	9	9	100	13	13	13	13	100
	Measurement and Data	8	8	8	8	100	12	12	12	12	100
	Geometry	4	4	4	4	100	10	10	10	10	100
4	Operations and Algebraic Thinking	5	5	5	5	100	10	10	10	10	100
	Numbers and Operations	13	13	13	13	100	17	17	17	17	100
	Measurement and Data	5	5	5	5	100	9	9	9	9	100
	Geometry	4	4	4	4	100	9	9	9	9	100
5	Operations and Algebraic Thinking	4	4	4	4	100	9	9	9	9	100
	Numbers and Operations	14	14	14	14	100	18	18	18	18	100
	Measurement and Data	5	5	5	5	100	9	9	9	9	100
	Geometry	4	4	4	4	100	9	9	9	9	100
6	Operations and Algebraic Thinking	7	7	7	7	100	11	11	11	11	100
	The Real and Complex Number Systems	12	12	12	12	100	16	16	16	16	100
	Geometry	4	4	4	4	100	9	9	9	9	100
	Statistics and Probability	4	4	4	4	100	9	9	9	9	100
7	Operations and Algebraic Thinking	5	5	5	5	100	10	10	10	10	100
	The Real and Complex Number Systems	11	11	11	11	100	15	15	15	15	100
	Geometry	6	6	6	6	100	10	10	10	10	100
	Statistics and Probability	5	5	5	5	100	10	10	10	10	100
8	Operations and Algebraic Thinking	13	13	13	13	100	17	17	17	17	100
	The Real and Complex Number Systems	4	4	4	4	100	9	9	9	9	100
	Geometry	6	6	6	6	100	10	10	10	10	100
	Statistics and Probability	4	4	4	4	100	9	9	9	9	100
HS	Operations and Algebraic Thinking	14	14	14	14	100	18	18	18	18	100

Grade	Instructional Area	Summative (State Blueprint)					Total Test (MAP Growth RIT Scores)				
		Target		Actual		%	Target		Actual		%
		Min	Max	Min	Max	Match	Min	Max	Min	Max	Match
	The Real and Complex Number Systems	4	4	4	4	100	8	8	8	8	100
	Geometry	8	8	8	8	100	12	12	12	12	100
	Statistics and Probability	4	4	4	4	100	9	9	9	9	100

3.2. Accuracy of Ability Estimations

In the following tables, the mean of MSE, bias, SE, and correlations are presented.

Table 3.3. Accuracy of Ability Estimations—Reading

			Mean		
Grade	Score Type	Correlation	Bias	SE	MSE
3	Summative	0.97	0.01	0.41	0.17
	MAP Growth RIT Score	0.98	0.01	0.32	0.10
	Literary Text	0.95	-0.03	0.50	0.25
	Informational Text	0.93	0.02	0.60	0.36
	Vocabulary	0.91	0.07	0.72	0.53
4	Summative	0.97	0.00	0.39	0.15
	MAP Growth RIT Score	0.98	0.01	0.31	0.09
	Literary Text	0.95	0.03	0.49	0.25
	Informational Text	0.94	-0.02	0.54	0.29
	Vocabulary	0.91	0.06	0.70	0.50
5	Summative	0.97	-0.04	0.40	0.16
	MAP Growth RIT Score	0.98	-0.01	0.32	0.10
	Literary Text	0.95	0.01	0.53	0.28
	Informational Text	0.94	-0.04	0.55	0.30
	Vocabulary	0.90	0.03	0.74	0.55
6	Summative	0.96	-0.02	0.40	0.16
	MAP Growth RIT Score	0.97	0.00	0.33	0.11
	Literary Text	0.93	0.01	0.60	0.36
	Informational Text	0.94	-0.02	0.52	0.27
	Vocabulary	0.90	0.03	0.72	0.52
7	Summative	0.96	0.02	0.41	0.17
	MAP Growth RIT Score	0.98	-0.01	0.33	0.11
	Literary Text	0.93	-0.01	0.61	0.38
	Informational Text	0.95	0.01	0.51	0.26
	Vocabulary	0.90	-0.05	0.74	0.55
8	Summative	0.96	0.02	0.38	0.15
	MAP Growth RIT Score	0.98	0.01	0.32	0.10
	Literary Text	0.92	0.01	0.63	0.40

			Mean		
Grade	**Score Type**	**Correlation**	**Bias**	**SE**	**MSE**
	Informational Text	0.95	-0.01	0.48	0.23
	Vocabulary	0.91	0.05	0.67	0.45
HS	Summative	0.95	0.21	0.54	0.34
	MAP Growth RIT Score	0.97	0.02	0.39	0.15
	Literary Text	0.91	0.09	0.72	0.53
	Informational Text	0.93	0.05	0.61	0.38
	Vocabulary	0.87	0.06	0.83	0.70

Table 3.4. Accuracy of Ability Estimations—Mathematics

			Mean		
Grade	**Score Type**	**Correlation**	**Bias**	**SE**	**MSE**
3	Summative	0.96	0.00	0.38	0.14
	MAP Growth RIT Score	0.98	0.02	0.29	0.08
	Operations and Algebraic Thinking	0.88	0.02	0.69	0.48
	Numbers and Operations	0.92	0.01	0.55	0.30
	Measurement and Data	0.91	0.04	0.58	0.34
	Geometry	0.89	0.01	0.66	0.44
4	Summative	0.98	-0.04	0.39	0.16
	MAP Growth RIT Score	0.98	0.03	0.29	0.09
	Operations and Algebraic Thinking	0.93	0.04	0.65	0.43
	Numbers and Operations	0.96	0.05	0.49	0.24
	Measurement and Data	0.92	0.00	0.72	0.51
	Geometry	0.91	0.01	0.72	0.51
5	Summative	0.97	-0.02	0.39	0.15
	MAP Growth RIT Score	0.98	-0.01	0.31	0.09
	Operations and Algebraic Thinking	0.92	-0.06	0.71	0.51
	Numbers and Operations	0.96	0.01	0.49	0.24
	Measurement and Data	0.92	-0.04	0.71	0.51
	Geometry	0.91	0.02	0.77	0.60
6	Summative	0.97	-0.03	0.37	0.14
	MAP Growth RIT Score	0.98	0.01	0.29	0.08
	Operations and Algebraic Thinking	0.92	0.01	0.63	0.39
	The Real and Complex Number Systems	0.95	0.02	0.50	0.25
	Geometry	0.90	0.01	0.74	0.55
	Statistics and Probability	0.90	-0.05	0.71	0.50
7	Summative	0.98	-0.02	0.38	0.14
	MAP Growth RIT Score	0.98	0.02	0.29	0.08
	Operations and Algebraic Thinking	0.92	0.01	0.70	0.49
	The Real and Complex Number Systems	0.96	0.03	0.50	0.25
	Geometry	0.92	0.04	0.71	0.51
	Statistics and Probability	0.92	-0.02	0.67	0.45

			Mean		
Grade	**Score Type**	**Correlation**	**Bias**	**SE**	**MSE**
8	Summative	0.97	-0.01	0.38	0.14
	MAP Growth RIT Score	0.98	0.00	0.31	0.09
	Operations and Algebraic Thinking	0.96	-0.01	0.52	0.27
	The Real and Complex Number Systems	0.92	0.03	0.74	0.55
	Geometry	0.93	-0.02	0.71	0.51
	Statistics and Probability	0.92	0.00	0.71	0.51
HS	Summative	0.96	-0.02	0.63	0.40
	MAP Growth RIT Score	0.98	-0.07	0.37	0.14
	Operations and Algebraic Thinking	0.95	-0.10	0.60	0.37
	The Real and Complex Number Systems	0.88	-0.15	0.96	0.95
	Geometry	0.93	-0.07	0.68	0.47
	Statistics and Probability	0.92	-0.10	0.78	0.62

3.2.1. Average SEM by Quantile

This table provides a comparison of the SEMs in the simulations across the population. It is used as an additional way to examine the interaction of the item-selection rules with the item pool. A higher SEM is an indication of a shallower pool for students within these abilities. For example, summative and MAP Growth RIT scores have consistently lower SEMs across different percentiles. Some of the instructional areas, such as Vocabulary, have a higher SEM, especially at the 95 percentiles.

Table 3.5. SEM Distribution of Summative Scores—Reading

			Percentile			
Grade	**Score Type**	**Overall**	**5**	**25**	**75**	**95**
3	Summative Only	0.40	0.30	0.40	0.40	0.50
	MAP Growth RIT Score	0.31	0.29	0.30	0.33	0.37
	Literary Text	0.46	0.42	0.44	0.51	0.62
	Informational Text	0.57	0.51	0.55	0.60	0.69
	Vocabulary	0.67	0.57	0.61	0.73	0.90
4	Summative Only	0.40	0.30	0.40	0.40	0.50
	MAP Growth RIT Score	0.31	0.29	0.30	0.32	0.37
	Literary Text	0.48	0.44	0.46	0.51	0.60
	Informational Text	0.53	0.47	0.51	0.56	0.64
	Vocabulary	0.68	0.56	0.61	0.76	1.07
5	Summative Only	0.40	0.30	0.40	0.40	0.50
	MAP Growth RIT Score	0.31	0.29	0.30	0.32	0.37
	Literary Text	0.51	0.46	0.49	0.54	0.64
	Informational Text	0.51	0.45	0.48	0.56	0.66
	Vocabulary	0.67	0.57	0.61	0.72	0.84
6	Summative Only	0.40	0.30	0.40	0.40	0.50
	MAP Growth RIT Score	0.31	0.29	0.30	0.32	0.38
	Literary Text	0.54	0.48	0.51	0.57	0.69

			Percentile			
Grade	**Score Type**	**Overall**	**5**	**25**	**75**	**95**
	Informational Text	0.49	0.44	0.47	0.53	0.63
	Vocabulary	0.66	0.58	0.62	0.72	1.01
7	Summative Only	0.40	0.40	0.40	0.40	0.50
	MAP Growth RIT Score	0.31	0.30	0.31	0.33	0.37
	Literary Text	0.58	0.54	0.56	0.62	0.77
	Informational Text	0.47	0.45	0.46	0.49	0.58
	Vocabulary	0.66	0.58	0.62	0.70	0.84
8	Summative Only	0.40	0.30	0.40	0.40	0.50
	MAP Growth RIT Score	0.31	0.29	0.30	0.32	0.36
	Literary Text	0.58	0.51	0.55	0.62	0.79
	Informational Text	0.46	0.43	0.45	0.48	0.56
	Vocabulary	0.63	0.57	0.60	0.68	0.80
HS	Summative Only	0.40	0.30	0.30	0.50	1.00
	MAP Growth RIT Score	0.33	0.29	0.30	0.39	0.51
	Literary Text	0.61	0.51	0.53	0.73	1.12
	Informational Text	0.48	0.43	0.45	0.59	0.82
	Vocabulary	0.76	0.55	0.66	0.86	1.11

Table 3.6. SEM Distribution of Summative Scores—Mathematics

			Percentile			
Grade	**Score Type**	**Overall**	**5**	**25**	**75**	**95**
3	Summative Only	0.40	0.30	0.40	0.40	0.40
	MAP Growth RIT Score	0.29	0.28	0.28	0.30	0.31
	Operations and Algebraic Thinking	0.65	0.58	0.61	0.67	0.79
	Number and Operations	0.54	0.51	0.52	0.58	0.63
	Measurement and Data	0.59	0.55	0.56	0.61	0.67
	Geometry	0.64	0.59	0.61	0.68	0.79
4	Summative Only	0.40	0.30	0.40	0.40	0.40
	MAP Growth RIT Score	0.29	0.28	0.29	0.30	0.31
	Operations and Algebraic Thinking	0.64	0.58	0.60	0.67	0.79
	Number and Operations	0.48	0.45	0.46	0.50	0.53
	Measurement and Data	0.68	0.63	0.64	0.71	0.83
	Geometry	0.67	0.61	0.64	0.70	0.81
5	Summative Only	0.40	0.40	0.40	0.40	0.40
	MAP Growth RIT Score	0.29	0.28	0.29	0.30	0.31
	Operations and Algebraic Thinking	0.68	0.62	0.66	0.71	0.81
	Number and Operations	0.48	0.43	0.46	0.50	0.54
	Measurement and Data	0.67	0.56	0.63	0.71	0.83
	Geometry	0.68	0.62	0.67	0.72	0.82
6	Summative Only	0.40	0.30	0.40	0.40	0.40
	MAP Growth RIT Score	0.29	0.28	0.29	0.30	0.31

			Percentile			
Grade	**Score Type**	**Overall**	**5**	**25**	**75**	**95**
	Operations and Algebraic Thinking	0.62	0.57	0.61	0.64	0.78
	Number and Operations	0.49	0.45	0.47	0.51	0.55
	Measurement and Data	0.68	0.63	0.65	0.72	0.83
	Geometry	0.66	0.61	0.63	0.69	0.81
7	Summative Only	0.40	0.40	0.40	0.40	0.40
	MAP Growth RIT Score	0.30	0.28	0.29	0.30	0.32
	Operations and Algebraic Thinking	0.64	0.59	0.62	0.68	0.79
	Number and Operations	0.52	0.47	0.51	0.54	0.60
	Measurement and Data	0.66	0.63	0.64	0.69	0.80
	Geometry	0.63	0.60	0.61	0.68	0.79
8	Summative Only	0.40	0.40	0.40	0.40	0.40
	MAP Growth RIT Score	0.30	0.28	0.29	0.30	0.32
	Operations and Algebraic Thinking	0.49	0.46	0.48	0.51	0.56
	Number and Operations	0.67	0.62	0.64	0.72	0.84
	Measurement and Data	0.66	0.62	0.64	0.70	0.80
	Geometry	0.66	0.62	0.64	0.70	0.81
HS	Summative Only	0.40	0.40	0.40	0.40	0.70
	MAP Growth RIT Score	0.30	0.29	0.30	0.33	0.43
	Operations and Algebraic Thinking	0.50	0.46	0.48	0.56	0.84
	Number and Operations	0.85	0.73	0.77	0.98	1.19
	Measurement and Data	0.60	0.55	0.58	0.63	0.86
	Geometry	0.70	0.64	0.66	0.80	1.07

3.3. Item Exposure Rates

The table below shows a summary of item exposure rates by item type (summative, MAP Growth, and field-test items) and 6 exposure-rate categories: 0–20%, 21–40%, 41–60%, 61–80%, 81–99%, and 100%. Note that the first category (0–20%) begins when at least one student receives the items. Because one student out of 1,000 (0.001) is rounded to 0, the range is presented as 0–20%.

Additionally, because the summative test allows items from adjacent grades and MAP Growth allows a wide grade band, as explained in Section 2.2. Item Pool Characteristics, the number of items used can be slightly higher than the item bank size.

Table 3.7. Item Exposure Rate by Item Status—Reading

				Item Exposure Rate, N					
Grade	**Item Type**	**Item Bank Size**	**# of Items Used**	**0–20%**	**21–40%**	**41–60%**	**61–80%**	**81–99%**	**100%**
3	Summative	390	232	181	37	9	3	0	2
	MAP Growth	539	728	728	0	0	0	0	0
	Field Test	5	5	0	0	0	0	0	5
4	Summative	367	326	284	30	5	4	1	2

				Item Exposure Rate, N					
Grade	**Item Type**	**Item Bank Size**	**# of Items Used**	**0–20%**	**21–40%**	**41–60%**	**61–80%**	**81–99%**	**100%**
	MAP Growth	664	627	627	0	0	0	0	0
	Field Test	6	5	0	0	0	0	0	5
5	Summative	349	304	264	22	13	3	0	2
	MAP Growth	492	515	513	2	0	0	0	0
	Field Test	7	5	0	0	0	0	0	5
6	Summative	342	308	275	18	7	4	4	0
	MAP Growth	635	553	553	0	0	0	0	0
	Field Test	7	49	42	2	3	0	0	2
7	Summative	362	351	318	23	3	3	2	2
	MAP Growth	342	493	493	0	0	0	0	0
	Field Test	6	5	0	0	0	0	0	5
8	Summative	414	324	276	38	4	4	0	2
	MAP Growth	446	575	575	0	0	0	0	0
	Field Test	14	12	5	0	4	1	2	0
HS	Summative	48	47	10	6	4	3	10	14
	MAP Growth	278	497	497	0	0	0	0	0
	Field Test	116	59	59	0	0	0	0	0

Table 3.8. Item Exposure Rate by Item Status—Mathematics

				Item Exposure Rate, N					
Grade	**Item Type**	**Item Bank Size**	**# of Items Used**	**0–20%**	**21–40%**	**41–60%**	**61–80%**	**81–99%**	**100%**
3	Summative	449	439	418	16	2	0	1	2
	MAP Growth	107	337	328	9	0	0	0	0
	Field Test	6	6	0	0	0	4	1	1
4	Summative	395	437	421	11	3	0	0	2
	MAP Growth	107	433	429	4	0	0	0	0
	Field Test	11	11	0	0	11	0	0	0
5	Summative	387	446	431	11	2	0	0	2
	MAP Growth	110	449	448	1	0	0	0	0
	Field Test	5	5	0	0	0	0	0	5
6	Summative	373	388	371	11	2	2	0	2
	MAP Growth	109	398	387	11	0	0	0	0
	Field Test	6	18	0	18	0	0	0	0
7	Summative	351	399	380	14	3	0	0	2

				Item Exposure Rate, N					
Grade	**Item Type**	**Item Bank Size**	**# of Items Used**	**0–20%**	**21–40%**	**41–60%**	**61–80%**	**81–99%**	**100%**
	MAP Growth	98	493	484	9	0	0	0	0
	Field Test	7	19	0	19	0	0	0	0
8	Summative	399	414	396	12	4	0	0	2
	MAP Growth	119	527	526	1	0	0	0	0
	Field Test	6	13	0	13	0	0	0	0
HS	Summative	63	63	11	21	5	18	2	6
	MAP Growth	136	524	512	12	0	0	0	0
	Field Test	145	145	145	0	0	0	0	0

3.3.1. Field-Test Items

Field-test items are embedded in the Spring 2024 test for possible operational use in future test administrations. After evaluating Maine demographic distributions, it has been determined to assign field-test items by gender and ethnicity. The adaptive test set a minimum of 250 students each for male and female subgroups as a guideline. The table below summarizes the numbers of students taking field-test items.

Table 3.9. Gender Assignment Results for Field-Test Items—Reading

		Mean		Female, %		Male, %	
Grade	**# of FT Items**	**Female, %**	**Male, %**	**Min**	**Max**	**Min**	**Max**
3	5	50	50	50	50	50	50
4	5	50	50	50	50	50	50
5	5	49	51	49	51	49	51
6	49	47	53	39	47	53	61
7	5	48	52	48	52	48	52
8	12	51	54	43	50	100	57
HS	59	49	51	45	44	56	55

Table 3.10. Gender Assignment Results for Field-Test Items—Mathematics

		Mean		Female, %		Male, %	
Grade	**# of FT Items**	**Female, %**	**Male, %**	**Min**	**Max**	**Min**	**Max**
3	6	50	50	50	50	50	50
4	11	49	51	49	51	49	51
5	5	50	50	50	50	50	50
6	18	50	50	50	50	50	50
7	19	49	51	49	50	50	51
8	13	45	55	45	55	45	55
HS	145	51	49	49	47	53	51

Table 3.11. Race/Ethnicity Assignment Results for Field-Test Items—Reading

		Mean%						
Grade	# of FT Items	American Indian	Asian	African American	Hispanic	Native Hawaiian or Pacific Islander	Caucasian	Two or More Races
3	5	5	1	1	2	0	87	4
4	5	5	1	1	2	0	87	4
5	5	3	1	1	4	0	88	4
6	49	7	4	3	7	1	85	7
7	5	3	1	1	3	0	87	4
8	12	3	1	1	4	0	90	3
HS	59	3	1	2	4	0	87	4

Table 3.12. Race/Ethnicity Assignment Results for Field-Test Items—Mathematics

		Mean%						
Grade	# of FT Items	American Indian	Asian	African American	Hispanic	Native Hawaiian or Pacific Islander	Caucasian	Two or More Races
3	6	3	1	1	3	0	88	4
4	11	3	1	1	4	0	87	4
5	5	4	1	1	3	0	85	6
6	18	4	1	1	3	0	85	6
7	19	3	1	1	3	0	89	2
8	13	3	1	1	3	0	88	4
HS	145	4	2	3	5	2	86	4

3.4. Score Precision and Test Reliability

Score precision is estimated through multiple indicators: standard deviation (SD) of estimated thetas across students, mean standard error of measure (SEM) associated with thetas, and reliability.

Table 3.13. Score Precision—Reading

Grade	Score Type	Average # Items	SD of Theta	Mean SEM	Reliability
3	Summative	27	1.68	0.41	0.94
	MAP Growth RIT Score	41	1.58	0.32	0.96
	Literary Text	19	1.65	0.49	0.91
	Informational Text	12	1.68	0.58	0.88
	Vocabulary	10	1.74	0.70	0.83
4	Summative	27	1.53	0.40	0.93
	MAP Growth RIT Score	41	1.51	0.31	0.96
	Literary Text	17	1.58	0.50	0.90
	Informational Text	14	1.57	0.54	0.88
	Vocabulary	10	1.68	0.71	0.81
5	Summative	27	1.54	0.41	0.93

Grade	Score Type	Average # Items	SD of Theta	Mean SEM	Reliability
	MAP Growth RIT Score	41	1.55	0.32	0.96
	Literary Text	16	1.65	0.53	0.89
	Informational Text	15	1.64	0.53	0.89
	Vocabulary	10	1.67	0.69	0.83
6	Summative	27	1.45	0.41	0.92
	MAP Growth RIT Score	41	1.48	0.32	0.95
	Literary Text	15	1.59	0.55	0.88
	Informational Text	17	1.54	0.51	0.89
	Vocabulary	9	1.66	0.70	0.81
7	Summative	27	1.44	0.42	0.91
	MAP Growth RIT Score	41	1.53	0.32	0.96
	Literary Text	13	1.63	0.61	0.86
	Informational Text	18	1.60	0.49	0.90
	Vocabulary	10	1.69	0.69	0.83
8	Summative	27	1.42	0.40	0.92
	MAP Growth RIT Score	41	1.51	0.31	0.96
	Literary Text	13	1.64	0.61	0.85
	Informational Text	18	1.58	0.48	0.91
	Vocabulary	10	1.62	0.66	0.83
HS	Summative	30	1.53	0.48	0.87
	MAP Growth RIT Score	42	1.56	0.36	0.94
	Literary Text	13	1.75	0.69	0.81
	Informational Text	18	1.65	0.56	0.87
	Vocabulary	11	1.71	0.79	0.77

Table 3.14. Score Precision—Mathematics

Grade	Score Type	Average # Items	SD of Theta	Mean SEM	Reliability
3	Summative	27	1.36	0.39	0.92
	MAP Growth RIT Score	45	1.32	0.29	0.95
	Operations and Algebraic Thinking	10	1.45	0.66	0.79
	Numbers and Operations	13	1.39	0.55	0.84
	Measurement and Data	12	1.44	0.59	0.83
	Geometry	10	1.47	0.66	0.80
4	Summative	27	1.82	0.39	0.95
	MAP Growth RIT Score	45	1.64	0.29	0.97
	Operations and Algebraic Thinking	10	1.74	0.65	0.86
	Numbers and Operations	17	1.69	0.49	0.92
	Measurement and Data	9	1.79	0.70	0.84
	Geometry	9	1.77	0.69	0.85
5	Summative	27	1.73	0.40	0.95
	MAP Growth RIT Score	45	1.73	0.30	0.97
	Operations and Algebraic Thinking	9	1.84	0.70	0.85

Grade	Score Type	Average # Items	SD of Theta	Mean SEM	Reliability
	Numbers and Operations	18	1.77	0.48	0.93
	Measurement and Data	9	1.87	0.69	0.86
	Geometry	9	1.88	0.72	0.85
6	Summative	27	1.52	0.40	0.93
	MAP Growth RIT Score	45	1.49	0.29	0.96
	Operations and Algebraic Thinking	11	1.58	0.63	0.84
	The Real and Complex Number Systems	16	1.54	0.49	0.90
	Geometry	9	1.68	0.71	0.82
	Statistics and Probability	9	1.63	0.68	0.82
7	Summative	27	1.71	0.40	0.94
	MAP Growth RIT Score	45	1.65	0.30	0.97
	Operations and Algebraic Thinking	10	1.77	0.66	0.86
	The Real and Complex Number Systems	15	1.72	0.53	0.91
	Geometry	10	1.80	0.68	0.86
	Statistics and Probability	10	1.76	0.66	0.86
8	Summative	27	1.59	0.40	0.94
	MAP Growth RIT Score	45	1.76	0.30	0.97
	Operations and Algebraic Thinking	17	1.81	0.50	0.92
	The Real and Complex Number Systems	9	1.95	0.70	0.87
	Geometry	10	1.89	0.68	0.87
	Statistics and Probability	9	1.88	0.69	0.86
HS	Summative	30	1.60	0.46	0.90
	MAP Growth RIT Score	47	1.75	0.32	0.97
	Operations and Algebraic Thinking	18	1.82	0.55	0.90
	The Real and Complex Number Systems	8	2.06	0.93	0.78
	Geometry	12	1.84	0.64	0.88
	Statistics and Probability	9	1.95	0.75	0.85

3.4.1. Classification Accuracy

Classification accuracy of the Maine summative scale scores measures the accuracy of achievement level categorizations. The “proficiency match” is the number of simulees correctly classified as proficient according to their true underlying ability (each grade and subject has 1,000 simulees).

The Maine summative scale has four achievement level categorizations. Achievement level classification accuracy is divided into three categories: (1) exact match, (2) adjacent match, and (3) no match. An exact match implies that a student was correctly classified into the correct achievement level according to their true underlying ability; an adjacent match implies that a student was classified into an achievement level one level below or above the correct achievement level according to their true underlying ability; otherwise, there is no match

between the observed achievement level classification and the expected classification based on the student's true underlying ability.

Table 3.15. Accuracy by Classification Type—Reading

Classification Type	Count
Grade 3	
Proficiency Match	939
Achievement Level	
Exact Match	838
Adjacent Match	162
No Match	0
Grade 4	
Proficiency Match	937
Achievement Level	
Exact Match	836
Adjacent Match	164
No Match	0
Grade 5	
Proficiency Match	946
Achievement Level	
Exact Match	860
Adjacent Match	140
No Match	0
Grade 6	
Proficiency Match	937
Achievement Level	
Exact Match	850
Adjacent Match	150
No Match	0
Grade 7	
Proficiency Match	921
Achievement Level	
Exact Match	834
Adjacent Match	165
No Match	1
Grade 8	
Proficiency Match	919
Achievement Level	
Exact Match	837
Adjacent Match	163
No Match	0
High School	
Proficiency Match	974

Classification Type	Count
Achievement Level	
Exact Match	830
Adjacent Match	170
No Match	0

Table 3.16. Accuracy by Classification Type—Mathematics

Classification Type	Count
Grade 3	
Proficiency Match	939
Achievement Level	
Exact Match	833
Adjacent Match	166
No Match	1
Grade 4	
Proficiency Match	930
Achievement Level	
Exact Match	856
Adjacent Match	144
No Match	0
Grade 5	
Proficiency Match	929
Achievement Level	
Exact Match	847
Adjacent Match	153
No Match	0
Grade 6	
Proficiency Match	912
Achievement Level	
Exact Match	839
Adjacent Match	160
No Match	1
Grade 7	
Proficiency Match	916
Achievement Level	
Exact Match	853
Adjacent Match	147
No Match	0
Grade 8	
Proficiency Match	925
Achievement Level	
Exact Match	841
Adjacent Match	159

Classification Type	Count
No Match	0
High School	
Proficiency Match	924
Achievement Level	
Exact Match	775
Adjacent Match	224
No Match	1

3.5. Item Sequence

When defining item positions, the plan was to start with summative items, gradually add MAP Growth items, and end with MAP Growth items. Field-test items were embedded across the test. The results show that math items were assigned positions according to the design. Due to constraints by passages and the number of items available in each passage, reading item positions shifted a little.

Table 3.17. Item Sequence—Reading, Grade 3

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	1000	0	0	1000
5	1000	0	0	1000
6	1000	0	0	1000
7	999	0	1	1000
8	986	0	14	1000
9	869	0	131	1000
10	649	0	351	1000
11	511	0	489	1000
12	968	29	3	1000
13	760	198	42	1000
14	556	407	37	1000
15	61	939	0	1000
16	997	3	0	1000
17	972	3	25	1000
18	918	0	82	1000
19	631	0	369	1000
20	897	0	103	1000
21	971	0	29	1000
22	945	0	55	1000
23	814	0	186	1000
24	825	0	175	1000
25	831	0	169	1000
26	273	22	705	1000

Item Sequence	Summative	MAP	Field Test	Total
27	375	46	579	1000
28	573	176	251	1000
29	474	290	236	1000
30	132	817	51	1000
31	24	974	2	1000
32	0	1000	0	1000
33	0	1000	0	1000
34	1000	0	0	1000
35	951	0	49	1000
36	890	0	110	1000
37	498	0	502	1000
38	217	689	94	1000
39	183	761	56	1000
40	151	820	29	1000
41	94	832	74	1000
42	3	996	1	1000
43	2	998	0	1000
44	0	1000	0	1000
45	0	1000	0	1000
46	0	1000	0	1000

Table 3.18. Item Sequence—Reading, Grade 4

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	1000	0	0	1000
5	1000	0	0	1000
6	1000	0	0	1000
7	1000	0	0	1000
8	918	0	82	1000
9	912	0	88	1000
10	717	0	283	1000
11	674	0	326	1000
12	827	173	0	1000
13	325	565	110	1000
14	215	767	18	1000
15	56	944	0	1000
16	1000	0	0	1000
17	1000	0	0	1000
18	900	0	100	1000
19	791	0	209	1000

Item Sequence	Summative	MAP	Field Test	Total
20	958	0	42	1000
21	995	0	5	1000
22	923	0	77	1000
23	839	0	161	1000
24	918	0	82	1000
25	958	0	42	1000
26	377	5	618	1000
27	547	5	448	1000
28	476	475	49	1000
29	233	723	44	1000
30	142	830	28	1000
31	11	989	0	1000
32	3	997	0	1000
33	1	999	0	1000
34	996	4	0	1000
35	997	3	0	1000
36	614	2	384	1000
37	287	1	712	1000
38	671	314	15	1000
39	674	324	2	1000
40	34	447	519	1000
41	8	453	539	1000
42	1	983	16	1000
43	1	999	0	1000
44	1	999	0	1000
45	0	999	1	1000
46	0	1000	0	1000

Table 3.19. Item Sequence—Reading, Grade 5

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	1000	0	0	1000
5	1000	0	0	1000
6	1000	0	0	1000
7	1000	0	0	1000
8	1000	0	0	1000
9	1000	0	0	1000
10	825	0	175	1000
11	889	0	111	1000
12	666	334	0	1000

Item Sequence	Summative	MAP	Field Test	Total
13	64	936	0	1000
14	25	940	35	1000
15	5	975	20	1000
16	999	1	0	1000
17	999	1	0	1000
18	926	1	73	1000
19	911	0	89	1000
20	991	0	9	1000
21	999	0	1	1000
22	804	0	196	1000
23	714	0	286	1000
24	942	0	58	1000
25	996	0	4	1000
26	293	0	707	1000
27	356	0	644	1000
28	347	653	0	1000
29	34	966	0	1000
30	2	997	1	1000
31	1	998	1	1000
32	0	1000	0	1000
33	0	1000	0	1000
34	998	2	0	1000
35	998	2	0	1000
36	754	1	245	1000
37	247	0	753	1000
38	919	46	35	1000
39	936	63	1	1000
40	273	64	663	1000
41	83	65	852	1000
42	4	955	41	1000
43	0	1000	0	1000
44	0	1000	0	1000
45	0	1000	0	1000
46	0	1000	0	1000

Table 3.20. Item Sequence—Reading, Grade 6

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	1000	0	0	1000
5	1000	0	0	1000

Item Sequence	Summative	MAP	Field Test	Total
6	1000	0	0	1000
7	1000	0	0	1000
8	817	0	183	1000
9	828	0	172	1000
10	582	0	418	1000
11	711	0	289	1000
12	671	153	176	1000
13	452	373	175	1000
14	230	739	31	1000
15	71	918	11	1000
16	968	0	32	1000
17	980	0	20	1000
18	992	0	8	1000
19	894	0	106	1000
20	951	0	49	1000
21	915	0	85	1000
22	879	0	121	1000
23	852	0	148	1000
24	885	0	115	1000
25	767	0	233	1000
26	452	0	548	1000
27	605	18	377	1000
28	704	251	45	1000
29	513	417	70	1000
30	232	681	87	1000
31	66	886	48	1000
32	22	966	12	1000
33	0	991	9	1000
34	936	2	62	1000
35	928	64	8	1000
36	921	64	15	1000
37	493	64	443	1000
38	257	264	479	1000
39	216	629	155	1000
40	177	727	96	1000
41	32	817	151	1000
42	1	976	23	1000
43	0	1000	0	1000
44	0	1000	0	1000
45	0	1000	0	1000
46	0	1000	0	1000

Table 3.21. Item Sequence—Reading, Grade 7

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	304	0	696	1000
5	696	0	304	1000
6	1000	0	0	1000
7	1000	0	0	1000
8	1000	0	0	1000
9	1000	0	0	1000
10	512	0	488	1000
11	287	0	713	1000
12	946	54	0	1000
13	126	172	702	1000
14	134	633	233	1000
15	4	994	2	1000
16	999	0	1	1000
17	1000	0	0	1000
18	999	0	1	1000
19	999	0	1	1000
20	999	0	1	1000
21	1000	0	0	1000
22	1000	0	0	1000
23	986	0	14	1000
24	986	0	14	1000
25	971	0	29	1000
26	555	0	445	1000
27	516	0	484	1000
28	861	122	17	1000
29	727	176	97	1000
30	368	614	18	1000
31	0	998	2	1000
32	0	1000	0	1000
33	0	1000	0	1000
34	1000	0	0	1000
35	865	0	135	1000
36	860	0	140	1000
37	712	0	288	1000
38	242	753	5	1000
39	170	830	0	1000
40	170	830	0	1000
41	6	830	164	1000

Item Sequence	Summative	MAP	Field Test	Total
42	0	994	6	1000
43	0	1000	0	1000
44	0	1000	0	1000
45	0	1000	0	1000
46	0	1000	0	1000

Table 3.22. Item Sequence—Reading, Grade 8

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	1000	0	0	1000
5	1000	0	0	1000
6	1000	0	0	1000
7	1000	0	0	1000
8	1000	0	0	1000
9	1000	0	0	1000
10	790	0	210	1000
11	962	0	38	1000
12	1000	0	0	1000
13	426	535	39	1000
14	65	782	153	1000
15	2	818	180	1000
16	922	0	78	1000
17	991	0	9	1000
18	994	0	6	1000
19	870	0	130	1000
20	913	0	87	1000
21	932	0	68	1000
22	967	0	33	1000
23	891	0	109	1000
24	761	0	239	1000
25	796	0	204	1000
26	213	4	783	1000
27	722	10	268	1000
28	848	57	95	1000
29	279	291	430	1000
30	57	462	481	1000
31	18	680	302	1000
32	0	1000	0	1000
33	0	1000	0	1000
34	1000	0	0	1000

Item Sequence	Summative	MAP	Field Test	Total
35	895	0	105	1000
36	983	0	17	1000
37	537	0	463	1000
38	85	647	268	1000
39	29	797	174	1000
40	26	952	22	1000
41	19	972	9	1000
42	7	993	0	1000
43	0	1000	0	1000
44	0	1000	0	1000
45	0	1000	0	1000
46	0	1000	0	1000

Table 3.23. Item Sequence—Reading, High School

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	1000	0	0	1000
5	1000	0	0	1000
6	1000	0	0	1000
7	1000	0	0	1000
8	1000	0	0	1000
9	1000	0	0	1000
10	1000	0	0	1000
11	0	0	1000	1000
12	0	0	1000	1000
13	0	0	1000	1000
14	0	0	1000	1000
15	0	0	1000	1000
16	0	0	1000	1000
17	0	0	1000	1000
18	1000	0	0	1000
19	1000	0	0	1000
20	1000	0	0	1000
21	1000	0	0	1000
22	1000	0	0	1000
23	1000	0	0	1000
24	1000	0	0	1000
25	1000	0	0	1000
26	941	59	0	1000
27	893	107	0	1000

Item Sequence	Summative	MAP	Field Test	Total
28	313	687	0	1000
29	152	848	0	1000
30	0	1000	0	1000
31	0	1000	0	1000
32	0	1000	0	1000
33	0	1000	0	1000
34	992	8	0	1000
35	992	8	0	1000
36	999	1	0	1000
37	1000	0	0	1000
38	1000	0	0	1000
39	1000	0	0	1000
40	1000	0	0	1000
41	1000	0	0	1000
42	848	152	0	1000
43	689	311	0	1000
44	113	887	0	1000
45	65	935	0	1000
46	3	997	0	1000
47	0	1000	0	1000
48	0	1000	0	1000
49	0	1000	0	1000

Table 3.24. Item Sequence—Mathematics, Grade 3

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	1000	0	0	1000
5	1000	0	0	1000
6	1000	0	0	1000
7	1000	0	0	1000
8	1000	0	0	1000
9	1000	0	0	1000
10	0	0	1000	1000
11	0	0	1000	1000
12	0	1000	0	1000
13	0	1000	0	1000
14	0	1000	0	1000
15	0	1000	0	1000
16	1000	0	0	1000
17	1000	0	0	1000

Item Sequence	Summative	MAP	Field Test	Total
18	1000	0	0	1000
19	1000	0	0	1000
20	1000	0	0	1000
21	1000	0	0	1000
22	1000	0	0	1000
23	1000	0	0	1000
24	1000	0	0	1000
25	1000	0	0	1000
26	0	0	1000	1000
27	0	0	1000	1000
28	0	1000	0	1000
29	0	1000	0	1000
30	0	1000	0	1000
31	0	1000	0	1000
32	0	1000	0	1000
33	0	1000	0	1000
34	1000	0	0	1000
35	1000	0	0	1000
36	1000	0	0	1000
37	1000	0	0	1000
38	1000	0	0	1000
39	1000	0	0	1000
40	1000	0	0	1000
41	1000	0	0	1000
42	0	0	1000	1000
43	0	1000	0	1000
44	0	1000	0	1000
45	0	1000	0	1000
46	0	1000	0	1000
47	0	1000	0	1000
48	0	1000	0	1000
49	0	1000	0	1000
50	0	1000	0	1000

Table 3.25. Item Sequence—Mathematics, Grade 4

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	1000	0	0	1000
5	1000	0	0	1000
6	1000	0	0	1000

Item Sequence	Summative	MAP	Field Test	Total
7	1000	0	0	1000
8	1000	0	0	1000
9	1000	0	0	1000
10	0	0	1000	1000
11	0	0	1000	1000
12	0	1000	0	1000
13	0	1000	0	1000
14	0	1000	0	1000
15	0	1000	0	1000
16	1000	0	0	1000
17	1000	0	0	1000
18	1000	0	0	1000
19	1000	0	0	1000
20	1000	0	0	1000
21	1000	0	0	1000
22	1000	0	0	1000
23	1000	0	0	1000
24	1000	0	0	1000
25	1000	0	0	1000
26	0	0	1000	1000
27	0	0	1000	1000
28	0	1000	0	1000
29	0	1000	0	1000
30	0	1000	0	1000
31	0	1000	0	1000
32	0	1000	0	1000
33	0	1000	0	1000
34	1000	0	0	1000
35	1000	0	0	1000
36	1000	0	0	1000
37	1000	0	0	1000
38	1000	0	0	1000
39	1000	0	0	1000
40	1000	0	0	1000
41	1000	0	0	1000
42	0	0	1000	1000
43	0	1000	0	1000
44	0	1000	0	1000
45	0	1000	0	1000
46	0	1000	0	1000
47	0	1000	0	1000
48	0	1000	0	1000

Item Sequence	Summative	MAP	Field Test	Total
49	0	1000	0	1000
50	0	1000	0	1000

Table 3.26. Item Sequence—Mathematics, Grade 5

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	1000	0	0	1000
5	1000	0	0	1000
6	1000	0	0	1000
7	1000	0	0	1000
8	1000	0	0	1000
9	1000	0	0	1000
10	0	0	1000	1000
11	0	0	1000	1000
12	0	1000	0	1000
13	0	1000	0	1000
14	0	1000	0	1000
15	0	1000	0	1000
16	1000	0	0	1000
17	1000	0	0	1000
18	1000	0	0	1000
19	1000	0	0	1000
20	1000	0	0	1000
21	1000	0	0	1000
22	1000	0	0	1000
23	1000	0	0	1000
24	1000	0	0	1000
25	1000	0	0	1000
26	0	0	1000	1000
27	0	0	1000	1000
28	0	1000	0	1000
29	0	1000	0	1000
30	0	1000	0	1000
31	0	1000	0	1000
32	0	1000	0	1000
33	0	1000	0	1000
34	1000	0	0	1000
35	1000	0	0	1000
36	1000	0	0	1000
37	1000	0	0	1000

Item Sequence	Summative	MAP	Field Test	Total
38	1000	0	0	1000
39	1000	0	0	1000
40	1000	0	0	1000
41	1000	0	0	1000
42	0	0	1000	1000
43	0	1000	0	1000
44	0	1000	0	1000
45	0	1000	0	1000
46	0	1000	0	1000
47	0	1000	0	1000
48	0	1000	0	1000
49	0	1000	0	1000
50	0	1000	0	1000

Table 3.27. Item Sequence—Mathematics, Grade 6

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	1000	0	0	1000
5	1000	0	0	1000
6	1000	0	0	1000
7	1000	0	0	1000
8	1000	0	0	1000
9	1000	0	0	1000
10	0	0	1000	1000
11	0	0	1000	1000
12	0	1000	0	1000
13	0	1000	0	1000
14	0	1000	0	1000
15	0	1000	0	1000
16	1000	0	0	1000
17	1000	0	0	1000
18	1000	0	0	1000
19	1000	0	0	1000
20	1000	0	0	1000
21	1000	0	0	1000
22	1000	0	0	1000
23	1000	0	0	1000
24	1000	0	0	1000
25	1000	0	0	1000
26	0	0	1000	1000

Item Sequence	Summative	MAP	Field Test	Total
27	0	0	1000	1000
28	0	1000	0	1000
29	0	1000	0	1000
30	0	1000	0	1000
31	0	1000	0	1000
32	0	1000	0	1000
33	0	1000	0	1000
34	1000	0	0	1000
35	1000	0	0	1000
36	1000	0	0	1000
37	1000	0	0	1000
38	1000	0	0	1000
39	1000	0	0	1000
40	1000	0	0	1000
41	1000	0	0	1000
42	0	0	1000	1000
43	0	1000	0	1000
44	0	1000	0	1000
45	0	1000	0	1000
46	0	1000	0	1000
47	0	1000	0	1000
48	0	1000	0	1000
49	0	1000	0	1000
50	0	1000	0	1000

Table 3.28. Item Sequence—Mathematics, Grade 7

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	1000	0	0	1000
5	1000	0	0	1000
6	1000	0	0	1000
7	1000	0	0	1000
8	1000	0	0	1000
9	1000	0	0	1000
10	0	0	1000	1000
11	0	0	1000	1000
12	0	1000	0	1000
13	0	1000	0	1000
14	0	1000	0	1000
15	0	1000	0	1000

Item Sequence	Summative	MAP	Field Test	Total
16	1000	0	0	1000
17	1000	0	0	1000
18	1000	0	0	1000
19	1000	0	0	1000
20	1000	0	0	1000
21	1000	0	0	1000
22	1000	0	0	1000
23	1000	0	0	1000
24	1000	0	0	1000
25	1000	0	0	1000
26	0	0	1000	1000
27	0	0	1000	1000
28	0	1000	0	1000
29	0	1000	0	1000
30	0	1000	0	1000
31	0	1000	0	1000
32	0	1000	0	1000
33	0	1000	0	1000
34	1000	0	0	1000
35	1000	0	0	1000
36	1000	0	0	1000
37	1000	0	0	1000
38	1000	0	0	1000
39	1000	0	0	1000
40	1000	0	0	1000
41	1000	0	0	1000
42	0	0	1000	1000
43	0	1000	0	1000
44	0	1000	0	1000
45	0	1000	0	1000
46	0	1000	0	1000
47	0	1000	0	1000
48	0	1000	0	1000
49	0	1000	0	1000
50	0	1000	0	1000

Table 3.29. Item Sequence—Mathematics, Grade 8

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	1000	0	0	1000

Item Sequence	Summative	MAP	Field Test	Total
5	1000	0	0	1000
6	1000	0	0	1000
7	1000	0	0	1000
8	1000	0	0	1000
9	1000	0	0	1000
10	0	0	1000	1000
11	0	0	1000	1000
12	0	1000	0	1000
13	0	1000	0	1000
14	0	1000	0	1000
15	0	1000	0	1000
16	1000	0	0	1000
17	1000	0	0	1000
18	1000	0	0	1000
19	1000	0	0	1000
20	1000	0	0	1000
21	1000	0	0	1000
22	1000	0	0	1000
23	1000	0	0	1000
24	1000	0	0	1000
25	1000	0	0	1000
26	0	0	1000	1000
27	0	0	1000	1000
28	0	1000	0	1000
29	0	1000	0	1000
30	0	1000	0	1000
31	0	1000	0	1000
32	0	1000	0	1000
33	0	1000	0	1000
34	1000	0	0	1000
35	1000	0	0	1000
36	1000	0	0	1000
37	1000	0	0	1000
38	1000	0	0	1000
39	1000	0	0	1000
40	1000	0	0	1000
41	1000	0	0	1000
42	0	0	1000	1000
43	0	1000	0	1000
44	0	1000	0	1000
45	0	1000	0	1000
46	0	1000	0	1000

Item Sequence	Summative	MAP	Field Test	Total
47	0	1000	0	1000
48	0	1000	0	1000
49	0	1000	0	1000
50	0	1000	0	1000

Table 3.30. Item Sequence—Mathematics, High School

Item Sequence	Summative	MAP	Field Test	Total
1	1000	0	0	1000
2	1000	0	0	1000
3	1000	0	0	1000
4	1000	0	0	1000
5	1000	0	0	1000
6	1000	0	0	1000
7	1000	0	0	1000
8	1000	0	0	1000
9	1000	0	0	1000
10	0	0	1000	1000
11	0	0	1000	1000
12	0	1000	0	1000
13	0	1000	0	1000
14	0	1000	0	1000
15	0	1000	0	1000
16	1000	0	0	1000
17	1000	0	0	1000
18	1000	0	0	1000
19	1000	0	0	1000
20	1000	0	0	1000
21	1000	0	0	1000
22	1000	0	0	1000
23	1000	0	0	1000
24	1000	0	0	1000
25	1000	0	0	1000
26	0	0	1000	1000
27	0	0	1000	1000
28	0	1000	0	1000
29	0	1000	0	1000
30	0	1000	0	1000
31	0	1000	0	1000
32	0	1000	0	1000
33	0	1000	0	1000
34	1000	0	0	1000
35	1000	0	0	1000

Item Sequence	Summative	MAP	Field Test	Total
36	1000	0	0	1000
37	1000	0	0	1000
38	1000	0	0	1000
39	1000	0	0	1000
40	1000	0	0	1000
41	1000	0	0	1000
42	1000	0	0	1000
43	1000	0	0	1000
44	0	0	1000	1000
45	0	0	1000	1000
46	1000	0	0	1000
47	0	0	1000	1000
48	0	1000	0	1000
49	0	1000	0	1000
50	0	1000	0	1000
51	0	1000	0	1000
52	0	1000	0	1000
53	0	1000	0	1000
54	0	1000	0	1000

4. Conclusion

The simulation study intends to provide important evidence for supporting test score interpretation and use arguments regarding student proficiency relative to the Common Core State Standards. The simulation demonstrates how well students receive comparable representations of content with sufficient psychometric measurement quality such that the state can infer that test scores have the same meaning across students' individualized test events. The findings from the simulations are as follows:

- The blueprint requirements are fulfilled without exception.
- Student ability estimates are relatively unbiased and highly precise.
- The item exposure rates are adequate.
- The reliability and standard of errors of measurement are reasonable.
- The summative, MAP Growth, and field-test items are presented in intended sequences.

A couple of notes are in order: The high school tests are set up as adaptive tests instead of as fixed form tests in the spring 2023, with limited item pool sizes. The classifications of students into the achievement level categories using the summative scores are adequate.

The results of the simulation indicate that the item pool and the test model setup are functioning properly for the Spring 2024 administration.

References

Han, K. T. (2016). Maximum likelihood score estimation method with fences for short-Length tests and computerized adaptive tests. *Applied Psychometric Measurement*, *40*(4), 289–301. https://doi.org/10.1177/0146621616631317

Robin, F., van der Linden, W. J., Eignor, D. R., Steffen, M., Stocking, M. L. (2005). A comparison of two procedures for constrained adaptive test construction (Research Report No. RR-04-39). Educational Testing Service (ETS). https://ris.utwente.nl/ws/files/5151649/EJ1111003.pdf

Samejima, F. (1994). Estimation of reliability coefficients using the test information function and its modifications. *Applied Psychological Measurement, 18*(3), 229–244. https://doi.org/10.1177/014662169401800304

Appendix D: Summary of *P* Values by Item Type

Table D.1. Summary of *P* Values by Item Type—Operational Items

Grade	Item Type	N	Mean	Median	SD	Min.	Max.	#Items by *P*-Value Range									
								≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.6	≤ 0.7	≤ 0.8	≤ 0.9	> 0.9
Reading																	
3	Multi-select	12	0.55	0.53	0.11	0.41	0.79	0	0	0	0	3	6	2	1	0	0
	Multiple Choice	156	0.47	0.46	0.12	0.12	0.88	0	1	9	39	45	38	20	1	3	0
	Composite	14	0.41	0.40	0.11	0.23	0.68	0	0	2	5	5	1	1	0	0	0
	Gap Match Multiple	9	0.46	0.43	0.16	0.29	0.77	0	0	1	3	1	2	1	1	0	0
	Hot Text	1	0.34	0.34		0.34	0.34	0	0	0	1	0	0	0	0	0	0
4	Multi-select	10	0.49	0.49	0.13	0.32	0.69	0	0	0	2	3	2	3	0	0	0
	Multiple Choice	211	0.47	0.46	0.13	0.12	0.93	0	1	15	41	77	48	16	8	4	1
	Composite	23	0.41	0.43	0.12	0.16	0.62	0	2	2	5	8	5	1	0	0	0
	Gap Match Multiple	11	0.46	0.44	0.13	0.29	0.70	0	0	1	4	2	1	2	1	0	0
	Hot Text	1	0.39	0.39	–	0.39	0.39	0	0	0	1	0	0	0	0	0	0
5	Multi-select	10	0.52	0.52	0.15	0.32	0.8	0	0	0	3	2	3	1	0	1	0
	Multiple Choice	232	0.48	0.47	0.13	0.05	0.86	1	2	9	38	89	55	24	11	3	0
	Composite	13	0.41	0.42	0.07	0.30	0.52	0	0	0	6	5	2	0	0	0	0
	Gap Match Multiple	12	0.46	0.45	0.19	0.20	0.78	0	0	3	0	5	1	1	2	0	0
	Gap Match Single	1	0.46	0.46	–	0.46	0.46	0	0	0	0	1	0	0	0	0	0
	Hot Text	4	0.49	0.49	0.04	0.44	0.54	0	0	0	0	3	1	0	0	0	0
6	Multi-select	22	0.46	0.45	0.10	0.33	0.7	0	0	0	6	9	5	1	1	0	0
	Multiple Choice	179	0.49	0.47	0.12	0.19	0.86	0	1	5	34	66	43	21	6	3	0
	Composite	12	0.46	0.44	0.09	0.35	0.62	0	0	0	4	5	2	1	0	0	0
	Gap Match Multiple	6	0.47	0.50	0.16	0.25	0.66	0	0	1	1	1	2	1	0	0	0
	Gap Match Single	1	0.41	0.41	–	0.41	0.41	0	0	0	0	1	0	0	0	0	0
	Hot Text	1	0.41	0.41	–	0.41	0.41	0	0	0	0	1	0	0	0	0	0

Grade	Item Type	N	Mean	Median	SD	Min.	Max.	#Items by *P*-Value Range									
								≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.6	≤ 0.7	≤ 0.8	≤ 0.9	> 0.9
7	Multi-select	12	0.41	0.38	0.08	0.29	0.50	0	0	1	6	3	2	0	0	0	0
	Multiple Choice	222	0.47	0.46	0.12	0.13	0.90	0	2	15	39	91	47	20	6	1	1
	Composite	6	0.39	0.41	0.06	0.30	0.44	0	0	1	1	4	0	0	0	0	0
	Gap Match Multiple	10	0.46	0.45	0.14	0.27	0.72	0	0	1	2	3	3	0	1	0	0
	Hot Text	1	0.47	0.47	–	0.47	0.47	0	0	0	0	1	0	0	0	0	0
8	Multi-select	26	0.46	0.42	0.10	0.34	0.73	0	0	0	6	14	2	3	1	0	0
	Multiple Choice	243	0.51	0.51	0.13	0.10	0.95	1	2	10	29	69	76	47	5	3	1
	Composite	17	0.46	0.48	0.13	0.11	0.63	0	1	1	2	5	7	1	0	0	0
	Gap Match Multiple	17	0.53	0.54	0.13	0.30	0.76	0	0	1	1	5	6	2	2	0	0
	Hot Text	3	0.47	0.50	0.14	0.32	0.60	0	0	0	1	1	1	0	0	0	0
HS	Multi-select	7	0.52	0.52	0.09	0.37	0.61	0	0	0	1	2	1	3	0	0	0
	Multiple Choice	31	0.50	0.46	0.15	0.19	0.84	0	1	1	2	17	2	5	2	1	0
	Composite	8	0.46	0.46	0.13	0.30	0.63	0	0	0	4	0	2	2	0	0	0
	Gap Match Multiple	1	0.50	0.50	–	0.50	0.50	0	0	0	0	0	1	0	0	0	0
	Gap Match Single	1	0.38	0.38	–	0.38	0.38	0	0	0	1	0	0	0	0	0	0
Mathematics																	
3	Multi-select	41	0.45	0.45	0.08	0.29	0.62	0	0	1	8	20	11	1	0	0	0
	Multiple Choice	233	0.48	0.48	0.10	0.22	0.78	0	0	7	36	104	63	17	6	0	0
	Composite	10	0.43	0.40	0.17	0.17	0.77	0	1	1	3	2	2	0	1	0	0
	Gap Match Multiple	39	0.50	0.53	0.12	0.17	0.70	0	1	3	1	12	15	7	0	0	0
	Gap Match Single	7	0.48	0.46	0.06	0.42	0.56	0	0	0	0	5	2	0	0	0	0
	Graphic Gap Match	27	0.47	0.46	0.11	0.20	0.75	0	1	1	2	14	6	2	1	0	0
	Hot Text	16	0.38	0.41	0.11	0.10	0.53	1	0	2	4	8	1	0	0	0	0
	Text Entry	49	0.51	0.51	0.07	0.37	0.69	0	0	0	5	15	24	5	0	0	0
4	Multi-select	37	0.45	0.43	0.10	0.27	0.69	0	0	2	10	13	10	2	0	0	0

Grade	Item Type	N	Mean	Median	SD	Min.	Max.	#Items by *P*-Value Range									
								≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.6	≤ 0.7	≤ 0.8	≤ 0.9	> 0.9
	Multiple Choice	207	0.49	0.48	0.09	0.27	0.91	0	0	3	26	100	62	10	5	0	1
	Composite	21	0.42	0.40	0.16	0.22	0.76	0	0	5	6	4	3	1	2	0	0
	Gap Match Multiple	28	0.47	0.49	0.11	0.20	0.62	0	1	1	3	12	8	3	0	0	0
	Graphic Gap Match	34	0.51	0.53	0.15	0.05	0.75	1	0	2	5	3	14	6	3	0	0
	Hot Text	25	0.42	0.42	0.08	0.27	0.64	0	0	1	9	12	2	1	0	0	0
	Text Entry	39	0.47	0.46	0.13	0.18	0.77	0	2	0	10	13	7	6	1	0	0
5	Multi-select	37	0.40	0.41	0.13	0.02	0.63	1	2	3	10	14	6	1	0	0	0
	Multiple Choice	225	0.47	0.46	0.09	0.16	0.72	0	1	6	45	94	67	11	1	0	0
	Composite	20	0.47	0.46	0.14	0.24	0.84	0	0	2	3	8	5	0	1	1	0
	Gap Match Multiple	27	0.47	0.45	0.13	0.15	0.72	0	1	0	4	13	5	2	2	0	0
	Gap Match Single	3	0.53	0.54	0.15	0.38	0.69	0	0	0	1	0	1	1	0	0	0
	Graphic Gap Match	30	0.48	0.49	0.10	0.31	0.74	0	0	0	6	13	7	2	2	0	0
	Hot Text	20	0.40	0.41	0.12	0.05	0.61	1	0	2	6	8	2	1	0	0	0
	Text Entry	46	0.46	0.49	0.12	0.08	0.67	1	1	2	6	16	17	3	0	0	0
6	Multi-select	29	0.42	0.39	0.13	0.18	0.71	0	1	6	9	7	3	2	1	0	0
	Multiple Choice	193	0.48	0.48	0.08	0.29	0.78	0	0	1	32	86	63	9	2	0	0
	Composite	20	0.45	0.44	0.16	0.14	0.81	0	1	1	6	6	3	1	1	1	0
	Gap Match Multiple	34	0.45	0.44	0.11	0.21	0.72	0	0	4	5	12	10	2	1	0	0
	Graphic Gap Match	11	0.46	0.48	0.11	0.32	0.64	0	0	0	5	2	3	1	0	0	0
	Hot Text	25	0.38	0.40	0.07	0.22	0.49	0	0	3	11	11	0	0	0	0	0
	Text Entry	45	0.46	0.45	0.10	0.24	0.81	0	0	2	8	24	7	3	0	1	0
7	Multi-select	33	0.43	0.46	0.13	0.13	0.70	0	3	1	8	14	5	1	1	0	0
	Multiple Choice	213	0.47	0.47	0.08	0.26	0.71	0	0	5	36	101	63	7	1	0	0
	Composite	12	0.46	0.46	0.13	0.27	0.69	0	0	2	1	5	3	1	0	0	0
	Gap Match Multiple	24	0.46	0.45	0.12	0.17	0.64	0	1	1	4	12	3	3	0	0	0

Grade	Item Type	N	Mean	Median	SD	Min.	Max.	#Items by *P*-Value Range									
								≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.6	≤ 0.7	≤ 0.8	≤ 0.9	> 0.9
	Gap Match Single	1	0.45	0.45	–	0.45	0.45	0	0	0	0	1	0	0	0	0	0
	Graphic Gap Match	4	0.53	0.54	0.06	0.45	0.59	0	0	0	0	1	3	0	0	0	0
	Hot Text	27	0.42	0.41	0.10	0.21	0.65	0	0	2	9	10	5	1	0	0	0
	Text Entry	35	0.44	0.43	0.08	0.23	0.60	0	0	1	8	19	7	0	0	0	0
8	Multi-select	28	0.44	0.45	0.10	0.25	0.68	0	0	4	3	13	7	1	0	0	0
	Multiple Choice	206	0.46	0.46	0.07	0.24	0.68	0	0	5	35	113	48	5	0	0	0
	Composite	17	0.38	0.42	0.12	0.14	0.58	0	2	2	4	7	2	0	0	0	0
	Gap Match Multiple	25	0.47	0.45	0.09	0.27	0.68	0	0	1	5	12	5	2	0	0	0
	Gap Match Single	1	0.47	0.47	–	0.47	0.47	0	0	0	0	1	0	0	0	0	0
	Graphic Gap Match	14	0.43	0.42	0.11	0.28	0.65	0	0	2	4	5	1	2	0	0	0
	Hot Text	45	0.43	0.43	0.08	0.29	0.64	0	0	3	11	21	9	1	0	0	0
	Text Entry	56	0.49	0.50	0.11	0.17	0.74	0	1	2	8	19	18	7	1	0	0
HS	Multi-select	14	0.42	0.43	0.19	0.04	0.74	1	1	1	3	2	4	1	1	0	0
	Multiple Choice	21	0.37	0.34	0.12	0.23	0.73	0	0	7	10	2	1	0	1	0	0
	Composite	10	0.38	0.35	0.14	0.17	0.61	0	1	2	3	2	1	1	0	0	0
	Gap Match Multiple	9	0.36	0.31	0.18	0.15	0.69	0	2	2	2	1	1	1	0	0	0
	Graphic Gap Match	1	0.61	0.61	–	0.61	0.61	0	0	0	0	0	0	1	0	0	0
	Hot Text	7	0.34	0.32	0.17	0.08	0.55	1	0	2	2	0	2	0	0	0	0
	Text Entry	1	0.20	0.20	–	0.20	0.20	0	1	0	0	0	0	0	0	0	0

Note. N/A = Not Applicable

Table D.2. Summary of *P* Values by Item Type—Field Test Items

Grade	Item Type	N	Mean	Median	SD	Min.	Max.	#Items by *P*-value Range									
								≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.6	≤ 0.7	≤ 0.8	≤ 0.9	> 0.9
Reading																	
3	Multiple Choice	4	0.36	0.35	0.11	0.24	0.50	0	0	1	2	1	0	0	0	0	0
	Composite	1	0.25	0.25	–	0.25	0.25	0	0	1	0	0	0	0	0	0	0
4	Multi-select	4	0.35	0.35	0.08	0.25	0.45	0	0	1	2	1	0	0	0	0	0
	Multiple Choice	1	0.39	0.39	–	0.39	0.39	0	0	0	1	0	0	0	0	0	0
5	Multi-select	4	0.55	0.53	0.18	0.34	0.78	0	0	0	1	1	1	0	1	0	0
	Multiple Choice	2	0.39	0.39	0.13	0.30	0.48	0	0	0	1	1	0	0	0	0	0
	Composite	1	0.19	0.19	–	0.19	0.19	0	1	0	0	0	0	0	0	0	0
6	Multi-select	2	0.49	0.49	0.00	0.49	0.49	0	0	0	0	2	0	0	0	0	0
	Multiple Choice	3	0.40	0.41	0.06	0.34	0.46	0	0	0	1	2	0	0	0	0	0
7	Multi-select	1	0.53	0.53	–	0.53	0.53	0	0	0	0	0	1	0	0	0	0
	Multiple Choice	4	0.46	0.49	0.16	0.23	0.62	0	0	1	0	2	0	1	0	0	0
8	Multi-select	3	0.60	0.51	0.20	0.46	0.83	0	0	0	0	1	1	0	0	1	0
	Multiple Choice	7	0.41	0.37	0.13	0.29	0.67	0	0	1	3	2	0	1	0	0	0
	Composite	1	0.21	0.21	–	0.21	0.21	0	0	1	0	0	0	0	0	0	0
	Gap Match Multiple	1	0.42	0.42	–	0.42	0.42	0	0	0	0	1	0	0	0	0	0
HS	Multi-select	13	0.58	0.60	0.08	0.42	0.70	0	0	0	0	2	5	6	0	0	0
	Multiple Choice	79	0.49	0.48	0.11	0.23	0.81	0	0	5	12	28	22	9	2	1	0
	Composite	17	0.40	0.39	0.12	0.22	0.61	0	0	3	6	5	1	2	0	0	0
	Gap Match Multiple	4	0.23	0.23	0.09	0.16	0.32	0	2	1	1	0	0	0	0	0	0
	Hot Text	4	0.29	0.30	0.16	0.13	0.43	0	2	0	0	2	0	0	0	0	0
Mathematics																	
3	Multi-select	1	0.12	0.12	–	0.12	0.12	0	1	0	0	0	0	0	0	0	0
	Multiple Choice	4	0.28	0.24	0.13	0.19	0.47	0	1	2	0	1	0	0	0	0	0
	Composite	1	0.21	0.21	–	0.21	0.21	0	0	1	0	0	0	0	0	0	0
4	Multiple Choice	7	0.36	0.30	0.14	0.24	0.60	0	0	3	2	0	2	0	0	0	0

Grade	Item Type	N	Mean	Median	SD	Min.	Max.	#Items by *P*-value Range									
								≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.6	≤ 0.7	≤ 0.8	≤ 0.9	> 0.9
	Gap Match Multiple	3	0.29	0.19	0.23	0.12	0.54	0	2	0	0	0	1	0	0	0	0
	Gap Match Single	1	0.17	0.17	–	0.17	0.17	0	1	0	0	0	0	0	0	0	0
5	Multi-select	1	0.27	0.27	–	0.27	0.27	0	0	1	0	0	0	0	0	0	0
	Multiple Choice	3	0.29	0.30	0.09	0.19	0.38	0	1	0	2	0	0	0	0	0	0
	Graphic Gap Match	1	0.06	0.06	–	0.06	0.06	1	0	0	0	0	0	0	0	0	0
6	Multi-select	5	0.12	0.09	0.09	0.03	0.27	3	1	1	0	0	0	0	0	0	0
	Multiple Choice	3	0.34	0.34	0.09	0.25	0.43	0	0	1	1	1	0	0	0	0	0
	Graphic Gap Match	2	0.07	0.07	0.01	0.06	0.08	2	0	0	0	0	0	0	0	0	0
	Hot Text	3	0.23	0.15	0.19	0.10	0.46	1	1	0	0	1	0	0	0	0	0
	Text Entry	5	0.08	0.04	0.08	0.01	0.20	3	2	0	0	0	0	0	0	0	0
7	Multi-select	7	0.13	0.12	0.07	0.03	0.26	3	3	1	0	0	0	0	0	0	0
	Graphic Gap Match	2	0.07	0.07	0.02	0.06	0.09	2	0	0	0	0	0	0	0	0	0
	Hot Text	3	0.25	0.13	0.21	0.12	0.49		2	0	0	1	0	0	0	0	0
	Text Entry	7	0.09	0.05	0.08	0.04	0.23	5	1	1	0	0	0	0	0	0	0
8	Multi-select	5	0.19	0.16	0.11	0.04	0.32	1	2	1	1	0	0	0	0	0	0
	Graphic Gap Match	1	0.17	0.17	–	0.17	0.17	0	1	0	0	0	0	0	0	0	0
	Hot Text	3	0.30	0.17	0.24	0.16	0.57	0	2	0	0	0	1	0	0	0	0
	Text Entry	4	0.10	0.09	0.03	0.07	0.13	2	2	0	0	0	0	0	0	0	0
HS	Multi-select	31	0.19	0.15	0.12	0.03	0.56	8	10	9	2	0	2	0	0	0	0
	Multiple Choice	31	0.37	0.37	0.12	0.18	0.63	0	2	8	10	7	3	1	0	0	0
	Composite	18	0.19	0.15	0.14	0.01	0.42	7	5	0	4	2	0	0	0	0	0
	Gap Match Multiple	23	0.21	0.15	0.20	0.04	0.78	5	11	3	1	0	1	1	1	0	0
	Gap Match Single	3	0.11	0.14	0.04	0.06	0.14	1	2	0	0	0	0	0	0	0	0
	Graphic Gap Match	2	0.17	0.17	0.24	0.00	0.34	1	0	0	1	0	0	0	0	0	0

Grade	Item Type	N	Mean	Median	SD	Min.	Max.	#Items by *P*-value Range									
								≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.6	≤ 0.7	≤ 0.8	≤ 0.9	> 0.9
	Hot Text	31	0.20	0.18	0.11	0.04	0.49	8	8	9	5	1	0	0	0	0	0
	Text Entry	6	0.10	0.07	0.08	0.02	0.23	4	1	1	0	0	0	0	0	0	0

Note. N/A = Not Applicable

Appendix E: Summary of Item-Total Correlation by Item Type

Table E.1. Summary of Item-Total Correlation by Item Type—Operational Items

Grade	Item Type	N	Mean	Median	SD	Min.	Max.	#Items by Item-Total Correlation Range									
								≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.6	≤ 0.7	≤ 0.8	≤ 0.9	> 0.9
Reading																	
3	Multiselect	12	0.35	0.40	0.20	0.01	0.69	2	1	2	1	3	2	1	0	0	0
	Multiple Choice	156	0.34	0.34	0.11	0.00	0.61	4	10	36	67	30	8	1	0	0	0
	Composite	14	0.46	0.46	0.08	0.32	0.57	0	0	0	3	5	6	0	0	0	0
	Gap Match Multiple	9	0.41	0.43	0.17	0.13	0.64	0	2	0	2	3	1	1	0	0	0
	Hot Text	1	0.43	0.43	–	0.43	0.43	0	0	0	0	1	0	0	0	0	0
4	Multiselect	10	0.41	0.42	0.13	0.20	0.58	0	1	1	1	5	2	0	0	0	0
	Multiple Choice	211	0.33	0.34	0.12	-0.16	0.68	7	16	52	87	38	8	3	0	0	0
	Composite	23	0.49	0.50	0.09	0.30	0.66	0	0	1	3	9	9	1	0	0	0
	Gap Match Multiple	11	0.34	0.41	0.17	-0.04	0.53	2	0	0	3	5	1	0	0	0	0
	Hot Text	1	0.52	0.52	–	0.52	0.52	0	0	0	0	0	1	0	0	0	0
5	Multiselect	10	0.42	0.44	0.15	0.14	0.64	0	1	0	3	3	2	1	0	0	0
	Multiple Choice	232	0.34	0.35	0.12	-0.37	0.70	5	17	47	107	41	13	1	1	0	0
	Composite	13	0.49	0.49	0.10	0.33	0.68	0	0	0	2	6	3	2	0	0	0
	Gap Match Multiple	12	0.36	0.37	0.11	0.17	0.54	0	1	2	4	4	1	0	0	0	0
	Gap Match Single	1	0.38	0.38	–	0.38	0.38	0	0	0	1	0	0	0	0	0	0
	Hot Text	4	0.40	0.36	0.18	0.24	0.65	0	0	1	1	1	0	1	0	0	0
6	Multiselect	22	0.41	0.42	0.11	0.12	0.60	0	1	3	6	8	4	0	0	0	0
	Multiple Choice	179	0.34	0.34	0.09	0.00	0.57	2	7	47	77	39	7	0	0	0	0
	Composite	12	0.52	0.53	0.08	0.41	0.71	0	0	0	0	4	7	0	1	0	0
	Gap Match Multiple	6	0.34	0.37	0.11	0.16	0.48	0	1	1	2	2	0	0	0	0	0
	Gap Match Single	1	0.50	0.50	–	0.50	0.50	0	0	0	0	0	1	0	0	0	0
	Hot Text	1	0.50	0.50	–	0.50	0.50	0	0	0	0	1	0	0	0	0	0
7	Multiselect	12	0.44	0.42	0.09	0.29	0.58	0	0	1	4	3	4	0	0	0	0

Grade	Item Type	N	Mean	Median	SD	Min.	Max.	#Items by Item-Total Correlation Range									
								≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.6	≤ 0.7	≤ 0.8	≤ 0.9	> 0.9
	Multiple Choice	222	0.33	0.34	0.11	0.00	0.66	7	17	59	83	47	3	6	0	0	0
	Composite	6	0.48	0.50	0.08	0.34	0.58	0	0	0	1	2	3	0	0	0	0
	Gap Match Multiple	10	0.42	0.43	0.11	0.23	0.59	0	0	1	3	4	2	0	0	0	0
	Hot Text	1	0.10	0.10	–	0.10	0.10	0	1	0	0	0	0	0	0	0	0
8	Multiselect	26	0.35	0.37	0.10	0.15	0.56	0	3	4	12	5	2	0	0	0	0
	Multiple Choice	243	0.35	0.35	0.11	-0.07	0.64	7	13	49	104	51	16	3	0	0	0
	Composite	17	0.44	0.46	0.11	0.14	0.58	0	1	1	1	9	5	0	0	0	0
	Gap Match Multiple	17	0.43	0.46	0.10	0.19	0.56	0	1	2	2	6	6	0	0	0	0
	Hot Text	3	0.47	0.44	0.08	0.41	0.56	0	0	0	0	2	1	0	0	0	0
HS	Choice Multiple	7	0.44	0.42	0.13	0.26	0.65	0	0	1	2	1	2	1	0	0	0
	Choice Single	31	0.33	0.32	0.13	-0.07	0.53	1	1	12	7	7	3	0	0	0	0
	Gap Match Multiple	8	0.52	0.53	0.06	0.43	0.61	0	0	0	0	3	4	1	0	0	0
	Gap Match Single	1	0.57	0.57	–	0.57	0.57	0	0	0	0	0	1	0	0	0	0
	Composite	1	0.32	0.32	–	0.32	0.32	0	0	0	1	0	0	0	0	0	0
Mathematics																	
3	Multiselect	41	0.37	0.36	0.11	0.06	0.64	1	1	5	24	6	2	2	0	0	0
	Multiple Choice	233	0.34	0.34	0.08	-0.09	0.57	2	2	60	129	35	5	0	0	0	0
	Composite	10	0.44	0.41	0.09	0.32	0.66	0	0	0	2	7	0	1	0	0	0
	Gap Match Multiple	39	0.36	0.35	0.07	0.23	0.52	0	0	9	19	10	1	0	0	0	0
	Gap Match Single	7	0.34	0.30	0.12	0.27	0.60	0	0	4	2	0	0	1	0	0	0
	Graphic Gap Match	27	0.35	0.37	0.07	0.19	0.46	0	1	5	16	5	0	0	0	0	0
	Hot Text	16	0.36	0.36	0.09	0.22	0.56	0	0	6	6	3	1	0	0	0	0
	Text Entry	49	0.40	0.40	0.06	0.27	0.57	0	0	4	19	23	3	0	0	0	0
4	Multiselect	37	0.38	0.36	0.09	0.24	0.53	0	0	7	16	8	6	0	0	0	0
	Multiple Choice	207	0.34	0.35	0.09	-0.08	0.75	4	5	50	108	34	5	0	1	0	0
	Composite	21	0.41	0.40	0.12	0.20	0.68	0	0	4	7	5	4	1	0	0	0

Grade	Item Type	N	Mean	Median	SD	Min.	Max.	#Items by Item-Total Correlation Range									
								≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.6	≤ 0.7	≤ 0.8	≤ 0.9	> 0.9
	Gap Match Multiple	28	0.38	0.37	0.09	0.22	0.58	0	0	7	11	7	3	0	0	0	0
	Graphic Gap Match	34	0.35	0.36	0.11	-0.07	0.52	1	1	4	19	8	1	0	0	0	0
	Hot Text	25	0.37	0.36	0.09	0.19	0.57	0	1	3	14	5	2	0	0	0	0
	Text Entry	39	0.38	0.37	0.06	0.20	0.49	0	1	2	21	15	0	0	0	0	0
5	Multiselect	37	0.39	0.39	0.11	-0.03	0.57	1	0	3	17	10	6	0	0	0	0
	Multiple Choice	225	0.34	0.34	0.08	0.07	0.65	1	7	62	104	47	2	2	0	0	0
	Composite	20	0.46	0.45	0.09	0.33	0.62	0	0	0	6	5	8	1	0	0	0
	Gap Match Multiple	27	0.35	0.34	0.10	0.16	0.64	0	2	5	14	4	0	2	0	0	0
	Gap Match Single	3	0.29	0.29	0.02	0.27	0.32	0	0	2	1	0	0	0	0	0	0
	Graphic Gap Match	30	0.38	0.38	0.09	0.18	0.52	0	2	3	12	11	2	0	0	0	0
	Hot Text	20	0.33	0.35	0.13	-0.16	0.50	1	0	4	12	3	0	0	0	0	0
	Text Entry	46	0.34	0.36	0.08	0.07	0.51	1	2	11	22	9	1	0	0	0	0
6	Multiselect	29	0.42	0.42	0.11	0.10	0.56	0	1	3	9	7	9	0	0	0	0
	Multiple Choice	193	0.34	0.34	0.08	-0.01	0.68	1	8	54	89	38	2	1	0	0	0
	Composite	20	0.45	0.46	0.09	0.22	0.65	0	0	2	1	12	4	1	0	0	0
	Gap Match Multiple	34	0.38	0.39	0.10	0.14	0.64	0	1	6	12	12	2	1	0	0	0
	Graphic Gap Match	11	0.40	0.42	0.07	0.29	0.48	0	0	1	4	6	0	0	0	0	0
	Hot Text	25	0.33	0.32	0.10	-0.01	0.51	1	0	6	11	6	1	0	0	0	0
	Text Entry	45	0.39	0.39	0.09	0.24	0.71	0	0	5	25	10	4	0	1	0	0
7	Multiselect	33	0.40	0.38	0.10	0.11	0.58	0	1	3	16	8	5	0	0	0	0
	Multiple Choice	213	0.33	0.34	0.08	-0.03	0.52	2	13	50	105	42	1	0	0	0	0
	Composite	12	0.47	0.50	0.10	0.26	0.63	0	0	1	4	1	5	1	0	0	0
	Gap Match Multiple	24	0.36	0.35	0.08	0.21	0.53	0	0	4	12	7	1	0	0	0	0
	Gap Match Single	1	0.49	0.49	–	0.49	0.49	0	0	0	0	1	0	0	0	0	0
	Graphic Gap Match	4	0.38	0.38	0.08	0.28	0.48	0	0	1	2	1	0	0	0	0	0

Grade	Item Type	N	Mean	Median	SD	Min.	Max.	#Items by Item-Total Correlation Range									
								≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.6	≤ 0.7	≤ 0.8	≤ 0.9	> 0.9
	Hot Text	27	0.36	0.37	0.09	0.13	0.56	0	2	3	14	7	1	0	0	0	0
	Text Entry	35	0.39	0.40	0.07	0.23	0.57	0	0	5	13	15	2	0	0	0	0
8	Multiselect	28	0.33	0.34	0.10	-0.03	0.54	1	0	8	14	4	1	0	0	0	0
	Multiple Choice	206	0.31	0.32	0.07	0.12	0.55	0	8	71	111	15	1	0	0	0	0
	Composite	17	0.41	0.43	0.19	-0.09	0.80	1	0	3	3	7	1	1	0	1	0
	Gap Match Multiple	25	0.39	0.37	0.09	0.25	0.57	0	0	4	11	5	5	0	0	0	0
	Gap Match Single	1	0.46	0.46	–	0.46	0.46	0	0	0	0	1	0	0	0	0	0
	Graphic Gap Match	14	0.36	0.35	0.06	0.29	0.48	0	0	2	9	3	0	0	0	0	0
	Hot Text	45	0.33	0.33	0.08	0.19	0.55	0	2	14	20	8	1	0	0	0	0
	Text Entry	56	0.39	0.38	0.08	0.25	0.65	0	0	9	24	18	4	1	0	0	0
HS	Multiselect	14	0.45	0.43	0.11	0.31	0.66	0	0	0	5	5	2	2	0	0	0
	Multiple Choice	21	0.28	0.28	0.10	0.07	0.43	2	1	11	4	3	0	0	0	0	0
	Composite	10	0.42	0.43	0.11	0.22	0.59	0	0	1	3	4	2	0	0	0	0
	Gap Match Multiple	9	0.42	0.40	0.06	0.29	0.50	0	0	1	3	4	1	0	0	0	0
	Graphic Gap Match	1	0.43	0.43	–	0.43	0.43	0	0	0	0	1	0	0	0	0	0
	Hot Text	7	0.38	0.38	0.13	0.17	0.61	0	1	0	4	1	0	1	0	0	0
	Text Entry	1	0.52	0.52	–	0.52	0.52	0	0	0	0	0	1	0	0	0	0

Note. N/A = Not Applicable

Table E.2. Summary of Item-Total Correlation by Item Type—Field Test Items

Grade	Item Type	N	Mean	Median	SD	Min.	Max.	#Items by Item-Total Correlation Range									
								≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.6	≤ 0.7	≤ 0.8	≤ 0.9	> 0.9
Reading																	
3	Multiple Choice	4	0.24	0.22	0.05	0.21	0.32	0	0	3	1	0	0	0	0	0	0
	Composite	1	0.35	0.35	–	0.35	0.35	0	0	0	1	0	0	0	0	0	0
4	Multiselect	1	0.19	0.19	–	0.19	0.19	0	1	0	0	0	0	0	0	0	0
	Multiple Choice	4	0.24	0.24	0.09	0.15	0.34	0	2	1	1	0	0	0	0	0	0
5	Multiselect	4	0.33	0.41	0.19	0.05	0.47	1	0	0	1	2	0	0	0	0	0
	Multiple Choice	2	0.31	0.31	0.08	0.25	0.37	0	0	1	1	0	0	0	0	0	0
	Composite	1	0.05	0.05	–	0.05	0.05	1	0	0	0	0	0	0	0	0	0
6	Multiselect	2	0.30	0.30	0.07	0.25	0.35	0	0	1	1	0	0	0	0	0	0
	Multiple Choice	3	0.25	0.23	0.11	0.16	0.37	0	1	1	1	0	0	0	0	0	0
7	Multiselect	1	0.26	0.26	–	0.26	0.26	0	0	1	0	0	0	0	0	0	0
	Multiple Choice	4	0.29	0.28	0.03	0.26	0.33	0	0	3	1	0	0	0	0	0	0
8	Multiselect	3	0.30	0.31	0.15	0.15	0.45	0	1	0	1	1	0	0	0	0	0
	Multiple Choice	7	0.20	0.21	0.08	0.10	0.31	1	2	3	1	0	0	0	0	0	0
	Composite	1	0.08	0.08	–	0.08	0.08	1	0	0	0	0	0	0	0	0	0
	Gap Match Multiple	1	0.35	0.35	–	0.35	0.35	0	0	0	1	0	0	0	0	0	0
HS	Multiselect	13	0.41	0.42	0.12	0.21	0.58	0	0	2	4	4	3	0	0	0	0
	Multiple Choice	79	0.33	0.36	0.12	0.01	0.55	5	8	12	28	25	1	0	0	0	0
	Composite	17	0.42	0.41	0.14	0.09	0.66	1	0	2	5	3	5	1	0	0	0
	Gap Match Multiple	4	0.39	0.38	0.10	0.28	0.53	0	0	1	2	0	1	0	0	0	0
	Hot Text	4	0.38	0.37	0.10	0.29	0.49	0	0	2	0	2	0	0	0	0	0
Mathematics																	
3	Multiselect	1	0.19	0.19	–	0.19	0.19	0	1	0	0	0	0	0	0	0	0
	Multiple Choice	4	0.14	0.12	0.06	0.08	0.22	1	2	1	0	0	0	0	0	0	0
	Composite	1	0.22	0.22	–	0.22	0.22	0	0	1	0	0	0	0	0	0	0
4	Multiselect	7	0.14	0.16	0.10	-0.02	0.26	2	3	2	0	0	0	0	0	0	0
	Multiple Choice	3	0.27	0.28	0.12	0.14	0.38	0	1	1	1	0	0	0	0	0	0

Grade	Item Type	N	Mean	Median	SD	Min.	Max.	#Items by Item-Total Correlation Range									
								≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.6	≤ 0.7	≤ 0.8	≤ 0.9	> 0.9
	Gap Match Single	1	0.42	0.42	–	0.42	0.42	0	0	0	0	1	0	0	0	0	0
5	Multiselect	1	0.27	0.27	–	0.27	0.27	0	0	1	0	0	0	0	0	0	0
	Multiple Choice	3	0.12	0.10	0.04	0.09	0.16	2	1	0	0	0	0	0	0	0	0
	Graphic Gap Match	1	0.39	0.39	–	0.39	0.39	0	0	0	1	0	0	0	0	0	0
6	Multiselect	5	0.23	0.24	0.07	0.11	0.29	0	1	4	0	0	0	0	0	0	0
	Multiple Choice	3	0.13	0.08	0.10	0.06	0.24	2	0	1	0	0	0	0	0	0	0
	Graphic Gap Match	2	0.27	0.27	0.20	0.13	0.41	0	1	0	0	1	0	0	0	0	0
	Hot Text	3	0.08	0.06	0.14	-0.05	0.22	2	0	1	0	0	0	0	0	0	0
	Text Entry	5	0.31	0.30	0.15	0.10	0.49	1	0	2	0	2	0	0	0	0	0
7	Multiselect	7	0.23	0.25	0.16	-0.01	0.42	2	1	1	1	2	0	0	0	0	0
	Graphic Gap Match	2	0.13	0.13	0.07	0.08	0.18	1	1	0	0	0	0	0	0	0	0
	Hot Text	3	0.15	0.13	0.07	0.09	0.22	1	1	1	0	0	0	0	0	0	0
	Text Entry	7	0.38	0.35	0.10	0.23	0.54	0	0	1	3	2	1	0	0	0	0
8	Multiselect	5	0.33	0.35	0.21	0.06	0.58	1	1	0	1	1	1	0	0	0	0
	Graphic Gap Match	1	0.35	0.35	–	0.35	0.35	0	0	0	1	0	0	0	0	0	0
	Hot Text	3	0.21	0.18	0.07	0.17	0.30	0	2	1	0	0	0	0	0	0	0
	Text Entry	4	0.41	0.41	0.05	0.35	0.47	0	0	0	2	2	0	0	0	0	0
HS	Multiselect	31	0.32	0.36	0.16	-0.02	0.60	3	4	6	6	9	3	0	0	0	0
	Multiple Choice	31	0.29	0.30	0.12	0.06	0.48	1	7	8	6	9	0	0	0	0	0
	Composite	18	0.43	0.44	0.11	0.10	0.60	0	1	0	4	8	4	1	0	0	0
	Gap Match Multiple	23	0.39	0.40	0.15	-0.02	0.66	1	1	3	6	6	4	2	0	0	0
	Gap Match Single	3	0.57	0.60	0.09	0.47	0.64	0	0	0	0	1	0	2	0	0	0
	Graphic Gap Match	2	0.25	0.25	0.13	0.16	0.35	0	1	0	1	0	0	0	0	0	0
	Hot Text	31	0.34	0.36	0.14	0.01	0.54	2	4	5	6	11	3	0	0	0	0
	Text Entry	6	0.38	0.38	0.12	0.19	0.56	0	1	0	3	1	1	0	0	0	0

Note. N/A = Not Applicable

The Maine Through Year Assessment

Reports Interpretive Guide

2024–2025 Maine Through Year Assessment
Reading and Mathematics

Contributors

The Maine Through Year Assessment is administered by the Maine Department of Education (Maine DOE).

23 State House Station
Augusta, Maine 04333-0023

207.624.6600

The assessment contractor is NWEA. NWEA can be reached by calling Partner Support at 855.430.1777.

Table of Contents

Part 1—Introduction

This guide is designed to help educators and administrators interpret the Maine Through Year Assessment reports in Reading and Mathematics in grades 3 through 8 and 2nd year of high school. For more resources and information on the Maine Through Year Assessment, visit the Maine DOE Through Year Assessment webpage or the NWEA Maine Connection page.

About the Maine Through Year Assessment

The Maine Through Year Assessment assesses all publicly funded Maine students in grades 3 through 8 and second year of high school in the content areas of Reading and Math. The Through Year Assessment consists of three administrations: fall, winter, and spring. The fall and spring administrations are required for all students; the winter administration is optional. Students with significant cognitive disabilities who qualify for the alternate assessment to the Maine Through Year Assessment participate in the MSAA. The Maine Through Year Assessment does not need to be submitted for any student who was assessed through the alternate assessment.

The Maine Through Year Assessment was developed to provide teachers, students, and parents with information on student learning strengths and needs throughout the year, as well as student progress in mastering college and career-ready skills based on Maine's accountability standards, the Common Core State Standards.

The Maine Through Year Assessment is administered online and is computer-adaptive, varying the difficulty of the next question based on a student's performance on prior assessment questions. As a result, the assessment adapts to the learner and is unique to every student. The assessment is composed of items, or questions, from NWEA's through-year item bank as well as MAP Growth™ items. The assessment is designed to be untimed; students should be provided as much time as they need to answer all assessment questions.

A variety of student supports and accessibility features are available to students to ensure that the assessment experience is as consistent as possible with the student's daily instruction. These accessibility features include, but are not limited to, supports and accommodations for students with support plans, Individual Language Acquisition Plans, 504 plans, and Individualized Education Programs (IEPs). Paper, large print, and braille assessments are available for qualifying students.

The Maine Through Year Assessment provides the opportunity for teachers to access and use immediate, actionable data within 72 hours to support classroom formative assessments and planning for instruction. The assessment reports student performance according to two different score types: the Maine-specific scale score and RIT scores. Each score type serves a different purpose.

- Maine-specific scale scores communicate information about a student's performance according to grade-level standards.
- RIT scores communicate overall performance in the content area so that comparisons can

be made between students and over time.

- Instructional Area RIT scores communicate areas of relative strength and relative need.

To create a more complete understanding of what your students know and can do, results from the Maine Through Year Assessment should be used alongside additional data sources, such as school assessments and classroom learning.

Maine-Specific Scale Score

The Maine-specific scale score, also known as the Maine scale score, is a four-digit score that measures a student's achievement according to Maine's accountability standards, the Common Core State Standards, at the student's enrolled grade.

The range of possible Maine scale scores is partitioned into four distinct achievement levels: Well Below State Expectations, Below State Expectations, At State Expectations, and Above State Expectations.

Achievement Level Descriptors

An achievement level is a range of scores that defines a specific level of student achievement, as articulated in the achievement level descriptors (ALDs). The ALDs are a plain-language description of what students must know to fall into each of the achievement levels established through cut scores. The ALDs firmly root the cut scores and achievement levels in the content that students are supposed to learn. In qualitative and quantitative terms, the ALDs and cut scores together define the difference between a student who is performing at, below, or above grade-level expectations.

- **Well Below State Expectations**: On this assessment, students at this achievement level **demonstrate limited understanding of the knowledge and skills** necessary at this grade level, as specified in the Common Core State Standards. The students ***need substantial academic support*** to be prepared for the next grade level and to be on track for college and career readiness.
- **Below State Expectations**: On this assessment, students at this achievement level **demonstrate partial understanding of the knowledge and skills** necessary at this grade level, as specified in the Common Core State Standards. The students ***need additional academic support*** to be prepared for the next grade level and to be on track for college and career readiness.
- **At State Expectations**: On this assessment, students at this achievement level **demonstrate the knowledge and skills** necessary at this grade level, as specified in the Common Core State Standards. The students ***are prepared*** for the next grade level and are on track for college and career readiness.
- **Above State Expectations**: On this assessment, students at this achievement level **demonstrate advanced understanding of the knowledge and skills** necessary at this grade level, as specified in the Common Core State Standards. The students ***are well prepared*** for the next grade level and are well prepared for college and career readiness.

The cut scores for these achievement levels were established and validated in summer 2023 by Maine educators, the Maine DOE, and the Maine Technical Advisory Committee.

Setting the Cut Scores

To establish the cut scores, a process called “embedded standard setting” helps to determine two points along the scale score range (known as cut scores) that define the score range for each achievement level. Maine educators and stakeholders from around the state participated in the embedded standard-setting process for the Maine Through Year Assessment facilitated by edCount and Creative Measurement. The cut score recommendations from this statewide committee were presented to the Maine Department of Education and were approved in late August 2023.

Table 1: Math Scale Score Ranges by Grade below and Table 2: Reading Scale Score Ranges by Grade below show the scale score ranges for each achievement level by content area. The At State Expectations cut scores demark the minimum level of achievement considered to be proficient for accountability purposes. For example, Grade 5 At State Expectations cut scores are 1500-1524 for Reading and 1500-1524 for Mathematics.

Table 1: Math Scale Score Ranges by Grade

Grade	Well Below State Expectations	Below State Expectations	At State Expectations	Above State Expectations
3	1400-1485	1486-1499	1500-1524	1525-1600
4	1400-1487	1488-1499	1500-1524	1525-1600
5	1400-1483	1484-1499	1500-1524	1525-1600
6	1400-1480	1481-1499	1500-1524	1525-1600
7	1400-1481	1482-1499	1500-1524	1525-1600
8	1400-1483	1484-1499	1500-1524	1525-1600
HS	1400-1488	1489-1499	1500-1524	1525-1600

Table 2: Reading Scale Score Ranges by Grade

Grade	Well Below State Expectations	Below State Expectations	At State Expectations	Above State Expectations
3	1400-1482	1483-1499	1500-1524	1525-1600
4	1400-1485	1486-1499	1500-1524	1525-1600
5	1400-1486	1487-1499	1500-1524	1525-1600
6	1400-1485	1486-1499	1500-1524	1525-1600
7	1400-1482	1483-1499	1500-1524	1525-1600
8	1400-1483	1484-1499	1500-1524	1525-1600
HS	1400-1488	1489-1499	1500-1524	1525-1600

In the interpretation of assessment results, it is not appropriate to compare scale scores across content areas. Each content area is scaled separately; therefore, the scale scores for one content area cannot be compared to those for another content area.

RIT Scores

In addition to the Maine-specific scale score, student reports show a RIT score. The RIT score is provided on reports for the overall Math and Reading content areas.

The RIT scale measures levels in academic difficulty, and extends across all grades, making it possible to compare a student's score at various points throughout their education. RIT scores do not indicate a student's mastery of grade-level standards, so there are no specific RIT scores that correspond to "approaches", "meets", "exceeds", or any other performance level adjectives typically used by state summative tests. Instead, the RIT scores help place student achievement in the context of national norms. Using RIT scores, both achievement and growth can be reported according to percentiles.

In addition, instructional area RIT scores are provided for reporting categories within both Reading and Math, for example Literary Text and Informational Text. Instructional area RIT scores help educators identify areas of relative strength and relative need.

Table 3: Reading Instructional Area Descriptions for Grades 3–8 and High School

Literary Text	Informational Text	Vocabulary
Students read literary texts closely to determine key ideas and details, inferences, theme, and literary elements. Students will also analyze author's purpose, text structure, points of view, and texts with similar topics/themes.	Students read informational texts closely to determine key ideas and details, inferences, central ideas, and to summarize main ideas. Students will also analyze and compare how texts are structured, various representation of ideas, claims and supporting evidence, and author's purpose and/or point of view.	Students will focus on using context, Greek and Latin affixes, and reference materials in order to find the meaning of words, including general academic and domain-specific vocabulary. Students will interpret figurative language, understand the relationship between words, and distinguish between connotations and denotations.

Table 4: Math Instructional Area Descriptions for Grades 3–5

Operations and Algebraic Thinking	Numbers and Operations	Measurement and Data	Geometry
Students represent and solve problems involving the four operations and build skills related to patterns. Students also gain understanding of factors, multiples, the properties of multiplication, as well as the relationship between multiplication and division.	Students compare the values of numbers and build place value understanding of whole numbers and decimals. Students also perform operations with whole numbers, fractions, and decimals to solve real-world and mathematical problems.	Students represent and interpret data. Students also solve problems involving measurement and conversion of measurements. Lastly, students understand concepts of area, perimeter, volume, and angles.	Students classify shapes by their properties and graph points on the coordinate plane to solve real-world and mathematical problems.

The Math Instructional Area Descriptions for Grades 6+ are slightly different than the Math Instructional Area Descriptions for Grades 3–5, as shown in Table 5: Math Instructional Area Descriptions for Grades 6+ below.

Table 5: Math Instructional Area Descriptions for Grades 6+

Operations and Algebraic Thinking	The Real and Complex Number Systems	Geometry	Statistics and Probability
Students solve real-life and mathematical problems using numerical and algebraic expressions and equations, as well as linear and quadratic functions.	Students use ratio reasoning and units to solve problems. Students also use properties of rational and irrational numbers and reason quantitatively.	Students solve real-world and mathematical problems involving length, angle measure, area, surface area, and volume. Students also apply and prove geometric theorems. Finally, students understand geometric constructions as well as congruency and similarity transformations.	Students use statistical measures to summarize distributions. Students also understand random sampling, comparative inferences, and probability models. Lastly, students investigate patterns of association as well as represent and interpret data and linear models.

Available Reports

The following reports are described in this The Maine Through Year Assessment Reports Interpretive Guide. Please note that the data used in these reports within this guide are mocked and do not reflect actual results.

Report Link	Aggregation Level	Summary
School Administrative Unit (SAU) Report on page 159	SAU	Shows the average Maine scale scores for schools in the SAU, the distribution of school average Maine scale scores across the achievement levels, and the distribution of student Maine scale scores in each school.
School and Teacher Report on page 163	School	Shows the average Maine scale scores for students in the school, the distribution of student Maine scale scores across the achievement levels, the average Maine scale scores and score distribution for each group in the school, and the individual Maine scale scores for each student in the school.
School and Teacher Report on page 163	Group	Shows the average scale scores for students in the group, the distribution of student scale scores across the achievement levels, and the individual scale scores for each student in the group. The Teacher Report has a tab that links to the RIT Report.
Dynamic Student Report on page 170	Individual student	Shows all the details for an individual student's test.

Report Link	Aggregation Level	Summary
Individual Student Report (Spring Only) on page 174	Individual Student	Shows all tests in all available content areas for a student in this academic year. Designed for parents and families.
Part 4—RIT Report on page 177	Varies—based on user type	Shows RIT score information for all students matching the search criteria, including RIT score, achievement percentile, and reporting category RIT.
Part 5—Demographic Report (Spring Only) on page 180	Varies—based on user type	Shows the average scale scores, average reporting category scores, and distribution of scale scores for demographic groups such as gender, ethnicity, and targeted group.
Part 6—Comparison Summary Report (Spring Only) on page 183	School	Shows aggregate comparison of multiple organizations by grades, subjects, and student demographics.
Part 7—Students Results File on page 187	SAU and State	Downloadable export of student-level data at SAU and state levels during the test window.

Part 2—Organization Reports (Spring Only)

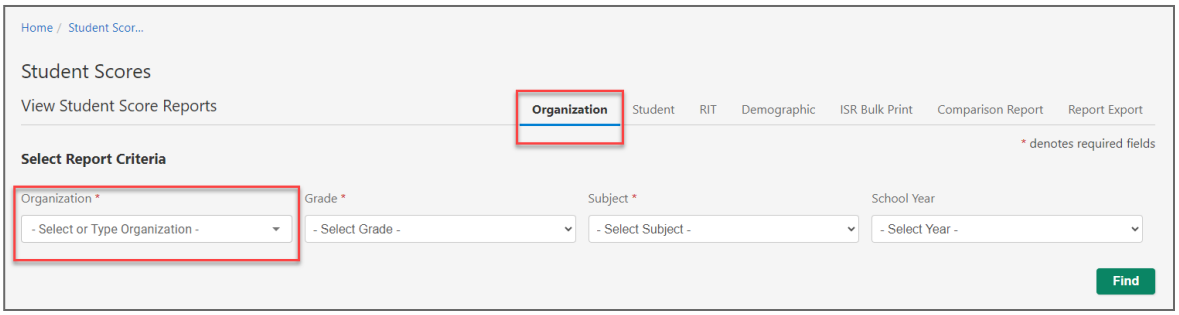
Note: Because the Organization Report uses the Maine-specific scale score, it is only available for the spring administrations.

School Administrative Unit (SAU) Report

The Organization Report for an SAU shows student performance by achievement level in each school in the SAU. Users with SAU-level access such as District Assessment Coordinators can access this report directly.

To generate an Organization Report at the SAU level:

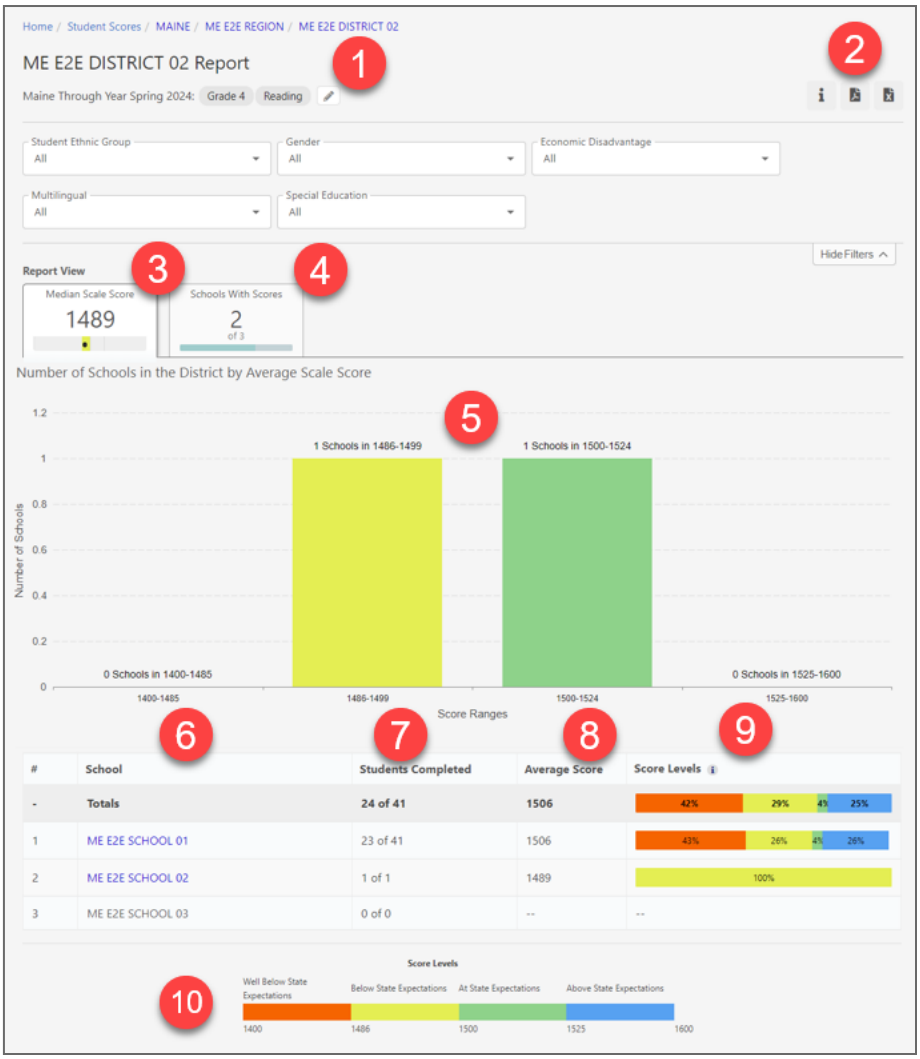
1. In Acacia, select **Menu** > **Reports** > **Student Scores**.
2. Verify that you are on the **Organization** tab. This is the default tab.
3. In the **Organization** field, select your SAU.



4. Select the other report criteria as desired, then select **Find**.
5. The report appears in the search results. Select the report to view it.

SAU Report: Histogram View

The histogram view of the SAU report contains bar graphs showing the number of schools with an average Maine scale score in each achievement level for the selected grade and content area.



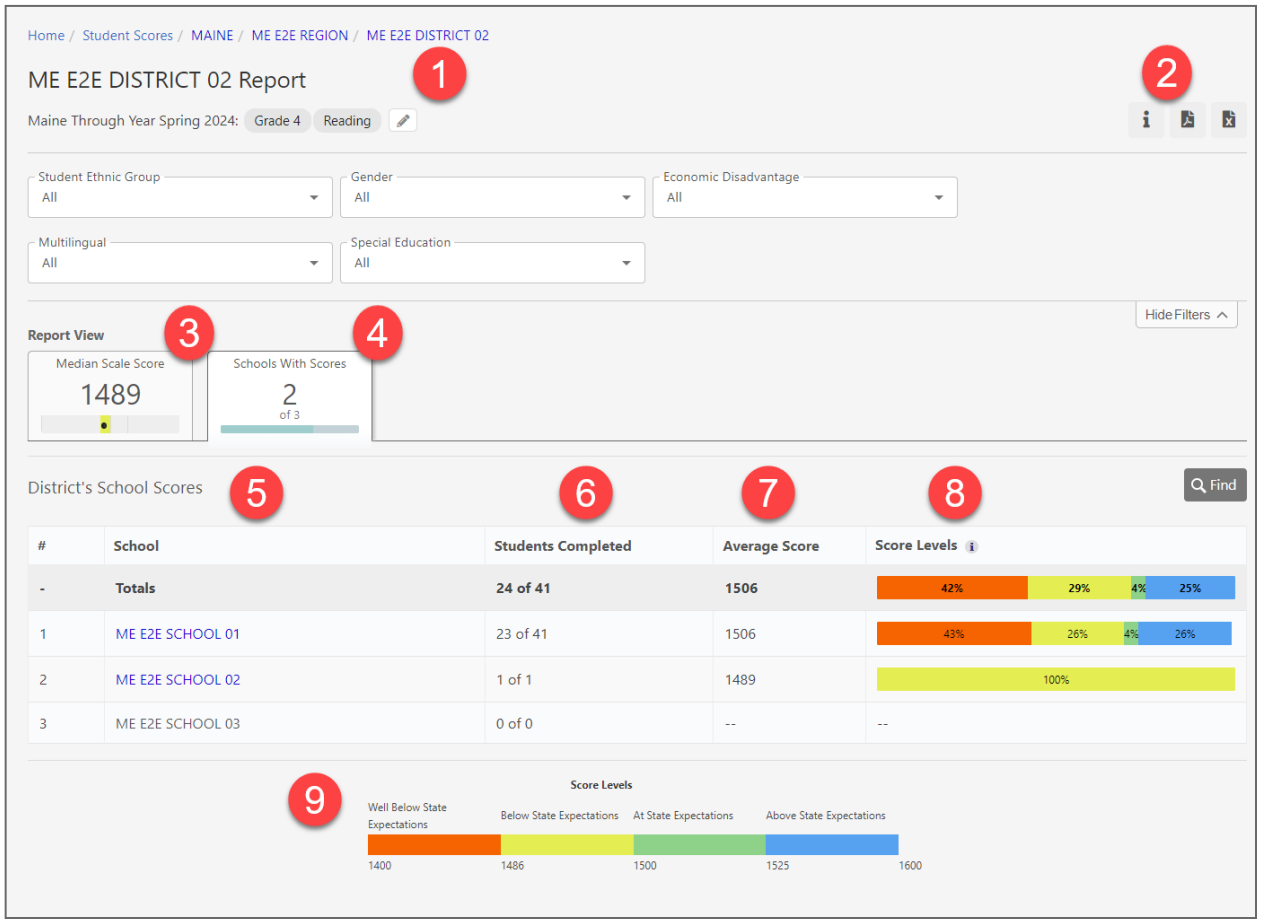
1. **Navigation header and report criteria**: Select the pencil icon to change the grade and content area, and use the menus to set demographic filters such as gender and ethnicity.
2. **Report actions**: Select the info icon to view report information. Select the PDF icon to print this report. Select the X file icon to download a CSV.
3. **Median* Scale Score**: The median Maine scale score for students in this grade at the district.
4. **Schools with Scores**: Select to switch to the list view. Refer to SAU Report: List View on page 162.
5. **Bar Graph**: Total number of schools with an average* Maine scale score lying in each achievement level.

6. **Schools**: Select an achievement level from the bar graph to see a list of schools with an average* Maine scale score in that achievement level. Select any school to view the school report. Refer to School Report: Histogram View on page 161.
7. **Students Completed**: Students with completed tests, out of the total number of registered students in the grade at this school.
8. **Average* Score**: The average, or mean, Maine scale score for students in this grade and content area.
9. **Score Levels**: The percentage of students at the school who scored in each achievement level.
10. **Score Levels (footer)**: The cut scores for each achievement level for this grade and content area.

***Medians and averages**: Note that all averages and medians are updated as students continue testing; wait until the end of the test window to make comparisons based on school, SAU, or state averages and medians.

SAU Report: List View

The list view shows each school in the SAU along with information about assessment results in the selected content area at that school. The list view also lists schools without any scores posted, which can help administrators track whether schools have begun testing.



1. **Navigation header and report criteria**: Select the pencil icon to change the grade and content area, and use the menus to set demographic filters such as gender and ethnicity.
2. **Report actions**: Select the info icon to view report information. Select the PDF icon to print this report. Select the X file icon to download a CSV.
3. **Median* Scale Score**: Select to switch to the histogram view. Refer to SAU Report: Histogram View on page 160.
4. **Schools with Scores**: Select to switch to the list view.
5. **Schools**: List of schools within the SAU.
6. **Students Completed**: Students in the grade and school with completed tests, out of the total number of registered students in the grade and school.
7. **Average* Score**: The average, or mean, Maine scale score for students in the grade and school with completed tests.

8. **Score Levels**: A visual representation of the distribution of Maine scale scores for students in the grade and school. It shows the percentage of students in each achievement level.
9. **Score Levels (footer)**: The cut scores for each achievement level for this grade and content area.

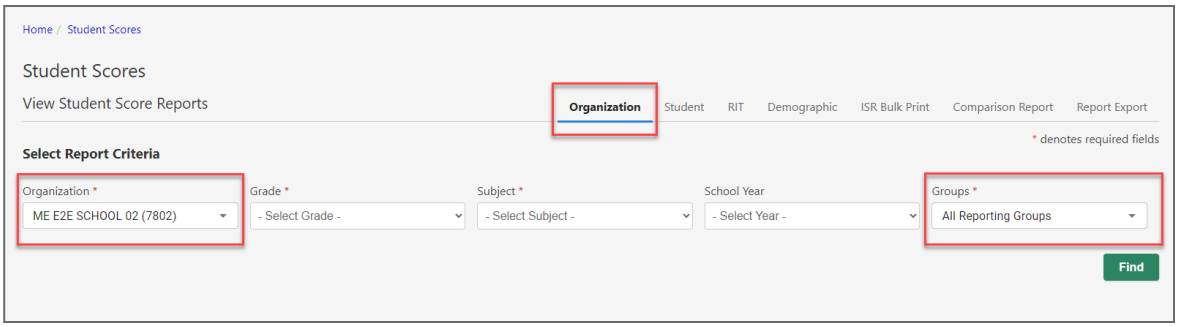
***Medians and averages**: Note that all averages and medians are updated as students continue testing; wait until the end of the test window to make comparisons based on school, SAU, or state averages and medians.

School and Teacher Report

The Organization Report for a school and for teachers shows student performance at the school or in a particular group. Users with school-level access can access these reports directly, while users with higher level access can drill down to this report from the SAU report, as well as access the report directly by choosing a school name.

To generate an Organization Report at the school level:

1. In Acacia, select **Menu** > **Reports** > **Student Scores**.
2. Verify that you are on the **Organization** tab. This is the default tab.
3. In the **Organization** field, select your school. Once a school is selected, the Groups field will become available.
4. In the **Groups** field, select a group, or choose **All Reporting Groups**.



5. Select the other report criteria as desired, then select **Find**.
6. The report appears in the search results. Select the report to view it.

School and Teacher Report: Histogram View

The histogram view of the school and teacher report contains bar graphs showing the number of students with a Maine scale score in each achievement level for the selected grade and content area.

1. **Navigation header and report criteria**: Select the pencil icon to change the grade and content area, and use the menus to set demographic filters such as gender and ethnicity.
2. **Report actions**: Select the info icon to view report information. Select the PDF icon to print this report.
3. **Median* Scale Score**: The median Maine scale score for students in this grade and school.
4. **Students Tested**: Select to switch to the list view. Refer to School and Teacher Report: List View on page 166.
5. **Median* RIT**: Select to switch to the RIT score view. Refer to School and Teacher Report: RIT View on page 168.
6. **Bar Graph**: Total number of students in each achievement level for this grade and school.

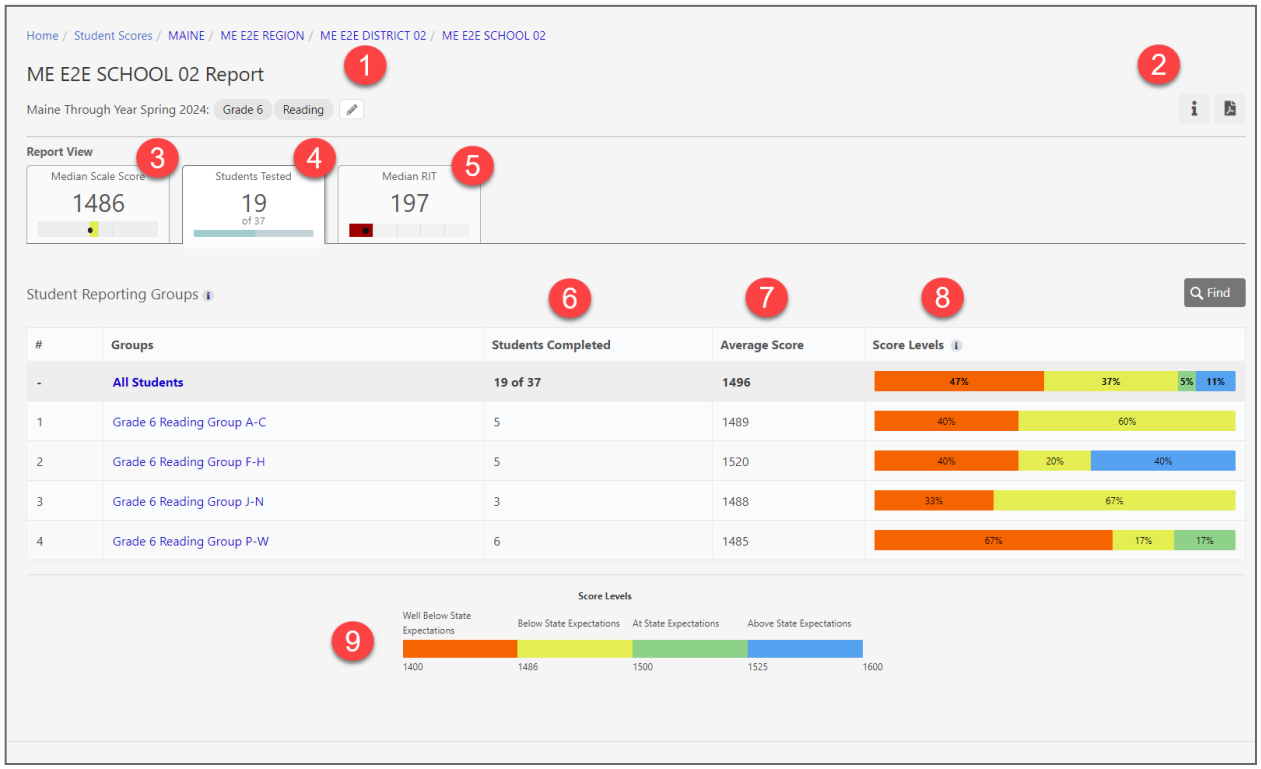
7. **Median* Score Comparison**: The median Maine scale scores for the grade at the school, region, SAU, and state level.
8. **Student Details**: Select an achievement level from the bar graph to see a list of students who scored in that achievement level. Select any column heading to sort. Choose any student to view their student report. Refer to Part 3—Student Reports on page 170.
9. **Current Achievement Level**: The achievement level associated with the student's Maine scale score.
10. **Score Range**: The student's Maine scale score and achievement level on the overall scale. The dot shows the student's Maine scale score, and the lines show the range. The range indicates that, if the student were tested again in similar circumstances, we would expect their score to be within the range.
11. **Score Levels (footer)**: The cut scores for each achievement level for this grade and content area.

***Medians and averages**: Note that all averages and medians are updated as students continue testing; wait until the end of the test window to make comparisons based on school, SAU, or state averages and medians.

School and Teacher Report: List View

The list view shows all reporting groups for the selected grade and content area at the school. For each reporting group, the number of students who have completed the test, the average score, and the distribution of scores across the achievement levels is listed.

At the top of the list of groups, select **All Students** to see information about all students in the selected grade at the school who were registered for the selected content area. Select a reporting group to view the teacher report for that group. Teachers will need to have been assigned to a reporting group to be able to have visibility.



1. **Navigation header and report criteria**: Select the pencil icon to change the grade and content area, and use the menus to set demographic filters such as gender and ethnicity.
2. **Report actions**: Select the info icon to view report information. Select the PDF icon to print this report.
3. **Median* Scale Score**: Select to switch to the histogram view. Refer to School and Teacher Report: Histogram View on page 164.
4. **Students Tested**: Students with completed tests, out of the total number of students in the grade and school.
5. **Median* RIT**: Select to switch to the RIT score view. Refer to School and Teacher Report: RIT View on page 168 on page 163.
6. **Students Completed**: Students in the reporting group with completed tests, out of the total number of students in the reporting group.

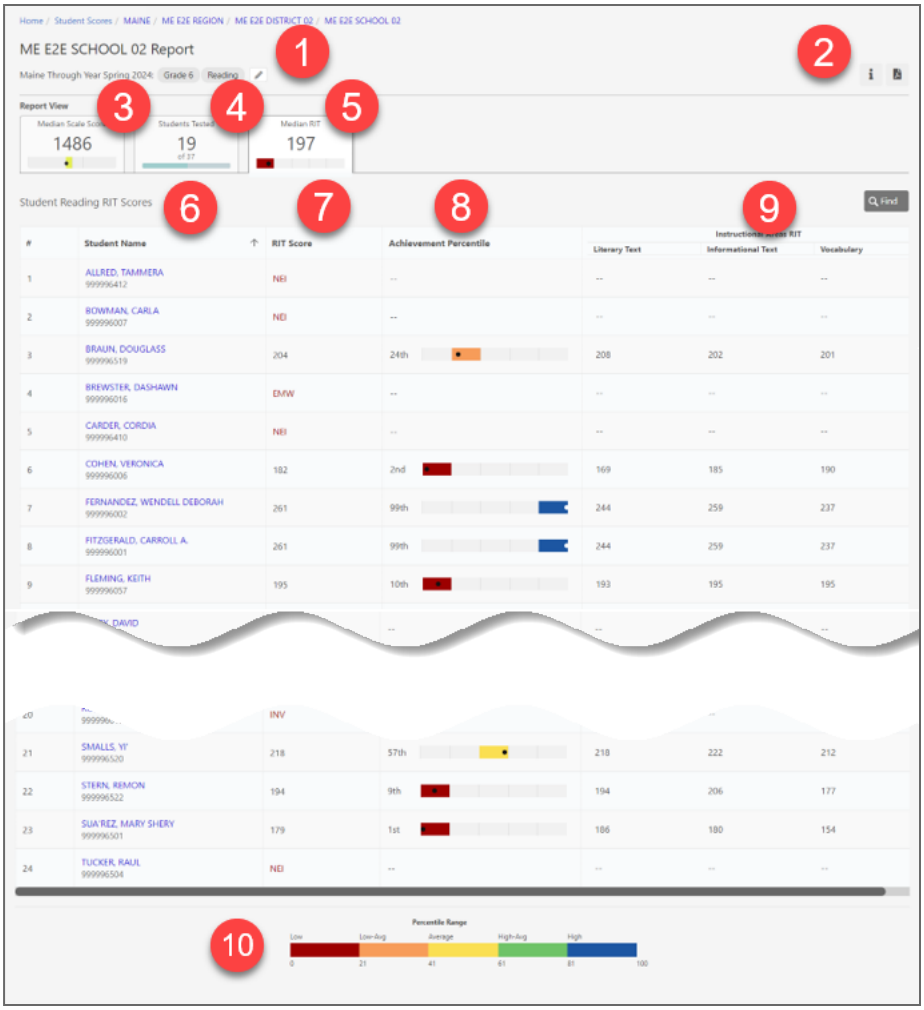
7. **Average* Score**: The average, or mean, Maine scale score for students in the reporting group with completed tests.
8. **Score Levels**: A visual representation of the distribution of Maine scale scores for students in the reporting group. It shows the percentage of students in each achievement level.
9. **Score Levels (footer)**: The cut scores for each achievement level for this grade and content area.

***Medians and averages**: Note that all averages and medians are updated as students continue testing; wait until the end of the test window to make comparisons based on school, SAU, or state averages and medians.

School and Teacher Report: RIT View

The list view shows all students in the selected grade and content area at the school.

The individual student RIT score data in this report is the same as the individual student data available in the RIT Report. Refer to Part 4—RIT Report on page 177.



1. **Navigation header and report criteria**: Select the pencil icon to change the grade and content area, and use the menus to set demographic filters such as gender and ethnicity.
2. **Report actions**: Select the info icon to view report information. Select the PDF icon to print this report.
3. **Median* Scale Score**: Select to switch to the histogram view. Refer to School and Teacher Report: Histogram View on page 174.
4. **Students Tested**: Select to switch to the list view. Refer to School and Teacher Report: List View on page 176.
5. **Median* RIT:** The median RIT score for students in this grade and content area. The dot represents the achievement percentile corresponding to the median RIT score, and the colored box behind the dot represents the quintile.
6. **Student Details**: Select any column heading to sort the student list.

7. **RIT Score**: The RIT score for this assessment. If the student has a not-tested code (NTC) instead of a score, the NTC will be displayed.
   For more information on RIT scores, refer to RIT Scores on page 156. For more information on NTCs, refer to Appendix A—Available NTCs on page 189.
8. **Achievement Percentile**: The dot represents the student's percentile ranking based on the RIT score. The colored box behind the dot represents the quintile.
9. **Instructional Area RIT**: The RIT score for each reporting category in this content area. Reporting category RIT scores are calculated based on student responses to the items in that category and cannot be averaged to generate an overall RIT.
10. **Percentile Range**: The range of percentiles that comprise each quintile.

***Medians and averages**: Note that all averages and medians are updated as students continue testing; wait until the end of the test window to make comparisons based on school, SAU, or state averages and medians.

Part 3—Student Reports

Dynamic Student Report

The Dynamic Student Report shows a student's achievement on the Maine Through Year Assessment. Separate reports are generated for Reading and Math.

In the fall and winter, the Dynamic Student Report shows the student's:

- Overall RIT score in the content area
- Instructional area RIT scores
- Item responses (correct, incorrect, or partially correct) by instructional area with general information about each item and the specific Common Core State Standard to which each item is aligned

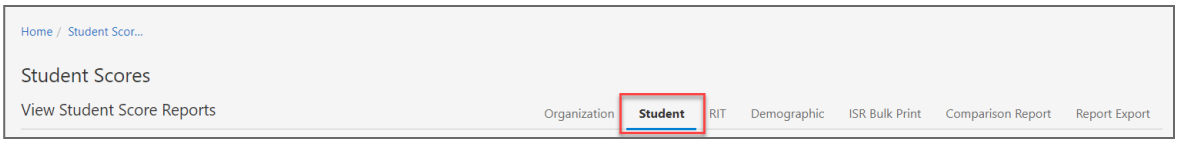
> **Note**: The Student Report available in Acacia is designed for use by educators. The Individual Student Report (ISR, available as a static PDF, is designed to communicate student performance on the Maine Through Year Assessment to families. Refer to Individual Student Report (Spring Only) on page 174 for details.

In the spring, the Dynamic Student Report also contains the student's Maine scale score and achievement level, as well as the average Maine scale score for the SAU/district.

The Dynamic Student Report is available online through Acacia.

To access the report for a student:

1. In Acacia, select **Menu** > **Reports** > **Student Scores**.
2. At the top right, select the **Student** tab.

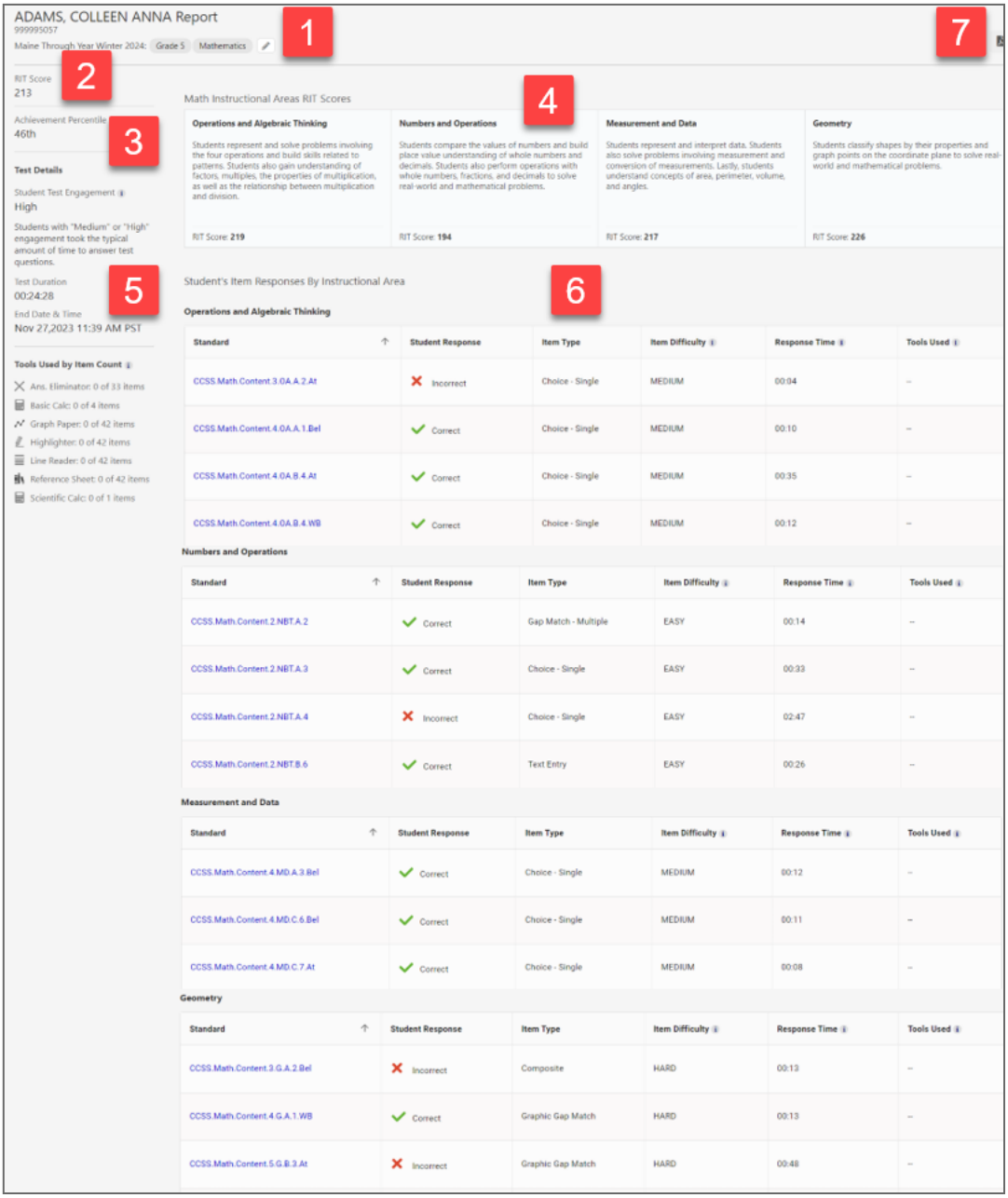


3. Enter the search criteria and select **Find**.
4. All reports matching the search criteria are listed. Select the report for the test administration and content area you wish to view.

   You can also drill down to a student through a report for any group, school, or SAU that you have access to.

Fall and Winter Dynamic Student Report

Figure 1: Sample Dynamic Student Report for Fall and Winter



1. **Header and report criteria**

   The header area of the report contains information about the student and their test, as well as navigation options. The student's name, state student ID, and the test administration window are listed. Below the state student ID, you can see the student's grade and switch between viewing different content areas.

2. **RIT Score**

   The student's RIT score for the content area is shown on the left, while the RIT score for each instructional area is in the Instructional Area RIT Scores section.

3. **Achievement Percentile**

   The student's achievement percentile compares the student to national norms, as reported in the 2020 MAP Growth norms. This achievement percentile is calculated using the RIT score.

   Note: The achievement percentile is calculated assuming a default number of weeks of instruction prior to testing. MAP Growth reports can have the number of weeks of instruction customized, so you may see different achievement percentiles for the same RIT score in MAP Growth reports.

4. **Instructional Area RIT Scores**

   This section shows the student's RIT score for each instructional area, and the instructional areas are briefly described.

5. **Test Details and Tools Used**

   Details about the student's test are listed here. The Student Test Engagement metric uses the presence of rapid responses to questions to infer whether students rushed through the test. Students with low engagement answered some questions very quickly, and the final score may not reflect the student's best effort. Students with medium or high engagement took the typical amount of time to answer test questions.

   In the Tools Used by Item Count section, select the info icon ℹ on the report to learn more about the tools the student used during the test.

6. **Student's Item Responses by Instructional Area**

   This section shows how the student responded to each assessment item in each Instructional Area. You can see whether the student's answer was correct, incorrect, or partially correct, the item's relative difficulty, the student's response time, and the tools used on that item. The available item types are:

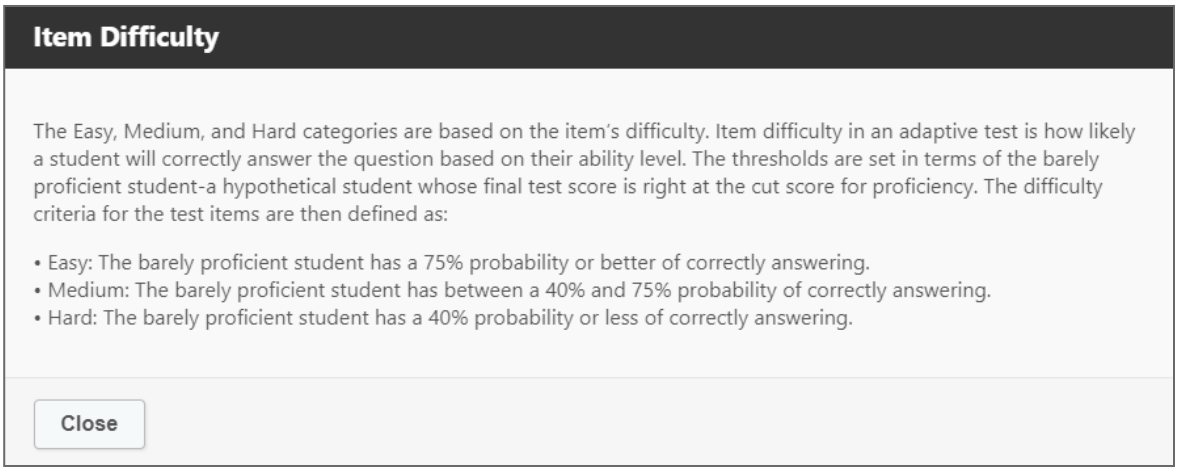
   - Choice – Single
   - Choice – Multiple
   - Composite
   - Gap Match
   - Graphic Gap Match
   - Hot Text
   - Text Entry

   The Common Core State Standards to which the item is aligned is also listed. Select the standard code to read the standard in full.

   Note: Each item assesses only one standard. Some items contribute to both the Maine scale score and RIT score; other items contribute only to the RIT score.

7. **Item Difficulty**

Clicking the information icon will open the following **Item Difficulty** description:



8. **Print**

   To print a PDF of the report, select the PDF icon on the bottom right.

Spring Dynamic Student Report

The spring Dynamic Student Report contains all of the student results data present in the fall and winter Dynamic Student Report, in addition to the two report features described below:

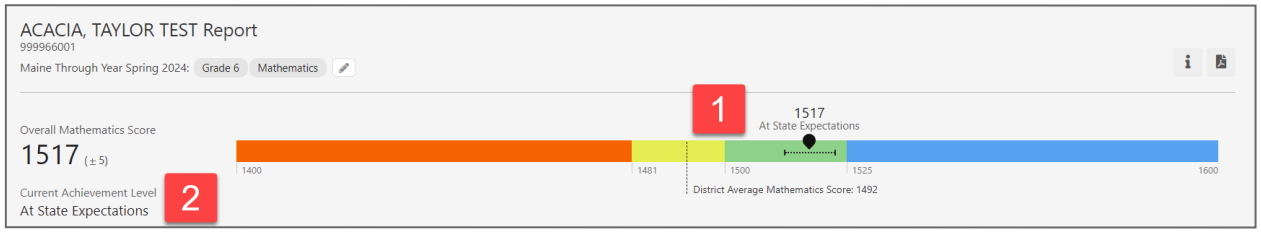
1. **Maine-Specific Scale Score**

   The student's Maine-specific scale score, also known as the Maine scale score, for this content area.

2. **Achievement Level**

   The student's current achievement level, determined by comparing their overall Maine scale score to the cut scores, is shown on the top left. Refer to Achievement Level Descriptors on page 154 and Setting the Cut Scores on page 155 for more details.

Figure 2: Sample Dynamic Student Report for Spring

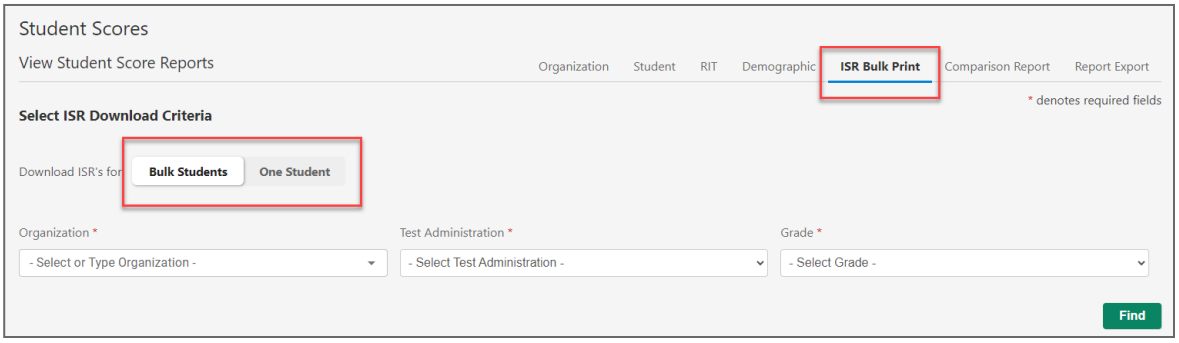


Individual Student Report (Spring Only)

The Individual Student Report (ISR) is designed to show a student's achievement on the Maine Through Year Reading and Mathematics Assessment to families. Educators can print these reports in batches, making it easy to distribute after testing is complete. The Individual Student Reports are generated for the spring assessment and will not be available for the fall and winter assessments.

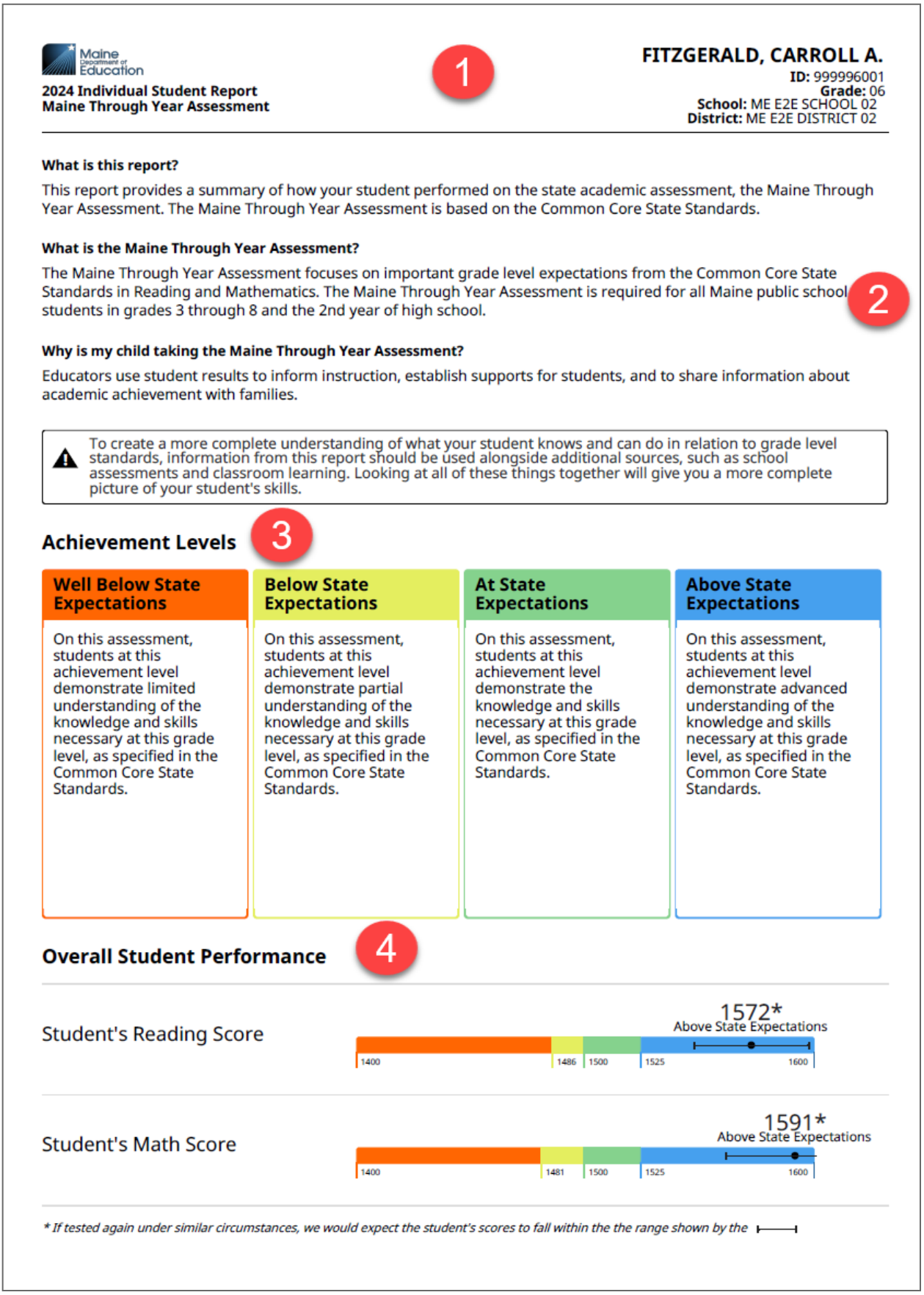
To generate the ISR for an individual student or a batch of students:

1. In Acacia, select **Menu** > **Reports** > **Student Scores**.
2. At the top right, select the **ISR Bulk Print** tab.
3. Select whether to download ISRs for **Bulk Students** or **One Student**.



4. Enter the search criteria and select **Find**.
5. All reports matching the search criteria are listed. For an individual student, select the PDF icon to download the report. For a batch of students, select the **Generate** button to download a zip file containing the ISRs for this batch of students.

Figure 3: Sample Individual Student Report



Maine Department of Education

2024 Individual Student Report
Maine Through Year Assessment

1

FITZGERALD, CARROLL A.
ID: 999996001
Grade: 06
School: ME E2E SCHOOL 02
District: ME E2E DISTRICT 02

What is this report?

This report provides a summary of how your student performed on the state academic assessment, the Maine Through Year Assessment. The Maine Through Year Assessment is based on the Common Core State Standards.

What is the Maine Through Year Assessment?

The Maine Through Year Assessment focuses on important grade level expectations from the Common Core State Standards in Reading and Mathematics. The Maine Through Year Assessment is required for all Maine public school students in grades 3 through 8 and the 2nd year of high school.

2

Why is my child taking the Maine Through Year Assessment?

Educators use student results to inform instruction, establish supports for students, and to share information about academic achievement with families.

⚠ To create a more complete understanding of what your student knows and can do in relation to grade level standards, information from this report should be used alongside additional sources, such as school assessments and classroom learning. Looking at all of these things together will give you a more complete picture of your student's skills.

Achievement Levels

3

Well Below State Expectations	Below State Expectations	At State Expectations	Above State Expectations
On this assessment, students at this achievement level demonstrate limited understanding of the knowledge and skills necessary at this grade level, as specified in the Common Core State Standards.	On this assessment, students at this achievement level demonstrate partial understanding of the knowledge and skills necessary at this grade level, as specified in the Common Core State Standards.	On this assessment, students at this achievement level demonstrate the knowledge and skills necessary at this grade level, as specified in the Common Core State Standards.	On this assessment, students at this achievement level demonstrate advanced understanding of the knowledge and skills necessary at this grade level, as specified in the Common Core State Standards.

Overall Student Performance

4

Student's Reading Score

1572*
Above State Expectations

1400 | 1486 | 1500 | 1525 | 1600

Student's Math Score

1591*
Above State Expectations

1400 | 1481 | 1500 | 1525 | 1600

** If tested again under similar circumstances, we would expect the student's scores to fall within the the range shown by the* ⊢⊣

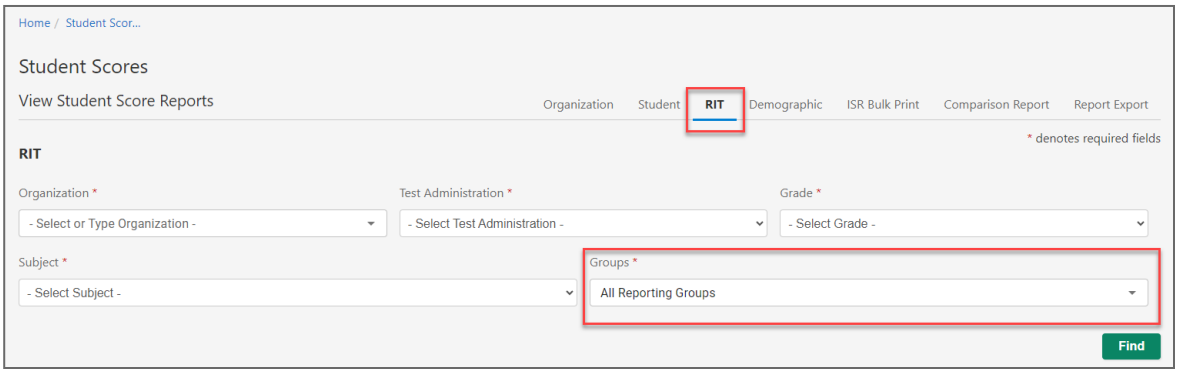
1. **Header**
   The header at the top of each page contains information about the student and the assessment. On the left is the assessment year and assessment name. On the right is the student's name, state student ID, grade, school, and district/SAU.
2. **Written Summary**
   The text at the top of page 150 provides context for families to help them understand the purpose of the Maine Through Year Assessment.
3. **Achievement Levels**
   Abbreviated achievement level descriptors (ALDs) are presented here. Refer to Achievement Level Descriptors on page 154 for more information.
4. **Overall Student Performance**
   The student's Maine-specific scale scores for Reading and Math. Refer to Maine-Specific Scale Score on page 154 for more information.

Part 4—RIT Report

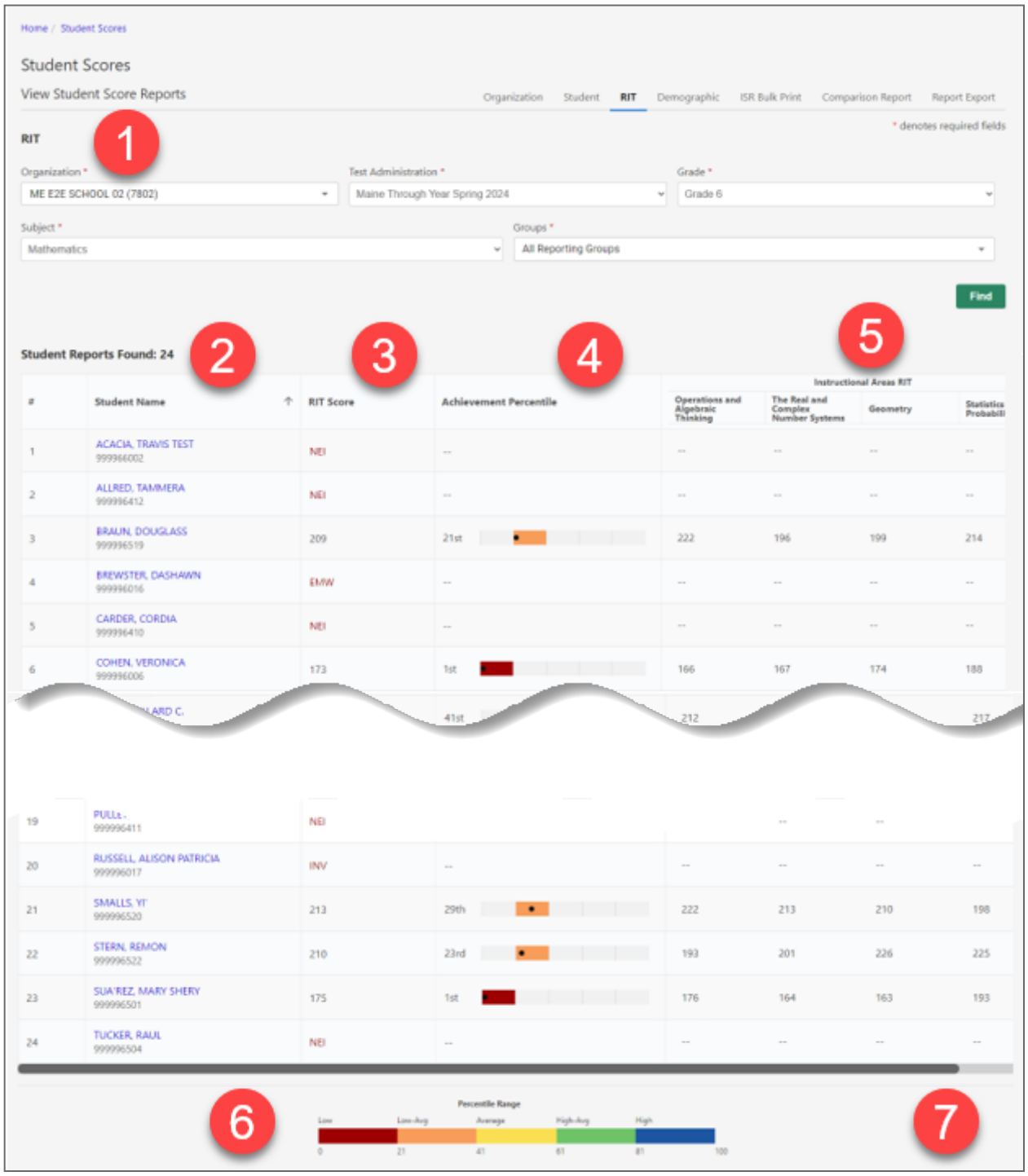
The RIT Report shows the RIT scores for students in a school, organized by grade level and student group. The student's overall RIT score in the content area is displayed, along with the achievement percentile and RIT score for each reporting category.

To generate a RIT report:

1. In Acacia, select **Menu** > **Reports** > **Student Scores**.
2. Select the **RIT** tab.
3. From the menus, select the organization, test administration, grade, subject, and groups as desired. To view all students in the organization, select **All Reporting Groups**.



4. Select **Find**.
5. A list of students matching the search criteria appears.
6. Select the student's name to view the Dynamic Student Report. Refer to Dynamic Student Report on page 170 for more details.



1. **Search Criteria**: Adjust the search criteria to view a different selection of students.
2. **Student Name**: The student's name and ID. Select the student's name to view the Dynamic Student Report for that student.
3. **RIT Score**: The RIT score for the student's test. If the student has a not-tested code (NTC), it will be listed instead of the RIT score.
   For more information on RIT scores, refer to RIT Scores on page 156. For more information on NTCs, refer to Appendix A—Available NTCs on page 189.

4. **Achievement Percentile**: The dot represents the student's percentile ranking based on the RIT score. The colored box behind the dot represents the quintile.
5. **Instructional Area RIT**: The RIT score for each instructional area in this content area. Instructional area RIT scores are calculated based on student responses to the items in that category and cannot be averaged to generate an overall RIT score.
6. **Percentile Range**: The percentile ranges are grouped into five categories: Low, Low-Average, Average, High-Average, and High. This chart shows the ranges of each of these categories. These categories do not correlate to achievement levels.
7. **Page Navigation**: If there are multiple pages of search results, use this section to move between pages of students.

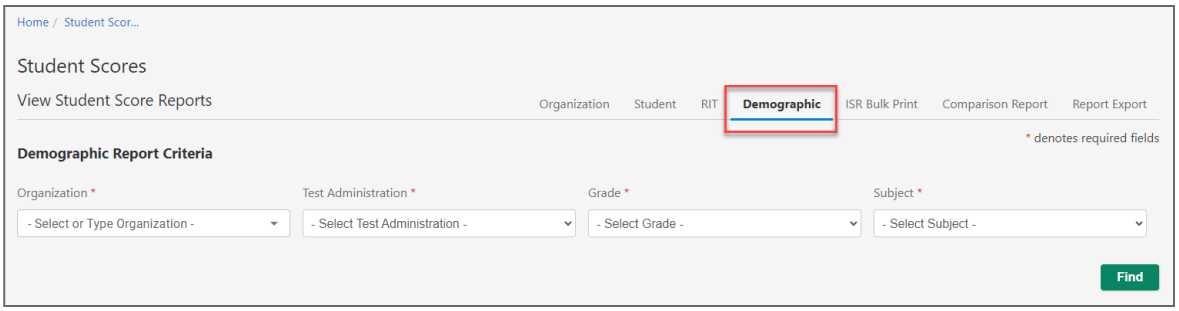
Part 5—Demographic Report (Spring Only)

The Demographic Report shows the average Maine scale score in the selected content area for students in various demographic or targeted groups. This report helps educators identify achievement trends for specific genders, ethnicities, or other groups such as multilingual learners or economically disadvantaged students. The Demographic Report will only be available with your Spring results and will not be available in the Fall and Winter. This report can be generated at the state, region, SAU, or school level.

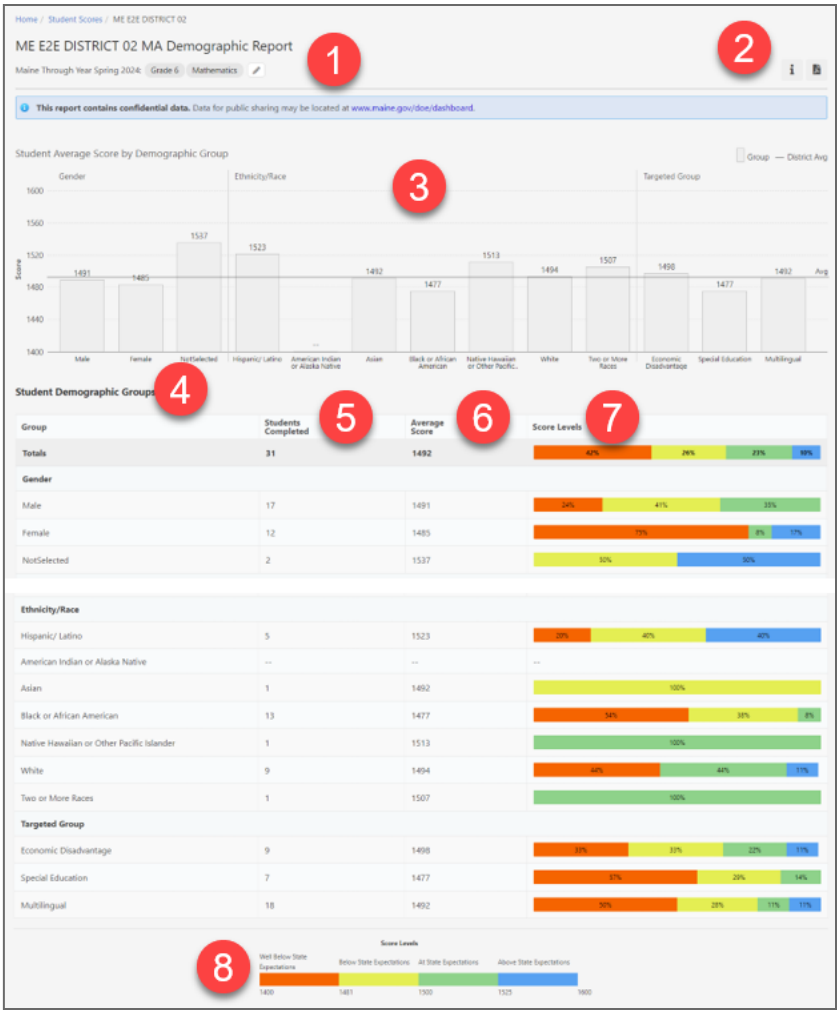
Note: Only Maine DOE has access to the state and regional Demographic Reports.

To generate a Demographic Report:

1. In Acacia, select **Menu** > **Reports** > **Student Scores**.
2. Select the **Demographic** tab.
3. From the menus, select the organization, test administration, grade, and content area as desired.



4. Select **Find**.
5. A list of available reports appears. Select the report you want to view.



1. **Navigation header and report criteria**: Select the pencil icon to change the grade and content area.
2. **Report actions**: Select theinfo icon to view report information. Select the PDF icon to print this report.
3. **Bar graph**: The average* Maine scale score for students in each demographic group. The solid line represents the district average for all students.
4. **Student Demographic Groups**: More details about the test results for each demographic group. The Totals line shows the information for all students.
5. **Students Completed**: The total number of students with completed test scores in this demographic group.
6. **Average* Score**: The average Maine scale score for students in this demographic group.
7. **Score Levels**: The distribution of scores across the achievement levels for students in this demographic group.

Note: The Multilingual (ML student group in the Demographic Report represents both students currently receiving ML-related services as well as those who have exited services and are in monitoring status. To disaggregate this group further and separate ML students currently receiving services from those in monitoring status, refer to Part 6—Comparison Summary Report (Spring Only) on page 183

8. **Score Levels (footer)**: The cut scores for each achievement level for this grade and content area.

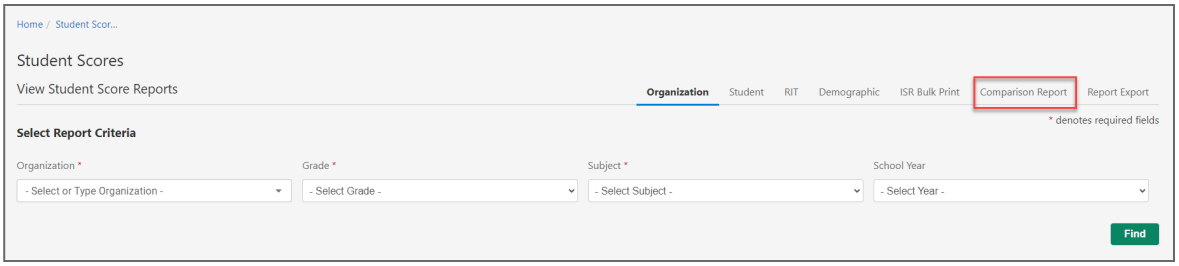
***Medians and averages**: Note that all averages and medians are updated as students continue testing; wait until the end of the test window to make comparisons based on school, SAU, or state averages and medians.

Part 6—Comparison Summary Report (Spring Only)

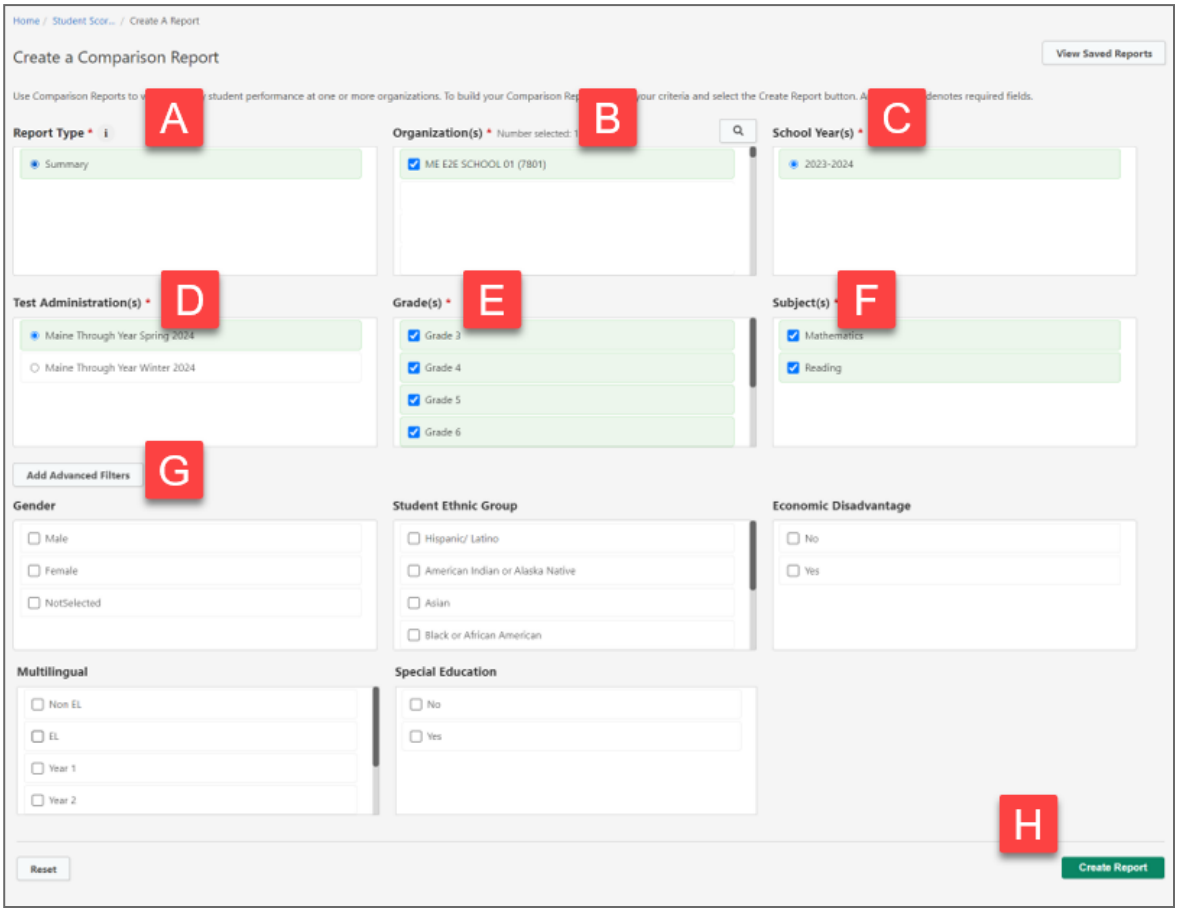
The Comparison Summary Report compares aggregate student performance at one or more organizations based on overall Maine scale score averages. The dynamic report creator supports multiple selections, pre-filtering student demographics, and the ability to rerun saved report queries quickly.

This report can be generated at the state, SAU, or school level. To create a Comparison Summary Report:

1. In Acacia, select **Menu** > **Reports** > **Student Scores**.
2. Select the **Comparison Report** tab.
3. From the menus, select the organization, grade, content area, and school year as desired.



4. The next screen will populate the report criteria selectors.



A. **Report Type**: Summary Report is the only option currently available.

B. **Organizations**: Multi-select search for the schools and SAUs to include in the report.

C. **School Year**: Single select from available school years.

D. **Test Administration**: Single select from available test administrations.

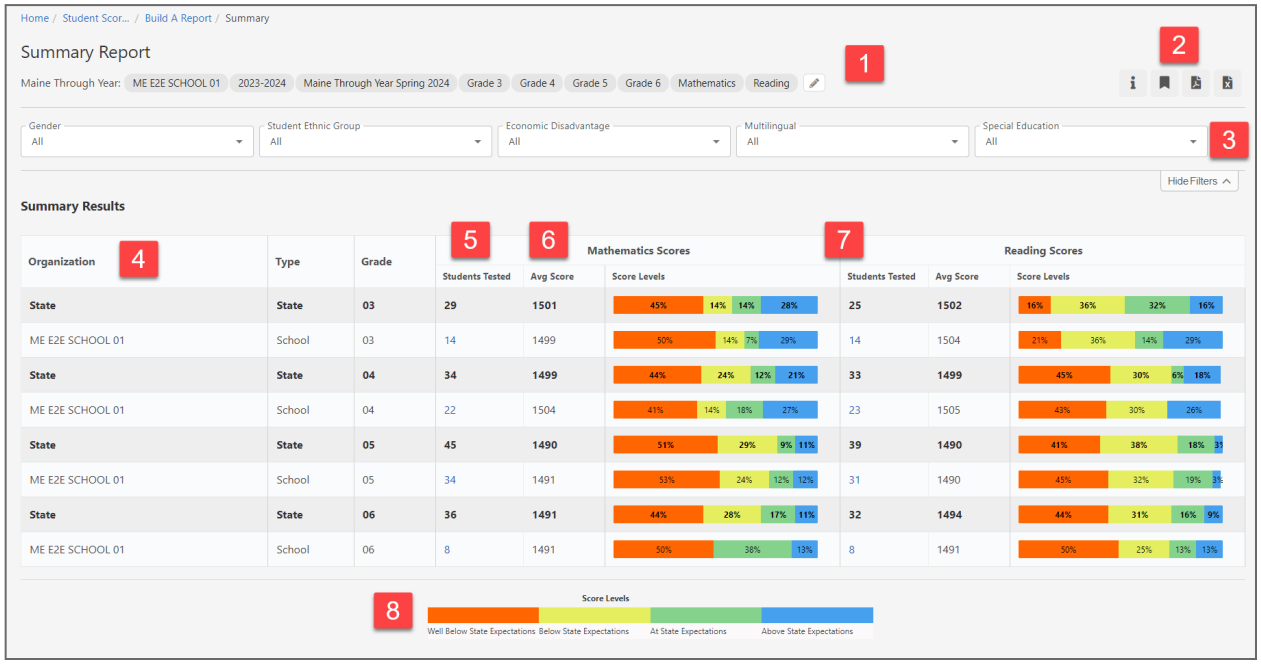
   Note: Only spring administrations will generate a report.

E. **Grades**: Multi-select from a list of grades.

F. **Subjects**: Multi-select from a list of subjects.

G. **Advanced Filter Options**: Select demographics to include.

H. **Create Report**: Once you have selected your criteria, select **Create Report**.



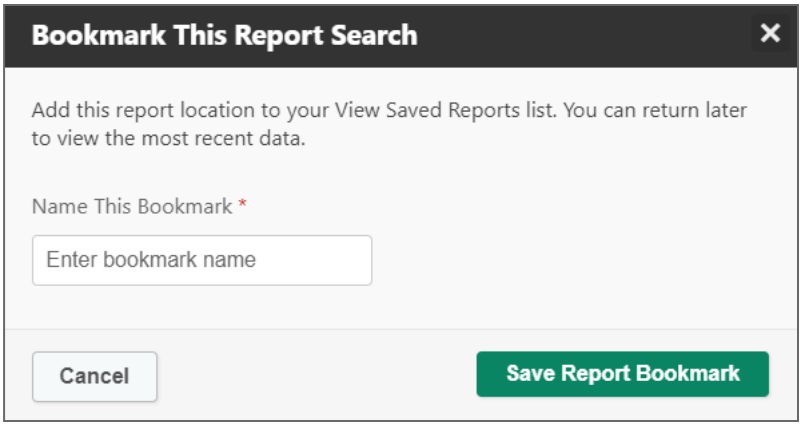
1. **Navigation Header and Report Criteria**: Displays the selected organization, school year, test administration, grade, and subject criteria. Use the pencil icon to update the grades and subjects shown.
2. **Report Actions**: Select the info icon to learn more about the data on the report. Select the bookmark icon to save this report. Select the PDF icon to print this report. Select the X file icon to download a CSV.
3. **Advanced Filters**: Displays the selected student demographics. Use the menus to change the demographic filters in the report. Select **Hide Filters/Show Filters** to hide or show the demographic filters.
4. **Organizational Identifiers**: Each row identifies an organization by name, organization type, and grade. State-level aggregation for each grade and subject criteria is displayed by default.
5. **Students Tested**: Count of students with reportable Maine scale scores in the grade and subject at the organization. Select any linked number in the Students Tested column to view the single-grade, single-subject organization reports.
6. **Average Score**: The mean Maine scale score of students for grade and subject.
7. **Score Levels**: A distribution graph of the percentage of students at the organization who scored in each achievement level for the grade and subject.
8. **Score Levels (footer)**: The achievement level names and colors for the selected test administration.

Bookmarking a Comparison Summary Report

The Bookmark feature allows users to quickly save Comparison Summary Reports. Throughout the administration window, the report will update with the latest student results data.

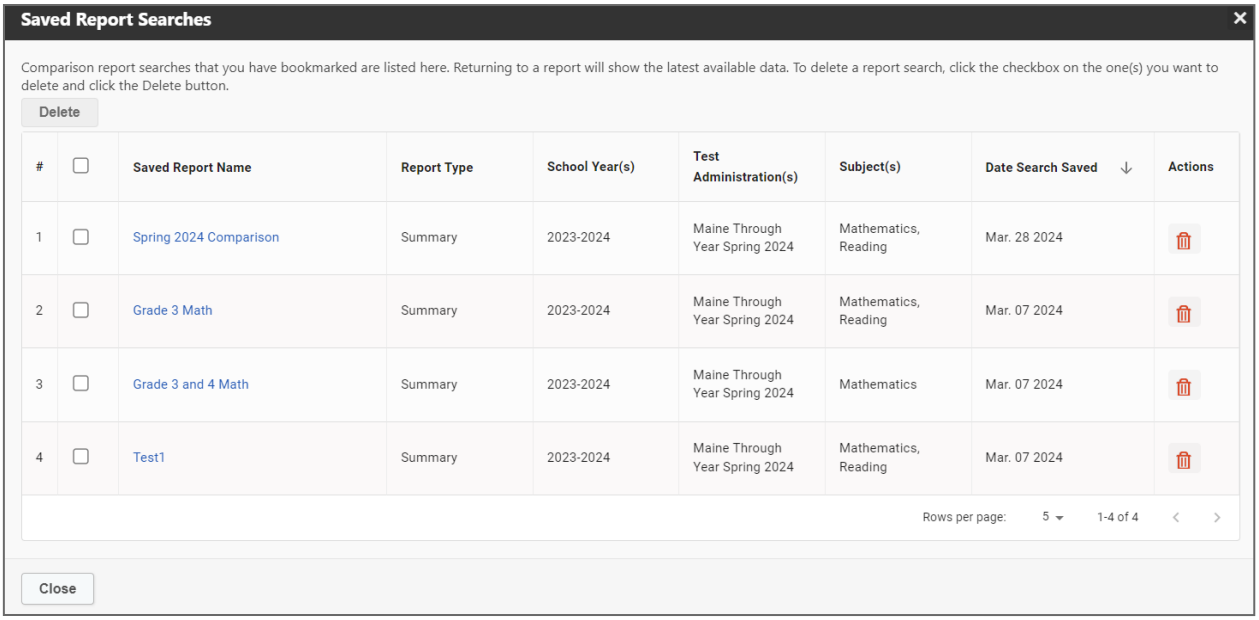
To bookmark a Comparison Summary Report:

1. On the **Summary Report** page, select the bookmark icon from the **Report Actions** section at the top of the report. The **Bookmark This Report Search** window opens.



2. Enter the name of the saved report criteria.
3. Select **Save Report Bookmark** to save the report search criteria.

Use the **View Saved Reports** button on the **Create a Comparison Report** page to view a list of saved reports.



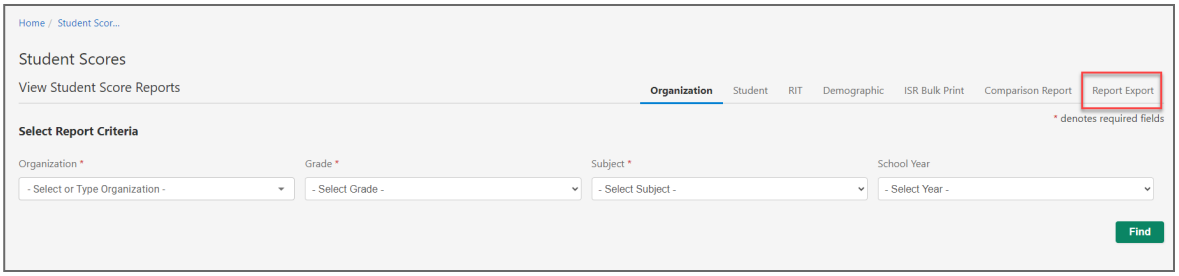
Part 7—Students Results File

The Students Results File allows a state representative or SAU administrator to download a CSV file of reportable student results for every assessment reported for the State of Maine or SAU within a test administration to facilitate further review and analysis. This point-in-time report gives accurate information at the time of processing and will update throughout the testing window.

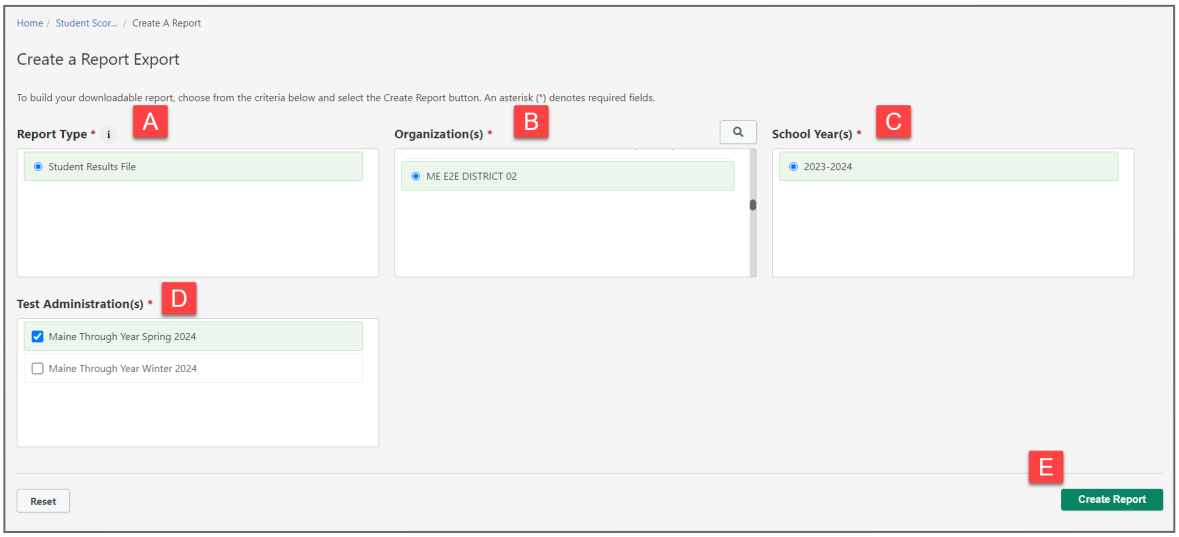
This report is available for the Fall, Winter, and Spring administrations. In the Fall and Winter, however, the Maine scale score columns will be blank.

This report can be generated at the state or SAU level. To create a Student Results File:

1. In Acacia, select **Menu** > **Reports** > **Student Scores**.
2. Select the **Report Export** tab.
3. From the menus, select the organization, grade, content area, and school year as desired.



4. The **Create a Report Export** screen appears. Choose from the criteria shown.



A. **Report Type**: Student Results File is the only option currently available.

B. **Organizations**: Single select the SAU to include in the report.

   Note: State representatives may also select Maine as the organization.

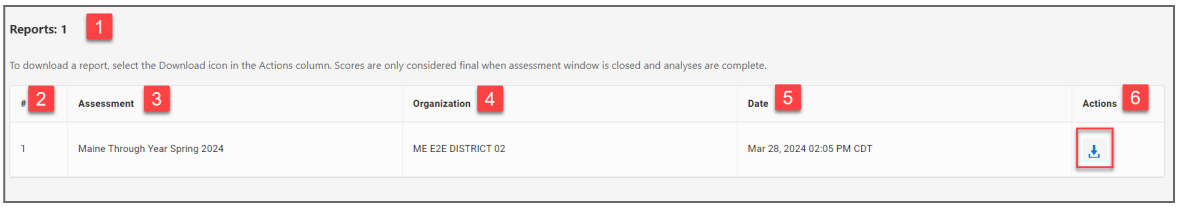
C. **School Year**: Single select from available school years.

D. **Test Administration**: Multi-select from available test administrations.

   Note: A separate Student Results File will be generated for each administration selected.

E. **Create Report**: Once you have selected your criteria, select Create Report.

5. After you select Create Report, the report download list appears.



A. **Report Criteria**: Indicates the number of report results for criteria selected.

B. **Row Count**: Number of each report row.

C. **Assessment**: Test administration name for the report criteria selected.

D. **Organization**: Name of the state or SAU included in the report.

E. **Date**: Date and time stamp when the report was most recently batch processed. The report updates every 4 hours during the assessment window until the final administration data validations are complete.

F. **Report Actions**: Allows the user to download the report to their desktop. If no download icon is available, then no report data is available.

Appendix A—Available NTCs

Not-tested codes (NTCs) are used solely by the Maine DOE to track special circumstances in which students' assessment data will not be included in an SAU's or school's aggregated data. Table 6: List of Reportable NTCs below lists the NTCs you may see on reports.

> Only Maine DOE will enter NTCs into the Acacia platform. SAUs should not enter NTCs, and any NTCs entered by SAUs will be removed.

Table 6: List of Reportable NTCs

Code	Description	Explanation of Use
INV	Invalid	Student's assessment was invalidated, such as due to a security breach. Requires Maine DOE approval, and Maine DOE will assign the NTC.
EMV	Emergency Medical Waiver	Student was not tested because of an approved emergency medical waiver (special considerations request). Requires Maine DOE approval, and Maine DOE will assign the NTC.
RMV	Removal	Used by NWEA behind-the-scenes as necessary for transferring assessment scores and resetting assessments. All RMV NTCs are removed by NWEA prior to the generation of the Student Score Data File.

Appendix B—Not Enough Items Code (NEI)

The attemptedness code of Not Enough Items (NEI) can be applied to both the Maine-specific scale score and RIT scores.

Note: Maine-specific scale scores will be available in for Spring administration only and RIT scores will be available for the Fall, Winter, and Spring administrations.

For the Maine-specific scale score, at least 25% of operational items must be completed to receive a Maine scale score. For the RIT scores, all operational items must be completed to receive RIT scores.

NEIs are not included in aggregate calculations for reporting groups, schools, SAUs, or the state.

NEIs are not included in the Demographic Report since NEIs are not included in the aggregate calculations.

Appendix G: Maine Through Year Summative Blueprints—Reading and Math

Reading – Grade 3

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Literary Text	12 - 14	12 - 16	45-50%	6 - 7
RL.3.1, RL.3.2, RL.3.3, RL.3.5, RL.3.6, RL.3.7, RL.3.9				
Informational Text	8 - 9	8 - 11	30-35%	4 - 5
RI.3.1, RI.3.2, RI.3.3, RI.3.5, RI.3.6, RI.3.7, RI.3.8, RI.3.9				
Vocabulary	5 - 7	5 - 9	20-25%	3 - 4
RI.3.4, RL.3.4, L.3.4 (a – d), L.3.5 (a, c), L.3.6				
TOTAL	27	30 - 31		14

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Text Type	Literary Text - Approximate %	Informational Text - Approximate %
Summative	55-60%	40-45%

Standard Grade Preferences	Targets
Summative	The assessment should contain at least 70% on-grade content overall and per instructional area (19 – 27 items). The assessment should contain no more than 30% adjacent grade(s) content overall and per instructional area (0 – 8 items).

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

Reading – Grade 4

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Literary Text	11 – 12	11 - 14	40-45%	6
RL.4.1, RL.4.2, RL.4.3, RL.4.5, RL.4.6, RL.4.7, RL.4.9				
Informational Text	9 – 11	9 - 13	35-40%	5 - 6
RI.4.1, RI.4.2, RI.4.3, RI.4.5, RI.4.6, RI.4.7, RI.4.8, RI.4.9				
Vocabulary	5 – 7	5 - 9	20-25%	3 - 4
RI.4.4, RL.4.4, L.4.4 (a – c), L.4.5 (a – c), L.4.6				
TOTAL	27	30 - 31		14

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Text Type	Literary Text - Approximate %	Informational Text - Approximate %
Summative	55-60%	40-45%

Standard Grade Preferences	Targets
Summative	The assessment should contain at least 70% on-grade content overall and per instructional area (19 – 27 items). The assessment should contain no more than 30% adjacent grade(s) content overall and per instructional area (0 – 8 items).

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

Reading – Grade 5

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Literary Text	9 - 11	9 - 13	35-40%	5 - 6
RL.5.1, RL.5.2, RL.5.3, RL.5.5, RL.5.6, RL.5.7, RL.5.9				
Informational Text	9 – 11	9 - 13	35-40%	5 - 6
RI.5.1, RI.5.2, RI.5.3, RI.5.5, RI.5.6, RI.5.7, RI.5.8, RI.5.9				
Vocabulary	5 – 7	5 - 9	20-25%	3 - 4
RI.5.4, RL.5.4, L.5.4 (a – c), L.5.5 (a – c), L.5.6				
TOTAL	27	30 - 31		14

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Text Type	Literary Text - Approximate %	Informational Text - Approximate %
Summative	50%	50%

Standard Grade Preferences	Targets
Summative	The assessment should contain at least 70% on-grade content overall and per instructional area (19 – 27 items). The assessment should contain no more than 30% adjacent grade(s) content overall and per instructional area (0 – 8 items).

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

Reading – Grade 6

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Literary Text	9 - 11	9 - 13	35-40%	5 - 6
RL.6.1, RL.6.2, RL.6.3, RL.6.5, RL.6.6, RL.6.9				
Informational Text	9 – 12	9 - 14	40-45%	6
RI.6.1, RI.6.2, RI.6.3, RI.6.5, RI.6.6, RI.6.7, RI.6.8, RI.6.9				
Vocabulary	5 – 7	5 - 9	20-25%	3 - 4
RI.6.4, RL.6.4, L.6.4 (a – d), L.6.5 (a – c), L.6.6				
TOTAL	27	30 - 31		14

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Text Type	Literary Text - Approximate %	Informational Text - Approximate %
Summative	45 - 50%	50 - 55%

Standard Grade Preferences	Targets
Summative	The assessment should contain at least 70% on-grade content overall and per instructional area (19 – 27 items). The assessment should contain no more than 30% adjacent grade(s) content overall and per instructional area (0 – 8 items).

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

Reading – Grade 7

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Literary Text	8 - 9	8 - 11	30-35%	4 - 5
RL.7.1, RL.7.2, RL.7.3, RL.7.5, RL.7.6, RL.7.9				
Informational Text	12 – 14	12 - 16	45-50%	6 - 7
RI.7.1, RI.7.2, RI.7.3, RI.7.5, RI.7.6, RI.7.7, RI.7.8, RI.7.9				
Vocabulary	5 – 7	5 - 9	20-25%	3 - 4
RI.7.4, RL.7.4, L.7.4 (a – d), L.7.5 (a – c), L.7.6				
TOTAL	27	30 - 31		14

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Text Type	Literary Text - Approximate %	Informational Text - Approximate %
Summative	40 - 45%	55 - 60%

Standard Grade Preferences	Targets
Summative	The assessment should contain at least 70% on-grade content overall and per instructional area (19 – 27 items). The assessment should contain no more than 30% adjacent grade(s) content overall and per instructional area (0 – 8 items).

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

Reading – Grade 8

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Literary Text	8 - 9	8 - 11	30-35%	4 - 5
RL.8.1, RL.8.2, RL.8.3, RL.8.5, RL.8.6, RL.8.9				
Informational Text	12 – 14	12 - 16	45-50%	6 - 7
RI.8.1, RI.8.2, RI.8.3, RI.8.5, RI.8.6, RI.8.7, RI.8.8, RI.8.9				
Vocabulary	5 – 7	5 - 9	20-25%	3 - 4
RI.8.4, RL.8.4, L.8.4 (a – d), L.8.5 (a – c), L.8.6				
TOTAL	27	30 - 31		14

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Text Type	Literary Text - Approximate %	Informational Text - Approximate %
Summative	40 - 45%	55 - 60%

Standard Grade Preferences	Targets
Summative	The assessment should contain at least 70% on-grade content overall and per instructional area (19 – 27 items). The assessment should contain no more than 30% adjacent grade(s) content overall and per instructional area (0 – 8 items).

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

Reading – 2nd Year of High School / Grade 10

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Literary Text	8	8 - 10	30-35%	3 - 5
RL.9-10.1, RL.9-10.2, RL.9-10.3, RL.9-10.5, RL.9-10.6, RL.9-10.9				
Informational Text	14	14 - 16	45-50%	4 - 7
RI.9-10.1, RI.9-10.2, RI.9-10.3, RI.9-10.5, RI.9-10.6, RI.9-10.7, RI.9-10.8, RI.9-10.9				
Vocabulary	8	8 - 10	20-25%	3 - 4
RI.9-10.4, RL.9-10.4, L.9-10.4 (a – d), L.9-10.5 (a – b), L.9-10.6				
TOTAL	30	33 - 34		12

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Text Type	Literary Text - Approximate %	Informational Text - Approximate %
Summative	40 - 45%	55 - 60%

Standard Grade Preferences	Targets
Summative	The assessment should contain 100% on-grade level content (30 items)

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

Math – Grade 3

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Operations and Algebraic Thinking	6	6 - 8	23 - 25%	4
3.OA.A.1, 3.OA.A.2, 3.OA.A.3, 3.OA.A.4, 3.OA.B.5, 3.OA.B.6, 3.OA.C.7, 3.OA.D.8, 3.OA.D.9				
Numbers and Operations	9	9 - 11	33 – 35%	4
3.NBT.A.1, 3.NBT.A.2, 3.NBT.A.3, 3.NF.A.1, 3.NF.A.2 (a-b), 3.NF.A.3 (a-d)				
Measurement and Data	8	8 - 10	28 – 30%	4
3.MD.A.1, 3.MD.A.2, 3.MD.B.3, 3.MD.B.4, 3.MD.C.5, 3.MD.C.6, 3.MD.C.7 (a-d), 3.MD.D.8				
Geometry	4	4 - 6	13-15%	6
3.G.A.1, 3.G.A.2				
TOTAL	27	30 - 31		18

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Standard Grade Preferences	Targets
Summative	The assessment should contain at least 70% on-grade content overall and per instructional area (19 – 27 items). The assessment should contain no more than 30% adjacent grade(s) content overall and per instructional area (0 – 8 items).

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

Math – Grade 4

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Operations and Algebraic Thinking	5	5 - 7	18 -20%	5
4.OA.A.1, 4.OA.A.2, 4.OA.A.3, 4.OA.B.4, 4.OA.B.5				
Numbers and Operations	13	13 - 15	48 - 50%	4
4.NBT.A.1, 4.NBT.A.2, 4.NBT.A.3, 4.NBT.B.4, 4.NBT.B.5, 4.NBT.B.6, 4.NF.A.1, 4.NF.A.2, 4.NF.B.3 (a-d), 4. NF.B.4 (a-c), 4.NF.C.5, 4.NF.C.6, 4.NF.C.7				
Measurement and Data	5	5 - 7	18 - 20%	4
4.MD.A.1, 4.MD.A.2, 4.MD.A.3, 4.MD.B.4, 4.MD.C.5, 4.MD.C.6, 4.MD.C.7				
Geometry	4	4 - 6	13 - 15%	5
4.G.A.1, 4.G.A.2, 4.G.A.3				
TOTAL	27	30 - 31		18

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Standard Grade Preferences	Targets
Summative	The assessment should contain at least 70% on-grade content overall and per instructional area (19 – 27 items). The assessment should contain no more than 30% adjacent grade(s) content overall and per instructional area (0 – 8 items).

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

Math – Grade 5

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Operations and Algebraic Thinking	4	4 - 6	13 - 15%	5
5.OA.A.1, 5.OA.A.2, 5.OA.B.3				
Numbers and Operations	14	14 - 16	53 – 55%	4
5.NBT.A.1, 5.NBT.A.2, 5.NBT.A.3 (a-b), 5.NBT.A.4, 5.NBT.B.5, 5.NBT.B.6, 5.NBT.B.7, 5.NF.A.1, 5.NF.A.2, 5.NF.B.3, 5. NF.B.4 (a-b), 5.NF.B.5 (a-b), 5.NF.B.6, 5.NF.B.7(a-c)				
Measurement and Data	5	5 - 7	18 – 20%	4
5.MD.A.1, 5.MD.B.2, 5.MD.C.3 (a-b), 5.MD.C.4, 5.MD.C.5 (a-c)				
Geometry	4	4 - 6	13 – 15%	5
5.G.A.1, 5.G.A.2, 5.G.B.3, 5.G.B.4				
TOTAL	27	30 - 31		18

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Standard Grade Preferences	Targets
Summative	The assessment should contain at least 70% on-grade content overall and per instructional area (19 – 27 items). The assessment should contain no more than 30% adjacent grade(s) content overall and per instructional area (0 – 8 items).

There are differences between the Grade 3, 4, 5 reporting categories and the Grade 6, 7, 8, and 2nd Year High School instructional areas. If off-grade items are selected, they will be items that align to the instructional area for the grade the test is intended for. For example, a 5th grade item selected on a 6th grade test will be aligned to a 6th grade instructional area which may differ from the instructional area the item aligns to on the 5th grade assessment.

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

Math – Grade 6

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Operations and Algebraic Thinking	7	7 - 9	25%	4
6.EE.A.1, 6.EE.A.2 (a-c), 6.EE.A.3, 6.EE.A.4, 6.EE.B.5, 6.EE.B.6, 6.EE.B.7, 6.EE.B.8, 6.EE.C.9				
The Real and Complex Number Systems	12	12 - 14	45%	4
6.RP.A.1, 6.RP.A.2, 6.RP.A.3 (a-d), 6.NS.A.1, 6.NS.B.2, 6.NS.B.3, 6.NS.B.4, 6.NS.C.5, 6.NS.C.6 (a-c), 6.NS.C.7 (a-d), 6.NS.C.8				
Geometry	4	4 - 6	15%	5
6.G.A.1, 6.G.A.2, 6.G.A.3, 6.G.A.4				
Statistics and Probability	4	4 - 6	15%	5
6.SP.A.1, 6.SP.A.2, 6.SP.A.3, 6.SP.B.4, 6.SP.B.5 (a-d)				
TOTAL	27	30 - 31		18

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Standard Grade Preferences	Targets
Summative	The assessment should contain at least 70% on-grade content overall and per instructional area (19 – 27 items). The assessment should contain no more than 30% adjacent grade(s) content overall and per instructional area (0 – 8 items).

There are differences between the Grade 3, 4, 5 reporting categories and the Grade 6, 7, 8, and 2nd Year High School instructional areas. If off-grade items are selected, they will be items that align to the instructional area for the grade the test is intended for. For example, a 5th grade item selected on a 6th grade test will be aligned to a 6th grade instructional area which may differ from the instructional area the item aligns to on the 5th grade assessment.

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

Math – Grade 7

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Operations and Algebraic Thinking	5	5 - 7	20%	5
7.EE.A.1, 7.EE.A.2, 7.EE.B.3, 7.EE.B.4 (a-b)				
The Real and Complex Number Systems	11	11 - 13	40%	4
7.RP.A.1, 7.RP.A.2 (a-d), 7.RP.A.3, 7.NS.A.1 (a-d), 7.NS.A.2 (a-d), 7.NS.A.3				
Geometry	6	6 - 8	20%	4
7.G.A.1, 7.G.A.2, 7.G.A.3, 7.G.B.4, 7.G.B.5, 7.G.B.6				
Statistics and Probability	5	5 - 7	20%	5
7.SP.A.1, 7.SP.A.2, 7.SP.B.3, 7.SP.B.4, 7.SP.C.5, 7.SP.C.6, 7.SP.C.7 (a-b) , 7.SP.C.8 (a-c)				
TOTAL	27	30 - 31		18

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Standard Grade Preferences	Targets
Summative	The assessment should contain at least 70% on-grade content overall and per instructional area (19 – 27 items). The assessment should contain no more than 30% adjacent grade(s) content overall and per instructional area (0 – 8 items).

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

Math – Grade 8

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Operations and Algebraic Thinking	13	13 - 15	48-53%	4
8.EE.A.1, 8.EE.A.2, 8.EE.A.3, 8.EE.A.4, 8.EE.B.5, 8.EE.B.6, 8.EE.C.7 (a-b), 8.EE.C.8 (a-c), 8.F.A.1, 8.F.A.3, 8.F.A.3, 8.F.B.3, 8.F.B.5				
The Real and Complex Number Systems	4	4 - 6	13-15%	5
8.NS.A.1, 8.NS.A.2				
Geometry	6	6 - 8	21-23%	4
8.G.A.1(a-c), 8.G.A.2, 8.G.A.3, 8.G.A.4, 8.G.A.5, 8.G.B.6, 8.G.B.7, 8.G.B.8, 8.G.C.9				
Statistics and Probability	4	4 - 6	13-15%	5
8.SP.A.1, 8.SP.A.2, 8.SP.A.3, 8.SP.A.4				
TOTAL	27	30 - 31		18

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Standard Grade Preferences	Targets
Summative	The assessment should contain at least 70% on-grade content overall and per instructional area (19 – 27 items). The assessment should contain no more than 30% adjacent grade(s) content overall and per instructional area (0 – 8 items).

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

Math – 2nd Year of High School / Grade 10

Instructional Areas	Summative			MAP Growth
	Item Count	Approximate Points	Approximate %	Item Count
On-Grade Summative Standards				
Operations and Algebraic Thinking	14	14 - 16	46-50%	4
HSA-APR.A.1, HSA-CED.A.1, HSA-CED.A.2, HSA-CED.A.3, HSA-CED.A.4, HSA-REI.A.1, HSA-REI.B.3, HSA-REI.B.4 (b), HSA-REI.C.5, HSA-REI.C.6, HSA-REI.C.7, HSA-REI.D.10, HSA-REI.D.11, HSA-REI.D.12, HSA-SSE.A.1, HSA-SSE.A.2, HSA-SSE.B.3 (a), HSF-BF.A.1 (a-b), HSF-BF.A.2, HSF-BF.B.3, HSF-IF.A.1, HSF-IF.A.2, HSF-IF.A.3, HSF-IF.B.4, HSF-IF.B.5, HSF-IF.B.6, HSF-IF.C.7(a, e), HSF-IF.C.8 (a), HSF-IF.C.9				
The Real and Complex Number Systems	4	4 - 6	13-15%	4
HSN-RN.A.1, HSN-RN.A.2, HSN-RN.B.3, HSN-Q.A.1, HSN-Q.A.2, HSN-Q.A.3				
Geometry	8	8 - 10	26-30%	4
HSG-C.A.1, HSG-C.A.2, HSG-C.A.3, HSG-CO.A.1, HSG-CO.A.2, HSG-CO.A.3, HSG-CO.A.4, HSG-CO.A.5, HSG-CO.B.6, HSG-CO.B.7, HSG-CO.B.8, HSG-CO.C.9, HSG-CO.C.10, HSG-CO.C.11, HSG-CO.D.12, HSG-CO.D.13, HSG-GMD.A.3, HSG-GMD.B.4, HSG-GPE.A.1, HSG-GPE.B.4, HSG-GPE.B.5, HSG-GPE.B.6, HSG-GPE.B.7, HSG-SRT.A.1 (a-b), HSG-SRT.A.2, HSG-SRT.A.3, HSG-SRT.B.4, HSG-SRT.B.5, HSG-SRT.C.6, HSG-SRT.C.7, HSG-SRT.C.8				
Statistics and Probability	4	4 - 6	13-15%	5
HSS-CP.A.1, HSS-CP.A.2, HSS-CP.A.4, HSS-ID.A.1, HSS-ID.A.2, HSS-ID.A.3, HSS-ID.B.5, HSS-ID.B.6 (a, c), HSS-ID.C.7				
TOTAL	30	33 - 34		17

Due to the adaptive nature of the test, some counts may reflect targets rather than requirements.

Standard Grade Preferences	Targets
Summative	The assessment should contain at least 70% on-grade content overall and per instructional area (21 – 27 items). The assessment should contain no more than 30% adjacent grade(s) content overall and per instructional area (0 – 9 items).

Achievement Level Coverage	Targets
Summative	No more than 40% of items should align to “Well Below” or “Below” (0 – 11 items). At least 60% of the items should align to “At” or “Above” (12 – 27 items).

August 2023 TAC Meeting

Maine Through Year Assessment
Technical Advisory Committee Meeting
August 18, 2023
9:00am – 5:00pm ET

Attendees

Name	Org	Role	Invited	Attended
Martha Thurlow	TAC	Senior Research Associate, National Center on Educational Outcomes/TIES Center at the University of Minnesota		
April Zenisky, EdD	TAC	Research Associate Professor and Director of Computer-Based Testing Initiatives, Center for Educational Assessment at UMass Amherst		
Nathan Dadey, PhD	TAC	Senior Associate, The National Center for the Improvement of Educational Assessment		
Richard Luecht	TAC	Professor of Educational Research Methodology, UNC Greensboro		
James Pellegrino	TAC	Liberal Arts and Sciences Distinguished Professor and Founding Co-director of the Learning Sciences Research Institute, University of Illinois Chicago		

*** Denotes not in attendance**

Agenda Topics

August 18, 2023, 9:00 am – 3:00 pm (ET)	Time	Presenter
Welcome • Opening Comments	9:00am – 9:10am	ME DOE
Introduction • Introductions	9:10am – 9:20am	Fred Valenzuela
NWEA Psychometrics • Calibration Results • Scaling Method • Technical Report Template Review and Discussion	9:20am – 11:30am	Kwang-lee / Yong Luo

August 2023 TAC Meeting

• Comparability Evidence • Score Comparisons		
Lunch	11:30am – 12:30pm	
Standard Setting Technical Report Embedded Standard Setting and Alignment Study Discussion and Cut Score Review and Discussion	12:30pm – 2:30pm	Dan Lewis
Break	2:30pm – 2:45pm	
Debrief • Notes and TAC Recommendation Review	2:45pm – 3:30pm	
TAC/DOE Closed Session	3:30pm	

Appendix I: August 2023 Content Bias Review

Demographic information about Maine educators who participated in the August 2023 Content Bias Review:

Mathematics Panel Demographic Information

Current Job Title	***N***	***%***
Classroom Teacher	3	75.0%
Other – Classroom and Coach	1	25.0%
School Administrative Unit	***N***	***%***
RSU 23	1	25.0%
RSU 2	1	25.0%
Kittery Schools	1	25.0%
RSU 10	1	25.0%
Gender	***N***	***%***
Female	4	100.0%
Male	0	0.0%
Race/Ethnicity	***N***	***%***
Hispanic or Latino	0	0.0%
American Indian or Alaska Native	0	0.0%
Asian	0	0.0%
Black or African American	0	0.0%
Native American	0	0.0%
White	4	100.0%
Years of Experience in Education	***N***	***%***
1 to 10 years	1	25.0%
11 to 20 years	2	50.0%
21 or more years	1	25.0%
Highest Level of Education	***N***	***%***
Bachelor's degree	0	0.0%
Master's degree	3	75.0%
Doctoral degree	1	25.0%
Experience Teaching Special Education Students	***N***	***%***
Yes	6	100.0%
No	0	0.0%
Experience Teaching Multilingual Learners	***N***	***%***
Yes	3	75.0%
No	1	25.0%

Experience Teaching Gifted Students	N	%
Yes	4	100.0%
No	0	0.0%
Experience Teaching a Class with a High Percentage of Economically Disadvantaged Students	**N**	**%**
Yes	4	100.0%
No	0	0.0%

Reading Panel Demographic Information

Current Job Title	N	%
Classroom Teacher	4	66.6%
Other – Write-in	2	16.7%
Curriculum Specialist		16.7%
School Administrative Unit	**N**	**%**
MSAD 41	1	16.7%
Well Ogunquit Community	1	16.7%
MSAD 58	1	16.7%
RSU 24	1	16.7%
Maine Virtual Academy	1	16.7%
Auburn School Dept	1	16.7%
Gender	**N**	**%**
Female	6	100.0%
Male	0	0.0%
Race/Ethnicity	**N**	**%**
Hispanic or Latino	0	0.0%
American Indian or Alaska Native	0	0.0%
Asian	0	0.0%
Black or African American	0	0.0%
Native American	0	0.0%
White	6	100.0%
Years of Experience in Education	**N**	**%**
1 to 10 years	1	16.7%
11 to 20 years	2	33.3%
21 or more years	3	50.0%
Highest Level of Education	**N**	**%**
Bachelor's degree	2	33.3%
Master's degree	2	33.3%

Doctoral degree	0	0.0%
Write In	2	33.3%
Experience Teaching Special Education Students	***N***	***%***
Yes	6	100.0%
No	0	0.0%
Experience Teaching Multilingual Learners	***N***	***%***
Yes	5	83.3%
No	1	16.7%
Experience Teaching Gifted Students	***N***	***%***
Yes	5	83.3%
No	1	16.7%
Experience Teaching a Class with a High Percentage of Economically Disadvantaged Students	***N***	***%***
Yes	5	83.3%
No	1	16.7%

nwea

Virtual Content and Bias Review Prework

August 8th and 9th

Measuring What Matters™

Overview of Prework

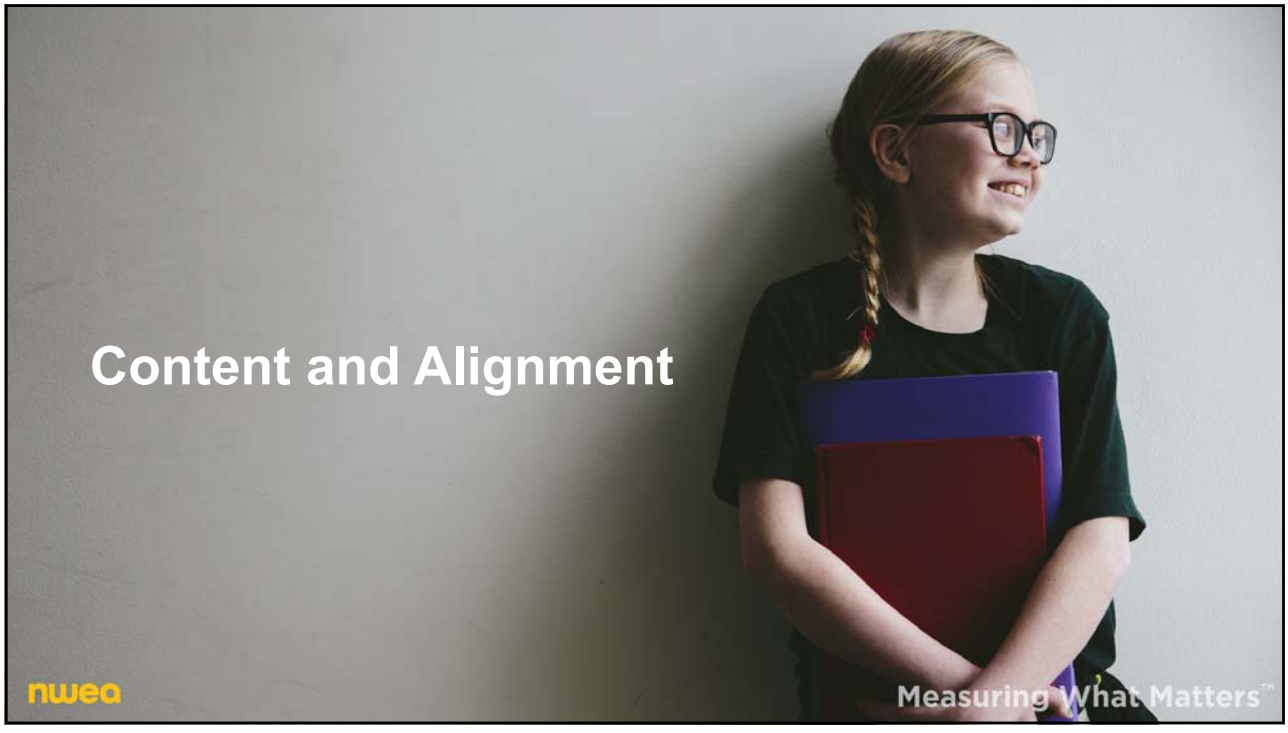
- You will independently review newly developed grade 10 items and provide feedback to determine if the items should be considered field test eligible.
- During your review you will decide if these items should be accepted as is, modified, or rejected. Modified items will be revised based on input from committee members. Accepted and revised items will be eligible for fielding testing on the Spring Summative Assessment.
- You will provide your initial feedback using the NWEA Review Portal where you can review passages and items and leave comments on individual items. NWEA content specialists will then compile all comments and mark items for discussion during the virtual Content and Bias Review meeting. All items marked as modify or reject will be discussed.

nwea

Important Dates

- Monday, July 17th - Independent educator review starts using the Review Portal.
- Monday, July 31st - Independent educator reviews end at 5:00 pm ET.
- Tuesday, August 1st to Monday, August 7th – NWEA content specialists download feedback from the Review Portal and review and analyze educator comments.
- Tuesday, August 8th and Wednesday, August 9th - Virtual Content and Bias Review from 9:00 am to 2:00 pm EST with a 30-minute break for lunch.

nwea | 3



What needs to be reviewed?

- Please review the items to determine if the following criteria are met:
  - Items are free of bias or sensitivity issues
  - Item and passage content is accurate and appropriate for the grade level
  - Items are aligned correctly to the standard and the Range ALD
  - Item DOK is correct
  - Items score correctly

nwea | 5

What are Range ALDs?

- Range Achievement Level Descriptors
  - Based on Common Core State Standards
  - Describe what a student should likely be able to do at a particular level regarding on-grade content.
- Range ALDs show the range of on-grade content for the grade from easiest or least cognitively challenging to the most difficult or most cognitively challenging.
- Range ALDs Demonstrate coherence within an individual standard and within the grade, in addition to coherence across grades.
- The Range ALDs were reviewed and revised in September 2022 by Maine educators.

nwea | 6

Using the Range ALDs?

- Consider the alignment to the RALDs when reviewing the items.
- We need a variety of RALDs to build grade appropriate adaptive assessments, so we need to distribute items across the range.

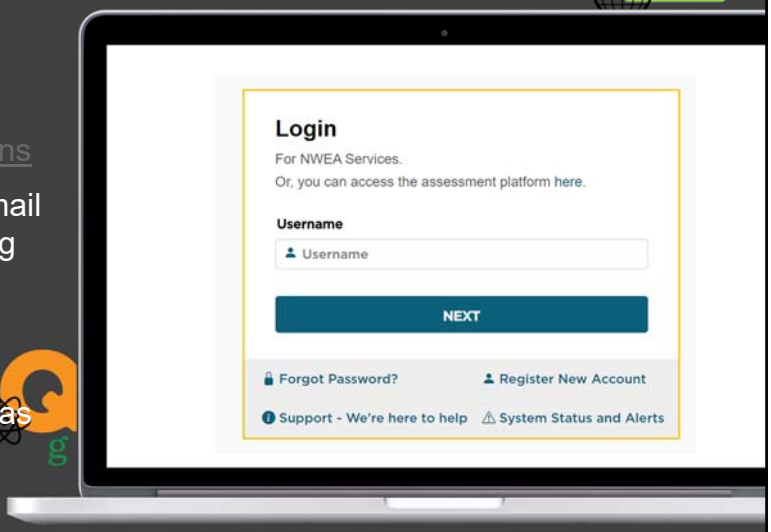
nwea | 7

Review Portal Login

- Please enter the following URL into the search:

https://review.nwea.org/sessions

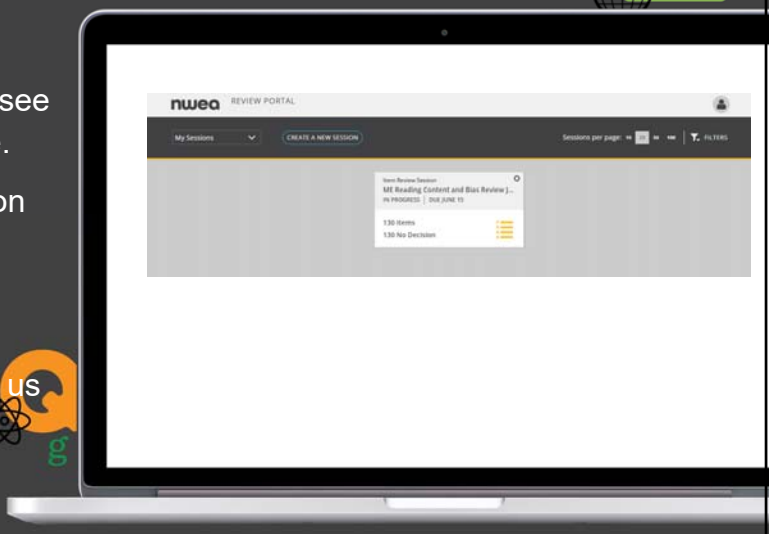
- You must log in with the email provided to us when signing up with NWEA and Maine.
- If you are unsure of which email address you used, please contact us as soon as possible.



nwea

Review Portal Sessions

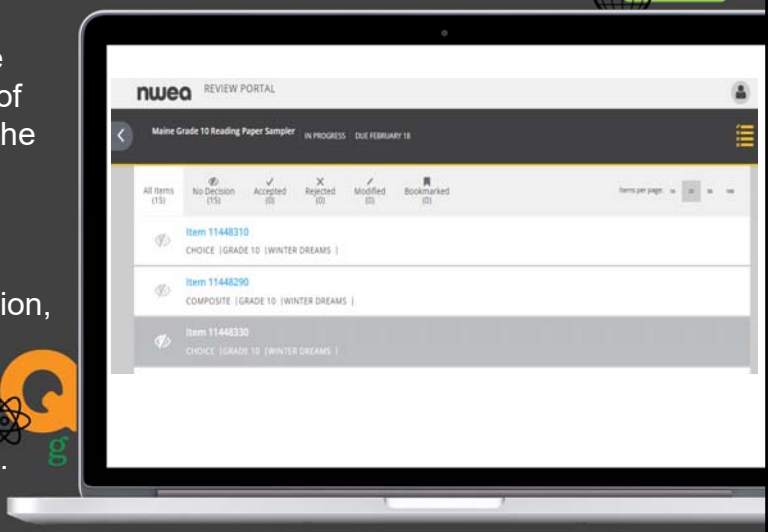
- Once you have logged into Review Portal, you should see a screen with a session tile.
- You will only see the session to which you have been assigned.
- If you do not see a session assigned to you, please let us know.



nwea

Review Portal Toolbar

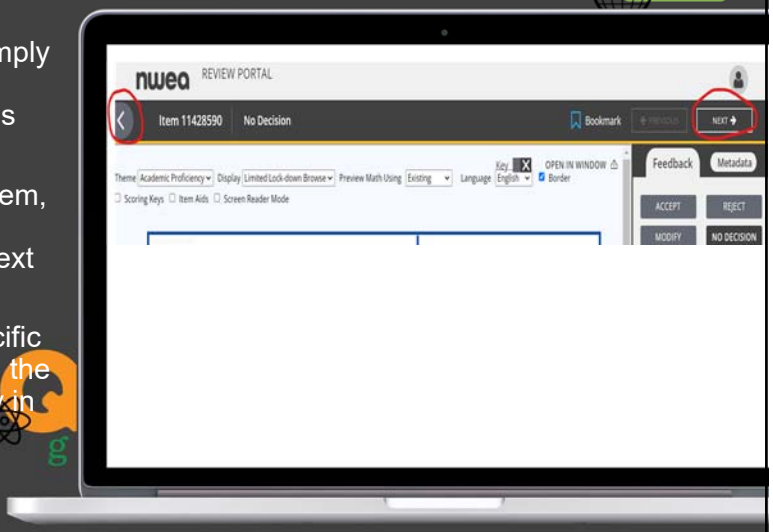
- When you click to open the session, you will see a list of items with a toolbar along the top.
- These will allow you to navigate to items that you need to review in No Decision, have marked as Accepted, rejected or Modified, and those which you have bookmarked for discussion.



nwea

Review Portal Item Navigation and Selection

- To select an item, you can simply click on the item from the list. This allows you to select items from any within the session.
- Once you have selected an item, you can either go back to the main list, or advance to the next item by selecting NEXT.
- If you are trying to find a specific item, use your find function in the search engine's tools (usually in the upper right-hand corner).



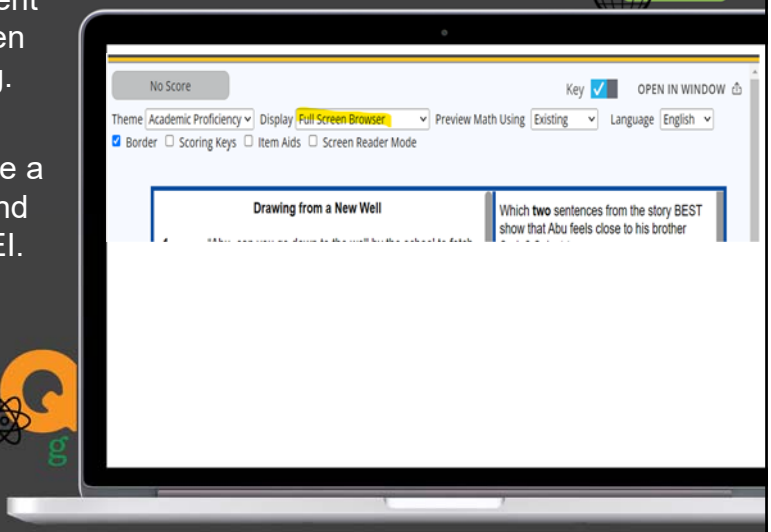
nwea

Review Portal Item Review

+ To view the items as the student will view them, select Full Screen Browser as your Display setting. This minimizes scrolling.

+ Review the item as if you were a student. Answer the question and interact with the item if it is a TEI.

+ Use the Item Aids and try the screen reader to see the accommodations available to students who receive accommodations.

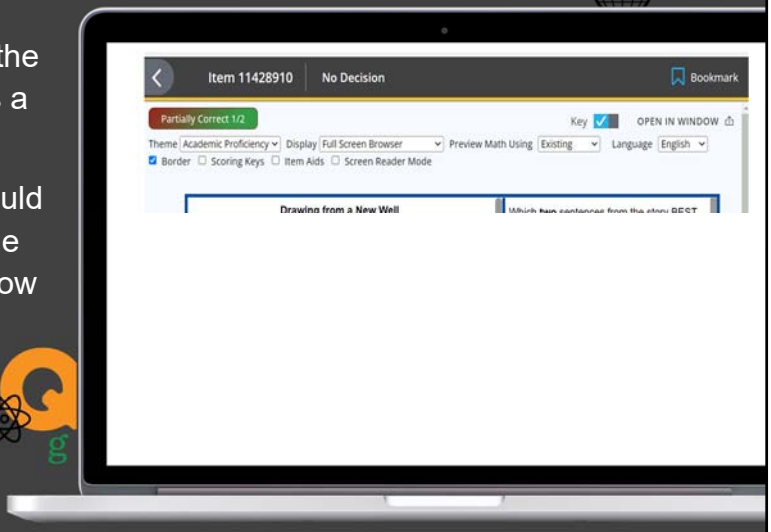


nwea

Review Portal Scoring

+ To check the scoring, move the toggle on the Key until there is a blue check.

+ Select the answers. You should see the scoring validation in the upper left corner change to show incorrect, partially correct or correct.



nwea

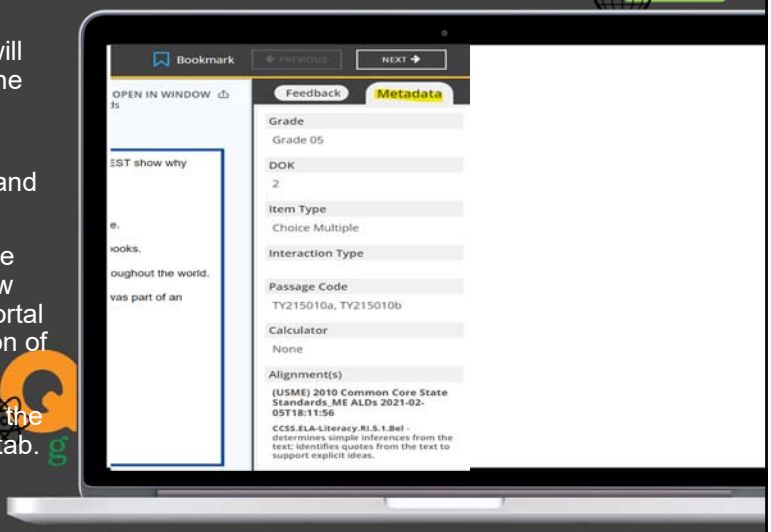
Review Portal Metadata Review

+ On the right side of the item, you will see a menu with two tabs. Choose the metadata tab.

+ Review the information in the tab paying careful attention to the DOK and the Alignment.

+ When reviewing math items, please refer to the PDF of the ALDs. In a few instances the ALDs in the Review Portal do not reflect the most current version of the ALD.

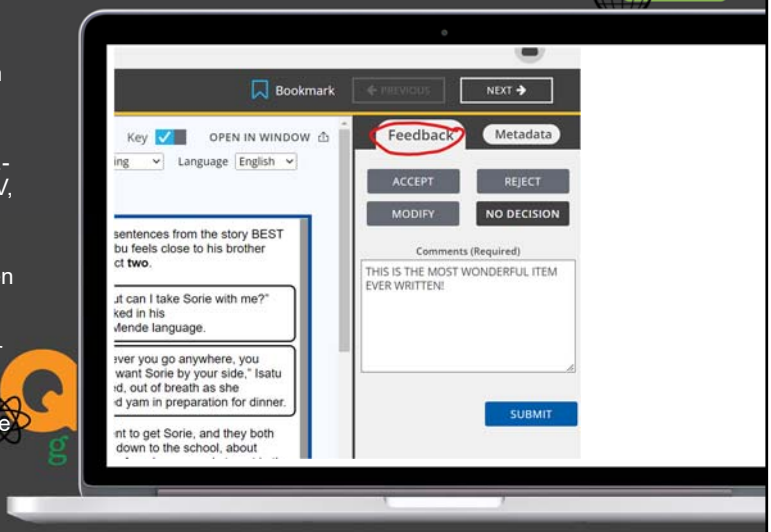
+ When you have finished reviewing the metadata, go back to the Feedback tab.



nwea

Review Portal Comments and Decision

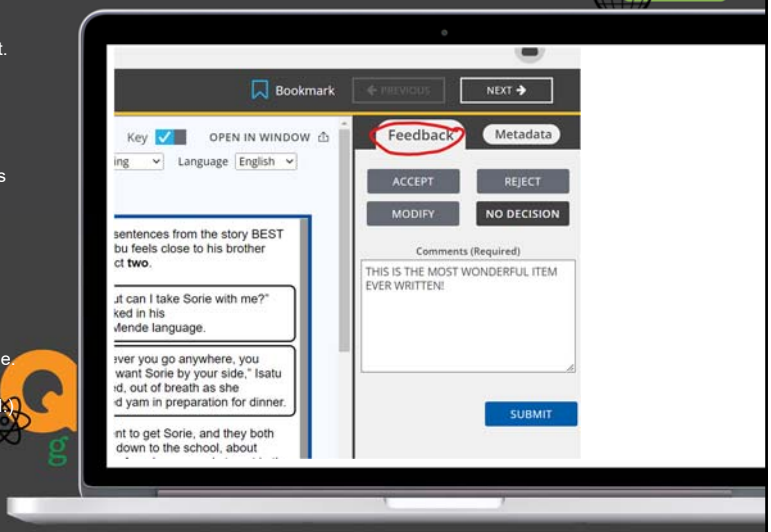
- On the feedback tab, you have the ability to leave specific notes about an item's content, alignment, and DOK.
- If you would like something to be changed, please be very specific. (Ex.- change to DOK 1, change ALD to DEV, change standard to RP.4, change answer choice C to "When I was …").
- Positive feedback is also helpful! When you like a certain item, we try to use similar items in future development. If you think students will find the topic or method engaging, please tell us! We will make every effort to find similar topics or use similar methods for future tests.



nwea

Review Portal Comments and Decision (Continued)

- Once you have left notes, please choose submit. This will capture your comments in the system.
- If you have suggested any edits to metadata or content, please select Modify.
- If you liked the item and do not feel any changes are needed, please select Accept. No notes are required for items that do not need modification.
- As a last resort, select Reject if there are no changes that would make the item align to the standards or the item is severely flawed. Please include a comment on why you think the item should be rejected and be as specific as possible. (Ex. - This item doesn't align to the ALD and I cannot find another ALD where it can be aligned.)
- When you have finished, you can go on to the next item.



nwea

Next steps

+ That was a lot of information! Please feel free to reach out with any questions or technology issues.

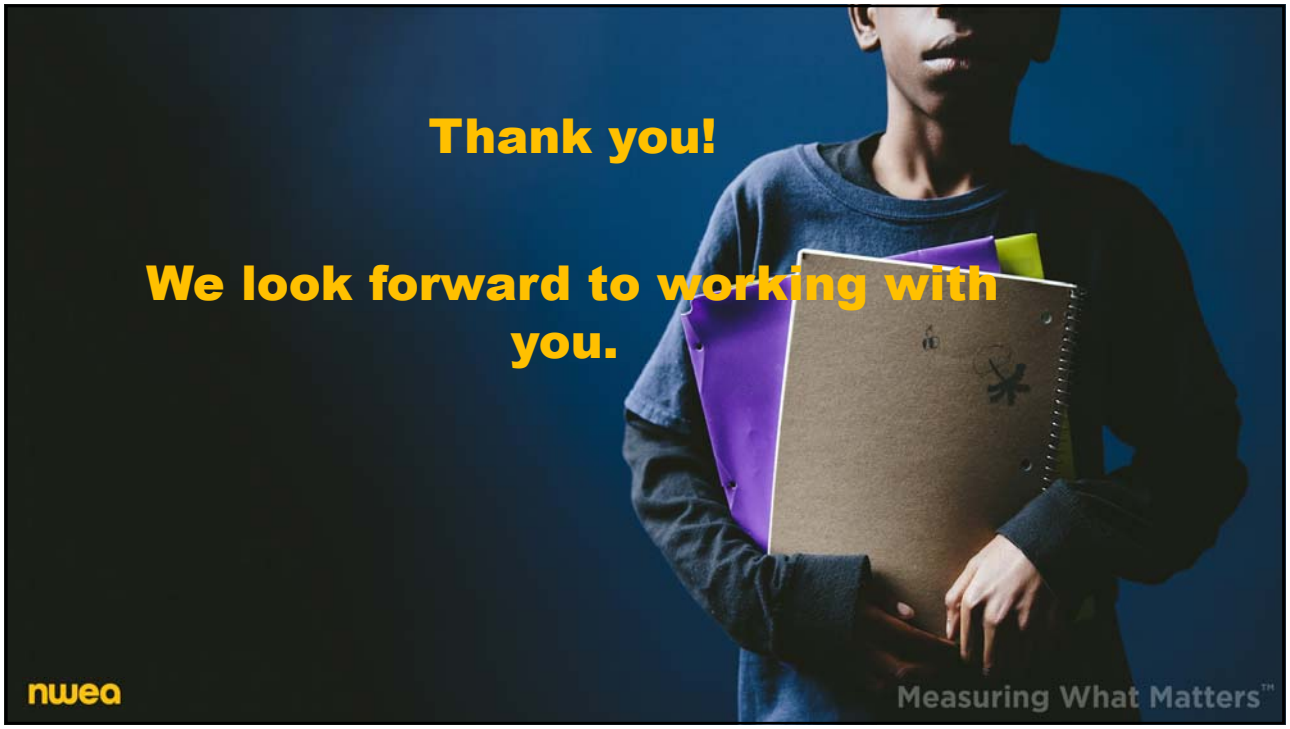
+ Please complete the review by 5:00 pm ET Monday, July 31st. This will allow the facilitators to analyze and prepare responses to your comments and to focus the discussion during our meeting time.

nwea | 18

Security

Materials

- Public materials include materials on the Maine website.
- Information shared in the Review Portal is considered secure.
- Access to the Review Portal will be restricted after Wednesday, August 9th at 3pm ET.
- You may discuss the process of the review, but please keep specific content confidential.
- Expect a survey at the end of the week to gather your feedback on the meeting. We use your input to improve future meetings.


Thank you!
We look forward to working with you.
nwea
Measuring What Matters™

Item Writing Guidelines/Specifications

Content of items should:

- Be aligned clearly to the identified standard
- Meet the indicated ALD level
- Assess meaningful content (ask questions worth asking)
- Be accessible to all students (avoid sensitive topics, social economic bias, color dependency, etc.)
- Be supported by credible sources when facts are used
- Be unique and not cue other items
- Context vocabulary items should be at least 1–2 grade-levels above with answer choices at or below (Reading).

Item stems should:

- State information clearly and concisely
- Use complete sentences with ending punctuation
- Use “which” before a noun and “what” before a verb or when options are not given (e.g., text entry)
- Avoid using “Which of the following . . . “
- Be stated as a question for multiple choice and choice multiple (multi-select)
- Be intentional when using qualifiers, such as “BEST,” “MOST LIKELY”

Item answer options should:

- Be consistent grammatically with the stem and parallel in form
- Be mutually exclusive (independent) unless appropriate for the content being assessed (e.g., domain and range)
- Be phrased positively (avoid “not”)
- Be free of clues to the correct answer
- Be ordered purposefully (by length, alphabetically, ascending or descending value, etc.)
- Be plausible but not justifiable (for choice multiple [multi-select], be sure answers are same level or correctness)

Avoid:

- Using “All (or none) of the above”
- Using words like “not” and “except”
- Using overused words, such as “a lot,” “very, nice,” “thing”
- Beginning sentences with “There is___,” “There was___,” or “There are___”
- Using contractions wherever possible, with the exception of contractions within dialogue
- Using extreme descriptors (e.g., “always,” “never,” “all”) unless appropriate for the standard
- Using vague modifiers (e.g., “best,” “worst”) unless appropriate for the standard
- Passive voice (when avoidable)

Style:

- Spell out acronyms.
- Use grade-level appropriate words.
- Model correct grammar, punctuation, capitalization, spelling.

- Use active voice and present tense when possible.
- Use smart quotations and apostrophes.
- Minimize scrolling when possible.
- For more specific details about style, see the Style Guide for the specific project.

Universal Design:

Create items accessible to all students based on Universal Design Applied to Large Scale Assessments (Thomson et al., 2002).

- Items should be free of unnecessary linguistic complexity.
- Information presented in items should be clear, concise, and relevant to the standard being assessed.
- Context and language should be fair and familiar to students at the grade level and should not give advantages or disadvantages to subgroups.
- Item should be free of stereotypes and potential disrespect regarding age, gender, race, ethnicity, language, religion, sexual orientation, social economic status, disability, or geographic region.
- Item should not challenge personal beliefs or values and should avoid emotionally charged topics.
- Avoid names and gender unless necessary. If names must be used, use a variety of genders and ethnicities.
- Graphics should be intentional and not merely decorative.
- Graphics should not color dependent.
- MathML should have equation tags compatible with text-to-speech and screen readers.
- Art should be tagged to be compatible with screen readers where possible.

Fact Checking

- Items are supported by credible sources when facts are used
- At least one valid source is used for generic factual statements (e.g., a rectangular table is x feet by y feet) and specific factual statements (e.g., a cheetah runs at x miles per hour).
- Specific factual statements are verified by additional reviewer.

Tools (Math)

Calculators

- Items at grades 3–5 do not include calculators on summative items.
- Basic calculators are assigned on an item-by-item basis for grade 6 summative items.
- Scientific calculators are assigned on an item-by-item basis for grades 7, 8, and 10 summative items.
- Graphing calculators are assigned on an item-by-item basis for grade 10 summative items.
- Items aligned to standards assessing computational skills in grades 6–8 and 10 will not allow calculators.

Rulers and Protractors

- Rulers are assigned on an item-by-item basis for items assessing measurement with a ruler.
- Protractors are assigned on an item-by-item basis for items assessing measurement with a protractor.
- Rulers and protractors are not assigned for items not requiring it.

Reference Sheets

- Reference sheets are not used in grades 3–5. Required formulas or conversion factors must be provided within the item.
- A universal reference sheet will be available for all items in grades 6–8 and HS assessments.

Technology enhanced items should:

- Follow the general item writing guidelines
- Clearly indicate the desired outcome in the stem
- Use direction lines specified in the style guide
- Enhance the content measurement through the technology
- Avoid technology for technology's sake
- Avoid redundancy in tasks

Depth of Knowledge (DOK)

- Items should be assigned one of the following DOK levels (Webb, 2009) as appropriate:
  - DOK 1: Recall
  - DOK 2: Skill & Concepts
  - DOK 3: Strategic Thinking
- Items at DOK 2 and 3 require conceptual and/or inferential thinking. DOK 3 items typically demand that students analyze and synthesize concepts from various parts of a text or from the text as a whole. DOK 4 items are not included in this assessment.

Item Types—Specific Considerations for Use

Multiple-Choice (Choice):

- There are 4 answer choice options.
- The correct response requires selecting one answer choice option.

Multi-Select (Choice Multiple):

- There are 5 to 8 answer choice options.
- The correct response requires selecting two or more answer choice options.
- Students must select all correct answers and no incorrect answers to earn the point.
- At grades 3–5, students will be directed on how many options to choose. At grades 6–8, students may be directed to "select all that apply."

Gap Match/Graphic Gap Match/Graphing

- The correct response requires moving answer choice options into gaps by selecting and moving the options, selecting the option and then selecting the gap, or using click-and-pop functionality.
- Graphing will allow for plotting points and eventually other graphs.

Hot Text

- The correct response requires selecting one or more answer choice options embedded within text, images, or tables.

Text Entry/Equation Editor

- The correct response requires entering the response in the response box.

Composite Items

- The item contains multiple parts/functionalities. For example, Part A and Part B.
- The parts are related but avoid redundancy.
- The correct response requires answering all parts.

Scoring

Dichotomous (1pt) Items

- Dichotomous items may be used for all assessable standards.
- Students earn 0 or 1 point for dichotomous items.
- Dichotomous items may include multiple choice, multi-select, gap match, graphic gap match/graphing, hot text, or text entry/equation editor.

Dichotomous Scoring & Item Types

Multiple-Choice (Choice)

- Multiple-choice items are 1-point items with a single correct answer.

Multi-Select (Choice Multiple)

- Multi-select 1-point items typically have 2 correct answers. The item may have 3 or more correct answers if it is appropriate for the standard.
- Students must select all correct answers and no incorrect answers to earn the point.

Gap Match/Graphic Gap Match/Graphing

- Gap match 1-point items typically have 2–6 correct selections (for example, sorting shapes into a table).
- Graphic gap match/graphing 1-point items typically have 2–6 correct selections that are placed on a graphic background (for example, creating a fraction with the fraction bar provided as art or plotting a point on a coordinate grid).
- Students must place all correct answers and no incorrect answers to earn the point.

Hot Text

- Hot text 1-point items typically have 2 correct answers. The item may have 3 or more correct answers if it is appropriate for the standard.
- Students must select all correct answers and no incorrect answers to earn the point.

Text Entry/Equation Editor

- Text entry/equation editor interactions can be worth 1 point.
- Text entry only allows numerical responses.
- The correct response allows equivalent numerical values based on the allowed characters.

Polytomous (2pt) Items

- Polytomous items are used when the standard allows depth beyond what is required in a dichotomous item. These items assess multiple aspects of the standard within the same item.
- Students may earn 0, 1, or 2 points for polytomous items.
- Polytomous items are usually composite items but may also include multi-select, gap match, graphic gap match/graphing, hot text, or text entry/equation editor.

- For polytomous items with a single interaction, selecting all correct responses without any incorrect responses results in a total of 2 points.
- For polytomous items with a single interaction, the item should require more cognitive processing than a dichotomous item that uses the same type of interaction.

Polytomous Scoring & Item Types

Composite

- Composite 2-point items have two parts, each part is 1 point and the parts sum to a total of 2 points.
- The parts may be multiple choice, multi-select, hot text, text entry/equation editor, and eventually gap match and graphic gap match/graphing.
- Each part follows the above rules for dichotomous scoring for its item type.
- Students may earn 0, 1, or 2 points. Getting both parts incorrect scores 0 points. Getting all parts correct scores 2 points. Depending on the standard and the item, sometimes 1 point can be gained by getting either Part 1 or Part 2 correct; however, other items require the student to get Part 1 correct to receive 1 point, and getting only Part 2 correct would result in a 0 score.

Multi-Select (Choice Multiple):

- Multi-select 2-point items typically have 3 or more correct answers. The item may have 2 correct answers if the item is cognitively demanding and worth 2 points for the standard the item is measuring.
- The item should allow for partial credit so that students can earn 0, 1, or 2 points.
- If the item is meant to allow partial credit for both 2 correct answers or 2 correct answers and 1 incorrect answer, a scoring line needs to be entered for each situation. At the lower grades, where students must choose the set number, it will limit how many scoring lines need to be entered.

Gap Match/Graphic Gap Match/Graphing

- Gap match/graphic gap match/graphing 2-point items typically require more than 4 selections, such as constructing line plots.
- Each point value needs to be defined. For example, if partial credit is to be given for answering 2 gaps or 3 gaps correctly out of a total of 4 gaps, a line needs to be entered in scoring for each. 2 gaps = 1 point, and 3 gaps = 1 point.

Hot Text

- Hot text 2-point items typically have 3 or more correct answers. The item may have 2 correct answers if the item is cognitively demanding and be worth 2 points for the standard the item is measuring. The item should allow for partial credit so that students can earn 0, 1, or 2 points.
- Each point value needs to be defined. If partial credit is to be given for answering 2 groups or 3 groups correctly, a line needs to be entered in scoring for each. 2 groups = 1 point, and 3 groups = 1 point.

Text Entry/Equation Editor

- Text entry/equation editor 2-point items are made up of 2 or more text entry/equation editor interactions that are best presented as a stand-alone item instead of labeled parts. Items will be set up and scored as a composite item, but the student sees 1 item.

References

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). University of Minnesota, National Center on Educational Outcomes. https://nceo.umn.edu/docs/onlinepubs/synth44.pdf

Webb, N. L. (2009). *Design of content alignment studies in mathematics and reading for 12th grade NAEP and other assessments to be used in preparedness research studies.* National Assessment Governing Board. https://www.nagb.gov/content/dam/nagb/en/documents/publications/design-document-final.pdf

Item Development: Content Review Checklists

ALL ITEMS	
☐	Item is properly aligned to assigned standard/objective/benchmark.
☐	Item meets client specifications (e.g., context), if applicable.
☐	Item is assigned to an appropriate cognitive level and difficulty level (if applicable).
☐	Item is grade-level (reading-level context, topic, etc.) appropriate. In general, on-grade level for ELA, one grade level below for other subject areas.
☐	Item adheres to principles of universal design for test items.
☐	Item content has been verified for correctness and clarity (correct, clear, and engaging).
☐	Item complies with style guide.
☐	If the item is based on a context or scenario, it is realistic (e.g., no 75-pound cats).
☐	Item is free of repetitious wording.
☐	Item is free from clues that could lead students to a particular option (e.g., word repeated in both stem and option; correct grammar between stem and only one option).
☐	Item is free from any bias or sensitivity issues.
☐	Item does not ask for opinions (e.g., “what do you think” or evidence for “your answer” in Part B).
☐	References to art, passages, or other stimuli are precise (e.g., “paragraphs 1 and 2” rather than “the beginning of the passage”).

ITEMS REQUIRING SOURCES (FACTUAL DATA)	
☐	Factual data and source are verifiable.
☐	The source is recorded within the item metadata.

ITEMS REQUIRING ART OR GRAPHICS	
☐	Art is necessary and appropriate for the item.
☐	Art is ordered (for items requiring art); Art specification is clear and concise; Art specification follows style guide (applicable for Content Reviews prior to art creation).
☐	Art is free of errors; art was created according to the request and meets specifications (applicable for Content Reviews post art creation).

MULTIPLE-CHOICE ITEMS	
☐	Item has one, and only one, possible correct answer. A second or third answer choice cannot be considered correct unless the item is designed to have two or more correct responses.

☐	Rationales are plausible and clearly detail the error or misassumption made by the student.
☐	Item has viable options; rationales are clear and concise.
☐	Item has parallel distractors (although excessive attention to parallel length is not necessary, so long as one distractor is not conspicuously long or short).
☐	Choices are ordered according to program requirements (e.g., numerical value, location of information in passage). If direct quotations are used as answer choices, the answer choices appear in the same order as they appear in the passage.

TEXT ENTRY/NUMERIC ENTRY	
☐	The correct answer is precise (there are limited variations). Text entry items should be limited to one or two words. Numeric entry items should be limited to a whole number, decimal fraction, or improper fraction. (Currently, mixed numbers cannot be scored properly, but this might change in the future.)
☐	Full range of mathematical responses are identified, if applicable.
☐	All variations/equivalencies of the correct answer are identified (misspellings, if applicable, fraction and decimal equivalents).

TECHNOLOGY-ENHANCED ITEMS	
☐	Directions are clear and concise and follow program style.
☐	Item elements for the specified item type are clear and remain within item type guidelines. (A multiple-select (MS) response has the characteristics of a MS and only those of a MS; item types are not blended or unclear.)
☐	All tools, symbols, and/or numbers required to answer the item are provided or are accessible.
☐	Item functions as designed (e.g., multiple answers can be chosen for a multiple-select response).
☐	Scoring table is clear and complete. Items with multiple correct answers include all possible combinations.
☐	Item scoring details include the correct answer or all possible correct answers.

ITEM STIMULI	
☐	Reading passages are accompanied by both quantitative and qualitative analyses that justify the grade-level placement.
☐	Passages or other stimuli are capable of supporting questions to specified standards or item types and are of sufficient quality.
☐	Passages or other stimuli meet the program’s specifications for bias and sensitivity issues.

Passage Quality Checklist

Title:	Author:	Grade Level or Band:
Lexile:	FK:	Word Count:

Selection Criteria	Comments		
1. Work worthy of study: a. Accurate content b. Lends itself to a close reading and analysis c. Provides ample opportunity for examining an author's craft: i. Clear and effective structure ii. Development of arguments, ideas, characters, plot, setting (etc.) are detailed and thorough rather than superficial iii. Relevant evidence, reasoning, and concrete details iv. Rich, varied language (style, syntax, diction, rhetorical devices, domain-specific terms)	☐Yes	☐No	☐Maybe
2. Free of bias and sensitivity concerns: a. Does not provoke an undue emotional response outside of highly individualized experiences b. Represents groups fairly, accurately, respectfully, and without stereotype c. Distinguishes traditional behaviors/values from stereotypes d. Presents differences and varieties without moral judgment e. Does not overgeneralize f. Characters are not depicted as victims of/dependent on dominant culture for help/success	☐Yes	☐No	☐Maybe
3. Engaging and appropriate for target readers: a. Topics, issues, or arguments are likely to be of interest; OR b. Text is engaging	☐Yes	☐No	☐Maybe
4. Ideal for assessment: a. Presents multiple opportunities for reading-related questions b. Appropriate for grade level given both text complexity and grade-specific standards c. Aligned to Common Core standards	☐Yes	☐No	☐Maybe
5. Complex text that feels complete: a. Appropriate for grade level or grade band based on quantitative and qualitative measures b. Does not require more prior knowledge than would be appropriate at the given grade c. Has the sense of a beginning, middle, and end. d. Does not require an elaborate contextual introduction e. Falls within word count guidelines for grade level or band (with allowance for +/-10%)	☐Yes	☐No	☐Maybe

Understanding Your Student's Spring 2024 Maine Through Year Assessment Score Report

Maine Scale Score

The colored, horizontal bars on your student's report show your student's Maine scale score. This score measures your student's performance on this assessment based on the important skills in reading and math for their grade level based on the Common Core State Standards.

The Grade 3 reading assessment measures skills such as:

- Answering questions about characters, settings, and events in stories.
- Identifying the lesson in folktales and fables from different cultures.
- Finding the main idea and key details of informational texts.
- Using text features like headings and keywords to find information quickly.

The Grade 3 math assessment measures skills such as:

- Understanding strategies for multiplication and division within 100.
- Developing an understanding of fractions.
- Developing an understanding of rectangular arrays and area.
- Describing and analyzing two-dimensional shapes.

The Maine scale score is associated with an achievement level that describes how the student's performance aligns with state expectations. The achievement levels are Well Below State Expectations, Below State Expectations, At State Expectations, and Above State Expectations.

Achievement Level Explorer Tool

The Achievement Level Explorer Tool provides detailed information on the specific academic standards assessed at each grade level and how those standards are broken down into different skills at each achievement level.

Access the Achievement Level Explorer Tool

Understanding Your Student's Spring 2024 Maine Through Year Assessment Score Report

Maine Scale Score

The colored, horizontal bars on your student's report show your student's Maine scale score. This score measures your student's performance on this assessment based on the important skills in reading and math for their grade level based on the Common Core State Standards.

The Grade 4 reading assessment measures skills such as:

- Summarizing the main points of stories, plays, poems, and informational text.
- Using details and examples from texts when explaining what the text says.
- Making inferences using evidence from the text.
- Identifying the message or lesson of a story, play, or poem using key details.
- Comparing and contrasting stories by different writers or from different cultures.

The Grade 4 math assessment measures skills such as:

- Multiplying and dividing multi-digit numbers.
- Understanding equal fractions.
- Adding and subtracting fractions with the same denominator.
- Multiplying fractions by whole numbers.
- Classifying shapes by their properties.

The Maine scale score is associated with an achievement level that describes how the student's performance aligns with state expectations. The achievement levels are Well Below State Expectations, Below State Expectations, At State Expectations, and Above State Expectations.

Achievement Level Explorer Tool

The Achievement Level Explorer Tool provides detailed information on the specific academic standards assessed at each grade level and how those standards are broken down into different skills at each achievement level.

Access the Achievement Level Explorer Tool

Understanding Your Student's Spring 2024 Maine Through Year Assessment Score Report

Maine Scale Score

The colored, horizontal bars on your student's report show your student's Maine scale score. This score measures your student's performance on this assessment based on the important skills in reading and math for their grade level based on the Common Core State Standards.

The Grade 5 reading assessment measures skills such as:

- Using details in the text to answer questions about stories, plays, poems, and informational text.
- Summarizing texts by talking about the main ideas and key details.
- Analyzing and comparing the structure of stories, poems, and informational text.
- Comparing and contrasting important points presented in two texts on the same topic.
- Using context clues and reference materials to figure out the meanings of words.

The Grade 5 math assessment measures skills such as:

- Dividing whole numbers with 2-digit divisors.
- Adding, subtracting, multiplying, and dividing fractions.
- Understanding place value with decimals up to the hundredths place.
- Adding, subtracting, multiplying, and dividing decimals.
- Developing an understanding of volume (how much space an object takes up).

The Maine scale score is associated with an achievement level that describes how the student's performance aligns with state expectations. The achievement levels are Well Below State Expectations, Below State Expectations, At State Expectations, and Above State Expectations.

Achievement Level Explorer Tool

The Achievement Level Explorer Tool provides detailed information on the specific academic standards assessed at each grade level and how those standards are broken down into different skills at each achievement level.

Access the Achievement Level Explorer Tool

Understanding Your Student's Spring 2024 Maine Through Year Assessment Score Report

Maine Scale Score

The colored, horizontal bars on your student's report show your student's Maine scale score. This score measures your student's performance on this assessment based on the important skills in reading and math for their grade level based on the Common Core State Standards.

The Grade 6 reading assessment measures skills such as:

- Using examples and quotes to explain the meaning of stories, plays, poems, and informational text.
- Identifying how the structure of a story, play, or poem fits together to develop the theme, setting, or plot.
- Determining the point of view of the narrator or speaker.
- Explaining how an author develops the point of view of the narrator or speaker in a text.

The Grade 6 math assessment measures skills such as:

- Understanding how ratios and rates relate to multiplication and division.
- Using ratios and rates to solve problems.
- Understanding dividing fractions fully.
- Extending number knowledge to positive and negative numbers.
- Writing, understanding, and using expressions and equations.

The Maine scale score is associated with an achievement level that describes how the student's performance aligns with state expectations. The achievement levels are Well Below State Expectations, Below State Expectations, At State Expectations, and Above State Expectations.

Achievement Level Explorer Tool

The Achievement Level Explorer Tool provides detailed information on the specific academic standards assessed at each grade level and how those standards are broken down into different skills at each achievement level.

Access the Achievement Level Explorer Tool

Understanding Your Student's Spring 2024 Maine Through Year Assessment Score Report

Maine Scale Score

The colored, horizontal bars on your student's report show your student's Maine scale score. This score measures your student's performance on this assessment based on the important skills in reading and math for their grade level based on the Common Core State Standards.

The Grade 7 reading assessment measures skills such as:

- Using details from the text to answer questions about stories, plays, poems, and informational text.
- Analyzing how plot, setting, and characters work together in stories.
- Judging if the reasons and evidence in an article support the author's points.
- Comparing and contrasting two texts on the same topic.
- Using context clues and word parts to figure out word meanings.

The Grade 7 math assessment measures skills such as:

- Understanding proportional relationships (how amounts change in relation to each other).
- Using positive and negative numbers in math expressions.
- Solving problems using scale drawings and building shapes.
- Finding the area, surface area, and volume of shapes.
- Using data to make conclusions.

The Maine scale score is associated with an achievement level that describes how the student's performance aligns with state expectations. The achievement levels are Well Below State Expectations, Below State Expectations, At State Expectations, and Above State Expectations.

Achievement Level Explorer Tool

The Achievement Level Explorer Tool provides detailed information on the specific academic standards assessed at each grade level and how those standards are broken down into different skills at each achievement level.

Access the Achievement Level Explorer Tool

Understanding Your Student's Spring 2024 Maine Through Year Assessment Score Report

Maine Scale Score

The colored, horizontal bars on your student's report show your student's Maine scale score. This score measures your student's performance on this assessment based on the important skills in reading and math for their grade level based on the Common Core State Standards.

The Grade 8 reading assessment measures skills such as:

- Referring directly to the text to answer questions about stories, plays, poems, and informational text.
- Analyzing how dialogue and events in a story drive the plot or provide information about characters.
- Determining the author's point of view and purpose.
- Using context, word roots, prefixes, suffixes, and reference materials to determine word meanings.

The Grade 8 math assessment measures skills such as:

- Writing, reasoning about, and solving equations and systems of equations.
- Understanding functions and using them to show relationships between amounts.
- Analyzing 2D and 3D shapes using distance, angles, similarity, and congruence.
- Understanding and applying the Pythagorean Theorem to find side lengths.

The Maine scale score is associated with an achievement level that describes how the student's performance aligns with state expectations. The achievement levels are Well Below State Expectations, Below State Expectations, At State Expectations, and Above State Expectations.

Achievement Level Explorer Tool

The Achievement Level Explorer Tool provides detailed information on the specific academic standards assessed at each grade level and how those standards are broken down into different skills at each achievement level.

Access the Achievement Level Explorer Tool

Understanding Your Student's Spring 2024 Maine Through Year Assessment Score Report

Maine Scale Score

The colored, horizontal bars on your student's report show your student's Maine scale score. This score measures your student's performance on this assessment based on the important skills in reading and math for their grade level based on the Common Core State Standards.

The high school reading assessment measures skills such as:

- Reading and comprehending fiction and non-fiction texts.
- Citing evidence from texts to support analysis and conclusions.
- Analyzing how authors use varying writing styles to shape meaning and tone.

The high school math assessment measures skills such as:

- Using algebra and models to understand relationships, represent situations, and solve problems.
- Using shapes and models to solve real-world problems and describing how points and shapes relate.
- Using different types of numbers and math operations to solve equations.
- Using functions to show number patterns and model real-world situations.
- Using statistics and probability to study data, describe chance, make inferences, and explain conclusions.

The Maine scale score is associated with an achievement level that describes how the student's performance aligns with state expectations. The achievement levels are Well Below State Expectations, Below State Expectations, At State Expectations, and Above State Expectations.

Achievement Level Explorer Tool

The Achievement Level Explorer Tool provides detailed information on the specific academic standards assessed at each grade level and how those standards are broken down into different skills at each achievement level.

Access the Achievement Level Explorer Tool

Maine Through Year Assessment Program Standard Setting Technical Report

Submitted to NWEA & the Maine Department of Education

by Creative Measurement Solutions LLC

October 30, 2023

Table of Contents

Table of Tables

Table of Figures

Executive Summary

The Maine Through Year Assessment (MTYA) program is Maine's statewide system of summative assessments in Mathematics and Reading in grades 3-8 and the second year of high school that was first administered in Spring 2023. The Maine Department of Education (DOE) contracted with NWEA to design and develop the MTYA, and NWEA contracted with edCount, LLC and Creative Measurement Solutions, LLC to design and implement their alignment study and standard setting.

Embedded Standard Setting (ESS) was employed to establish the MTYA cut scores. ESS transforms standard setting from a standalone workshop to a set of processes actively integrated throughout the assessment development lifecycle. ESS processes directly contribute to the valid interpretation and use of test scores and improve test quality and the strength of validity arguments by maintaining a consistent focus on optimizing the evidentiary relationship between test items and the Common Core State Standards (CCSS) reflected by the associated achievement level descriptors (ALDs).

ESS processes, described in Section 1, include:

- ALD development, an articulation of the intended interpretations of the MTYA across the achievement levels.
- The ESS Alignment Study, a review of a representative sampling of MTYA items by Maine educators, in which they provide independent alignments of these items to the CCSS and Maine achievement levels and review and resolve items with alignments that are inconsistent with data.
- ESS analyses and the estimation of cut scores, in which the educators' alignments of items to Maine achievement levels are employed to identify optimal cut scores.
- Post ESS Alignment Study workshop activities leading to the adoption of cut scores, including cut score refinement to support an integrated, vertically articulated system of cross-grade cut scores meeting workshop panelists' and other stakeholders' expectations, and in consideration of the Maine DOE policy goals.
- Documentation of validity evidence supporting the MTYA-adopted cut scores, including those forms of evidence commonly cited in the measurement literature, and those used to satisfy federal peer review requirements.

This report summarizes the standard setting design, processes, analyses, and results leading to the adoption of cut scores and documentation of their validity. Findings from each of these activities provide evidence that the ESS processes work together to promote the coherence of the MTYA. Specifically:

- Range ALDs were developed to align to the CCSS and were reviewed and refined by Maine educators, as described in Section 2.

- Results from the ESS Alignment Study demonstrated the efficacy of panelists' Consensus Level alignments of items to ALDs. High correlations with empirical difficulty, weighted Kappa values, and panelist agreement rates demonstrated a strong panelist understanding of their role and judgment tasks. Results from the multi-round workshop showed iterative improvement in the consensus Item-ALD alignments and associated efficacy measures, as described in Section 3.
- ESS analyses produced cut scores that optimally reflect panelist judgments by minimizing inconsistencies between those judgments and empirical data, as described in Section 4.
- Post workshop vertical articulation produced a well-articulated, cross-grade system of cut scores in Mathematics and Reading that reflect panelist and other stakeholder expectations for impact data, using methods supported by the MTYA Technical Advisory Committee, as described in Section 5.
- Thorough documentation of validity evidence supporting the MTYA-adopted cut scores demonstrated strong adherence to principles of test score validation as articulated in the measurement literature and the guidelines for federal peer review, as described in Section 6.

Summary

Together, these findings support the validity of the cut scores adopted for the Maine Through Year Assessment Program. Linkages from ALDs to test scores are consistent with the tenets of Principled Assessment Design, support intended score interpretations, and inform decision making.

Section 1. Introduction

The Maine Through Year Assessment (MTYA) program is Maine's statewide system of summative assessments in Mathematics and Reading in grades 3-8 and the second year of high school that was first administered in Spring 2023. The grades 3-8 assessments were adaptive and the second year of high school assessments were fixed forms in this administration. The Maine Department of Education (DOE) contracted with NWEA to design and develop the MTYA, and NWEA contracted with edCount, LLC and Creative Measurement Solutions, LLC to design and implement their alignment study and standard setting. Documentation for development of the MTYA program and the alignment study is provided by NWEA (2023) and edCount (2023), respectively.

The MTYA standard setting design is a systematic approach grounded in Principled Assessment Design (PAD), which was employed by NWEA (2023) for the development of the MTYA. The MTYA achievement level descriptors (ALDs) serve as the foundation of the standard setting design. Under PAD, ALDs are developed early in the test development lifecycle to support domain definition (e.g., explication of the construct of interest), item development, and standard setting. The Embedded Standard Setting (ESS) methodology was selected because it is the natural extension of PAD to standard setting (Lewis & Cook, 2020). ESS processes are embedded throughout the PAD lifecycle as illustrated in Figure 1. The iterative nature of the ESS processes, represented by the green feedback arrows in Figure 1 supports the coherence of various assessment components and artifacts including ALDs, item development, Item-ALD alignment, empirical data, and cut scores (and therefore, score interpretation). Thus, adherence to these iterative processes supports validity of the assessments and score interpretation.

Four achievement levels were established for the MTYA:

- Level 1: Well Below State Expectations
- Level 2: Below State Expectations
- Level 3: At State Expectations
- Level 4: Above State Expectations

Three cut scores were adopted, defining the four levels of achievement:

- The Below State Expectations cut score separates the Well Below State Expectations and Below State Expectations levels.
- The At State Expectations cut score separates the Below State Expectations and At State Expectations levels.
- The Above State Expectations cut score separates the At State Expectations and Above State Expectations levels.

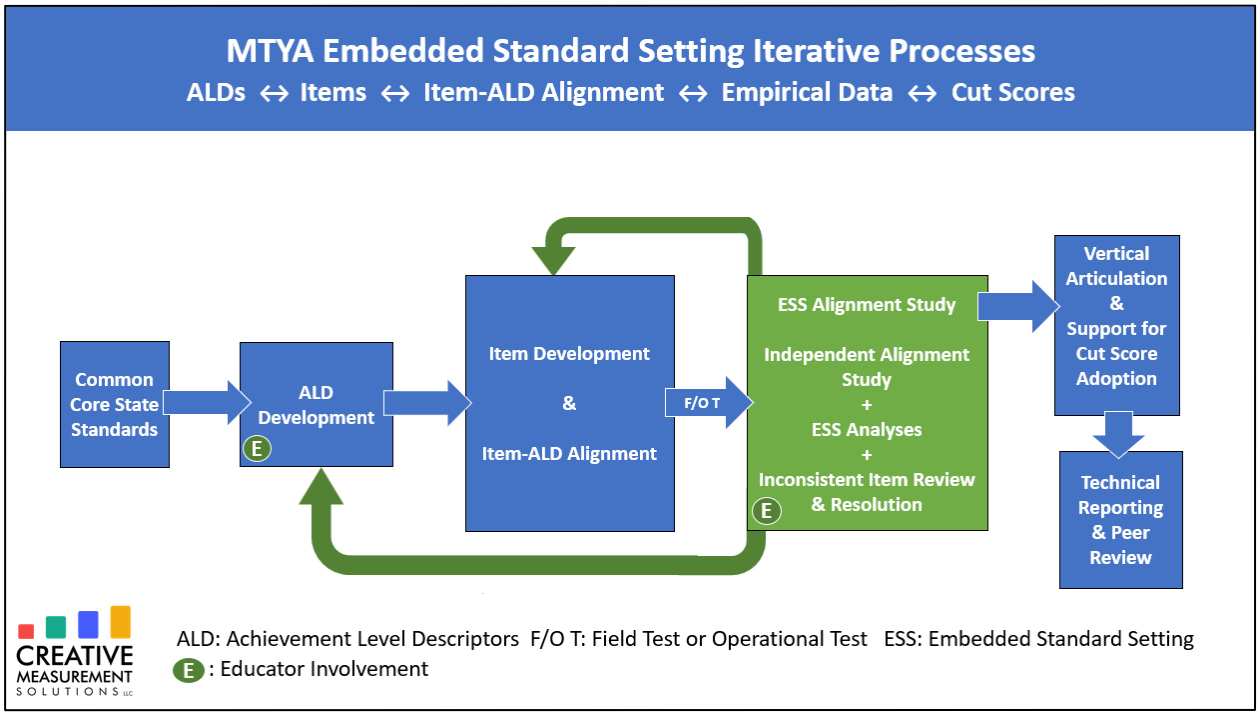


Figure 1. Maine Through Year Assessment Embedded Standard Setting Iterative Processes

Overview of Embedded Standard Setting and the MTYA Standard Setting Design

Embedded Standard Setting (ESS; Lewis & Cook, 2020) is the logical extension of Principled Assessment Design to standard setting. ESS transforms standard setting from a standalone workshop, which typically occurs after test administration and just prior to score reporting, to a set of processes that are an active part of the assessment development lifecycle. ESS processes directly contribute to the valid interpretation and use of test scores and improve test quality and the strength of validity arguments. They do so by maintaining a consistent focus on optimizing the evidentiary relationship between test items and the Common Core State Standards (CCSS) that are reflected by the associated ALDs.

ESS is based on three big ideas:

1. ALDs are the fundamental component of standard setting. That is, the ALDs operationalize the policy goals of the sponsoring agency (as specified in the General ALDs) by articulating the knowledge, skills, and attributes (KSAs) of students in each achievement level. The process of developing ALDs from the CCSS is represented by the first two boxes on the left in Figure 1.
2. The alignment of test items to achievement levels (Item-ALD alignment) by subject matter experts (SMEs) is effectively the same judgment made during traditional item-based standard setting workshops (e.g., Bookmark, ID Matching, Modified Angoff Yes/No, etc.). Thus, the Item-ALD alignments resulting from SME alignment study judgments obviate the need for the judgments traditionally made by participants in a standard setting workshop.
3. When empirical data on test items are available, ESS cut scores emerge organically and analytically by optimizing the coherence of the SME Item-ALD alignments and empirical data. That is, ESS cut scores are estimated by optimizing the evidentiary relationship between test items and the CCSS articulated in the ALDs. In this case, data from the Spring 2023 MTYA test administration are used to support the estimation of ESS cut scores.

ESS is not a single activity—it is a set of iterative processes and analyses, as illustrated in Figure 1, that occur throughout the assessment development lifecycle. ESS advances the principled notion of assessment design based on evidentiary reasoning by requiring the alignment of each assessment item (and each within-item score point for polytomous items) to an achievement level by the explicit linkage of the item (score point) to a specific ALD evidence statement (e.g., claim or measurement target). Thus, the evidentiary chain runs not just from the CCSS to the test items, but first from the CCSS to the ALDs and then from the ALDs to the test items, providing more precise interpretability of the measurement target evidenced by the items.

While ESS was developed to provide a practical approach to standard setting for assessments adhering to a PAD framework, its methods add value that extend well beyond the estimation of cut scores, supporting the coherence of the various assessment system elements described next and illustrated in Figure 1.

Embedded Standard Setting and Assessment System Coherence

Assessment system coherence refers to the interrelationship between the steps and processes engaged during assessment design and development working to preserve the chain of interpretability from the CCSS to ALD development to the realization of their interpretable operationalization through empirically identified cut scores and meaningful classifications. Assessment system coherence is manifested when the various assessment components form an internally consistent system. For example:

- ALDs should clearly and comprehensively articulate the CCSS and reflect the content and rigor to fulfill the intent of the General ALDs and the MTYA system.

- Items should provide evidence for the attributes of students specified by the CCSS and measurement targets in the various achievement levels.
- Items should be explicitly aligned to specific achievement levels because they provide evidence for the associated claims and measurement targets.
- Empirical data should support SME's Item-ALD alignments.
- Cut scores should have empirical data supporting the evidentiary relationship between assessment items and the CCSS; that is, examinees in each achievement level should have an appropriate likelihood of success on the items aligned to the claims and measurement targets in the associated level.

Assessment system coherence is supported by the application of PAD when the application appropriately employs the ESS iterative processes illustrated in Figure 1. A comprehensive application of PAD should, in fact, work to guarantee such coherence, and the ESS iterative processes ensure that the PAD process continues to do its work until coherence is achieved.

Assessment system coherence results from the understanding that initial drafts of the various assessment elements—ALDs, the assessment items, SMEs' Item-ALD alignments, and cut scores—often require iterative improvement and are only considered "final" once coherence is sufficiently supported by evidence. Cut scores are then imbued with the interpretations the assessment was developed to provide and are ready for adoption by the sponsoring agency. By explicitly incorporating iterative processes in the assessment development lifecycle, we acknowledge that we not only are comfortable revisiting the various assessment elements when and if anomalies manifest, but explicitly plan for, manage, and document the iterative activities that provide evidence for assessment system coherence.

Next, we provide an overview of each element of the MTYA standard setting design.

Coordination of Embedded Standard Setting Iterative Processes

The ESS design coordinates various activities that occur throughout the assessment development lifecycle. These processes include:

- ALD Development
- The ESS Alignment Study
- ESS Analyses and the Estimation of Cut Scores
- Post Workshop Activities Leading to the Adoption of Cut Scores
- Documentation of Validity Evidence Supporting the MTYA-Adopted Cut Scores

ALD Development

The MTYA General ALDs articulate the overarching goals of the assessment program and are provided in Table 1. The MTYA Range ALDs provide more detailed, content-based information about the knowledge, skills, and other attributes of students across the range of each level. The first draft of Maine's ALDs were developed by NWEA content specialists but were carefully evaluated by Maine educators to ensure that the ALDs reflected the appropriate standards and expectations for Maine student achievement in each grade and subject. And, for standards that span multiple grades such as high school CCSS, they confirmed that the expectations were consistent with the performance expectations for Maine's second year of high school students.

Based on specific feedback from Maine educators, the draft ALDs were edited and refined by NWEA's content specialists. Thus, the ALDs articulate the content, knowledge, skills, and attributes that students are expected to have in each level. The Maine ALDs used to support the establishment of cut scores are available at Achievement Level Explorer (nwea.org). A more detailed summary of the MTYA ALD development activity is described in Section 2.

Table 1. MTYA General Descriptors

Well-Below State Expectations	**Below State Expectations**	**At State Expectations**	**Above State Expectations**
On this assessment, students at this achievement level **demonstrate limited understanding of the knowledge and skills** necessary at this grade level, as specified in the Common Core State Standards. The students ***need substantial academic support*** to be prepared for the next grade level and to be on track for college and career readiness.	On this assessment, students at this achievement level **demonstrate partial understanding of the knowledge and skills** necessary at this grade level, as specified in the Common Core State Standards. The students ***need additional academic support*** to be prepared for the next grade level and to be on track for college and career readiness.	On this assessment, students at this achievement level **demonstrate the knowledge and skills** necessary at this grade level, as specified in the Common Core State Standards. The students ***are prepared*** for the next grade level and are on track for college and career readiness.	On this assessment, students at this achievement level **demonstrate advanced understanding of the knowledge and skills** necessary at this grade level, as specified in the Common Core State Standards. The students ***are well prepared*** for the next grade level and are well prepared for college and career readiness.

The ESS Alignment Study

The ESS Alignment Study was first proposed by Schneider & Lewis (2021) as an innovative and efficient application of ESS by integrating ESS methodology within a traditional independent alignment study. That is, if alignment study participants conduct Item-ALD alignment as part of an alignment study, then the results may be used to support the requirements of ESS for the estimation of cut scores. A three-day workshop was held to support peer review requirements for an independent alignment study as well as the ESS Item-ALD alignment requirements. The

first two days of the ESS Alignment Study workshop were conducted as a traditional independent alignment study by edCount, LLC (edCount, 2023). The alignment study panelists were Maine educators who independently aligned a representative sample of MTYA items in grades 3-8 and high school to the CCSS and to the Maine achievement levels.

ESS analyses were conducted on the Item-ALD alignments resulting from Day 2 of the workshop to identify ESS Inconsistent items—items with empirical difficulties that were not consistent with each grades' consensus Item-ALD alignments. On Day 3 of the ESS Alignment Study workshop, panelists reviewed the inconsistent items in order of difficulty. Panelists discussed each inconsistent item in terms of its measurement attributes and content characteristics and maintained or updated their previously documented Item-ALD alignments. The results of Day 3 were analyzed to identify the consensus (modal) Item-ALD alignment for each item. A detailed description of the ESS Alignment Study is provided in Section 3 of this report and in the MTYA Alignment Evaluation Report (edCount, 2023).

ESS Analyses and the Estimation of Cut Scores

The consensus Item-ALD alignments resulting from the ESS Alignment Study workshop were analyzed using Creative Measurement's proprietary ESS software (EmStanS; Lewis & Lee, 2021) to estimate the initial cut scores. The results are used to quantify and characterize the efficacy of the panelists' Item-ALD alignments and to estimate the initial ESS cut scores. A detailed description of the results of the ESS analyses is provided in Section 4.

Post Workshop Procedures Leading to the Adoption of Cut Scores

Initial cut scores resulting from a standard setting process are commonly smoothed to support a well-articulated system of cross-grade cut scores. That is, changes in impact data across levels within grade and within level across grades should be reasonable and meet stakeholder expectations. However, because panelists make within-grade judgments to produce initial cut scores, cross-grade smoothing is typically required. Considerations and methods leading to the adoption of a well-articulated cross-grade system of cut scores based on stakeholder review and in consideration of policy goals is described in detail in Section 5.

Documentation of Validity Evidence Supporting the MTYA-Adopted Cut Scores

Forms of evidence commonly used to support the validity of cut scores are cited in the measurement literature and in federal peer review guidelines. Validity evidence includes application of an appropriate and defensible standard setting method, documentation of the qualifications of participants in the standard setting process, and evidence that participants understood the tasks required of them. A detailed summary of validity evidence for the MTYA standard setting process and adopted cut scores is provided in Section 6.

Section 2. Achievement Level Descriptor Development and Validity

Overview of Task

In an ALD workshop, Maine educators were asked to review the Range ALDs initially developed by NWEA and based on the CCSS for use in Maine. Each participant reviewed Range ALDs for grades 3-8 and second year of high school in either Reading or Mathematics. The review's purpose was to give Maine educators an opportunity to study the draft ALDs and share their feedback with NWEA content specialists.

Panelists

The number of committee members for each content area was limited to three educators. For this reason, educators with expertise in all grade levels were recruited to participate. The state identified approximately 140 curriculum coordinators. The DOE emailed these educators with a link to a survey generated by NWEA that allowed them to indicate their interest and availability. Seven educators with positions as district administrators or curriculum specialists responded. All seven educators were invited to participate. Of these seven educators, two declined and one did not complete the prework or attend the workshop. The four participants represented three different regions of the state including Southern Maine, Southern-Central Maine, and Down East Maine. One educator represented a virtual academy.

Process

Maine educators were asked to complete prework for the ALD Workshop. They were provided with a guide that defined Range ALDs, explained how they are organized, and described how they are used. The guide also outlined the process for the review and listed three statements to consider when evaluating the ALD progression for each standard. Each educator was given a version of the ALDs with two columns for feedback. The first column was used to indicate whether they approved the ALD or would like to discuss the ALD at the workshop. The second column was used for comments.

NWEA content specialists compiled the feedback into one document and used it to determine which standards to discuss at the workshop. The NWEA content specialists also discussed the feedback with their content team before the workshop and had suggested revisions prepared to share with the educators. The workshop was held on the evenings of September 12-13, 2022. All standards marked by educators for discussion were addressed at the workshop. Four NWEA content specialists attended the workshop. Each content area had a content specialist that facilitated and another to help encourage discussion and record notes.

Results

Both the Reading and Mathematics ALDs had ALD progressions updated based on feedback from the Maine educators. These updates included:

- reassigning ALD statements to another level within the progression,
- removing ALD statements,
- revising ALD statements, and
- crafting new ALD statements.

Panelists recommended ALD updates for the following standards:

Reading

Grades 3-8 and the and the second year of high school. Standard updates for: RL.3.1, RL.3.2, RL.3.3, RI.3.1, RI.3.4, RI.3.6, L.3.4.b, L.3.6, RL.4.5, RL.4.6, RI.4.4, L.4.4.b, L.4.6, RL.5.2, RL.5.5, RL.5.6, RI.5.2, RI.5.4, L.5.4.a L.5.4.b, RL.6.2, RI.6.3, RI.6.8, L.6.4.b, RI.7.2, RI.7.8, L.7.4.c, RL.8.1, RI.8.8 , L.8.4.b, L.8.4.c, RI.9-10.8, and L.9-10.4.c.

Mathematics

Grades 3-8. Standard updates for: 3.MD.C.7.a, 7.EE.B.3, 7.SP.B, 7.SP.C.6, 8.EE.C.8, 8.F.B.4, 8.G.C.8, and 8.G.C.9

Second year of high school. Standard updates for: CC.9-12.N.RN.1, CC.9-12.N.Q.2, CC.9-12.A.SSE.1, CC.9-12.A.SSE.3.a, CC.9-12.A.CED.1, CC.9-12.A.CED.2, CC.9-12.A.CED.3, CC.9-12.A.REI.4.b, CC.9-12.A.REI.6, CC.9-12.A.REI.7, CC.9-12.A.REI.11, CC.9-12.A.REI.12, CC.9-12.F.IF.2, CC.9-12.F.IF.7.a, CC.9-12.F.IF.7.e, CC.9-12.F.IF.8.a, CC.9-12.F.BF.3, CC.9-12.F.LE.2, CC.9-12.F.LE.5, CC.9-12.G.CO.9, CC.9-12.G.CO.10, CC.9-12.G.CO.11, CC.9-12.G.C.2, CC.9-12.S.ID.5, CC.9-12.S.ID.6.a, CC.9-12.S.CP.1, and CC.9-12.S.CP.4.

Standards removed from ALDs in high school grade 10. The Maine educators believed the following high school standards were not appropriate for a grade 10 assessment in Maine and descriptors associated with the following standards were removed: CC.9-12.A.SSE.3.b, CC.9-12.A.SSE.3.c, CC.9-12.A.REI.4.a, CC.9-12.F.IF.7.b, CC.9-12.F.IF.8.b, CC.9-12.F.LE.1.c, CC.9-12.G.C.5, and CC.9-12.S.CP.3.

Validity

The validity of MTYA ALDs is supported by their coherence with the assessments and by the appropriateness of the expectations expressed by the ALDs. Two elements support the validity of the MTYA ALDs. First, panelists supporting the ESS Alignment Study workshop completed evaluations at the close of the workshop. The following three evaluation items reflect panelist endorsement of the ALDs:

1. It was easy to align items to specific achievement levels using the ALDs.
2. The ALDs helped illustrate the standards being tested in my grade and subject.
3. The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.

The evaluation results provided in Appendix E indicate:

Mathematics: Median cross-grade panelist ratings of 3, 4, and 4 (out of 4) support statements 1, 2, and 3, respectively.

Reading: Median cross-grade panelist ratings of 3, 3, and 3 (out of 4) support statements 1, 2, and 3, respectively.

Thus, panelists' evaluation responses provide moderate to strong support for the ALDs. That is, Statement 1 indicates that the ALDs tend to be sufficiently comprehensive, such that panelists were able to identify a clear association between the evidence statements in one achievement level and each reviewed test item. Statements 2 and 3 support the efficacy of the purposes of the ALDs—to explicate and articulate the CCSS and to reflect the content and rigor associated with the performance expectations for Maine students in the associated grades and subjects.

Second, the validity of the ALDs is supported by iterative review and refinement. In this case, when ESS workshop panelists' Item-ALD alignments varied from empirical data, their variances were noted and the information necessary to refine the ALDs was captured. The workshop panelists' consensus Item-ALD alignments were supported by empirical data for the vast majority of cases, providing validity evidence in support of their judgments and the ALDs. The remaining cases can be reviewed to determine whether ALD refinement is appropriate to support coherence and score interpretation.

Summary

The processes used to develop, refine, and evaluate the MTYA ALDs to comprehensively reflect the performance expectations for Maine students in each grade and subject provide evidence in support of the validity of the Range ALDs for their intended purposes.

Section 3. The ESS Alignment Study

The ESS Alignment Study was first proposed by Schneider & Lewis (2021) as an innovative and efficient application of ESS by integrating ESS methodology within a traditional independent alignment study. That is, if alignment study participants conduct Item-ALD alignment as part of an alignment study, then the results may be used to support ESS alignment requirements.

A proposal for the ESS Alignment Study design was provided in a detailed standard setting design document to the Maine DOE, reviewed by their Technical Advisory Committee, and approved for implementation. Creative Measurement Solutions (CMS) partnered with edCount to design and facilitate the ESS Alignment Study. Three-day workshops were held July 18-20, 2023 for Reading and July 25-27, 2023 for Mathematics, to support peer review requirements for an independent alignment study and for standard setting. The first two days of each ESS Alignment Study workshop were conducted as a traditional independent alignment study, as designed by edCount, LLC and reported in their technical documentation (edCount, 2023). The third day of each workshop was devoted to enhancing the efficacy of panelists' consensus Item-ALD alignments and thus, the validity of the resulting ESS cut scores, via the review and resolution of alignments that were inconsistent with data.

The ESS Alignment Study examined Mathematics and Reading items from a single fixed form in high school and three representative adaptive forms in grades 3-8:

- one adaptive form targeted to a typical lower achieving student—one at about the 16th percentile,
- one adaptive form targeted to a typical average achieving student, at about the 50th percentile, and
- one adaptive form targeted to a typical high achieving student, at about the 84th percentile

edCount's alignment study activities on Days 1 and 2 engaged panelists in two "rounds" of alignment activities. In the first round, panelists individually and independently aligned items from the representative test forms to the CCSS and to achievement levels. In the second round, panelists shared their individual alignment judgments, provided rationales when their judgments differed, and following discussion they updated their alignments based on their updated understanding.

At the close of Day 2, CMS estimated the consensus Item-ALD alignment for each item. That is, the modal (the most frequently occurring) achievement level selected by panelists for each item was identified as the "Initial Consensus Alignment." We refer to it as "initial" because a "Final Consensus Level" was estimated at the close of Day 3. Items with no modal achievement due to ties were reviewed again on Day 3.

Terminology: Initial Consensus Level. The *Initial Consensus Level* for an item is the achievement level to which the item was aligned by a plurality of panelists in the given grade and subject at the close of Day 2 of each workshop.

Terminology: Final Consensus Level. The *Final Consensus Level* for an item is the achievement level to which the item was aligned by a plurality of panelists in the given grade and subject at the close of Day 3 of each workshop.

Following the second day of each workshop, ESS analyses were conducted to identify ESS-Inconsistent items. Educators engaged in the review and resolution of these items on Day 3 of each workshop in a process designed by CMS, facilitated by edCount, and reported in detail in this section.

Next, we describe the recruitment and characteristics of educators participating in the ESS Alignment Study. We then provide a detailed account of the processes engaged by panelists on Day 3 of the ESS Alignment Study—the review and resolution of ESS-Inconsistent items.

The following information on the recruitment and characteristics of panelists supporting the ESS Alignment Study is cited from edCount's Alignment Evaluation for the Maine Through Year Assessment (2023).

Characteristics of Panelists

Fourteen panels of educators consisting of five panelists each (for a total of 70 educators) were recruited. One Reading panelist and six Mathematics panelists had to withdraw prior to the start of the rating process due to last-minute conflicts or technical difficulties, bringing the total number of study participants to 63, with no panel having fewer than 4 participants. Panelists were recruited to ensure representation of content area expertise and experience teaching and assessing students with disabilities and English learners. The majority of panelists (85.7%) had at least 10 years of education-related experience.

Each panelist completed a questionnaire that provided information about the panelist's gender, ethnicity, teaching, and professional experiences (see Exhibit 4). Demographic information for all panelists is summarized below.

Self-Reported Panelist Demographics

The panels were composed primarily of content area teachers, with over three-quarters of the panelists (76.2%) currently serving in this role. Panels also included educators serving in a wide variety of roles, including school administrators (6.3%), instructional coaches (3.2%), education consultants (3.2%), interventionists (3.2%), and one panelist each serving as a district administrator, department chair, curriculum coordinator, literacy specialist, and special

educator. Most panelists had experience working with students with disabilities (93.7%), working with a high percentage of economically disadvantaged students (93.7%), teaching content area courses (87.3%), working with gifted students (84.1%), and working with multilingual learners (57.1%). When asked to rate their level of understanding of the CCSS, all panelists indicated that they were familiar with the standards, with the majority (79.4%) indicating they were very familiar.

Panelists represented rural (63.5%), suburban (28.6%), and urban (7.9%) school districts and represented regions from all parts of Maine, including the Southern (33.3%), Midcoast (28.6%), Eastern (25.4%), Western (6.3%) and Northern (6.3%) regions. The majority of panelists identified as female, while five panelists (7.9%) identified as male, and one panelist preferred not to specify gender. Two panelists (3.2%) identified as Two or More Races, one panelist identified as Asian, one panelist preferred not to specify race, and the remaining panelists (96.8%) identified as White. Nearly three-quarters of the panelists (73.0%) had attained a master's degree or higher.

Over a quarter of panelists (28.6%) had a professional certification or endorsement in the content area associated with their panel. In addition, a majority of participants had an elementary (73.0%) certification or endorsement, while 46% had a middle school certification or endorsement, and 25.4 % had one for high school. Panelists also indicated administrator (23.8%), exceptional children/children with disabilities (11.1%), literacy coach (9.5%), English as a Second Language (6.3%), and curriculum instructional specialist (4.8%) among their listed certifications. On average, panelists had 17 years of experience in their educational area of expertise.

Exhibit 1 (in edCount Report): Panelist Self-Reported Demographic Characteristics

	Reading		Mathematics		Overall	
	N	%	N	%	N	%
Current Role						
Content Area Teacher	24	70.6	24	82.8	48	76.2
Instructional Coach	1	2.9	1	3.4	2	3.2
School Administrator	3	8.8	1	3.4	4	6.3
District Administrator	0	0.0	1	3.4	1	1.6
Other	6	17.6	2	6.9	8	12.7
Highest Degree Earned						
Bachelor's	7	20.6	10	34.5	17	27.0

	Reading		Mathematics		Overall	
	N	%	N	%	N	%
Master's	26	76.5	16	55.2	42	66.7
Specialist	0	0.0	2	6.9	2	3.2
Doctorate	1	2.9	1	3.4	2	3.2
Location in State						
Southern	11	32.4	10	34.5	21	33.3
Midcoast	10	29.4	8	27.6	18	28.6
Western	4	11.8	0	0.0	4	6.3
Eastern	7	20.6	9	31.0	16	25.4
Northern	2	5.9	2	6.9	4	6.3
Certifications and Endorsements (multiple responses possible per panelist)						
Elementary	24	70.6	22	75.9	46	73.0
Middle School	15	44.1	14	48.3	29	46.0
High School	9	26.5	7	24.1	16	25.4
English Language Arts	10	29.4	1	3.4	11	17.5
Mathematics	1	2.9	8	27.6	9	14.3
Administrator	10	29.4	5	17.2	15	23.8
Exceptional Children/Children with Disabilities	5	14.7	2	6.9	7	11.1
English as a Second Language	3	8.8	1	3.4	4	6.3
Curriculum Instructional Specialist	2	5.9	1	3.4	3	4.8
Literacy Coach	5	14.7	1	3.4	6	9.5
Years of Experience in Education						
5 or less	2	5.9	1	3.4	3	4.8
6–9	4	11.8	2	6.9	6	9.5
10–20	16	47.1	21	72.4	37	58.7
More than 20	12	35.3	5	17.2	17	27.0

Panelists were also asked to indicate which, if any, of the following statements described their past or current work experience (see Exhibit 5). Responses to these statements are summarized by content area below.

For both content areas, a majority of panelists indicated that they conduct or have conducted professional development for teachers in their content area and that they have participated in reviews of assessment items in their content area. Four panelists in each content area indicated they have taught future teachers in a higher education setting. Lastly, an average of approximately three-quarters of panelists across both content areas indicated that the statement, "I have/had a leadership role in curriculum planning in my school, district, or university position." applied to them.

Exhibit 2 (in edCount Report): Additional Panelist Experience

Statements	Reading		Mathematics		Overall	
	N	%	N	%	N	%
I conduct/have conducted professional development for teachers in my content area.	25	73.5	18	62.1	43	68.3
I have/had a leadership role in curriculum planning in my school, district, or university position.	25	73.5	22	75.9	47	74.6
I have taught future teachers in a higher education setting.	4	11.8	4	13.8	8	12.7
I have participated in reviews of assessment items in my content area.	19	55.9	20	69.0	39	61.9

ESS Alignment Study Day 3: Inconsistent Item Review and Resolution

ESS, with its foundations in principled design and emphasis on assessment system coherence, does more than support defensible and interpretable cut scores. ESS identifies cut scores that optimize the consistency of items' hypothesized Item-ALD alignments, in this case items' *Initial Consensus Levels*, and their empirical IRT scale locations. In so doing, the process also identifies ESS-Inconsistent items—items for which the Initial Consensus Level is not consistent with empirical data—providing an opportunity to further evaluate inconsistent items and resolve their inconsistencies.

ESS cut scores were estimated at the close of Day 2 based on each item's Initial Consensus Level and IRT RP location. The resulting cut scores were used to identify an empirical achievement level for each item, as described next.

Establishing ESS Empirical Item-ALD Alignments

After ESS cut scores are estimated, items are classified into empirical achievement levels based on the following criterion:

Level 1 (Well Below State Expectations): An item's Empirical level is Level 1 if the item's Item Response Theory (IRT) RP location is below the ESS Level 2 cut score.

Level 2 (Below State Expectations): An item's Empirical level is Level 2 if the item's IRT RP location is at or above the ESS Level 2 cut score but below the Level 3 cut score.

Level 3 (At State Expectations): An item's Empirical level is Level 3 if the item's IRT RP location is at or above the ESS Level 3 cut score but below the Level 4 cut score.

Level 4 (Above State Expectations): An item's Empirical level is Level 4 if the item's IRT RP location is at or above the ESS Level 4 cut score.

Terminology: Empirical Level. An item's *Empirical Level* refers to the Empirical Item-ALD alignment determined by the item's IRT RP location relative to the ESS cut scores.

An item was classified as ESS-Inconsistent when the Initial Consensus Level did not agree with the Empirical Level.

Terminology: ESS-Inconsistent Item. An item is *ESS-Inconsistent*, or simply inconsistent, when the Initial Consensus Level does not agree with the item's Empirical Level.

On Day 3, the inconsistent items were subject to review and resolution (R&R). Specifically, panelists engaged in a facilitated discussion of the inconsistent items, providing them an opportunity to reconsider their previous alignment judgments; develop a deeper shared understanding of the items, the ALDs, and their relationship; and to potentially resolve the inconsistencies between their Initial Consensus Levels and the Empirical Levels.

Terminology: Day 3 Activities. Day 3 workshop activities are also referred to as ESS-Inconsistent Item Review and Resolution, ESS Review and Resolution, or simply R&R.

The ESS Review and Resolution process consisted of a facilitated discussion of each inconsistent item. The discussion was intended to provide a better understanding of the items' content characteristics and measurement attributes in relation to the ALD evidence statements. Panelists made a final independent Item-ALD alignment for each item following discussion. Results of the R&R were used to (a) identify panelists' Day 3 Final Consensus Levels as summarized in the section "Review & Resolution Analyses" (b) form the basis for recommendations in the section "Recommendations", and (c) support the estimation of initial ESS cut scores as discussed in the section "Initial ESS Cut Score Estimation."

Training

Facilitator Training. Prior to the ESS Alignment Study, edCount Facilitators were provided training on key concepts and activities to support facilitation of the Day 3 R&R. Topics included an overview of the ESS procedures, key concepts relating to item consistency, important considerations for the review and resolution of inconsistent items, and an overview of the information facilitators would share and the focus of their facilitated discussions. Inconsistent Item Review & Resolution Facilitator Training Slides can be found in Appendix A.

Panelist Training. The first day of the three-day ESS Alignment workshop began with a plenary opening session detailing the study processes, item rubrics, and key concepts for the workshop. This included training on the ALDs and how panelists should document their alignment judgments. edCount's training slides are found in their Alignment Evaluation for the Maine Through Year Assessment (2023) and are replicated in Appendix B of this document.

Creative Measurement Solutions provided additional training at the Day 3 opening plenary session, including an overview of the Embedded Standard Setting procedure, key concepts relating to item consistency and alignment, important considerations for the review and resolution of inconsistent items, and an overview of the Day 3 panelist rating form. Day 3 plenary opening session presentation slides are provided in Appendix C.

ESS Review and Resolution Materials

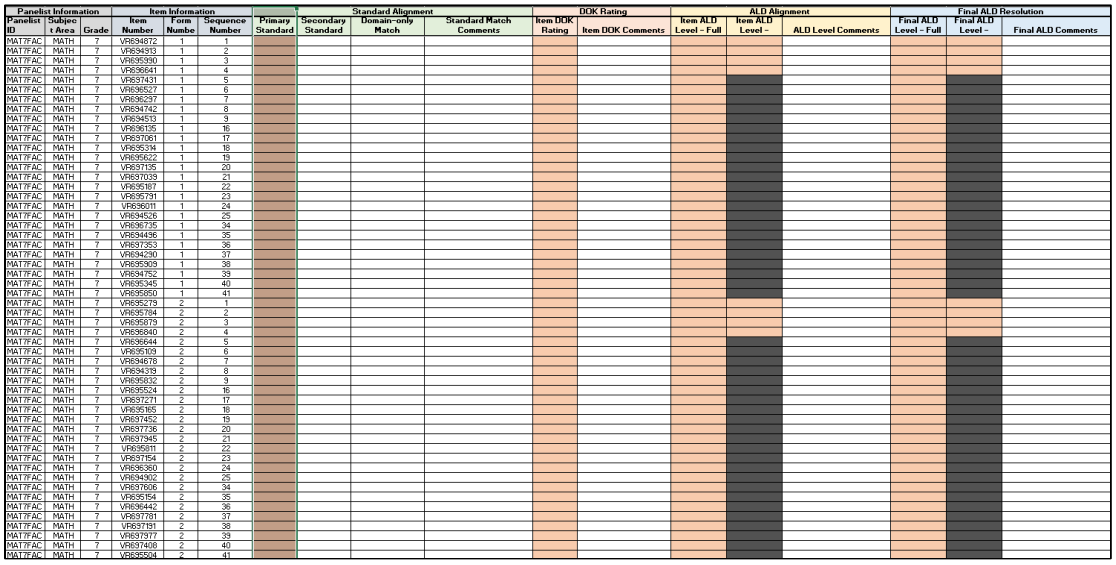
Panelist Materials

The following materials were provided to each panelist in a cloud folder.

- Guide to Accessing Resources and Data Entry
- Panelist Rating Form
- Common Core State Standards for ELA or Mathematics
- Achievement Level Descriptors (ALDs) for Reading or Mathematics

- Item Scoring Rubrics

Panelist Rating Form. Panelists worked from the same Day 1 and 2 rating forms, but with additional columns to document their final Item-ALD alignments for the reviewed items. An example of the Panelists' Day 3 R&R Rating form is shown in Figure 2. The additional columns, under the "Final ALD Resolution" header are provided to document their Day 3 judgments while preserving their initial alignments. On Day 3, as on Days 1 and 2, items were presented on the Panelist Rating Forms in test sequence order, not order of difficulty.



Panelist Information			Item Information			Final ALD Resolution		
Panelist ID	Subject Area	Grade	Item Number	Form Number	Sequence Number	Final ALD Level - Full Credit	Final ALD Level - 1 Point Correct	Final ALD Comments
MAT7FAC	MATH	7	VR694872	1	1			
MAT7FAC	MATH	7	VR694913	1	2			
MAT7FAC	MATH	7	VR695990	1	3			
MAT7FAC	MATH	7	VR696641	1	4			
MAT7FAC	MATH	7	VR697431	1	5			
MAT7FAC	MATH	7	VR696527	1	6			
MAT7FAC	MATH	7	VR696297	1	7			
MAT7FAC	MATH	7	VR694742	1	8			
MAT7FAC	MATH	7	VR694513	1	9			
MAT7FAC	MATH	7	VR696135	1	16			
MAT7FAC	MATH	7	VR697061	1	17			
MAT7FAC	MATH	7	VR695314	1	18			
MAT7FAC	MATH	7	VR695622	1	19			
MAT7FAC	MATH	7	VR697135	1	20			
MAT7FAC	MATH	7	VR697039	1	21			
MAT7FAC	MATH	7	VR695187	1	22			
MAT7FAC	MATH	7	VR695791	1	23			
MAT7FAC	MATH	7	VR696011	1	24			
MAT7FAC	MATH	7	VR694526	1	25			
MAT7FAC	MATH	7	VR696735	1	34			
MAT7FAC	MATH	7	VR694496	1	35			
MAT7FAC	MATH	7	VR697353	1	36			
MAT7FAC	MATH	7	VR694290	1	37			

Figure 2. Panelist R&R Rating Form: Full and Detailed Views

Facilitator Materials

Facilitator Form. Facilitators worked from a form similar in appearance to the panelists' forms but with additional information to guide discussion. A key difference in the Panelist's Rating Form from the Facilitator Form is that the items on the Facilitator Forms are ordered by difficulty. This was done to reflect the ordering that informed the ESS cut score analysis and to provide panelists with the context for the inconsistency of the items. For this reason, Item Information on the Facilitator Form included the Panelist Form Row, which allowed the facilitator to appropriately direct panelists' attention for each item when it was discussed.

An additional section of the facilitator form, labeled "Consistency Information," identified the item's order of difficulty, Empirical Level, Initial Consensus Level, and the consistency status. Consistent items were masked in gray to focus attention only on the inconsistent items that would be discussed. A final column allowed the facilitator to select the discussion status for each item. Setting the status to "Discussed" via a drop-down menu would mask that item in gray to indicate its status and to focus attention on items not yet discussed. An example of the Facilitators' R&R form is shown in Figure 3.

Item Information						Standard Alignment	Final ALD Resolution		Consistency Information				
Item Number	Panelist Row	Form Number	Sequence Number	Score Point	Max Points	Primary Standard	Final ALD Level - Item/Score Point	Final ALD Comments	Order of Difficulty	Empirical ALD Alignment Level	Consensus ALD Alignment Level	Consistency Status	R&R Discussion Status
VR702675_1	16	1	21	1	2	RL.3.1			1	Below	Below	Consistent	
VR702675_1	79	3	23	1	2	0			1	Below	Below	Consistent	
VR699667_1	4	1	2	1	2	RI.3.1			3	Below	Below	Consistent	
VR701589	43	2	14	1	1	RL.3.1			4	Below	Below	Consistent	
VR699670_1	7	1	5	1	2	RI.3.1			5	Below	Well Below	Inconsistent	
VR701587	41	2	12	1	1	RL.3.1			6	Below	Below	Consistent	
VR701540	39	2	10	1	1	RI.3.2			7	Below	Below	Consistent	
VR701594	47	2	26	1	1	RL.3.1			8	Below	Below	Consistent	
VR701538	38	2	9	1	1	RI.3.2			9	Below	Below	Consistent	
VR701536	36	2	7	1	1	RI.3.1			10	Below	Below	Consistent	
VR701591	44	2	23	1	1	L.3.4			11	Below	Below	Consistent	
VR701588	42	2	13	1	1	RL.3.1			12	Below	Below	Consistent	
VR701535	35	2	6	1	1	RI.3.1			13	Below	Well Below	Inconsistent	
VR701586	40	2	11	1	1	RL.3.1			14	Below	Below	Consistent	

Figure 3. Facilitator Form

Facilitator Guide and Scripts. Facilitators were provided with a guide and a step-by-step script for leading the discussion of each item. The Inconsistent Item Review & Resolution Facilitator Guide and Script is presented in Appendix D.

ESS Review and Resolution Proceedings

ESS Review and Resolution activities followed a similar structure for all grades and subjects, beginning with training in an opening plenary session, followed by a full day of R&R activities, described next.

Plenary Opening Training Session. A plenary opening session provided the context and background information for the day's activities. This included:

- An introduction to Embedded Standard Setting
- An introduction to consistent and inconsistent items
- Reasons why items may be inconsistent
- Ways that inconsistencies can be resolved

A review of the post Day 2 grade 8 Reading results prompted an addition to the opening session discussion. Specifically, only a modest number of items in grade 8 Reading were observed with Initial Consensus Levels other than At State Expectations. The main points of the added discussion topic follow:

- Three forms were systematically selected in grades 3-8 targeted to low-, moderate-, and high-achieving students. The three forms were targeted to students at the 16th, 50th, and 84th percentiles. High school had a fixed form.
- The diversity in the difficulty of items on the three forms in grades 3-8 was expected to result in Item-ALD alignments to all four levels.
- Items are presented in order of difficulty for the Day 3 R&R activity and thus, one might expect the easiest items to be associated with lower achievement levels and the hardest items to be associated with higher achievement levels.
- Panelists were encouraged to pay special attention to the Item-ALD alignments of items appearing at the beginning (the easiest items) and at the end (the hardest items) of their ordered inconsistent item review. However, ultimately it was the association of the items' content characteristics that must guide their decision making, not these expectations. Thus, panelists were asked to consider their Initial Consensus Levels, but not to automatically change them to comply with expectations.

The opening plenary session closed with a review of the day's R&R activities. The following information was conveyed:

Your Grade Group Facilitator will guide you, one by one, through each item identified as inconsistent. For each inconsistent item you will:

- …examine the ALD evidence statements associated with your panel's [initial] consensus alignments and those of the empirical level,
- discuss which evidence statement is the best match to the content characteristics of the item, and consider
    - maintaining or adjusting your alignment based on your new perspective following discussion, or
    - recommending refinements to the ALD evidence statements necessary for them to better support the [initial] consensus alignments.

Review & Resolution Activities. Following the opening plenary session the panelists convened into their grade-specific breakout rooms. The list of inconsistent items in each grade was presented in order of difficulty, beginning with the easiest inconsistent item. Facilitators followed the Facilitator Script (see Appendix D) to support discussion of each inconsistent item. Facilitated discussion of each inconsistent item was conducted and the workshop was concluded after all inconsistent items were discussed, and panelists documented their final Item-ALD alignments. The specific steps supporting the R&R included:

- For each item, the facilitator:
    - Identified the item number and, for polytomous items, the associated score point,
    - Identified the row associated with the item on the panelist rating form (which was not in order of difficulty) and confirmed all panelists were observing the correct item
    - Reviewed the item in the item viewer
    - Identified the item's
        - Consensus, or panel's modal, alignment to the CCSS
        - Initial Consensus Level
        - Empirical Level
    - Reviewed the ALD evidence statements associated with the Initial Consensus and Empirical Levels
    - Facilitated a discussion of the varying rationales for panelists' differing individual Item-ALD alignments

When discussion ceased to provide new information, facilitators directed panelists to enter their final Item-ALD alignment in the appropriate field on the panelist rating form. Panelists could maintain or modify their initial ratings. When all panelists confirmed they had entered their final alignments for the current item, they repeated these steps for the next item until all items had been discussed and final panelist alignments were documented.

We discuss the results derived from the panelists' final Item-PLD alignment ratings next.

Calculating Final Educator Consensus Levels

Final Consensus Levels were calculated at the conclusion of the R&R workshop. For items classified as consistent following Day 2, their Initial Consensus Levels were preserved as the Final Consensus Levels. For inconsistent items reviewed on Day 3, Final Consensus Levels were calculated as the mode of the panelists' final Item-ALD alignments. There were no modal ties.

Terminology: Final Consensus Level. The *Final Consensus Level* is the Item-ALD alignment selected by a plurality of the panelists on Day 3.

Review & Resolution Analyses

The following analyses are based on the items' Final Consensus Levels.

Agreement Rates. Panelists' agreement rates were calculated for the Initial Consensus Levels after Day 2 and the Final Consensus Levels after Day 3. The agreement rate is the percentage of panelists that selected the Initial or Final Consensus Alignment.

Based on the number of panelists (4 or 5), agreement rates were 40%, 60%, 80%, or 100% for the five-educator panels and 50%, 75%, or 100% for the four-educator panels. These percentages were associated with three levels of agreement:

- Plurality Agreement (40-50%)
- Clear Majority Agreement (60-80%)
- Consensus Agreement (100%).

Mathematics Agreement Rates. Agreement rates post Day 2 and post R&R are presented in Tables 2 and 3, respectively. Table 2 indicates that nearly all items on Day 2 reached their consensus alignments with either a Clear Majority (22% of items) or Full Consensus (77% of items). Only four items (less than 1%) reached only Plurality Agreement.

Table 3 indicates that on Day 3, no Plurality Agreement items remained and the shift from Clear Majority (14% of items) toward Consensus (86% of items) was strong. Table 4 indicates the number and percentage of items shifting agreement rate levels in each grade. Across all grades, 49 items (9%) shifted from Plurality or Clear Majority to Consensus.

Table 2. Post Day 2 Initial Consensus Level Agreement Rates: Math

Grade Level	Plurality (40-50%)	Clear Majority (60-80%)	Consensus (100%)	Total
Grade 3	0 (0%)	10 (11%)	78 (89%)	88 (100%)
Grade 4	2 (2%)	27 (30%)	62 (68%)	91 (100%)
Grade 5	0 (0%)	27 (31%)	60 (69%)	87 (100%)
Grade 6	0 (0%)	15 (17%)	74 (83%)	89 (100%)
Grade 7	0 (0%)	36 (40%)	53 (60%)	89 (100%)
Grade 8	2 (2%)	7 (8%)	80 (90%)	89 (100%)
High School	0 (0%)	3 (9%)	31 (91%)	34 (100%)
All Grades	4 (1%)	125 (22%)	438 (77%)	567 (100%)

Table 3. Post R&R Final Consensus Level Agreement Rates: Math

Grade Level	Plurality (40-50%)	Clear Majority (60-80%)	Consensus (100%)	Total
Grade 3	0 (0%)	7 (8%)	81 (92%)	88 (100%)
Grade 4	0 (0%)	15 (16%)	76 (84%)	91 (100%)
Grade 5	0 (0%)	15 (17%)	72 (83%)	87 (100%)
Grade 6	0 (0%)	6 (7%)	83 (93%)	89 (100%)
Grade 7	0 (0%)	33 (37%)	56 (63%)	89 (100%)
Grade 8	0 (0%)	4 (4%)	85 (96%)	89 (100%)
High School	0 (0%)	0 (0%)	34 (100%)	34 (100%)
All Grades	0 (0%)	80 (14%)	487 (86%)	567 (100%)

Table 4. Agreement Rate Differentials—Post R&R - Post Day 2: Math

Grade Level	Plurality (40-50%)	Clear Majority (60-80%)	Consensus (100%)
Grade 3	0 (0%)	-3 (-3%)	3 (3%)
Grade 4	-2 (-2%)	-12 (-14%)	14 (15%)
Grade 5	0 (0%)	-12 (-14%)	12 (14%)
Grade 6	0 (0%)	-9 (-10%)	9 (10%)
Grade 7	0 (0%)	-3 (-3%)	3 (3%)
Grade 8	-2 (-2%)	-3 (-4%)	5 (6%)
High School	0 (0%)	-3 (-9%)	3 (9%)
All Grades	-4 (-1%)	-45 (-8%)	49 (9%)

Reading Agreement Rates. Post Day 2 and post Day 3 Reading agreement rates are presented in Tables 5 and 6, respectively. All items on Day 2 reached either Clear Majority (38% of items) or Consensus (62% of items). On Day 3, though one item dropped to Plurality Agreement, the overall shift was from Clear Majority (25% of items) to Consensus (75% of items). Table 7

indicates the number (percentage) of inconsistent items shifting agreement rate levels. Across all grades, 70 items (13%) shifted from Plurality or Clear Majority to Consensus.

Table 5. Post Day 2 Initial Consensus Level Agreement Rates: Reading

Grade Level	Plurality (40-50%)	Clear Majority (60-80%)	Consensus (100%)	Total
Grade 3	0 (0%)	51 (66%)	26 (34%)	77 (100%)
Grade 4	0 (0%)	16 (23%)	55 (77%)	71 (100%)
Grade 5	0 (0%)	28 (37%)	48 (63%)	76 (100%)
Grade 6	0 (0%)	34 (40%)	51 (60%)	85 (100%)
Grade 7	0 (0%)	28 (30%)	64 (70%)	92 (100%)
Grade 8	0 (0%)	17 (18%)	76 (82%)	93 (100%)
High School	0 (0%)	28 (68%)	13 (32%)	41 (100%)
All Grades	0 (0%)	202 (38%)	333 (62%)	535 (100%)

Table 6. Post R&R Final Consensus Level Agreement Rates: Reading

Grade Level	Plurality (40-50%)	Clear Majority (60-80%)	Consensus (100%)	Total
Grade 3	0 (0%)	38 (49%)	40 (51%)	78 (100%)
Grade 4	0 (0%)	8 (11%)	63 (89%)	71 (100%)
Grade 5	0 (0%)	15 (20%)	61 (80%)	76 (100%)
Grade 6	0 (0%)	25 (29%)	60 (71%)	85 (100%)
Grade 7	1 (1%)	12 (13%)	80 (86%)	93 (100%)
Grade 8	0 (0%)	10 (11%)	83 (89%)	93 (100%)
High School	0 (0%)	25 (61%)	16 (39%)	41 (100%)
All Grades	1 (0%)	133 (25%)	403 (75%)	536 (100%)

Table 7. Agreement Rate Differentials—Post R&R - Post Day 2: Reading

Grade Level	Plurality (40-50%)	Clear Majority (60-80%)	Consensus (100%)
Grade 3	0 (0%)	-13 (-17%)	14 (17%)
Grade 4	0 (0%)	-8 (-12%)	8 (12%)
Grade 5	0 (0%)	-13 (-17%)	13 (17%)
Grade 6	0 (0%)	-9 (-11%)	9 (11%)
Grade 7	1 (1%)	-16 (-17%)	16 (17%)
Grade 8	0 (0%)	-7 (-7%)	7 (7%)
High School	0 (0%)	-3 (-7%)	3 (7%)
All Grades	1 (0%)	-69 (-13%)	70 (13%)

Final Consensus Levels by Achievement Level. The number of items aligned to each achievement level, based on the Final Consensus Levels are provided next.

Mathematics. The distribution of Mathematics Final Consensus Levels by achievement level are provided in Tables 8 and 9 for the post-Day 2 and post-R&R results, respectively, and indicate that the distribution of achievement levels across all grades is somewhat balanced. Table 8 indicates that post Day 2, the distribution of items across achievement levels for all grade levels is 17%, 31%, 39%, and 12% in Well Below, Below, At, and Above State Expectations, respectively.

Table 9 indicates that post R&R, we see only slight changes in these proportions, with 16%, 36%, 38%, and 10% items in all grades distributed in Well Below, Below, At, and Above State Expectations, respectively. High school had no changes from post Day 2 to post R&R. Note that overall item counts increased after Day 3 as some ties, not associated with any achievement level post Day 2, were resolved post R&R. Table 10 provides the differential in these distributions for Mathematics by grade and overall.

Table 8. Distribution of Post Day 2 Initial Consensus Levels: Mathematics

Grade Level	Well Below	Below	At	Above	Total
Grade 3	10 (11%)	20 (23%)	34 (39%)	24 (27%)	88 (100%)
Grade 4	7 (8%)	21 (23%)	51 (57%)	11 (12%)	90 (100%)
Grade 5	18 (21%)	28 (32%)	34 (39%)	7 (8%)	87 (100%)
Grade 6	19 (21%)	28 (31%)	34 (38%)	8 (9%)	89 (100%)
Grade 7	9 (10%)	35 (39%)	35 (39%)	10 (11%)	89 (100%)
Grade 8	31 (36%)	32 (37%)	22 (25%)	2 (2%)	87 (100%)
High School	3 (9%)	13 (38%)	12 (35%)	6 (18%)	34 (100%)
All Grades	97 (17%)	177 (31%)	222 (39%)	68 (12%)	564 (100%)

Table 9. Distribution of Post R&R Final Consensus Levels: Mathematics

Grade Level	Well Below	Below	At	Above	Total
Grade 3	4 (5%)	26 (30%)	35 (40%)	23 (26%)	88 (100%)
Grade 4	3 (3%)	16 (18%)	60 (66%)	12 (13%)	91 (100%)
Grade 5	12 (14%)	33 (38%)	37 (43%)	5 (6%)	87 (100%)
Grade 6	16 (18%)	46 (52%)	23 (26%)	4 (4%)	89 (100%)
Grade 7	5 (6%)	47 (53%)	34 (38%)	3 (3%)	89 (100%)
Grade 8	46 (52%)	25 (28%)	17 (19%)	1 (1%)	89 (100%)
High School	3 (9%)	13 (38%)	12 (35%)	6 (18%)	34 (100%)
All Grades	89 (16%)	206 (36%)	218 (38%)	54 (10%)	567 (100%)

Table 10. Distribution Differentials—Post R&R - Post Day 2: Mathematics

Grade Level	Well Below	Below	At	Above
Grade 3	-6 (-7%)	6 (7%)	1 (1%)	-1 (-1%)
Grade 4	-4 (-4%)	-5 (-6%)	9 (10%)	1 (1%)
Grade 5	-6 (-7%)	5 (6%)	3 (3%)	-2 (-2%)
Grade 6	-3 (-3%)	18 (20%)	-11 (-12%)	-4 (-4%)
Grade 7	-4 (-4%)	12 (13%)	-1 (-1%)	-7 (-8%)
Grade 8	15 (17%)	-7 (-8%)	-5 (-6%)	-1 (-1%)
High School	0 (0%)	0 (0%)	0 (0%)	0 (0%)
All Grades	-8 (-1%)	29 (5%)	-4 (-1%)	-14 (-2%)

Reading. The distribution of Reading Final Consensus Levels by achievement level are provided in Tables 11 and 12 for the post Day 2 and post R&R results, respectively.

No obvious patterns emerge across grades, but Table 11 indicates that for all grades, a preponderance of items (64%) were aligned to At State Expectations post Day 2 and most of the remaining items (29%) were aligned to Below State Expectations, placing 92% (with rounding) in the two middle achievement levels and 8% in the lowest and highest levels (5% in Well Below and 3% in Above State Expectations). post Day 2, grade 8 had 98% of all items in At State Expectations and no items falling into Well Below or Above State Expectations.

Changes from Day 2 to Day 3 were modest, overall. Table 12 indicates a small decrease From Day 2 in the percentage of items At State Expectations with small increases of items aligned to Below and Above State Expectations. Table 12 indicates the percentage of items aligned to Well Below State Expectations (3%), Below State Expectations (30%), At State Expectations (61%), and Above State Expectations (5%).

The Opening Session discussion of the diversity in item difficulty in grades 3-8 with encouragement to more deeply consider their Initial Consensus Levels for the easiest and hardest inconsistent items resulted in an improved spread of items across all levels in grade 8 Reading, reducing the proportion of items aligned to At State Expectations from 98% (Table 11) in their Initial Consensus Levels to 77% (Table 12) in their Final Consensus Levels, with 2%, 16%, and 4% in Well Below, Below, and Above State Expectations, respectively. Table 13 provides the differential in the distribution of Reading items by level post Day 2 to post R&R.

Table 11. Distribution of Post Day 2 Initial Consensus Levels: Reading

Grade Level	Well Below	Below	At	Above	Total
Grade 3	5 (6%)	36 (47%)	36 (47%)	0 (0%)	77 (100%)
Grade 4	2 (3%)	15 (21%)	50 (70%)	4 (6%)	71 (100%)
Grade 5	1 (1%)	18 (24%)	54 (71%)	3 (4%)	76 (100%)
Grade 6	8 (9%)	22 (26%)	53 (62%)	2 (2%)	85 (100%)
Grade 7	11 (12%)	49 (53%)	32 (35%)	0 (0%)	92 (100%)
Grade 8	0 (0%)	2 (2%)	91 (98%)	0 (0%)	93 (100%)
High School	0 (0%)	11 (27%)	24 (59%)	6 (15%)	41 (100%)
All Grades	27 (5%)	153 (29%)	340 (64%)	15 (3%)	535 (100%)

Table 12 Distribution of Post R&R Final Consensus Levels: Reading

Grade Level	Well Below	Below	At	Above	Total
Grade 3	1 (1%)	42 (54%)	31 (40%)	4 (5%)	78 (100%)
Grade 4	0 (0%)	9 (13%)	57 (80%)	5 (7%)	71 (100%)
Grade 5	2 (3%)	7 (9%)	59 (78%)	8 (11%)	76 (100%)
Grade 6	5 (6%)	22 (26%)	58 (68%)	0 (0%)	85 (100%)
Grade 7	8 (9%)	52 (56%)	31 (33%)	2 (2%)	93 (100%)
Grade 8	2 (2%)	15 (16%)	72 (77%)	4 (4%)	93 (100%)
High School	0 (0%)	14 (34%)	21 (51%)	6 (15%)	41 (100%)
All Grades	18 (3%)	161 (30%)	329 (61%)	29 (5%)	537 (100%)

Table 13. Distribution Differentials—Post R&R - Post Day 2: Reading

Grade Level	Well Below	Below	At	Above
Grade 3	-4 (-5%)	6 (8%)	-5 (-6%)	4 (5%)
Grade 4	-2 (-3%)	-6 (-8%)	7 (10%)	1 (1%)
Grade 5	1 (1%)	-11 (-14%)	5 (7%)	5 (7%)
Grade 6	-3 (-4%)	0 (0%)	5 (6%)	-2 (-2%)
Grade 7	-3 (-3%)	3 (3%)	-1 (-1%)	2 (2%)
Grade 8	2 (2%)	13 (14%)	-19 (-20%)	4 (4%)
High School	0 (0%)	3 (7%)	-3 (-7%)	0 (0%)
All Grades	-9 (-2%)	8 (1%)	-11 (-2%)	14 (3%)

Final Consensus Level Resolution Types. The Final Consensus Level resolutions are classified in Table 14 to Table 17 according to the following types:

- Final Consensus Level agrees with Empirical Alignment
- Final Consensus Level agrees with Initial Consensus Alignment
- Final Consensus Level is aligned to a New Level (neither Empirical nor Initial Consensus Level)

Counts and percentages of items classified as Consistent following Day 2 are included in the resolution tallies.

Reading. The distribution of Final Consensus Level Resolution types for Reading is provided in Table 14. For Reading, over half of all items in each grade, and two thirds of items across all grades, were consistent (i.e., Initial Consensus Level agreed with the Empirical Level). Of the inconsistent items, the R&R resulted in an additional 23% in agreement with Empirical, 12% of items retaining their initial Consensus Level, and 2% aligned to a new level altogether.

For Final Consensus Level resolutions associated with a New Level, the New Level can be ordered. That is, the New Level can be:

- Higher than Initial Consensus and Empirical Levels,
- Lower than Initial Consensus and Empirical Levels, or
- Between the Initial Consensus and Empirical Levels

Of those inconsistent items located to a New Level, Table 15 indicates that two-thirds are ordered between the Initial Consensus Level and the Empirical Level and one-third are lower than both the Empirical and Initial Consensus Levels.

Table 14. Final Consensus Level Resolution Types, Reading

Grade Level	Consistent Post Day 2	Agree with Empirical Level	Agree with Initial Consensus Level	New Level	Total
Grade 3	51 (65%)	24 (31%)	3 (4%)	0 (0%)	78 (100%)
Grade 4	46 (65%)	11 (15%)	12 (17%)	2 (3%)	71 (100%)
Grade 5	44 (58%)	23 (30%)	6 (8%)	3 (4%)	76 (100%)
Grade 6	57 (67%)	14 (16%)	14 (16%)	0 (0%)	85 (99%)
Grade 7	48 (52%)	27 (29%)	16 (17%)	2 (2%)	93 (100%)
Grade 8	71 (76%)	17 (18%)	3 (3%)	2 (2%)	93 (99%)
High School	25 (61%)	7 (17%)	9 (22%)	0 (0%)	41 (100%)
All Grades	342 (64%)	123 (23%)	63 (12%)	9 (2%)	537 (100%)

Table 15. New Level Order, Relative to Consensus and Empirical Levels, Reading

Grade Level	Higher	Lower	Between	Total
Grade 3	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Grade 4	0 (0%)	0 (0%)	2 (100%)	2 (100%)
Grade 5	0 (0%)	1 (33%)	2 (67%)	3 (100%)
Grade 6	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Grade 7	0 (0%)	0 (0%)	2 (100%)	2 (100%)
Grade 8	0 (0%)	2 (100%)	0 (0%)	2 (100%)
High School	0 (0%)	0 (0%)	0 (0%)	0 (0%)
All Grades	0 (0%)	3 (33%)	6 (67%)	9 (100%)

Mathematics. The distribution of Final Consensus Level resolution types for Mathematics is given in Table 16. Across all grades, slightly less than half of all items were consistent (i.e., the Initial Consensus Level agrees with the Empirical Level). Of the inconsistent items, the Final Consensus Level resolution placed 21% in agreement with the Empirical Level, 28% of items retaining their Initial Consensus Levels, and 2% were aligned to a New Level. Table 17 indicates that, among the items aligned to a New Level, 71% were ordered between the Initial Consensus

Level and the Empirical Level, 14% were ordered higher than the Empirical Level, and 14% were ordered lower than the Initial Consensus Levels. Details of the distribution of the New Levels relative to the Consensus and Empirical Levels for Mathematics are found in Table 17.

Table 16. Resolution Types, Mathematics

Grade Level	Consistent Post Day 2	Agree with Empirical Level	Agree with Initial Consensus Level	New Level	Total
Grade 3	43 (49%)	16 (18%)	28 (32%)	1 (1%)	88 (100%)
Grade 4	58 (64%)	20 (22%)	13 (14%)	0 (0%)	91 (100%)
Grade 5	40 (46%)	14 (16%)	30 (34%)	3 (3%)	87 (99%)
Grade 6	40 (45%)	24 (27%)	22 (25%)	3 (3%)	89 (100%)
Grade 7	40 (45%)	22 (25%)	26 (29%)	1 (1%)	89 (100%)
Grade 8	43 (48%)	17 (19%)	24 (27%)	5 (6%)	89 (100%)
High School	13 (38%)	4 (12%)	16 (47%)	1 (3%)	34 (100%)
All Grades	277 (49%)	117 (21%)	159 (28%)	14 (2%)	567 (100%)

Table 17. New Level Order, Relative to Consensus and Empirical Levels, Mathematics

Grade Level	Higher	Lower	Between	Total
Grade 3	0 (0%)	0 (0%)	1 (100%)	1 (100%)
Grade 4	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Grade 5	2 (67%)	0 (0%)	1 (33%)	3 (100%)
Grade 6	0 (0%)	1 (33%)	2 (67%)	3 (100%)
Grade 7	0 (0%)	0 (0%)	1 (100%)	1 (100%)
Grade 8	0 (0%)	1 (20%)	4 (80%)	5 (100%)
High School	0 (0%)	0 (0%)	1 (100%)	1 (100%)
All Grades	2 (14%)	2 (14%)	10 (71%)	14 (100%)

Post Review & Resolution Panelist Survey Results

Following the R&R activities, panelists completed an online evaluation to gather feedback on the quality of the R&R workshop process, materials, and associated artifacts. Panelists were asked to rate their agreement—Strongly Agree, Agree Somewhat, Disagree Somewhat, or Strongly Disagree—with a series of statements about the R&R activity and their experiences. The evaluation was completed by 61 of the 63 panelists (96.8%). The results are presented for all grades by subject in Tables 18 and 19 for Mathematics and Reading, respectively. Evaluation results by grade and subject are provided in Appendix E.

The evaluations reflect high levels of understanding of their task, and satisfaction with the process and outcomes of the alignment study, as reflected by median levels of agreement at Strongly Agree (4) or Agree Somewhat (3) for all evaluation statements.

Table 18. Evaluation Results: Mathematics – All Grades

All Grades (N=28)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	1	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	1	4	4
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	1	4	4
It was easy to align items to specific achievement levels using the ALDs.	1	4	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	1	4	4
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	1	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	1	4	4
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	1	4	4
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	1	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	1	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	1	4	3

Table 19. Evaluation Results: Reading – All Grades

All Grades (N=28)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	1	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	1	4	4
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	1	4	4
It was easy to align items to specific achievement levels using the ALDs.	1	4	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	1	4	4
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	1	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	1	4	3
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	1	4	4
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	1	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	1	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	1	4	3

Recommendations

The R&R workshop was conducted to investigate, resolve, and perhaps understand the rationale for items with Initial Consensus Levels that did not agree with the Empirical Levels. Post R&R, a negligible number of items reached only Plurality Agreement, while 14% of Mathematics items (Table 3) and 25% of Reading items (Table 6) reached Majority Agreement but not Consensus (100% agreement). The vast majority of items in both Reading and Mathematics achieved 100% (Consensus Agreement) in the Final Consensus Levels.

The nature of the R&R resolutions informs specific recommendations beyond identifying Item-ALD alignments of record. Our recommendations are based on the following logic: Inconsistent items have two independent Item-ALD alignments that disagree—the Initial Consensus Level and Empirical Level. When such a situation arises in a well-researched area of measurement—the scoring of constructed response items with complex rubrics where two raters provide

different ratings—the standard practice for high-stakes assessments is to seek resolution in a third independent review by a highly qualified rater. The workshop panels provided that third review (though not independent) in the R&R on Day 3. Specific recommendations depending on resolution type follow:

Panelists' Initial Consensus Level agrees with the Empirical Level. These items were consistent following Day 2 and required no resolution during the R&R workshop. The agreement between the Empirical Level and Panelists' Initial Consensus Level provides evidence that content characteristics and measurement attributes of the item fit the evidence statements of the Empirical Level.

Recommended Actions: The Initial Consensus Level = Empirical Level is the alignment of record.

Panelists' Final Consensus Level agrees with the Initial Consensus Level. A Final Consensus Level in agreement with the Initial Consensus Level supports the panelists' initial judgments—the item's content characteristics and measurement attributes correctly reflect the evidence statements of the initially identified ALD; however, the lack of support by empirical data indicates that some element of the item's construction or issues related to students' opportunity to learn (OTL) may have an effect on the item's difficulty.

Recommended Actions: The initial Consensus Level is preserved as the alignment of record, but the item is flagged for review.

Panelists' Final Consensus Level agrees with the Empirical Level. A Final Consensus Level that confirms the Empirical Level resolves the inconsistency and results in the alignment of record as the Final Consensus Level = the Empirical Level. That is, following analysis and discussion during the R&R, panelists assert that the content characteristics and measurement attributes of the item better fit the evidence statements of the Empirical Level than those of their Initial Consensus alignment.

Recommended Actions: The Final Consensus Level = Empirical Level is the alignment of record.

Panelists' Final Consensus Level disagrees with both the Initial Consensus Level and the Empirical Level. A Final Consensus Level that agrees with neither the Initial Consensus Level nor the Empirical Level reflects a refinement in the panelists' thinking. After weighing all available evidence independently and following discussion, panelists identified a level in disagreement with their Initial Consensus Level and with the Empirical Level. However, the lack of support by empirical data indicates that some element of the item's construction or issues related to students' opportunity to learn (OTL) may affect the item's difficulty.

Recommended Actions: The Final Consensus Level is the alignment of record, but the item is flagged for review.

Summary

Item review and resolution has a potentially powerful and beneficial effect on score interpretation. As inconsistencies are resolved, it brings intended and observed interpretations of the cut scores into coherence with the ALDs, Item-ALD alignments, and empirical data. Applying the results of the R&R workshop increased the percentage of consistent items from 64% post Day 2 to 87% post R&R for Reading (Table 14) and from 49% post Day 2 to 70% post R&R for Mathematics (Table 16). While 13% of Reading items and 30% of Mathematics items remain inconsistent, this is not a cause for concern. Inconsistent items that affect score interpretation are not a reflection of the quality of the Maine Through Year Assessments. They exist under any item-based standard setting methodology (i.e., Bookmark, ID Matching, Yes-No Angoff, etc.) but go undetected under these other approaches. ESS minimizes the effect of item inconsistency on score interpretation and offers opportunities to further mitigate it through iterative review and revision.

Section 4. Embedded Standard Setting Analyses

ESS analyses use data from the spring 2023 MTYA administration and Panelists' Final Consensus Levels to provide four key outcomes:

First, initial ESS cut scores emerge analytically and organically by optimizing the coherence of the Final Consensus Levels and the empirical data. We use the initial ESS cut scores and items' IRT RP50 locations (see the section "Methodological Choice: IRT Response Probability RP50" to establish an ESS Empirical Level for each item, defined as follows:

ESS Empirical Item-ALD Alignments. After initial ESS cut scores are estimated, items are classified into the following Empirical Levels if the item's IRT RP50 location is:

Level 1 (Well Below): Below the Level 2 cut score

Level 2 (Below): At or above the Level 2 cut score but below the Level 3 cut score

Level 3 (At): At or above the Level 3 cut score but below the Level 4 cut score

Level 4 (Above): At or above the Level 4 cut score

Terminology. Empirical Level: An item's *Empirical Level* refers to the ESS Empirical Item-ALD alignment determined by the item's IRT RP50 location relative to the ESS cut scores.

Second, ESS analyses provide the information necessary to evaluate the efficacy of the Final Consensus Levels. Evaluation criteria include:

a) the correlation of empirical item difficulty (IRT RP50 location) and the ordinality of the panelists' Final Consensus Levels (Level 1 = 1, Level 2 = 2, Level 3 = 3, Level 4 = 4),

b) agreement rates between panelists' Final Consensus Levels and the Empirical Levels derived from the ESS cut scores, and

c) weighted Kappa values that quantify the degree to which the panelists' Final Consensus Levels are concordant with the Empirical Levels.

Third, ESS analyses provide impact data—the proportion of students in each achievement level—is estimated.

Fourth, lists of ESS-Inconsistent items are produced. These are items with Final Consensus Levels that are not supported by empirical data. Each of these outcomes are described in this section. But first, we describe a few concepts and metrics used to contextualize agreement rate and Kappa.

Data

NWEA conducted analyses based on the Spring 2023 MTYA administration. Details are provided in the NWEA Maine Through Year Assessment Program Technical Report (NWEA, 2023). Table 20 provides the number of students included in the impact data reported in this section.

Table 20. Impact Data N-Counts: Mathematics and Reading

	Grade						
Subject	3	4	5	6	7	8	HS
Mathematics	12,151	12,140	11,927	12,083	12,250	12,627	12,318
Reading	12,085	12,060	11,863	12,042	12,182	12,574	12,141

Methodological Choice: IRT Response Probability RP50

The Maine Through Year Assessment standard setting design document states that both RP50 and RP67 response probabilities will be considered. RP50 was selected. The rationale is based on several factors and precedent for adaptive assessments. First, the use of RP50 for adaptive assessments has precedent; it was used for the Smarter Balanced standard setting in 2014 (Cizek; 2014). The Smarter Balanced and the Maine Through Year Assessment (in grades 3-8) are both adaptive assessments and RP50 corresponds to most adaptive algorithms with respect to the targeting of item difficulty and student ability.

RP50 relaxes the performance expectations associated with RP67. This was deemed appropriate, given the precedent cited above for adaptive assessments and the following three contextual elements:

Opportunity to Learn. The following citations were listed in the Cognitive Challenge Surveys produced by edCount (2023) in the ESS Alignment Study:

- Mathematics Grade 7: “We often do not get through ALL the curriculum, but the assessment had a decent coverage.”
- Mathematics Grade 8: “There was a good amount on Stats and Prob, more than what you would see covered in an instructional year.”

COVID effects. Recent research indicates that student achievement has not yet returned to its pre-COVID levels (NWEA, 2021).

Rigorous Achievement Level Descriptors. The ALDs were cited to be rigorous. The mathematics grade 7 Cognitive Challenge Survey produced by edCount in the ESS Alignment Study (edCount, 2023) indicated that “I think the ALDs are a bit high. The Below feels more like the At compared to the standards.”

Thus, RP50 was adopted as the response probability value.

Initial ESS Cut Score Estimation

Embedded Standard Settidng (Lewis & Cook, 2020) cut scores are estimated by optimizing the coherence between the Final Consensus Levels and empirical data. That is, cut scores emerge organically and analytically from the empirically tested Item-ALD alignments by optimizing the evidentiary relationship between items and the claims and measurement targets articulated in the ALDs.

ESS cut scores were estimated using the ESS-Weight algorithm described by Lewis & Cook (2020) and Lewis, Lee, and Choi (2021). ESS-Weight can be expressed mathematically as:

$$\text{ESS-Weight} \equiv \arg\min_{c} \sum_{i=1}^{n} I(\text{Inconsistent}) \cdot |D_i| \text{ , where}$$

$$D_i \equiv RP_i - c \text{ ,}$$

where RP_i is the IRT RP location for item i and c is a cut score candidate. In this notation the second term in the summand, $|D_i|$, is clearly seen as the weight for item i, $W_i = |D_i|$. |D*i*| can be interpreted as the scale distance required to shift the location of an Inconsistent item, *i*, to the threshold of the panelists' Consensus Level. The binary indicator function, $I(\text{Inconsistent})$, for an item or score point is set to 1 if the item or score point is ESS-Inconsistent and 0 otherwise.

ESS-Distance. Given a cut score, we define the Distance of an ESS-Inconsistent item as the minimum number of scale score points that the item's IRT location must shift to place the item at a border of the Final Consensus Level. The greater the ESS-Distance of an inconsistent item, the greater the magnitude of inconsistency.

Terminology: ESS-Distance (Distance). The *Distance* associated with an item is the minimum number of scale score points that the item's IRT location must shift to place the item at a border of the Final Consensus Level.

An item, or within-item score point, is called ESS-Inconsistent if $L_i^{(SME)} \neq L_i^{(RP|c)}$, that is, the SME Item-ALD aligned level for item or score point i is not equal to the ESS Empirical Item-ALD level based on the prompt's IRT RP location relative to cut score candidate c.

Essentially Consistent. We say that an item is *Essentially Consistent* if the absolute value of its Distance is less than or equal to 1 Standard Error of Measurement (SEM) of the test. This arbitrary, but not capricious, metric is useful when SMEs are engaged in the exercise of resolving inconsistent items. That is, Essentially Consistent items have Distances that are so

inconsequential that a SME would be unlikely to be able to identify a content-based rationale for the inconsistency.

Terminology: Essentially Consistent Item. We say that an item is *Essentially Consistent* if the absolute value of its Distance is less than or equal to 1 Standard Error of Measurement (SEM) of the test. Essentially Consistent items have Distances that are so inconsequential that a SME would be unlikely to be able to identify a content-based rationale for the inconsistency; thus, attempts to resolve the inconsistency are not likely to be successful.

Initial Cut Scores. ESS cut scores are estimated by identifying the minimum value of ESS-Weight for all cut score candidates across the test scale. The cut scores produced from the RP50 ESS-Weight algorithm are provided in Table 21. These cut scores are referred to as "initial" cut scores because they may be adjusted during vertical articulation. The initial cut scores are used to evaluate the efficacy of the panelists' Final Consensus Levels, described next.

Table 21. Initial ESS Cut Scores

Content Area	Grade	Below State Expectations Cut Score	At State Expectations Cut Score	Above State Expectations Cut Score
Mathematics	3	-2.8	-0.75	0.53
	4	-3.4	-1.95	2.02
	5	-1.73	0.11	1.63
	6	-1.5	1.09	2.66
	7	-2.73	0.27	1.96
	8	-0.76	0.79	2.29
	HS	-1.36	-0.47	0.62
Reading	3	-2.87	-0.05	2.62
	4	-3.2	-0.96	1.43
	5	-2.2	-0.99	1.52
	6	-1.92	-0.63	2.43
	7	-1.34	0.64	2.05
	8	-1.42	-0.77	2.09
	HS	-2.07	-0.56	0.95

Efficacy of Educators' Final Consensus Levels

In this section we examine criteria used to analyze the efficacy of the educators' Consensus Levels. First, we estimate the correlation of the Final Consensus Level ordinality (Level 1 = 1, Level 2 = 2, Level 3 = 3, Level 4 = 4) and the IRT RP50 location for each item and within-item score point. Then, we provide two metrics summarizing the concordance between the Final Consensus Levels and the Empirical Levels including agreement rates and weighted Kappa values.

We also provide detailed crosstab tables displaying the Final Consensus Levels crossed with the Empirical Levels. Two crosstabs are provided for each grade and subject. The second crosstab in each table is the result of reclassifying Essentially Consistent items as consistent. The classifications in the second crosstab are used in the tables summarizing classification agreement and weighted Kappa. A detailed description of each of these follows.

Correlations

Table 22 lists the correlations of items' Final Consensus Level ordinality and RP50 location by grade for each subject. The column labeled "Unadjusted" is the standard Pearson correlation coefficient. However, because the RP50 location is a continuous variable and Consensus Level ordinality is an ordinal variable, the maximum correlation under perfect alignment is constrained to less than 1. We adjust for this to better interpret the magnitude of the correlation by estimating the "Maximum Correlation," defined as the correlation between the perfectly ordered Empirical Levels and the RP50 locations. The ratio of the Unadjusted to Maximum Correlation is reported as the Adjusted Correlation.

The Adjusted Correlations for Mathematics and Reading ranged from 0.25 to 0.83 with a median of 0.58 and 0.21 to 0.96 with a median of 0.76, respectively.

Table 22. Correlations

Subject	Grade	Correlations		
		Unadjusted	Maximum	Adjusted
Math	3	0.57	0.91	0.62
	4	0.7	0.85	0.83
	5	0.37	0.94	0.39
	6	0.57	0.93	0.61
	7	0.39	0.89	0.44
	8	0.54	0.94	0.58
	HS	0.23	0.93	0.25
Reading	3	0.84	0.87	0.96
	4	0.16	0.78	0.21
	5	0.6	0.79	0.76
	6	0.63	0.84	0.75
	7	0.6	0.88	0.69
	8	0.71	0.81	0.87
	HS	0.78	0.90	0.87

Classification Agreement and Weighted Kappa

Classification Agreement. Classification agreement is described in the following terms:

Agree: The Empirical Level agrees with the Final Consensus Level

Disagree Adjacent: The Empirical Level disagrees with the Final Consensus Level, but they are adjacent levels.

Disagree Discrepant: The Empirical Level disagrees with the Final Consensus Level, and they are not adjacent levels.

Classification agreement is graphically represented as a crosstab as illustrated in Table 23.

Table 23. Classification Agreement Crosstab

		Consensus Level			
		Level 1	**Level 2**	**Level 3**	**Level 4**
Empirical Level	**Level 1**	Agree	Disagree: Adjacent	Disagree: Discrepant	Disagree: Discrepant
	Level 2	Disagree: Adjacent	Agree	Disagree: Adjacent	Disagree: Discrepant
	Level 3	Disagree: Discrepant	Disagree: Adjacent	Agree	Disagree: Adjacent
	Level 4	Disagree: Discrepant	Disagree: Discrepant	Disagree: Adjacent	Agree

Weighted Kappa. In addition to classification agreement rates, we also provide the weighted Kappa statistic using quadratic weighting. The Kappa statistic is a value from 0 to 1 that indicates how two types of independent classifications of the same phenomenon compare to random classifications. Higher values indicate stronger agreement between the two independent classifications. The quadratic weighting penalizes disagreements that are discrepant more than disagreements that are adjacent. To aid in the interpretation of the Kappa values, Table 24 displays the recommended ranges suggested by Landis and Koch (1977).

Table 24. Recommended Kappa Interpretations

Kappa Value	Strength of Agreement
0	None
<0.21	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost Perfect

Table 25 provides the classification agreement and weighted Kappa values by subject and grade. Classification agreement rates for Mathematics and Reading ranged from 41% to 82% with a median of 65% and from 73% to 95% with a median of 82%, respectively. Weighted Kappa values for Mathematics and Reading ranged from 0.16 to 0.81 with a median of 0.54 and from 0.20 to 0.93 with a median of 0.72, respectively. The Kappa values ranged from slight to almost perfect for both Mathematics and Reading according to the recommended interpretations provided in Table 24.

Table 25. Classification Agreement and Weighted Kappa

Subject	Grade	Agreement Rate	Weighted Kappa
Math	3	64%	0.54
	4	82%	0.81
	5	57%	0.42
	6	69%	0.61
	7	65%	0.41
	8	71%	0.55
	HS	41%	0.16
Reading	3	95%	0.93
	4	73%	0.20
	5	82%	0.61
	6	82%	0.72
	7	78%	0.69
	8	91%	0.86
	HS	80%	0.79

Crosstabs

The crosstabs are provided in Tables 27 to 40, which cross the items' Consensus Levels and the Empirical Levels established from ESS cut scores under two scenarios: ESS-Weight without reclassification of "essentially consistent" items and ESS-Weight after reclassifying "essentially consistent" items as consistent. Recall that an essentially consistent item has an ESS-Distance with absolute value less than the SEM.

The crosstabs reflect the *classification agreement* between the items' Empirical Levels and the Consensus Levels; classification agreement rates and Cohen's Kappa are provided with the crosstabs.

Table 26 displays the standard errors of measurement of the assessments (SEM_{MTYA}) and the standard errors of the ESS cut scores (SE_{ESS}) estimated with the bootstrap methods described by Lewis, Lee, and Choi (2021, May).

Table 26. Standard Errors by Grade and Content Area

Content Area	Grade	SEM_{MTYA}	SE_{ESS}		
			Below	At	Above
Math	3	0.41	0.56	0.18	0.19
	4	0.40	0.33	0.15	0.25
	5	0.41	0.28	0.24	0.27
	6	0.40	0.17	0.31	0.45
	7	0.41	0.29	0.25	0.36
	8	0.41	0.28	0.20	0.33
	HS	0.43	0.38	0.26	0.49
Reading	3	0.41	0.51	0.10	0.16
	4	0.41	0.67	0.28	0.29
	5	0.41	0.62	0.18	0.25
	6	0.41	0.42	0.15	0.13
	7	0.40	0.31	0.12	0.31
	8	0.41	0.29	0.16	0.18
	HS	0.36	0.28	0.15	0.22

Table 27. Crosstabs of Target and Empirical Levels: Mathematics Grade 3

Based on ESS-Weight cut scores

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	0 0%	2 2.3%	0 0%	1 1.1%	3 3.4%
	Level 2	2 2.3%	17 19.3%	3 3.4%	2 2.3%	24 27.3%
	Level 3	2 2.3%	6 6.8%	21 23.9%	8 9.1%	37 42%
	Level 4	0 0%	1 1.1%	11 12.5%	12 13.6%	24 27.3%
	Total	4 4.5%	26 29.5%	35 39.8%	23 26.1%	88 100%

Agree	50	57%
Disagree	38	43%
Adjacent Disagreement	32	36%
Discrepant Disagreement	6	7%
Weighted Kappa	0.51	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	0 0%	2 2.3%	0 0%	1 1.1%	3 3.4%
	Level 2	2 2.3%	19 21.6%	1 1.1%	2 2.3%	24 27.3%
	Level 3	2 2.3%	6 6.8%	25 28.4%	4 4.5%	37 42%
	Level 4	0 0%	1 1.1%	11 12.5%	12 13.6%	24 27.3%
	Total	4 4.5%	28 31.8%	37 42%	19 21.6%	88 100%

Agree	56	64%
Disagree	32	36%
Adjacent Disagreement	26	30%
Discrepant Disagreement	6	7%
Weighted Kappa	0.54	

Table 28. Crosstabs of Target and Empirical Levels: Mathematics Grade 4

Based on ESS-Weight cut scores

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	2 2.2%	1 1.1%	0 0%	0 0%	3 3.3%
	Level 2	1 1.1%	13 14.3%	3 3.3%	0 0%	17 18.7%
	Level 3	0 0%	2 2.2%	50 54.9%	6 6.6%	58 63.7%
	Level 4	0 0%	0 0%	7 7.7%	6 6.6%	13 14.3%
	Total	3 3.3%	16 17.6%	60 65.9%	12 13.2%	91 100%

Agree	71	78%
Disagree	20	22%
Adjacent Disagreement	20	22%
Discrepant Disagreement	0	0%
Weighted Kappa	0.75	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	3 3.3%	0 0%	0 0%	0 0%	3 3.3%
	Level 2	1 1.1%	16 17.6%	0 0%	0 0%	17 18.7%
	Level 3	0 0%	2 2.2%	50 54.9%	6 6.6%	58 63.7%
	Level 4	0 0%	0 0%	7 7.7%	6 6.6%	13 14.3%
	Total	4 4.4%	18 19.8%	57 62.6%	12 13.2%	91 100%

Agree	75	82%
Disagree	16	18%
Adjacent Disagreement	16	18%
Discrepant Disagreement	0	0%
Weighted Kappa	0.81	

Table 29. Crosstabs of Target and Empirical Levels: Mathematics Grade 5

Based on ESS-Weight cut scores

		Consensus Level				
		Level 1	Level 2	Level 3	Level4	Total
Empirical Level	Level 1	3 3.4%	4 4.6%	3 3.4%	1 1.1%	11 12.6%
	Level 2	6 6.9%	18 20.7%	10 11.5%	0 0%	34 39.1%
	Level 3	3 3.4%	10 11.5%	22 25.3%	1 1.1%	36 41.4%
	Level 4	0 0%	1 1.1%	2 2.3%	3 3.4%	6 6.9%
	Total	12 13.8%	33 37.9%	37 42.5%	5 5.7%	87 100%

Agree	46	53%
Disagree	41	47%
Adjacent Disagreement	33	38%
Discrepant Disagreement	8	9%
Weighted Kappa	0.36	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	5 5.7%	2 2.3%	3 3.4%	1 1.1%	11 12.6%
	Level 2	6 6.9%	20 23%	8 9.2%	0 0%	34 39.1%
	Level 3	3 3.4%	10 11.5%	22 25.3%	1 1.1%	36 41.4%
	Level 4	0 0%	1 1.1%	2 2.3%	3 3.4%	6 6.9%
	Total	14 16.1%	33 37.9%	35 40.2%	5 5.7%	87 100%

Agree	50	57%
Disagree	37	43%
Adjacent Disagreement	29	33%
Discrepant Disagreement	8	9%
Weighted Kappa	0.42	

Table 30. Crosstabs of Target and Empirical Levels: Mathematics Grade 6

Based on ESS-Weight cut scores

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	6 6.7%	10 11.2%	1 1.1%	0 0%	17 19.1%
	Level 2	8 9%	31 34.8%	5 5.6%	1 1.1%	45 50.6%
	Level 3	2 2.2%	4 4.5%	15 16.9%	3 3.4%	24 27%
	Level 4	0 0%	1 1.1%	2 2.2%	0 0%	3 3.4%
	Total	16 18%	46 51.7%	23 25.8%	4 4.5%	89 100%

Agree	52	58%
Disagree	37	42%
Adjacent Disagreement	32	36%
Discrepant Disagreement	5	6%
Weighted Kappa	0.5	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	14 15.7%	2 2.2%	1 1.1%	0 0%	17 19.1%
	Level 2	8 9%	31 34.8%	5 5.6%	1 1.1%	45 50.6%
	Level 3	2 2.2%	4 4.5%	16 18%	2 2.2%	24 27%
	Level 4	0 0%	1 1.1%	2 2.2%	0 0%	3 3.4%
	Total	24 27%	38 42.7%	24 27%	3 3.4%	89 100%

Agree	61	69%
Disagree	28	31%
Adjacent Disagreement	23	26%
Discrepant Disagreement	5	6%
Weighted Kappa	0.61	

Table 31. Crosstabs of Target and Empirical Levels: Mathematics Grade 7

Based on ESS-Weight cut scores

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	1 1.1%	4 4.5%	1 1.1%	0 0%	6 6.7%
	Level 2	3 3.4%	30 33.7%	12 13.5%	1 1.1%	46 51.7%
	Level 3	0 0%	12 13.5%	21 23.6%	1 1.1%	34 38.2%
	Level 4	1 1.1%	1 1.1%	0 0%	1 1.1%	3 3.4%
	Total	5 5.6%	47 52.8%	34 38.2%	3 3.4%	89 100%

Agree	53	60%
Disagree	36	40%
Adjacent Disagreement	32	36%
Discrepant Disagreement	4	4%
Weighted Kappa	0.3	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	4 4.5%	1 1.1%	1 1.1%	0 0%	6 6.7%
	Level 2	3 3.4%	32 36%	10 11.2%	1 1.1%	46 51.7%
	Level 3	0 0%	12 13.5%	21 23.6%	1 1.1%	34 38.2%
	Level 4	1 1.1%	1 1.1%	0 0%	1 1.1%	3 3.4%
	Total	8 9%	46 51.7%	32 36%	3 3.4%	89 100%

Agree	58	65%
Disagree	31	35%
Adjacent Disagreement	27	30%
Discrepant Disagreement	4	4%
Weighted Kappa	0.41	

Table 32. Crosstabs of Target and Empirical Levels: Mathematics Grade 8

Based on ESS-Weight cut scores

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	37 41.6%	6 6.7%	2 2.2%	1 1.1%	46 51.7%
	Level 2	5 5.6%	15 16.9%	6 6.7%	0 0%	26 29.2%
	Level 3	4 4.5%	4 4.5%	7 7.9%	0 0%	15 16.9%
	Level 4	0 0%	0 0%	2 2.2%	0 0%	2 2.2%
	Total	46 51.7%	25 28.1%	17 19.1%	1 1.1%	89 100%

Agree	59	66%
Disagree	30	34%
Adjacent Disagreement	23	26%
Discrepant Disagreement	7	8%
Weighted Kappa	0.53	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	38 42.7%	5 5.6%	2 2.2%	1 1.1%	46 51.7%
	Level 2	5 5.6%	18 20.2%	3 3.4%	0 0%	26 29.2%
	Level 3	4 4.5%	4 4.5%	7 7.9%	0 0%	15 16.9%
	Level 4	0 0%	0 0%	2 2.2%	0 0%	2 2.2%
	Total	47 52.8%	27 30.3%	14 15.7%	1 1.1%	89 100%

Agree	63	71%
Disagree	26	29%
Adjacent Disagreement	19	21%
Discrepant Disagreement	7	8%
Weighted Kappa	0.55	

Table 33. Crosstabs of Target and Empirical Levels: Mathematics High School

Based on ESS-Weight cut scores

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	0 0%	1 2.9%	2 5.9%	0 0%	3 8.8%
	Level 2	3 8.8%	5 14.7%	3 8.8%	3 8.8%	14 41.2%
	Level 3	0 0%	5 14.7%	5 14.7%	2 5.9%	12 35.3%
	Level 4	0 0%	2 5.9%	2 5.9%	1 2.9%	5 14.7%
	Total	3 8.8%	13 38.2%	12 35.3%	6 17.6%	34 100%

Agree	11	32%
Disagree	23	68%
Adjacent Disagreement	16	47%
Discrepant Disagreement	7	21%
Weighted Kappa	0.13	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	0 0%	1 2.9%	2 5.9%	0 0%	3 8.8%
	Level 2	3 8.8%	7 20.6%	1 2.9%	3 8.8%	14 41.2%
	Level 3	0 0%	5 14.7%	6 17.6%	1 2.9%	12 35.3%
	Level 4	0 0%	2 5.9%	2 5.9%	1 2.9%	5 14.7%
	Total	3 8.8%	15 44.1%	11 32.4%	5 14.7%	34 100%

Agree	14	41%
Disagree	20	59%
Adjacent Disagreement	13	38%
Discrepant Disagreement	7	21%
Weighted Kappa	0.16	

Table 34. Crosstabs of Target and Empirical Levels: Reading Grade 3

Based on ESS-Weight cut scores

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	0 0%	1 1.3%	0 0%	0 0%	1 1.3%
	Level 2	1 1.3%	39 50%	1 1.3%	0 0%	41 52.6%
	Level 3	0 0%	2 2.6%	30 38.5%	1 1.3%	33 42.3%
	Level 4	0 0%	0 0%	0 0%	3 3.8%	3 3.8%
	Total	1 1.3%	42 53.8%	31 39.7%	4 5.1%	78 100%

Agree	72	92%
Disagree	6	8%
Adjacent Disagreement	6	8%
Discrepant Disagreement	0	0%
Weighted Kappa	0.89	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	0 0%	1 1.3%	0 0%	0 0%	1 1.3%
	Level 2	1 1.3%	40 51.3%	0 0%	0 0%	41 52.6%
	Level 3	0 0%	2 2.6%	31 39.7%	0 0%	33 42.3%
	Level 4	0 0%	0 0%	0 0%	3 3.8%	3 3.8%
	Total	1 1.3%	43 55.1%	31 39.7%	3 3.8%	78 100%

Agree	74	95%
Disagree	4	5%
Adjacent Disagreement	4	5%
Discrepant Disagreement	0	0%
Weighted Kappa	0.93	

Table 35. Crosstabs of Target and Empirical Levels: Reading Grade 4

Based on ESS-Weight cut scores

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	0 0%	0 0%	0 0%	0 0%	0 0%
	Level 2	0 0%	3 4.2%	5 7%	1 1.4%	9 12.7%
	Level 3	0 0%	6 8.5%	48 67.6%	3 4.2%	57 80.3%
	Level 4	0 0%	0 0%	4 5.6%	1 1.4%	5 7%
	Total	0 0%	9 12.7%	57 80.3%	5 7%	71 100%

Agree	52	73%
Disagree	19	27%
Adjacent Disagreement	18	25%
Discrepant Disagreement	1	1%
Weighted Kappa	0.2	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	0 0%	0 0%	0 0%	0 0%	0 0%
	Level 2	0 0%	3 4.2%	5 7%	1 1.4%	9 12.7%
	Level 3	0 0%	6 8.5%	48 67.6%	3 4.2%	57 80.3%
	Level 4	0 0%	0 0%	4 5.6%	1 1.4%	5 7%
	Total	0 0%	9 12.7%	57 80.3%	5 7%	71 100%

Agree	52	73%
Disagree	19	27%
Adjacent Disagreement	18	25%
Discrepant Disagreement	1	1%
Weighted Kappa	0.2	

Table 36. Crosstabs of Target and Empirical Levels: Reading Grade 5

Based on ESS-Weight cut scores

Empirical Level	Consensus Level 1	Consensus Level 2	Consensus Level 3	Consensus Level4	Total
Level 1	0 0%	1 1.3%	1 1.3%	0 0%	2 2.6%
Level 2	2 2.6%	2 2.6%	0 0%	0 0%	4 5.3%
Level 3	0 0%	4 5.3%	54 71.1%	3 3.9%	61 80.3%
Level 4	0 0%	0 0%	4 5.3%	5 6.6%	9 11.8%
Total	2 2.6%	7 9.2%	59 77.6%	8 10.5%	76 100%

Agree	61	80%
Disagree	15	20%
Adjacent Disagreement	14	18%
Discrepant Disagreement	1	1%
Weighted Kappa	0.59	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

Empirical Level	Consensus Level 1	Consensus Level 2	Consensus Level 3	Consensus Level 4	Total
Level 1	0 0%	1 1.3%	1 1.3%	0 0%	2 2.6%
Level 2	2 2.6%	2 2.6%	0 0%	0 0%	4 5.3%
Level 3	0 0%	4 5.3%	55 72.4%	2 2.6%	61 80.3%
Level 4	0 0%	0 0%	4 5.3%	5 6.6%	9 11.8%
Total	2 2.6%	7 9.2%	60 78.9%	7 9.2%	76 100%

Agree	62	82%
Disagree	14	18%
Adjacent Disagreement	13	17%
Discrepant Disagreement	1	1%
Weighted Kappa	0.61	

Table 37. Crosstabs of Target and Empirical Levels: Reading Grade 6

Based on ESS-Weight cut scores

Empirical Level	Consensus Level 1	Consensus Level 2	Consensus Level 3	Consensus Level 4	Total
Level 1	1 1.2%	3 3.5%	1 1.2%	0 0%	5 5.9%
Level 2	0 0%	5 5.9%	52 61.2%	0 0%	57 67.1%
Level 3	0 0%	5 5.9%	52 61.2%	0 0%	57 67.1%
Level 4	0 0%	0 0%	1 1.2%	0 0%	1 1.2%
Total	5 5.9%	22 25.9%	58 68.2%	0 0%	85 100%

Agree	67	79%
Disagree	18	21%
Adjacent Disagreement	17	20%
Discrepant Disagreement	1	1%
Weighted Kappa	0.66	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

Empirical Level	Consensus Level 1	Consensus Level 2	Consensus Level 3	Consensus Level 4	Total
Level 1	2 2.4%	2 2.4%	1 1.2%	0 0%	5 5.9%
Level 2	4 4.7%	16 18.8%	2 2.4%	0 0%	22 25.9%
Level 3	0 0%	5 5.9%	52 61.2%	0 0%	57 67.1%
Level 4	0 0%	0 0%	1 1.2%	0 0%	1 1.2%
Total	6 7.1%	23 27.1%	56 65.9%	0 0%	85 100%

Agree	70	82%
Disagree	15	18%
Adjacent Disagreement	14	16%
Discrepant Disagreement	1	1%
Weighted Kappa	0.72	

Table 38. Crosstabs of Target and Empirical Levels: Reading Grade 7

Based on ESS-Weight cut scores

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	1 1.1%	6 6.5%	0 0%	0 0%	7 7.5%
	Level 2	6 6.5%	43 46.2%	6 6.5%	0 0%	55 59.1%
	Level 3	1 1.1%	3 3.2%	23 24.7%	2 2.2%	29 31.2%
	Level 4	0 0%	0 0%	2 2.2%	0 0%	2 2.2%
	Total	8 8.6%	52 55.9%	31 33.3%	2 2.2%	93 100%

Agree	67	72%
Disagree	26	28%
Adjacent Disagreement	25	27%
Discrepant Disagreement	1	1%
Weighted Kappa	0.62	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	3 3.2%	4 4.3%	0 0%	0 0%	7 7.5%
	Level 2	6 6.5%	45 48.4%	4 4.3%	0 0%	55 59.1%
	Level 3	1 1.1%	3 3.2%	25 26.9%	0 0%	29 31.2%
	Level 4	0 0%	0 0%	2 2.2%	0 0%	2 2.2%
	Total	10 10.8%	52 55.9%	31 33.3%	0 0%	93 100%

Agree	73	78%
Disagree	20	22%
Adjacent Disagreement	19	20%
Discrepant Disagreement	1	1%
Weighted Kappa	0.69	

Table 39. Crosstabs of Target and Empirical Levels: Reading Grade 8

Based on ESS-Weight cut scores

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	0 0%	3 3.2%	0 0%	0 0%	3 3.2%
	Level 2	2 2.2%	9 9.7%	3 3.2%	0 0%	14 15.1%
	Level 3	0 0%	3 3.2%	69 74.2%	0 0%	72 77.4%
	Level 4	0 0%	0 0%	0 0%	4 4.3%	4 4.3%
	Total	2 2.2%	15 16.1%	72 77.4%	4 4.3%	93 100%

Agree	82	88%
Disagree	11	12%
Adjacent Disagreement	11	12%
Discrepant Disagreement	0	0%
Weighted Kappa	0.79	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	2 2.2%	1 1.1%	0 0%	0 0%	3 3.2%
	Level 2	2 2.2%	10 10.8%	2 2.2%	0 0%	14 15.1%
	Level 3	0 0%	3 3.2%	69 74.2%	0 0%	72 77.4%
	Level 4	0 0%	0 0%	0 0%	4 4.3%	4 4.3%
	Total	4 4.3%	14 15.1%	71 76.3%	4 4.3%	93 100%

Agree	85	91%
Disagree	8	9%
Adjacent Disagreement	8	9%
Discrepant Disagreement	0	0%
Weighted Kappa	0.86	

Table 40. Crosstabs of Target and Empirical Levels: Reading High School

Based on ESS-Weight cut scores

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	0 0%	0 0%	0 0%	0 0%	0 0%
	Level 2	0 0%	12 29.3%	3 7.3%	0 0%	15 36.6%
	Level 3	0 0%	2 4.9%	16 39%	1 2.4%	19 46.3%
	Level 4	0 0%	0 0%	2 4.9%	5 12.2%	7 17.1%
	Total	0 0%	14 34.1%	21 51.2%	6 14.6%	41 100%

Agree	33	80%
Disagree	8	20%
Adjacent Disagreement	8	20%
Discrepant Disagreement	0	0%
Weighted Kappa	0.79	

Based on ESS-Weight cut scores, counting essentially consistent items as consistent

		Consensus Level				
		Level 1	Level 2	Level 3	Level 4	Total
Empirical Level	Level 1	0 0%	0 0%	0 0%	0 0%	0 0%
	Level 2	0 0%	12 29.3%	3 7.3%	0 0%	15 36.6%
	Level 3	0 0%	2 4.9%	16 39%	1 2.4%	19 46.3%
	Level 4	0 0%	0 0%	2 4.9%	5 12.2%	7 17.1%
	Total	0 0%	14 34.1%	21 51.2%	6 14.6%	41 100%

Agree	33	80%
Disagree	8	20%
Adjacent Disagreement	8	20%
Discrepant Disagreement	0	0%
Weighted Kappa	0.79	

Quality Control Procedures

ESS analyses were conducted using EmStanS (2021), CMS' proprietary ESS software, which has been reviewed extensively for quality control to support the accurate estimation of cut scores based on the ESS-Weight algorithm. In addition, an independent review of all analyses was conducted by a second analyst to confirm the accuracy of all ESS output. Materials for the ESS-Alignment Study were verified with respect to the number of items, item classifications, and were confirmed to agree with those of edCount, as reported in their Alignment Evaluation technical report (edCount, 2023).

Summary

The results generally support the efficacy of the educators Final Consensus Levels. As shown in Table 25, except for high school, the Mathematics Kappa values were moderate to almost perfect and except for grade 4 Reading, the Reading Kappa values were substantial to almost perfect.

The ESS process supports an iterative approach to Item-ALD alignment. While the efficacy of the panelists' Consensus Levels is largely supported by the data reviewed in this section, additional review may be warranted. The tables in Appendix F: Detailed ESS Item Maps and Appendix G: Rosters of Inconsistent and Essentially Consistent Items provide detailed information on the items and their consistency status.

Appendix F provides item-level information for all items in grades 3-8 and high school including:

- item ID (ID)
- order of difficulty (OOD)
- item RP50 location (LOC)
- Consensus Level
- ESS-Count and ESS-Weight associated with the item location for each cut score
- Initial Empirical Level associated with the Initial ESS cut scores
- Final Empirical Level associated with the adopted, vertically articulated cut scores

Appendix G provides rosters of inconsistent items and essentially consistent items for each grade and subject. The information in the tables in Appendix G includes:

- grade and content area (GCA)
- item ID
- order of difficulty (OOD)
- Consensus Level
- Empirical level
- Level Difference: ordinal difference in Consensus and Empirical levels
- ESS-Distance (Distance)
- Absolute Distance (Absolute value of Distance)

Section 5. Post Workshop Procedures Leading to the Adoption of Cut Scores

In this section, we discuss considerations with respect to the vertical articulation of the initial cut scores to support a coherent cross-grade assessment system for each subject. Under ideal circumstances the estimation of initial ESS cut scores for each grade in a subject results in impact data that is reasonable and supports the Maine DOE policy goals for the MTYA program. That is, the proportion of students in each achievement level should be appropriate for each subject when viewed across levels within a grade and within each level across the grades. The appropriateness of data should be informed by theory and the expectations of SMEs knowledgeable of the population and subject area.

When data are not as suggested by theory and as expected by SMEs, then some statistical moderation and smoothing, referred to as vertical articulation, may be necessary to achieve this result. It is common to refine cut scores to support their vertical articulation either during a standard setting workshop or by policymakers and their technical advisors following a standard setting. The rationale and precedent for policymakers' adjustment of recommended cut scores following a standard setting workshop are provided in Appendix H: Considerations in the Use of Standard Errors for the Adjustment of Cut Scores.

The MTYA program is not on a common, cross-grade (vertical) scale. In this case, coherence is observed through patterns of impact data. That is, the progression of the percentage of students in each achievement level within and across grades should be sensible based on theory and expectations.

Methodological Choice on the Presentation of Impact Data to Workshop Panelists

The presentation of impact data in traditional standard setting workshops is commonplace. However, some prominent educational researchers currently recommend not presenting impact data to standard setting panelists. First, a preview of the standard setting chapter (Ferrara, Davis-Becker, & Kannan, 2023) of the Fifth edition of *Educational Measurement* (NCME, 2023) details the recommendation to forgo the presentation of impact data to standard setting workshop panelists:

> "…In some standard setting implementations, feedback is presented between two rounds of judgments. In others, three rounds of judgments may be included to offer different types of feedback staggered across rounds (Hambleton & Pitoniak, 2006). In such cases, designers and researchers have debated about when to present impact data in the workshop, because it undermines the content-based cognitive-judgmental task. **We argue that impact data is a policy consideration that is outside of workshop panelists' expertise and authority and should not be presented to panelists** [emphasis added]. Clauser et al. (2009) and Morgan and Michaelides (2005) have shown

> that presenting impact data to panelists may significantly alter panelists' judgments. Benchmarked standard setting enables impact data to influence cut score recommendations indirectly (Ferrara et al., 2021) **and policy review and adjustments after the workshop maintains the separation between content and policy expertise** [emphasis added]. (See Cizek, 2022 for a description of content and policy considerations that can pull panelists in two different directions) …"

In addition, Cizek (2022), a well-respected researcher and editor of the seminal book on standard setting (Setting Performance Standards, 2001 [1st edition]; 2012 [2nd edition]), noted that

> "…it is difficult to promote a cut score as "criterion referenced" or an assessment program as "standards referenced" when normative data form the basis for adjusting (and approving) the ultimate performance standards."

Cizek refers to standard setting schizophrenia—the initial focus of panelists on standards-referenced judgments and then being presented with sometimes inconsistent impact data that causes them to adjust their standards-based judgments, diminishing the adherence of their judgments to the performance standards, which are typically expressed as achievement level descriptors (ALDs).

Cizek, in a personal communication (March 1, 2022, shared with permission) in response to the assertion that impact data should not be presented to standard setting, noted that

> " …we are … in agreement on … the appropriate folks to consider impact data. i {sic} would still add that i (sic) think asking ANY group to endorse a cut score without knowing the impact of that decision isn't a good idea, but ESS would--appropriately--shift that policy decision away from the content experts and into the hands of the person(s) with authority to make the decisions and with knowledge of all the other relevant factors…"

Vertical Articulation When Panelists Do Not Observe Impact Data

ESS Alignment Study panelists did not view impact data, and this will have to be accounted for in the post workshop vertical articulation activities. To understand how this is accounted for, we first discuss the three phases during and following the traditional standard setting workshop that are affected by impact data. We refer to the three phases as (a) the moderation phase, (b)

the smoothing phase, and (c) the policy phase of cut score adoption, and we describe each next.

The Moderation Phase

The moderation phase of a traditional standard setting workshop occurs in the final round of grade-specific judgments when panelists first see impact data—the percentage of students in each achievement level. That is, panelists view the consequences of the latest round of judgments in the form of impact data so it can be considered prior to their final judgments. This often occurs in Round 3 of a standard setting workshop following Round 1, in which panelists make individual and independent judgments, and Round 2, in which panelists discuss the range of, and rationales for, their Round 1 individual judgments.

We refer to this as the moderation phase because the impact data may cause panelists to moderate their judgments. That is, when impact data reflects overly rigorous content expectations, panelists may moderate their judgments to be less rigorous. For a Bookmark workshop, this would entail moving bookmarks earlier in the ordered item booklet and when impact data does not reflect sufficiently rigorous content expectations, panelists moderate their judgments in the opposite direction.

Model language used in practice by the author in traditional standard setting workshops sets the context for the appropriate use of impact data in that moderating context, as follows:

> "We are going to share consequences data with you now—the percent of students who would be in each achievement level based on your current judgments. We present this as a reality check, not to get you to just chase numbers. That is, it is difficult to know when the data are right, but it is easy to observe when they seem wrong. If the data presented is well out of the bounds of your expectations, you will know it. In this case, you might reconsider your judgments (for Bookmark, your bookmark placement; for ID Matching, your cut score page; for Yes-No Angoff, your alignment of items to levels, etc.)."

Panelists' reactions to impact data that is well outside of their expectations typically results in a commensurate adjustment to their judgments. For example, panelists sometimes ask Bookmark Standard Setting facilitators just how far they would have to move a bookmark to obtain a specific impact data target (personal observation of the author).

Thus, the moderation phase supports the alignment of panelists' cut score recommendations with their impact data expectations.

The Smoothing Phase

The smoothing phase of a traditional standard setting workshop is the vertical articulation activity, which is a cross-grade, within-subject meeting that typically occurs after the final round of grade-specific judgments. The vertical articulation activity is focused on the cross-grade coherence of impact data (and of scale scores when an assessment is on a vertical scale). It is common for one or two representatives from each grade—occasionally all panelists—to gather to view and discuss the patterns of cross-grade impact data and, if necessary, to consider adjustments to their grade-specific cut scores to support cross-grade coherence. Details about the theory and methods used to achieve vertical articulated cut scores are provided by Lewis and Haug (2005) and Cizek and Agger (2012).

The introduction to their task includes an explication of, and rationale for, the goal of the vertical articulation activity—to have a coherent cross-grade system of cut scores. The challenges to vertical articulation are typically described as grade-to-grade impact data shifts or patterns that appear anomalous and do not correspond to expectations. We refer to this as the *smoothing phase* because the impact data and motivating rational provided by the workshop facilitators tend to cause panelists to compromise and adjust their judgments to support smoother grade-to-grade patterns of impact data.

For example, facilitators may observe that the percentage of students meeting or exceeding the At State Expectations cut score is much lower in grade 4 than in grades 3 and 5. We do not expect identical percentages in each grade, but we also do not expect large differences unless suggested by theory. Unless there is some rationale for why these grade-to-grade differences should be expected, the face validity of the assessment may be compromised; stakeholders may question the validity of the test results or, in this example, the efficacy of instruction in grade 4, when it is more likely that the grade-specific judgments in grades 3, 4, and 5 were applied with different levels of rigor.

A compromise may be suggested by panelists or the facilitator such that all three grades compromise to make the adjustments necessary to have less disparate impact data for those three grades, resulting in a smoothing of the impact data. Thus, the smoothing phase supports the articulation of impact data from grade to grade.

The Policy Phase

The policy phase of the processes leading to the adoption of cut scores occurs after the recommendation of cut scores by panelists in a standard setting workshop. The policy phase is engaged by policymakers who, along with their technical advisors, consider the recommended cut scores not only with respect to their support for federal accountability requirements, but with respect to (a) how they support the theory of action or the stated uses and goals of the assessment program, (b) how they support instruction, learning, and student growth, and (c) how they will be perceived by other assessment program stakeholders such as parent, teachers, students, the press, and/or legislators.

It is not uncommon for policymakers to adjust standard setting panelists' recommended cut scores prior to adoption to support one or more of the stated considerations. A discussion of the rationale for and precedent associated with the adjustment of standard setting panelists' recommended cut scores in the policy phase is provided in Appendix H.

Integration of the Moderation, Smoothing, and Policy Phases

Impact data was, by design, not presented to ESS Alignment Study panelists and as recommended by leading researchers (Ferrara, Becker-Davis, & Kannan, in press; Cizek, 2022) adjustments based on impact data must be made by policymakers. Thus, the moderation and smoothing phases must be integrated into the policy phase to achieve a well-articulated system of adopted cut scores and impact data.

Next, we describe how moderation and smoothing were integrated into the policy phase to achieve adopted cut scores that support the goals of the testing program and meet the expectations of the ESS Alignment Study panelists and other stakeholders.

Moderating the recommended cut scores. The moderation phase results in cut score adjustments that support the concordance of resultant impact data with ESS Alignment Study panelists' expectations for reasonable and defensible cut scores. In this case, the ESS Alignment Study panelists expressed their expectations via survey. Each panelist estimated the following values:

- What is the lowest percentage of students At or Above State Expectations that you would expect and support?
- What is the highest percentage of students At or Above State Expectations that you would expect and support?
- What is the lowest percentage of students Above State Expectations that you would expect and support?
- What is the highest percentage of students Above Expectations that you would expect and support?
- What is the lowest percentage of students Well Below State Expectations that you would expect and support?
- What is the highest percentage of students Well Below State Expectations that you would expect and support?

Panelists individual responses were averaged to estimate each grade's range of expected and defensible impact data. That is, each grade's recommended cut scores should result in impact data between that grade's minimum and maximum expected values. Panelists' responses to these survey questions were used to moderate cut scores that are outside panelists' expectations. The following moderation rules were applied:

- If an initial ESS cut score resulted in impact data that was lower than the panelists' minimum expectation, then the cut score was adjusted so that it resulted in the minimum defensible expectation.
- If an initial ESS cut score resulted in impact data that was higher than the panelists' maximum expectation, then the cut score was adjusted so that it resulted in the maximum defensible expectation.
- If an initial ESS cut score resulted in impact data that was between the panelists' minimum and maximum defensible expectations, then the cut score did not need to be moderated—it met defensible expectations.

The minimum and maximum of the panelists' average defensible expectations for each grade and subject are provided in tabular form at the end of Appendix C and graphically in Appendix I.

Smoothing the moderated cut scores. The moderated cut scores were then evaluated as they would be in the smoothing phase of a standad setting workshop. When smoothing was necessary to support within- and cross-grade band vertical articulation, the following guidelines were considered:

1. Relatively smooth transitions in the percentage of students at or above each cut score across grades should be observed. That is, there should be no "saw tooth" patterns in the percentage of students at or above each cut score unless supported by theory or expectations.
2. The Level 2, 3, and 4 cut scores should be sufficiently different within a grade to have a reasonable proportion of students in each level. This is desirable because an achievement level that is obtained by a trivial proportion of students is likely to be unreliable; that is, it is likely that all students in such a level are within an SEM of either of the adjacent levels.
3. The percentage of students at the various levels should reasonably correspond to expectations.
4. Refinements in cut scores to support smoothing should be as modest as possible.

Refinements were made to the moderated cut scores to support these guidelines. The impact data reflecting the initial cut scores and the results of the iterative moderation and smoothing activity for each subject and grade are provided graphically in Appendix I.

Vetting the moderated and smoothed cut scores with respect to stakeholder expectations and policy goals. Several iterations of moderation and smoothing were applied and reviewed by various stakeholders including the MTYA Technical Advisory Committee composed of nationally recognized experts in educational assessment, who supported the described moderation and smoothing methodology.

Adopted and Vertically Articulated Cut Scores

The resulting vertically-articulated cut scores for Mathematics and Reading, presented in Table 41 and Table 43, respectively, were adopted for operational use. The associated impact data is presented in Tables 42 and 44 and and Figures 4 and 5 for Mathematics and Reading, respectively. We note that the high school cut scores are considered provisional because the second year of high school assessments were administered as an operational field test, while the grade 3-8 assessments were administered as fully operational assessments in the Spring 2023 administration.

Table 41. Adopted, Vertically Articulated Cut Scores: Math

Grade	Math Cut Scores		
	Well Below State Expectations	Below State Expectations	At State Expectations
3	-2	-0.9	1.1
4	-1.9	-0.7	1.8
5	-1.7	-0.2	2.2
6	-1.8	-0.1	2.1
7	-2.2	-0.5	1.8
8	-1.9	-0.4	1.9
HS	-1.9	-1.3	0.1

Table 42. Impact Data Associated with Adopted, Vertically Articulated Cut Scores: Math

Grade	Math: Percent in Level			
	Well Below State Expectations	Below State Expectations	At State Expectations	Above State Expectations
3	17.3%	21.1%	43.9%	17.7%
4	18.6%	24.5%	44.0%	12.9%
5	18.5%	30.7%	40.0%	10.8%
6	18.8%	36.4%	35.9%	8.9%
7	20.1%	36.0%	35.4%	8.5%
8	20.5%	39.1%	33.5%	6.9%
HS	25.0%	32.0%	35.5%	7.5%

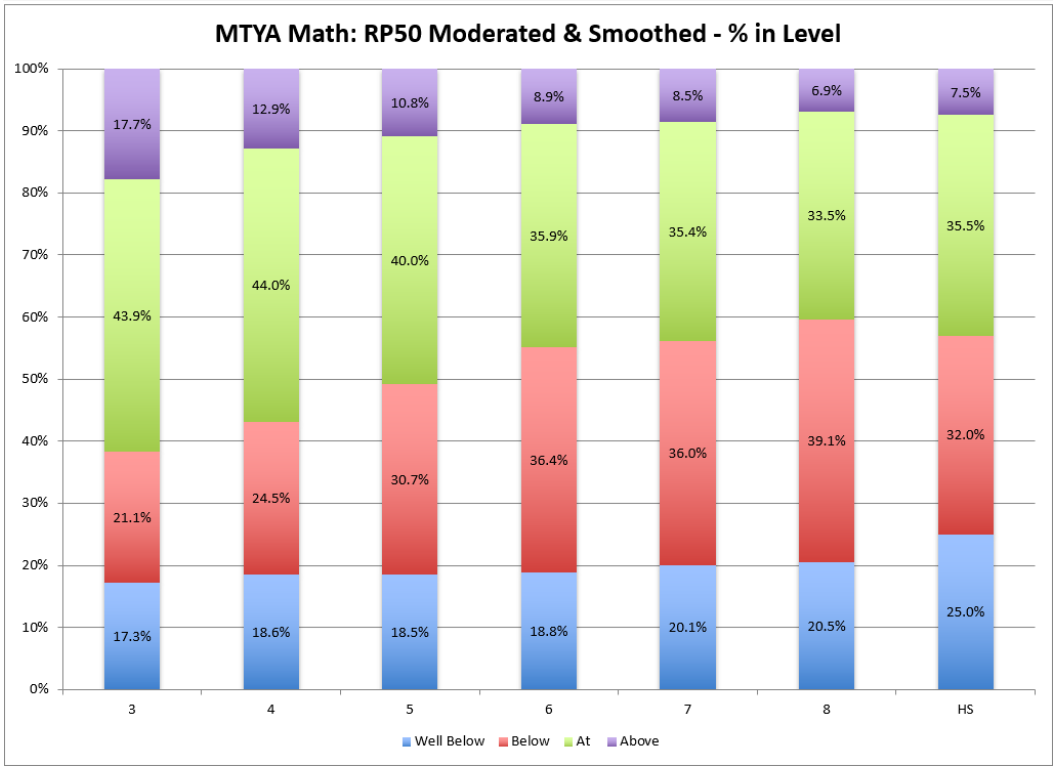


Figure 4. Stacked Bar Chart: Impact Data for Adopted Cut Scores: Math

Table 43. Adopted, Vertically Articulated Cut Scores: Reading

Grade	Reading Cut Scores		
	Below State Expectations	At State Expectations	Above State Expectations
3	-2	-0.6	1.5
4	-1.4	-0.3	1.7
5	-1.4	-0.4	1.5
6	-1.5	-0.4	1.5
7	-1.5	-0.2	1.7
8	-1.3	0	2
HS	-0.8	-0.1	1.5

Table 44. Impact Data Associated with Adopted, Vertically Articulated Cut Scores: Reading

Grade	Reading: Percent in Level			
	Well Below State Expectations	Below State Expectations	At State Expectations	Above State Expectations
3	12.6%	27.1%	47.3%	13.0%
4	12.2%	23.9%	48.5%	15.4%
5	12.8%	18.6%	53.0%	15.6%
6	10.4%	22.5%	53.5%	13.6%
7	11.4%	24.9%	50.4%	13.3%
8	10.1%	24.2%	53.4%	12.3%
HS	13.3%	24.7%	49.7%	12.3%

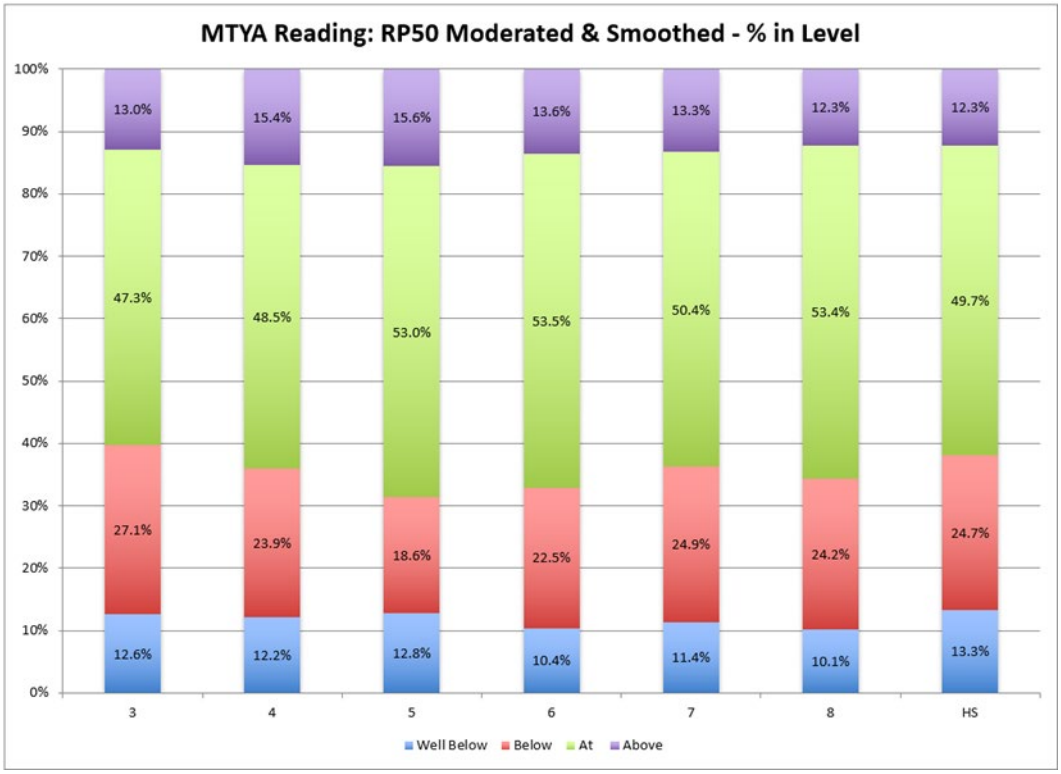


Figure 5. Stacked Bar Chart: Impact Data for Adopted Cut Scores: Reading

Section 6. Standard Setting Validity Evidence

In this section, we summarize the types of evidence traditionally used to support the validity of a standard setting process and provide evidence in support of the standard setting process leading to the adoption of cut scores for the MTYA program. Two perspectives on validity evidence are provided. First, the measurement literature provides validity criteria for the evaluation of standard setting processes (e.g., AERA/APA/NCME, 2014; Cizek, 2001; Cizek & Bunch, 2007; Kane, 2001; Hambleton, 2001). These criteria were reviewed and those relevant to Embedded Standard Setting are provided in Table 45. Second, the USDOE (2018) provides peer review guidelines with respect to standard setting. The USDOE guidelines provide evaluation criteria they refer to as Critical Elements, and examples of evidence for each Critical Element are listed in Table 47.

Descriptions of the criteria described in the measurement literature and traditionally used to support the validity of a standard setting process are provided next.

Standard Setting Validity Criteria from the Measurement Literature

There are several forms of standard setting validity evidence including procedural, internal, and external validity. Table 45 provides examples of specific evidence used to evaluate the relevant forms of validity evidence that have been suggested in the literature and which are appropriate for the evaluation of cut scores established under Embedded Standard Setting methodology.

Table 45. Forms of Standard Setting Validity Evidence from the Literature

Validity Type	Validity Evidence
Procedural	Support for the participants' qualifications
	Evidence that the participants understood the test and its intended use
	Evidence that the participants understood the construct reflected by the ALDs and how items provide evidence for ALD evidence statements
	Evidence that panelists were properly trained on the judgment task and were prepared to make the judgments
	Evidence that the standard setting method was appropriately selected based on the test and the intended use of the cut scores
	Evidence that the standard setting method was implemented as designed and if not, that the modifications were justified and appropriate
	A design that incorporates iterative processes
Internal	The efficacy of Item-ALD alignment hypotheses is supported by data
	Standard errors associated with cut scores are reasonable
External	Cut scores result in reasonable impact data
	Placement level expectations are reasonable and consistent with expectations

Next, we summarize each form of validity and provide the associated evidence in support of the validity of the cut scores.

Procedural Validity

Support for the Maine Educators' qualifications. Educators participating in the ESS Alignment Study were recruited and selected based on their qualifications—all had appropriate experience necessary to support the standard setting judgments. A summary of Maine educators' self-reported qualifications appears next:

The panels were composed primarily of content area teachers, with over three-quarters of the panelists (76.2%) currently serving in this role. Panels also included educators serving in a wide variety of roles, including school administrators (6.3%), instructional coaches (3.2%), education consultants (3.2%), interventionists (3.2%), and one panelist each serving as a district administrator, department chair, curriculum coordinator, literacy specialist, and special educator. Most panelists had experience working with students with disabilities (93.7%), working with a high percentage of economically disadvantaged students (93.7%), teaching content area courses (87.3%), working with gifted students (84.1%), and working with multilingual learners (57.1%). When asked to rate their level of understanding of the CCSS, all panelists indicated that they were familiar with the standards, with the majority (79.4%) indicating they were very familiar.

Over a quarter of panelists (28.6%) had a professional certification or endorsement in the content area associated with their panel. In addition, a majority of participants had an elementary (73.0%) certification or endorsement, while 46% had a middle school certification or endorsement, and 25.4% had one for high school. Panelists also indicated administrator (23.8%), exceptional children/children with disabilities (11.1%), literacy coach (9.5%), English as a Second Language (6.3%), and curriculum instructional specialist (4.8%) among their listed certifications. On average, panelists had 17 years of experience in their educational area of expertise

Additional details are provided in the section "Characteristics of Panelists."

Next, we provide evidence for the following three validity criteria from the literature: Evidence that the participants understood the test and its intended use; Evidence that the participants understood the construct reflected by the ALDs and how items provide evidence for ALD claims and targets; Evidence that panelists were properly trained on the judgment task and prepared to make the judgments. ESS Alignment Study panelists received training and participated in discussion of the purpose of the test, the CCSS, which were the basis for the assessments, the rubrics used to score items, and reviewed, on an item-by-item basis, three exemplar test forms in grades 3-8 and the operational test form in high school (see edCount, 2023). Panelists' familiarity with the test was further supported by first, individually and independently judging each item with respect to the CCSS, ALD level, and depth of knowledge

to which the item is best aligned followed by panel discussion of the rationales for any differences in their individual alignments (see edCount, 2023).

Evidence that the training and participation in the alignment activities resulted in an understanding of the test and its intended use is supported by the evaluation responses which indicate a modal rating of “strongly agree” to each of the relevant evaluation statements, as shown in Table 46. The full set of evaluation questions and summaries is provided in Appendix E.

Table 46. Evaluation Summaries, Mathematics and Reading

Mathematics Overall

Overall (N=28)	Median
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	4
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	4

Reading Overall

Overall (N=33)	Median
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	4
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	4

Evidence that the standard setting method was appropriately selected based on the test and the intended use of the cut scores. The Maine Through Year Assessment standard setting methodology was influenced by the early decision to use PAD as the framework for the assessment development process. Embedded Standard Setting is the natural extension of Principled Assessment Design to standard setting (Lewis & Cook, 2020) because it capitalizes on the PAD requirements to develop (a) ALDs early in the test development process (see Section 2) to support domain definition and (b) test items that provide evidence for the claims and measurement targets expressed in the ALDs and CCSS. The optimal deployment of ESS is for testing programs developed from inception using a PAD framework, as was the Maine Through Year Assessment. Figure 1 shows how ESS is integrated with the PAD test development framework by coordinating the relevant output of various test development activities—ALD development, Item-ALD alignment, educator item reviews, and empirical data analyses. ESS has a sound theoretical framework that is articulated by Lewis & Cook (2020). The fundamental requirements for the application of ESS are well-articulated ALDs and items that are aligned to specific achievement level evidence statements.

The Maine Through Year Assessment Technical Advisory Committee reviewed the standard setting design and provided recommendations to improve the design and supported its application for this purpose.

Evidence that the standard setting method was implemented as designed and if not, that the modifications were justified and appropriate. The processes documented in the current technical report closely follows the design specified in the Maine Through Year Assessment detailed standard setting design document, which was reviewed and approved by Maine DOE leadership and reviewed and supported by their Technical Advisory Committee. No significant deviations occurred. Section 1 of the current document describes the various assessment development processes and documentation coordinated to support the estimation of ESS cut scores, as follows: (see the section Coordination of Embedded Standard Setting Iterative Processes).

- Achievement Level Descriptor (ALD) Development
- The ESS Alignment Study
- ESS Analyses and the Estimation of Cut Scores
- Post Workshop Activities Leading to the Adoption of Cut Scores
- Documentation of Validity Evidence Supporting the MTYA-Adopted Cut Scores

These processes and their documentation were all described in detail in the Maine Through Year Assessment detailed standard setting design document (CMS, 2023) and are summarized by the nearly identical flowcharts in Figure 1 in the design document and in the current document. A few modest deviations to the standard setting methodology occurred, as described next:

First, 14 panels of educators consisting of five panelists each (for a total of 70 educators) were recruited, as specified in the design document. However, one reading panelist and six mathematics panelists had to withdraw prior to the start of the rating process due to last-minute conflicts or technical difficulties, bringing the total number of study participants to 63, with no panel having fewer than 4 participants

The only other deviation was based on the post Day 2 grade 8 reading results, in which few or no items had Initial Consensus Levels other than At State Expectations. This prompted an addition to the Day 3 opening session. The main points of the discussion follow:

- Three forms were systematically selected in grades 3 through 8 to support low-, moderate-, and high-achieving students. The three forms were targeted for students at the 16th, 50th, and 84th percentiles.
- The diversity in the difficulty of items on the three forms was expected to result in Item-ALD alignments to all four levels.
- Items are presented in order of difficulty for the Day 3 R&R activity and thus, one might expect the easiest items to be associated with lower achievement levels and the hardest items to be associated with higher achievement levels.
- Panelists were encouraged to pay special attention to the Item-ALD alignments of items appearing at the beginning (the easiest items) and at the end (the hardest items) of their ordered inconsistent item review. However, ultimately it was the association of the items' content characteristics that must guide their decision making, not these expectations. Thus, panelists were asked to consider their Initial Consensus Levels, but not to automatically change them to comply with expectations.

No other nontrivial deviations to the standard setting design occurred.

A design that incorporates iterative processes. Iteration is built into two levels—macro and micro—of the Maine Through Year Assessment standard setting design. At the macro level of the overall design, the green feedback arrows in Figure 1, demonstrate the iterative nature of the Embedded Standard Setting design—feedback from the ESS Alignment Study is iterated to support ALD refinement and Item-ALD Alignment.

At a micro-level, the ESS Alignment Study provides three "rounds" of iteration on the Item-ALD alignments necessary to support standard setting. On Days 1 and 2 of the workshop, panelists make their first round of individual and independent alignment judgments and then make their second round of alignment judgments following panel discussion of the rationales for their differing independent alignments. On Day 3, the third and final "round" of the iterative alignment activities, panelists are provided a third opportunity to update their alignments based on feedback from empirical data generated from ESS analyses. These iterative activities are described in Section 3.

Internal Validity

The efficacy of Item-ALD alignment hypotheses is supported by data. Adjusted correlations between panelists' Item-ALD Alignments (Consensus Levels) and empirical item difficulty (RP50 locations) are reported in Table 22. The median adjusted correlation across grades was 0.58 and 0.76 for Mathematics and Reading, respectively. Classification agreement and weighted Kappa are reported in Table 25. The median classification agreement rate across grades was 65% and 82% for Mathematics and Reading, respectively. The median weighted Kappa across grades was 0.54 and 0.72 for Mathematics and Reading, respectively, which, according to the suggested interpretations by Landis and Koch (1977) are moderate to substantial, respectively.

These data support the efficacy of the workshop panelists' Item-ALD alignments.

The standard errors associated with cut scores are reasonable*.* The standard errors of the ESS cut scores (SEESS) are estimated using the Bootstrapping techniques described by Lewis, Lee, and Choi (2021, May). These standard errors are reported in Table 26. The results indicate that the standard error of the cut scores, SE_{ESS} were generally considerably lower than the SEM of the Maine Through Year Assessments.

Table 26 indicates that, for Math, the median SEM of the Maine Through Year Assessments SEM_{MTYA} is 0.41 compared to the median SEs of the Mathematics cut scores which are 0.29, 0.24, and 0.33 for the Below, At, and Above State Expectations cut scores, respectively.

Table 26 indicates that, for Reading, the median SEM of the Maine Through Year Assessments SEM_{MTYA} is 0.41 compared to the median SEs of the Reading cut scores which are 0.42, 0.15, and 0.22 for the Below, At, and Above State Expectations cut scores, respectively.

Thus, the standard errors associated with the cut scores are reasonable.

External Validity

Cut scores result in reasonable impact data*;* Placement level expectations are reasonable and consistent with expectations. ESS Alignment Study panelists' impact data expectations were captured on the Day 3 evaluation and are provided in Appendix E. Specifically, panelists provided the minimum and maximum percentage of students they would expect and support in each achievement level. Panelists' average minimum and maximum expectations were used to guide the moderation phase of vertical articulation as described in the section Integration of the Moderation, Smoothing, and Policy Phases. Thus, the resulting vertically articulated cut scores result in impact data deemed reasonable and defensible by the Maine educators participating in the ESS Alignment Study.

Impact data were reviewed by the Maine DOE, their Technical Advisory Committee, and other stakeholders prior to the adoption of cut scores. The impact data were considered defensible and as expected by stakeholders relative to changes from prior Maine statewide assessment cut scores in Mathematics and Reading.

Thus, the adopted cut scores result in reasonable impact data in accordance with expectations.

Standard Setting Validity Criteria from Peer Review

Federal peer review accountability guidelines associated with standard setting are provided in Critical Element Section 6.2—Achievement Standards Setting (USDOE, 2018). The single Critical Element cited in this section follows:

> "The State used a technically sound method and process that involved panelists with appropriate experience and expertise for setting Academic achievement standards…, such that cut scores are developed for every grade…, content domain…and/or composite for which achievement level scores are reported."

Note that this Critical Element calls out the technical foundations of the method and the qualifications of the participants. The MTYA program Technical Advisory Committee reviewed and approved the technical foundations of ESS, which are carefully described in the Lewis & Cook (2020) publication, and panelist recruitment was conducted to support panelist qualifications.

The USDOE guidelines provide examples of evidence that may be included in the standard setting technical report to support this Critical Element for the assessments of interest. Table 47 provides these examples.

Table 47. Examples of Evidence Supporting Peer Review Critical Element 6.2

Peer Review Examples of Evidence
A description of the standards-setting method and process used by the State;
The rationale for the method selected;
Documentation that the method used for setting cut scores allowed panelists to apply their knowledge and experience in a reasonable manner and supported the establishment of reasonable and defensible cut scores;
Documentation of the process used for setting cut scores and developing achievement level descriptors aligned to the State's standards;
A description of the process for selecting panelists;
Documentation that the standards-setting panels consisted of panelists with appropriate experience and expertise
If available, a summary of statistical descriptions and analyses that provides evidence of the reliability of the cut scores and the validity of recommended interpretations;
A technical report providing a description of the method used, the diversity of the panelists involved and their qualifications, quality control procedures, the use of impact data, and panelist evaluation results;

Peer Review Validity Evidence

A description of the standards-setting method and process used by the State. An overview of the standards-setting method and process used by the Maine DOE for the Maine Through Year

Assessments is provided in Section 1 of the current document under the header Overview of Embedded Standard Setting and the MTYA Standard Setting Design. Details are provided in the remaining sections of this technical report.

The rationale for the method selected. The standard setting methodology was influenced by NWEA's application of PAD as the framework for the MTYA development process. Embedded Standard Setting is the natural extension of PAD to standard setting (Lewis & Cook, 2020) because it capitalizes on the PAD requirements to develop (a) ALDs early in the test development process to support domain definition and (b) test items that provide evidence for the claims and measurement targets expressed in the ALDs and CCSS. The optimal deployment of ESS is for testing programs developed from inception using a PAD framework, as was the MTYA program. Figure 1 shows how ESS is integrated with the PAD test development framework by coordinating the relevant output of various test development activities—ALD development, Item-ALD alignment, educator item reviews in the ESS Alignment Study, and ESS analyses. ESS has a sound theoretical framework that is articulated by Lewis & Cook (2020).

The Maine Through Year Assessment Technical Advisory Committee reviewed the standard setting design and provided recommendations to improve the design and supported its application for this purpose.

Documentation that the method used for setting cut scores allowed panelists to apply their knowledge and experience in a reasonable manner and supported the establishment of reasonable and defensible cut scores*.* This technical report provides documentation that the ESS methodology allowed panelists to apply their knowledge and experience in a reasonable manner to establish reasonable and defensible cut scores.

In support of the application of panelists' knowledge and experience in a reasonable manner, a fundamental aspect of the efficacy of the Embedded Standard Setting methodology is that the cognitive judgment task reduces to the alignment of test items to specific achievement level evidence statements (also referred to as claims and measurement targets). These judgments capitalize on the content area expertise of the teachers supporting the development of the ALDs (see Section 2) and the alignment of test items to ALDs in the ESS Alignment Study (see Section 3). Panelists in these two assessment development activities contributed to the fundamental ESS requirements—well articulated ALDs and the Item-ALD alignment of test items.

First, ALDs were developed by Maine educators and NWEA content specialists to reflect the performance expectations for Maine students as described in Section 2. The efficacy of their ALD work was supported by ESS Alignment Study panelists who responded with median responses of 3 (out of 4) in both Mathematics and Reading to the following statement on the evaluation: The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor (see the evaluation excerpt below).

Overall	Min	Max	Median
Math			
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	1	4	3
Reading			
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	1	4	3

Second, panels of qualified educators participated in the ESS Alignment Study, an iterative process to analyze and align items and to remediate items with Initial Consensus Level alignments that were not consistent with empirical data. Their sole judgment was the alignment of test items to the CCSS and the achievement levels based on the ALDs, a judgment for which they were trained and well qualified.

In support of the application of panelists' knowledge and experience to establish reasonable and defensible cut scores, the moderation phase of vertical articulation assured the adoption of cut scores with impact data that was within panelists' average minimum and maximum defensible expectations (as described in the section Moderation Phase of this report).

Documentation of the process used for setting cut scores and developing achievement level descriptors aligned to the State's standards. This technical report documents the process used for setting cut scores and Section 2 documents the development of ALDs aligned to the CCSS.

A description of the process for selecting panelists. Panelist recruitment and selection was conducted by NWEA. NWEA produced a survey asking about interest in upcoming educator meetings (including the ESS Alignment Study) and including questions regarding educator experience, demographics, and other pertinent information. This survey was sent to Maine educators by both NWEA and the Maine DOE. Maine DOE confirmed the credentials of educators expressing interest and availability for all days of the ESS Alignment Study, notifying NWEA so that NWEA could reach out to confirm participation by the Maine educators.

Documentation that the standards-setting panels consisted of panelists with appropriate experience and expertise. Educators participating in the ESS Alignment Study were recruited and selected based on their qualifications—all had appropriate experience to support their judgments. A summary of Maine educators' self-reported qualifications appears next.

The panels were composed primarily of content area teachers, with over three-quarters of the panelists (76.2%) currently serving in this role. Panels also included educators serving in a wide variety of roles, including school administrators (6.3%), instructional coaches (3.2%), education consultants (3.2%), interventionists (3.2%), and one panelist each serving as a district administrator, department chair, curriculum coordinator, literacy specialist, and special

educator. Most panelists had experience working with students with disabilities (93.7%), working with a high percentage of economically disadvantaged students (93.7%), teaching content area courses (87.3%), working with gifted students (84.1%), and working with multilingual learners (57.1%). When asked to rate their level of understanding of the CCSS, all panelists indicated that they were familiar with the standards, with the majority (79.4%) indicating they were very familiar.

Nearly three-quarters of the panelists, 73%, had attained a master's degree or higher. Over a quarter of panelists, 28.6%, had a professional certification or endorsement in the content area associated with their panel. In addition, 73% of participants had an elementary certification or endorsement, while nearly half, 46%, had a middle school certification or endorsement, and 25.4% had one for high school. The panelists had a variety of certifications including the following: 23.8% administrator, 11.1% exceptional children/children with disabilities, 9.5% literacy coach, 6.3% English as a Second Language, and 4.8% curriculum instructional specialist. On average, panelists had 17 years of experience in their educational area of expertise.

Additional details are provided in the section Characteristics of Panelists.

A summary of statistical descriptions and analyses that provides evidence of the reliability of the cut scores and the validity of recommended interpretations*.* The ESS process culminates in the activities that follow the Inconsistent Item Review and Resolution activities. Evidence of the reliability of the cut scores is provided in the section "The standard errors associated with cut scores are reasonable." That section indicates that the median standard errors of the ESS cut scores are generally lower than the SEM of the MTYA, supporting the reliability of the cut scores.

The validity of recommended interpretations is the primary goal of Embedded Standard Setting. That is, ESS supports the primary test interpretations—that students in a given achievement level are well described by the associated achievement level evidence statements and claims. ESS supports such score interpretations by maintaining a consistent focus on the evidentiary chain of reasoning from the CCSS to the Range ALDs to the test items and the associated Item-ALD alignments.

Evidence of the ESS focus on the validity of recommendations is the culminating activity—the R&R workshop—in which items whose Initial Consensus alignments are not supported by empirical data are independently reviewed to resolve the inconsistency. As a result of the workshop, 95% of the inconsistent items (462 of 485) were resolved so that their alignment of record is either supported by the panelists and empirical data (49% of the inconsistent items) or confirm the panelists' Initial Consensus Alignments (46% of the inconsistent items). The primary purpose of ESS is to support assessment system coherence, and when the assessment system elements are sufficiently concordant, valid cut scores are estimated that support score interpretation.

A technical report providing a description of the method used, the diversity of the panelists involved and their qualifications, quality control procedures, the use of impact data, and panelist evaluation results. This technical report describes the method used to set cut scores for the MTYA. The diversity of the panelists involved, and their qualifications are provided in the section "Characteristics of Panelists." Quality control procedures were conducted to assure the accuracy of materials for the ESS Alignment Study and resulting cut scores, as described in the section "Quality Control Procedures." The use of impact data is described in the section "Methodological Choice on the Presentation of Impact Data to Workshop Panelists." Panelist evaluation results are provided in Appendix E.

Summary

The data and evidence provided here provide support for the validity of the adopted Maine Through Year Assessment cut scores.

This technical report, while not formally structured in terms of a validity argument, presents one in terms of the singular focus on the following evidentiary chain of reasoning articulated throughout the report:

1. ALDs should explicate and articulate the content standards of interest—the CCSS—and map to intended interpretations as described in "Section 2. Achievement Level Descriptor Development and Validity."
2. Items should map to ALDs to operationalize and provide evidence for the claims and measurement targets articulated in the ALD evidence statements, as described in "Section 3. The ESS Alignment Study."
3. Cut scores should map to the appropriate items. This requirement is supported by the ESS estimation of cut scores that optimize the coherence of the Item-ALD alignments and empirical data, as described in "Section 4. Embedded Standard Setting Analyses."

These evidentiary linkages are supported by Principled Assessment Design (PAD) and Embedded Standard Setting (ESS) processes. Inconsistent items, which moderate score interpretation, were identified and most were resolved during the Inconsistent Item Review & Resolution workshop. The presence of inconsistent items is not a reflection of the quality of the Maine Through Year Assessments. They exist under any item-based standard setting methodology (i.e., Bookmark, ID Matching, Yes-No Angoff, etc.) but go undetected under these other approaches. ESS minimizes the number of inconsistent items and offers opportunities to further mitigate them through iterative review and revision.

The ongoing application of PAD and ESS in this way provides evidence supporting the mapping of MTYA scores to the intended score interpretations.

Section 7. 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.

Cizek, G. J. (1996). Standard setting guidelines. Educational Measurement: Issues and Practice, 15(1), 12–21.

Cizek, G. J. (2014). Observation and Report on Smarter Balanced Standard Setting. Available for download at Standard-setting-observation-and-report.pdf.

Cizek, G. (2001). Setting Performance Standards, Routledge, New York.

Cizek, G. (2012). Setting Performance Standards, 2nd Ed. Routledge, New York.

Cizek, G., & Agger, C. (2012). Vertically Moderated Standard Setting. In G. Cizek (Ed.), *Setting achievement standards: Foundations, methods, and innovations* (2nd ed.). Routledge.

Cizek, G., & Bunch, M. (2007). A guide to establishing and evaluating achievement standards on tests. Thousand Oaks, CA: Sage Publications.

edCount (2023). Alignment Evaluation for the Maine Through Year Assessment Program. Author.

Ferrara, S., Davis-Becker, S., Kannan, P. (2023). Standard Setting. Chapter in Educational Measurement, 5th Edition. Oxford University Press, USA.

Ferrara, S. & Lewis, D. M. (2012). The Item Descriptor Matching Standard Setting Procedure. Chapter in Setting Performance Standards: Concepts, Methods, and Perspectives, Second Edition. (ed: G. J. Cizek), Lawrence Erlbaum.

Hambleton, R. K. (2001). Setting achievement standards on educational assessments and criteria for evaluating the process. In G. J. Cizek (Ed.), *Setting achievement standards: Concepts, methods, and perspectives* (pp. 89–116). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Jaeger, R. M. (1989). Selection of Judges for Standard Setting: What Kinds? How Many? Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA.

Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. In G. J. Cizek (Ed.), *Setting achievement standards: Concepts, methods, and perspectives* (pp. 53–88). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174.

Lewis, D. & Cook, R. (2020). Embedded Standard Setting: Aligning standard setting methodology with contemporary assessment design principles. *Educational Measurement: Issues and Practices, 39*(1), 8-21.

Lewis, D. M., & Haug, C. (2005). Aligning policy and methodology to achieve consistent across-grade achievement standards. *Applied Measurement in Education*, 18(1), 11-34.

Lewis, D. & Lee, S. (2021). EmStanS [Embedded Standard Setting computer software]. Creative Measurement Solutions LLC.

Lewis, D., Lee, S., & Choi, S. (2021, May). An Investigation of Two Embedded Standard Setting Cut Score Estimation Algorithms: ESS-Count & ESS-Weight. In D. Lewis (Organizer), Embedded Standard Setting: Research & Advances. Symposium at the 2021 annual meeting of the National Council on Measurement in Education, virtual.

NCME (2023). Educational Measurement, 5th Edition (Cook, L. & Pitoniak, M., eds). Oxford University Press, USA.

NWEA (2023). Maine Through Year Assessment Program Technical Report. Author: Portland.

NWEA (2021). Learning during COVID-19: An update on student achievement and growth at the start of the 2021-22 school year. Report by Karyn Lewis & Megan Kuhfeld, NWEA.

Schneider, M. C. & Lewis, D. (2021, May) Embedded ID Matching: ESS-Enhancements to ID Matching to Reduce Panelists' Cognitive Load. In D. Lewis (Organizer), Embedded Standard Setting: Research & Advances. Symposium at the 2021 annual meeting of the National Council on Measurement in Education, virtual.

USDOE (2018). A State's Guide to the U.S. Department of Education's Assessment Peer Review Process. Author, Washington D.C.

Appendices

The following Appendices are published in a separate document:

Maine Through Year Assessment Program Standard Setting Technical Report: Appendices

Appendix A: Inconsistent Item Review & Resolution Facilitator Training Slides

Appendix B: edCount Alignment Evaluation Training Slides

Appendix C: Inconsistent Item Review & Resolution Orientation Slides

Appendix D: Inconsistent Item Review & Resolution Facilitator Guide and Script

Appendix E: ESS Alignment Workshop Day 3 Evaluation Survey Results

Appendix F: Detailed ESS Item Maps

Appendix G: Rosters of Inconsistent and Essentially Consistent Items

Appendix H: Considerations in the Use of Standard Errors for the Adjustment of Cut Scores

Appendix I: The Moderation and Smoothing of Panelists' Cut Scores

Maine Through Year Assessment Program Standard Setting Technical Report: Appendices

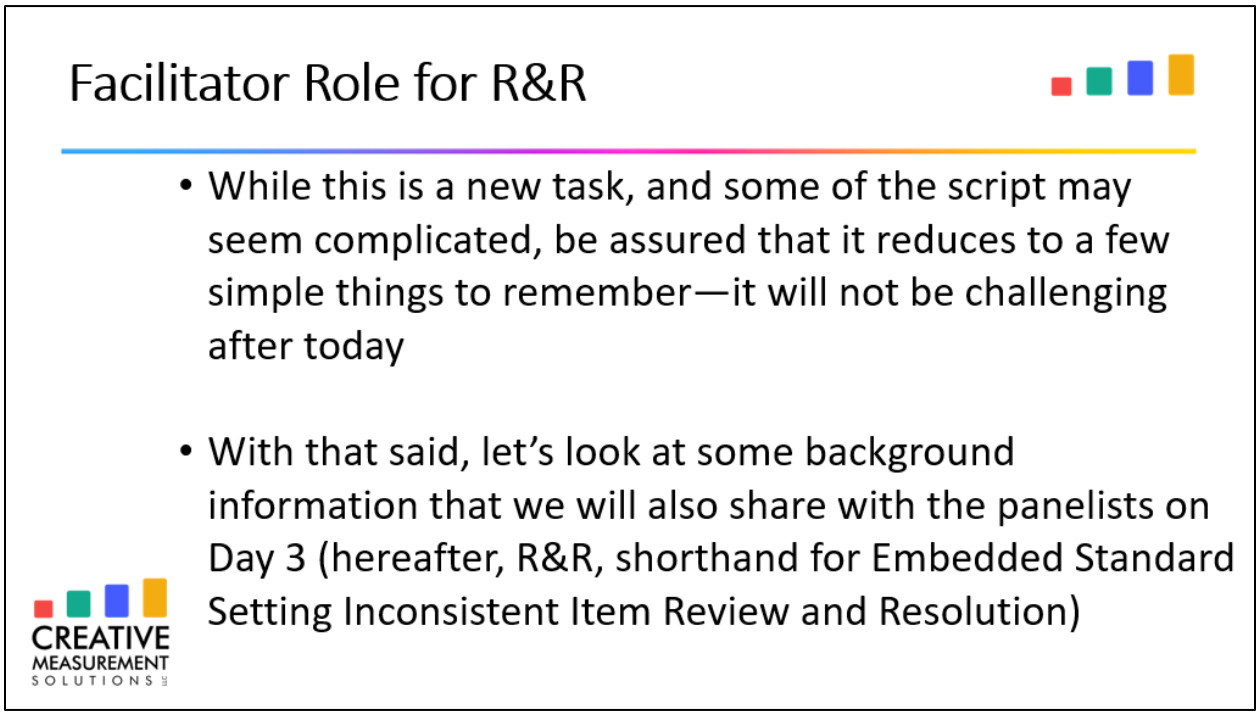
Submitted to NWEA & Maine DOE

Daniel Lewis, Creative Measurement Solutions LLC

September 30, 2023

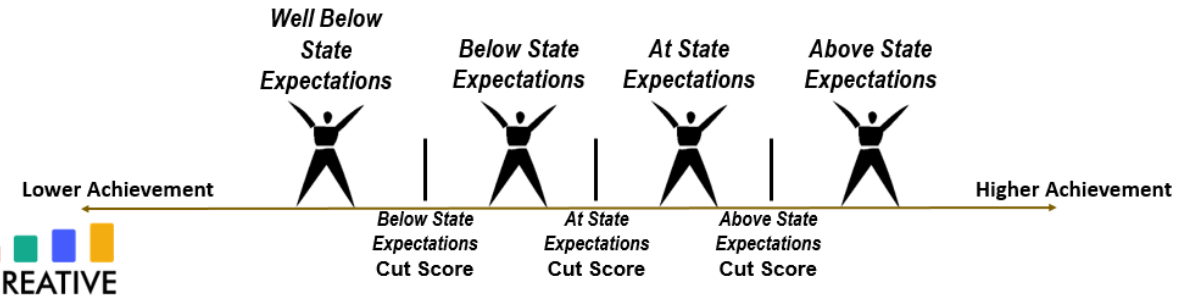
Table of Contents

Appendix A: Inconsistent Item Review & Resolution Facilitator Training Slides



What is Standard Setting?

- Standard setting is the process used to recommend cut scores on a test scale to differentiate specified levels of proficiency



CREATIVE MEASUREMENT SOLUTIONS LLC

3

Item Alignment and Standard Setting

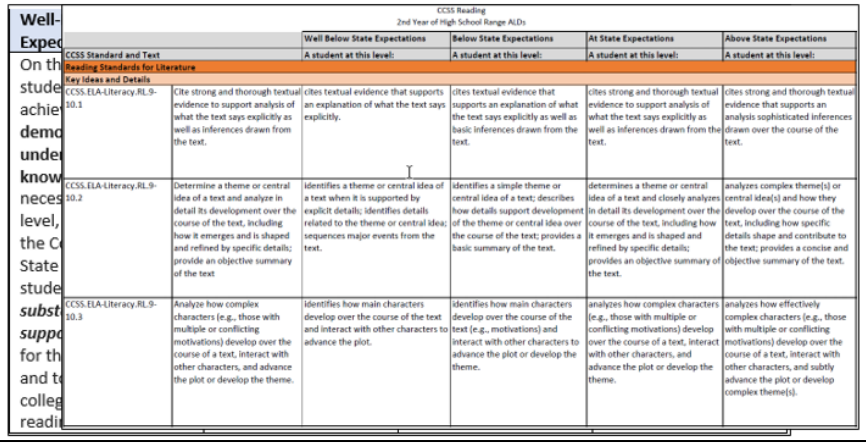
- You have been aligning items to standards and achievement levels
- This is the core activity of many standard setting methods including
  - Bookmark method
  - Item Descriptor Matching method
  - Modified Angoff method
  - ✓ Embedded Standard Setting

CREATIVE MEASUREMENT SOLUTIONS LLC

Embedded Standard Setting (ESS)

ESS is dependent on various elements that are developed throughout the assessment lifecycle:

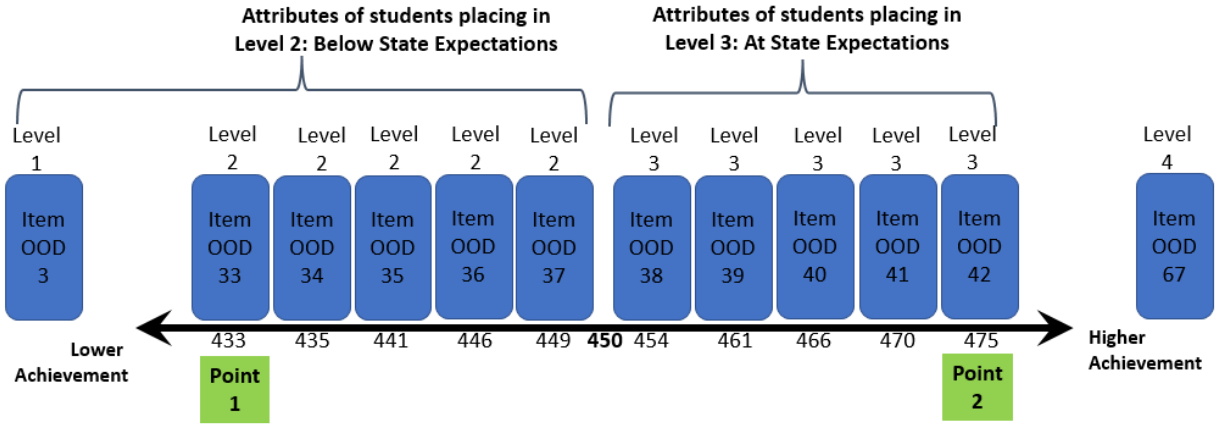
- ✓ ALDs
  - ✓ Policy ALDs
  - ✓ Range ALDs
- ✓ Item-ALD Alignments
- ✓ Empirical data



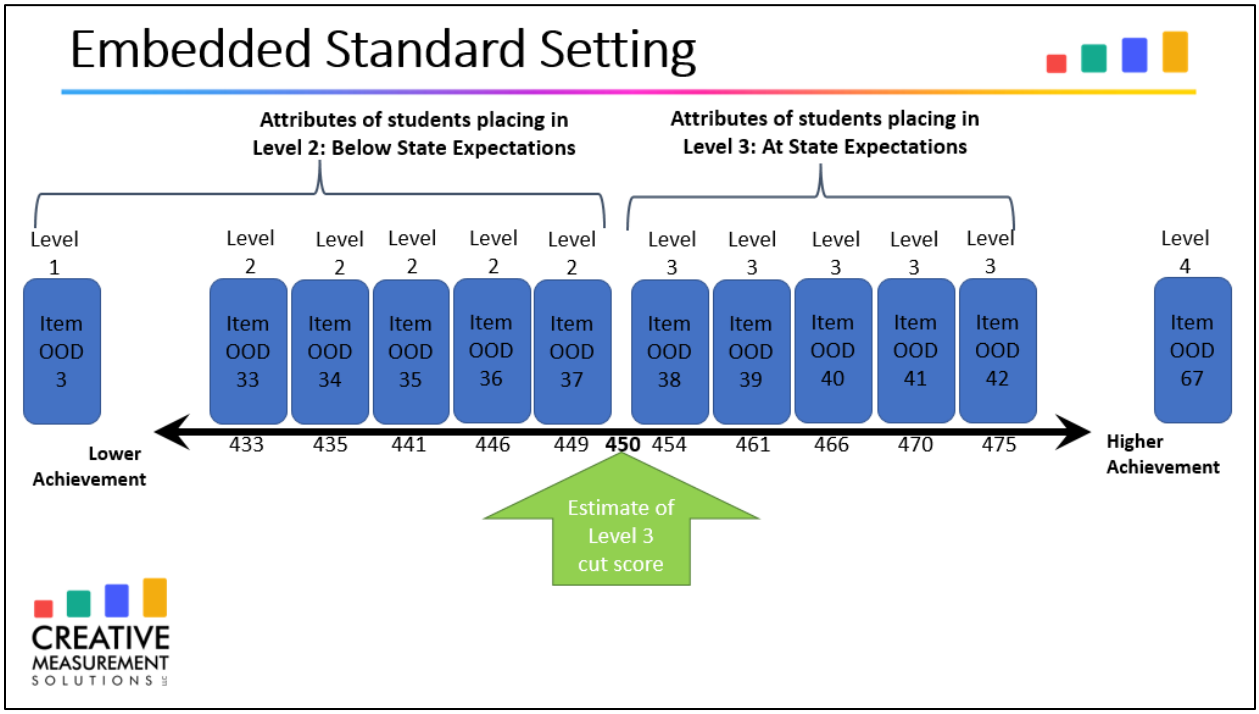
CCSS Reading
2nd Year of High School Range ALDs

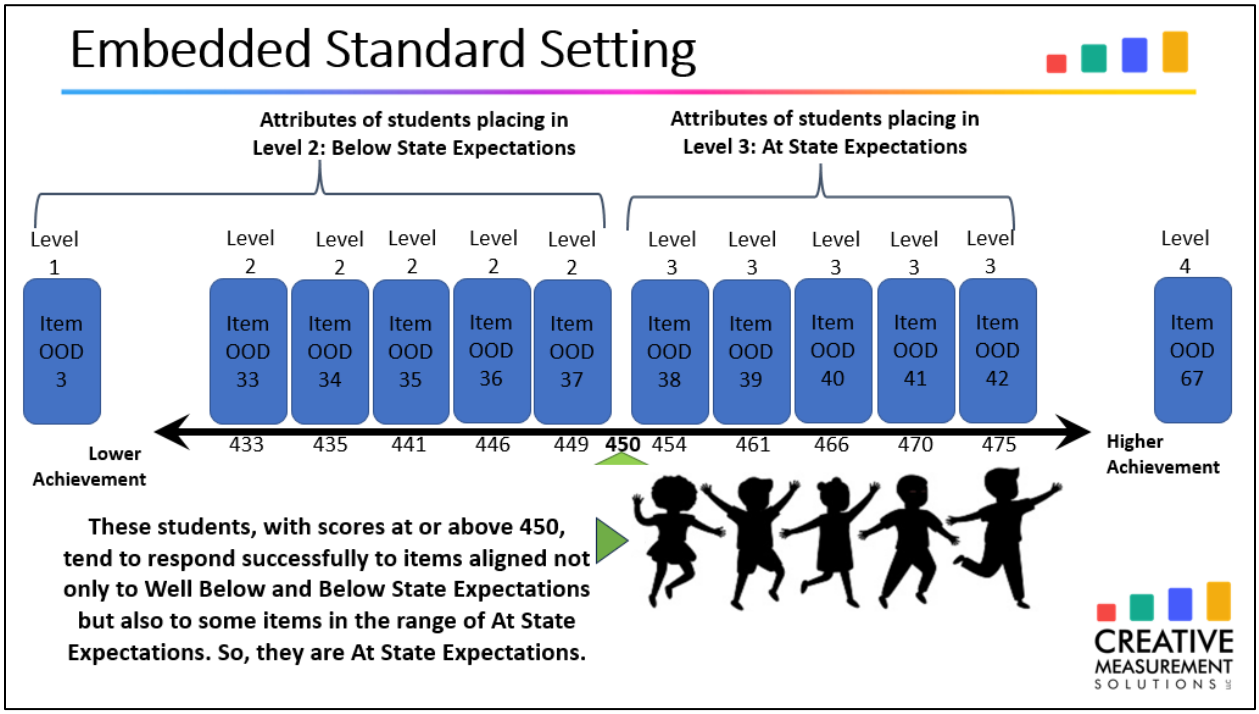
CCSS Standard and Text	Well Below State Expectations A student at this level:	Below State Expectations A student at this level:	At State Expectations A student at this level:	Above State Expectations A student at this level:
Reading Standards for Literature				
Key Ideas and Details				
CCSS.ELA-Literacy.RL.9-10.1	Cite strong and thorough textual evidence to support analysis of what the text says explicitly as well as inferences drawn from the text.	cites textual evidence that supports an explanation of what the text says explicitly.	cites textual evidence that supports an explanation of what the text says explicitly as well as basic inferences drawn from the text.	cites strong and thorough textual evidence to support analysis of what the text says explicitly as well as inferences drawn from the text.	cites strong and thorough textual evidence that supports an analysis sophisticated inferences drawn over the course of the text.
CCSS.ELA-Literacy.RL.9-10.2	Determine a theme or central idea of a text and analyze in detail its development over the course of the text, including how it emerges and is shaped and refined by specific details; provide an objective summary of the text	identifies a theme or central idea of a text when it is supported by explicit details; identifies details related to the theme or central idea; sequences major events from the text.	identifies a simple theme or central idea of a text; describes how details support development of the theme or central idea over the course of the text; provides a basic summary of the text.	determines a theme or central idea of a text and closely analyzes in detail its development over the course of the text, including how it emerges and is shaped and refined by specific details; provides an objective summary of the text.	analyzes complex theme(s) or central idea(s) and how they develop over the course of the text, including how specific details shape and contribute to the text; provides a concise and objective summary of the text.
CCSS.ELA-Literacy.RL.9-10.3	Analyze how complex characters (e.g., those with multiple or conflicting motivations) develop over the course of a text, interact with other characters, and advance the plot or develop the theme.	identifies how main characters develop over the course of the text and interact with other characters to advance the plot.	identifies how main characters develop over the course of the text (e.g., motivations) and interact with other characters to advance the plot or develop the theme.	analyzes how complex characters (e.g., those with multiple or conflicting motivations) develop over the course of a text, interact with other characters, and advance the plot or develop the theme.	analyzes how effectively complex characters (e.g., those with multiple or conflicting motivations) develop over the course of a text, interact with other characters, and subtly advance the plot or develop complex theme(s).

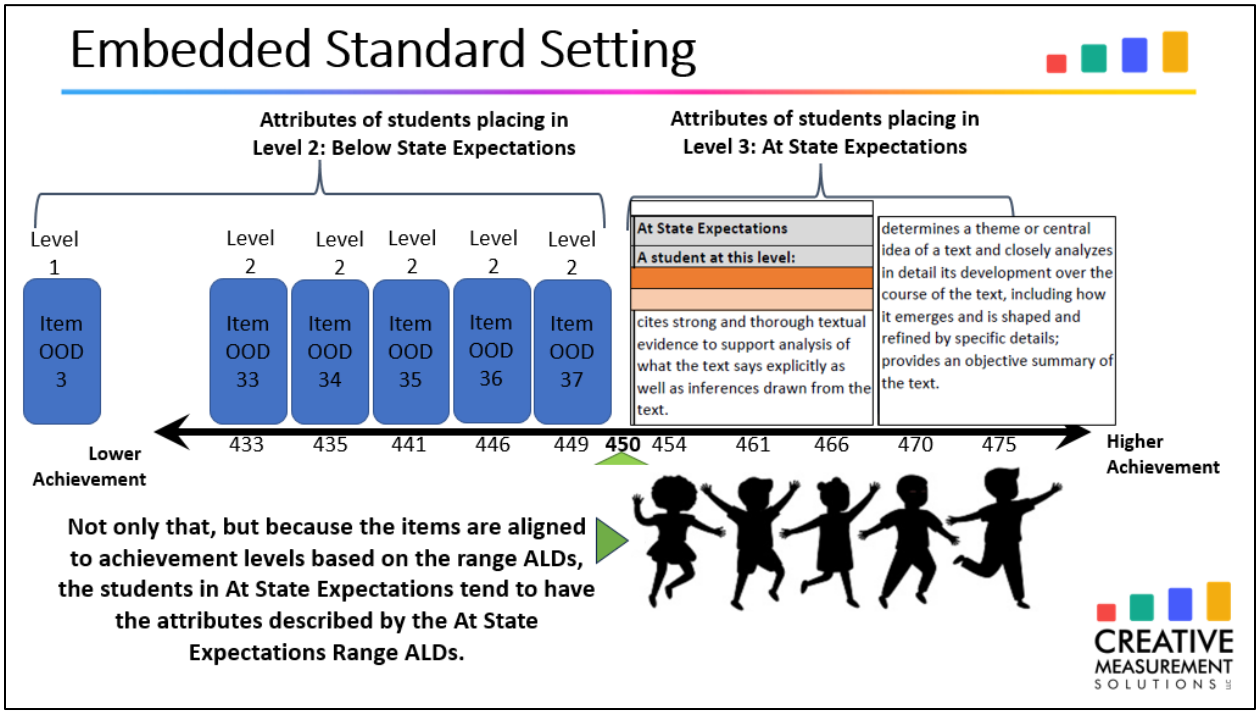
Embedded Standard Setting

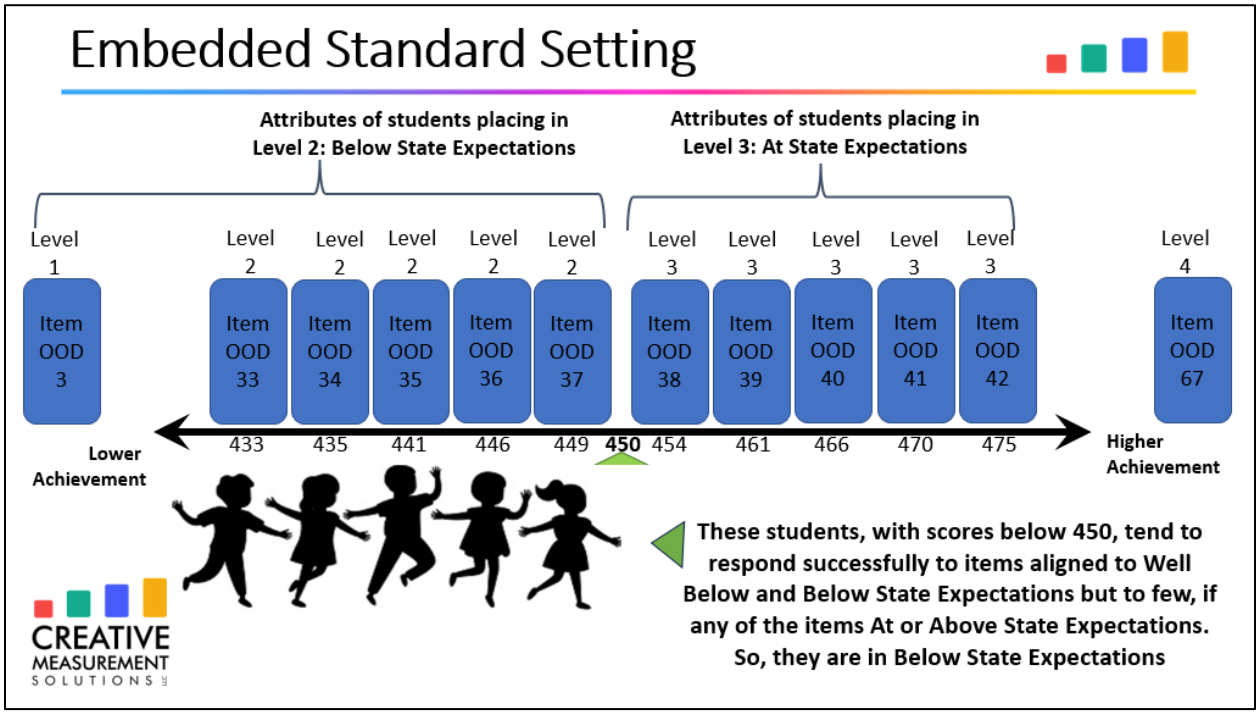


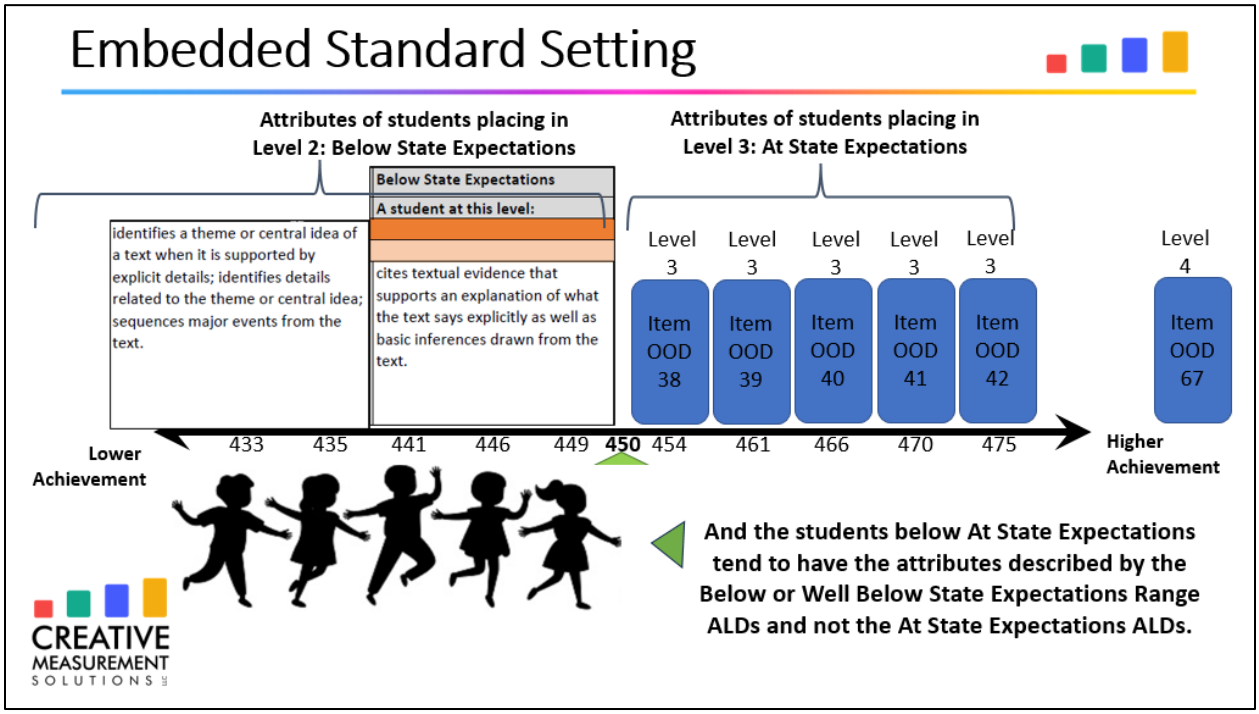
Because they are ordered by difficulty, the 2-point items are represented twice—once for each score point. This is because it should be easier to achieve a 1 on these items than a 2.

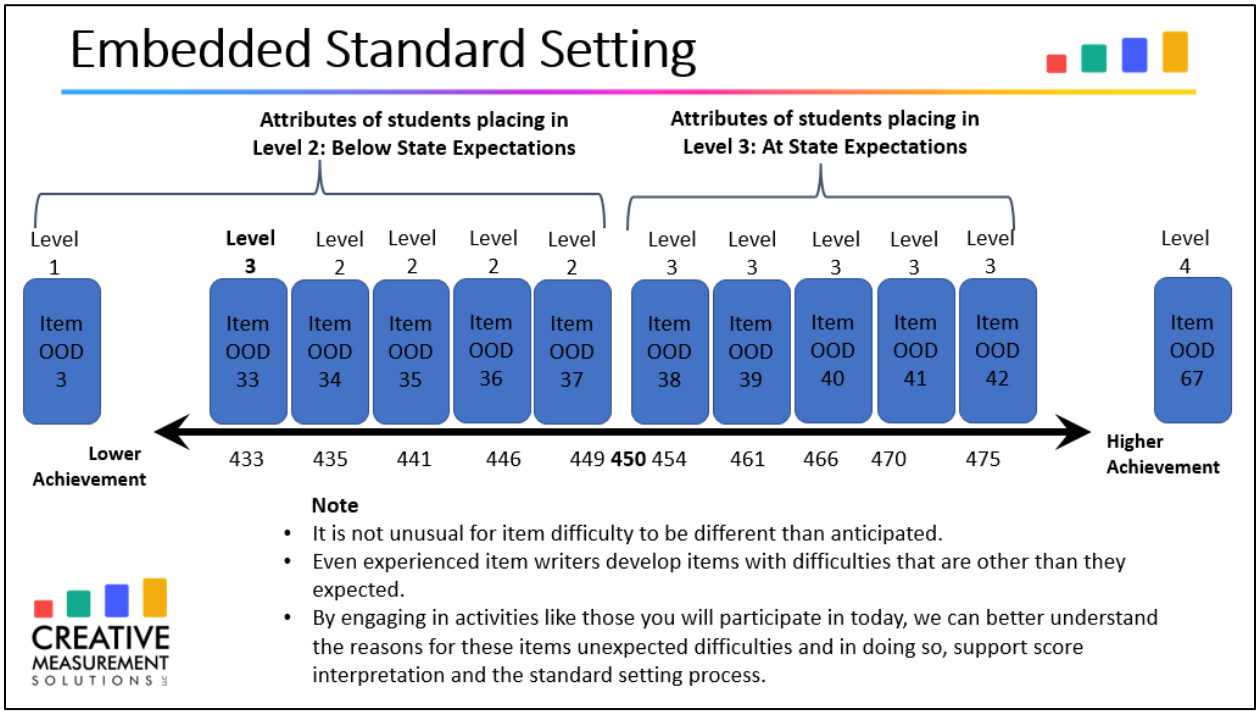

Embedded Standard Setting
Attributes of students placing in Level 2: Below State Expectations
Attributes of students placing in Level 3: At State Expectations
Level 1
Item OOD 3
Level 2
Item OOD 33
Level 2
Item OOD 34
Level 2
Item OOD 35
Level 2
Item OOD 36
Level 2
Item OOD 37
Level 3
Item OOD 38
Level 3
Item OOD 39
Level 3
Item OOD 40
Level 3
Item OOD 41
Level 3
Item OOD 42
Level 4
Item OOD 67
433 435 441 446 449 450 454 461 466 470 475
Lower Achievement
Higher Achievement
Estimate of Level 3 cut score
CREATIVE MEASUREMENT SOLUTIONS


Embedded Standard Setting
Attributes of students placing in Level 2: Below State Expectations
Attributes of students placing in Level 3: At State Expectations
Level 1
Item OOD 3
Level 2
Item OOD 33
Level 2
Item OOD 34
Level 2
Item OOD 35
Level 2
Item OOD 36
Level 2
Item OOD 37
Level 3
Item OOD 38
Level 3
Item OOD 39
Level 3
Item OOD 40
Level 3
Item OOD 41
Level 3
Item OOD 42
Level 4
Item OOD 67
433 435 441 446 449 450 454 461 466 470 475
Lower Achievement
Higher Achievement
These students, with scores at or above 450, tend to respond successfully to items aligned not only to Well Below and Below State Expectations but also to some items in the range of At State Expectations. So, they are At State Expectations.
CREATIVE MEASUREMENT SOLUTIONS


Embedded Standard Setting
Attributes of students placing in Level 2: Below State Expectations
Attributes of students placing in Level 3: At State Expectations
Level 1
Item OOD 3
Level 2
Item OOD 33
Level 2
Item OOD 34
Level 2
Item OOD 35
Level 2
Item OOD 36
Level 2
Item OOD 37
At State Expectations
A student at this level:
cites strong and thorough textual evidence to support analysis of what the text says explicitly as well as inferences drawn from the text.
determines a theme or central idea of a text and closely analyzes in detail its development over the course of the text, including how it emerges and is shaped and refined by specific details; provides an objective summary of the text.
433 435 441 446 449 450 454 461 466 470 475
Lower Achievement
Higher Achievement
Not only that, but because the items are aligned to achievement levels based on the range ALDs, the students in At State Expectations tend to have the attributes described by the At State Expectations Range ALDs.
CREATIVE MEASUREMENT SOLUTIONS


Embedded Standard Setting
Attributes of students placing in Level 2: Below State Expectations
Attributes of students placing in Level 3: At State Expectations
Level 1
Item OOD 3
Level 2
Item OOD 33
Level 2
Item OOD 34
Level 2
Item OOD 35
Level 2
Item OOD 36
Level 2
Item OOD 37
Level 3
Item OOD 38
Level 3
Item OOD 39
Level 3
Item OOD 40
Level 3
Item OOD 41
Level 3
Item OOD 42
Level 4
Item OOD 67
433 435 441 446 449 450 454 461 466 470 475
Lower Achievement
Higher Achievement
These students, with scores below 450, tend to respond successfully to items aligned to Well Below and Below State Expectations but to few, if any of the items At or Above State Expectations. So, they are in Below State Expectations
CREATIVE MEASUREMENT SOLUTIONS


Embedded Standard Setting
Attributes of students placing in Level 2: Below State Expectations
Attributes of students placing in Level 3: At State Expectations
Below State Expectations
A student at this level:
identifies a theme or central idea of a text when it is supported by explicit details; identifies details related to the theme or central idea; sequences major events from the text.
cites textual evidence that supports an explanation of what the text says explicitly as well as basic inferences drawn from the text.
Level 3
Item OOD 38
Level 3
Item OOD 39
Level 3
Item OOD 40
Level 3
Item OOD 41
Level 3
Item OOD 42
Level 4
Item OOD 67
433 435 441 446 449 450 454 461 466 470 475
Lower Achievement
Higher Achievement
And the students below At State Expectations tend to have the attributes described by the Below or Well Below State Expectations Range ALDs and not the At State Expectations ALDs.
CREATIVE MEASUREMENT SOLUTIONS


Embedded Standard Setting
Attributes of students placing in Level 2: Below State Expectations
Attributes of students placing in Level 3: At State Expectations
Level 1
Item OOD 3
Level 3
Item OOD 33
Level 2
Item OOD 34
Level 2
Item OOD 35
Level 2
Item OOD 36
Level 2
Item OOD 37
Level 3
Item OOD 38
Level 3
Item OOD 39
Level 3
Item OOD 40
Level 3
Item OOD 41
Level 3
Item OOD 42
Level 4
Item OOD 67
433 435 441 446 449 450 454 461 466 470 475
Lower Achievement
Higher Achievement
Note
It is not unusual for item difficulty to be different than anticipated.
Even experienced item writers develop items with difficulties that are other than they expected.
By engaging in activities like those you will participate in today, we can better understand the reasons for these items unexpected difficulties and in doing so, support score interpretation and the standard setting process.
CREATIVE MEASUREMENT SOLUTIONS

Item-ALD Alignment: Consistent Items

- You have been aligning items to standards and achievement levels
- For each item (or score point), a "consensus" alignment was established – the level selected by the most panelists.
- When consensus alignment agrees with the empirical level, we consider that item to be consistent.
  - These are items with difficulties estimated from Maine student data that are consistent with their expected difficulties, given the consensus alignment.
  - The way these items work empirically is *consistent* with your judgments of how they should work.

CREATIVE MEASUREMENT SOLUTIONS LLC

Item-ALD Alignment: Inconsistent Items

- For some items, consensus alignments do not agree with their empirical levels
  - They are not of the difficulty expected given the consensus alignment.
  - How the items work empirically is *inconsistent* with your expectations of how they should be working.
  - Example
    - An item aligned to Above State Expectations should be harder than items At or Below State Expectations.
    - If it is not, it is inconsistent with expectations.

CREATIVE MEASUREMENT SOLUTIONS LLC

Reasons Items May Be Inconsistent

- Items may be inconsistent for several reasons:
  - Relevant item attributes may better reflect the attributes associated with the empirically identified performance level.
  - Language in the ALDs' evidence statements make it challenging to establish a clear best fit
  - Opportunity to Learn (OTL) may make some items easier or harder than expected
    - Some aspect of the item is affecting the difficulty in unintended ways

CREATIVE MEASUREMENT SOLUTIONS LLC

Resolving Inconsistencies

- By discussing and comparing the ALD evidence statements associated with alignments, inconsistencies can be resolved by:
  - Identifying items where the best fitting evidence statement is associated with the empirical level, not the aligned level
  - Refining ALD evidence statements to more precisely reflect attributes that distinguish between achievement levels.
    - When multiple items support such refinements
  - Accepting the empirical alignment because the reasons for the inconsistency reflect the effects of Opportunity to Learn or unintended difficulty factors within the item.

CREATIVE MEASUREMENT SOLUTIONS LLC

Up Next

- Your Grade Group Facilitator will guide you one by one through each inconsistent item. For each inconsistent item you will:
  - …examine the ALD evidence statements associated with your panel's consensus alignment(s) and those of the empirical level,
  - discuss which evidence statement is the best match to the content characteristics or measurement attributes of the item, and consider
    - maintaining or adjusting your alignment based on your new perspective following discussion, or
    - Recommend refinements to the ALD evidence statements necessary for them to better support the consensus alignments.

CREATIVE MEASUREMENT SOLUTIONS LLC

Back to Basics: Facilitator Role for R&R

- While this is a new task, and some of the script may seem complicated, be assured that it reduces to a few simple things to remember—it will not be challenging after today

- With that said, let's look at some background information that we will also share with the panelists on Day 3 (hereafter, R&R, shorthand for Embedded Standard Setting Inconsistent Item Review and Resolution)

CREATIVE MEASUREMENT SOLUTIONS LLC

Back to Basics: Facilitator Role for R&R

CREATIVE MEASUREMENT SOLUTIONS LLC

	A	B	C	D	E	F	G	H	I	J	S	U	V	W	X	Y	Z
1	Panelist Information			Item Information						Standard Alignment	Final ALD Resolution		Consistency Information				
2	Panelist ID	Subject Area	Grade	Item Number	Panelist Row	Form Number	Sequence Number	Score Point	Max Points	Primary Standard	Final ALD Level - Item/Score Point	Final ALD Comments	Order of Difficulty	Empirical ALD Alignment Level	Consensus ALD Alignment Level	Consistency Status	R&R Discussion Status
3	RD3FAC	Reading	3	VR702675_1	16	1	21	1	2				1	Well Below	Below	Inconsistent	
4	RD3FAC	Reading	3	VR702675_1	79	3	23	1	2				1	Well Below	Not Aligned	Inconsistent	
5	RD3FAC	Reading	3	VR699667_1	4	1	2	1	2				4	Well Below	Well Below	Consistent	
6	RD3FAC	Reading	3	VR701589	43	2	14	1	1				4	Well Below	Well Below	Consistent	
7	RD3FAC	Reading	3	VR699670_1	7	1	5	1	2				5	Well Below	Below	Inconsistent	
8	RD3FAC	Reading	3	VR701587	41	2	12	1	1				6	Well Below	Well Below	Consistent	
9	RD3FAC	Reading	3	VR701540	39	2	10	1	1				7	Well Below	Below	Inconsistent	
10	RD3FAC	Reading	3	VR701594	47	2	26	1	1				8	Well Below	Well Below	Consistent	
11	RD3FAC	Reading	3	VR701538	38	2	9	1	1				9	Below	Above	Inconsistent	
12	RD3FAC	Reading	3	VR701536	36	2	7	1	1				10	Below	Below	Consistent	
13	RD3FAC	Reading	3	VR701591	44	2	23	1	1				11	Below	Well Below	Inconsistent	
14	RD3FAC	Reading	3	VR701588	42	2	13	1	1				12	Below	Below	Consistent	
15	RD3FAC	Reading	3	VR701535	35	2	6	1	1				13	Below	At	Inconsistent	
16	RD3FAC	Reading	3	VR701586	40	2	11	1	1				14	Below	Above	Inconsistent	
17	RD3FAC	Reading	3	VR701592	45	2	24	1	1				15	Below	Not Aligned	Inconsistent	
18	RD3FAC	Reading	3	VR702677	51	2	30	1	1				16	Below	Above	Inconsistent	
19	RD3FAC	Reading	3	VR701053	11	1	9	1	1				17	Below	At	Inconsistent	
20	RD3FAC	Reading	3	VR701593	46	2	25	1	1				18	Below	Well Below	Inconsistent	
21	RD3FAC	Reading	3	VR702674	15	1	20	1	1				19	Below	Above	Inconsistent	
22	RD3FAC	Reading	3	VR702674	49	2	28	1	1				19	Below	Well Below	Inconsistent	
23	RD3FAC	Reading	3	VR702674	78	3	22	1	1				19	Below	At	Inconsistent	
24	RD3FAC	Reading	3	VR699668	5	1	3	1	1				22	Below	Below	Consistent	
25	RD3FAC	Reading	3	VR701056	14	1	19	1	1				23	Below	Above	Inconsistent	
26	RD3FAC	Reading	3	VR702671_1	48	2	27	1	2				24	Below	At	Inconsistent	
27	RD3FAC	Reading	3	VR699686_1	30	2	1	1	2				25	Below	Above	Inconsistent	

Reading Grade 3 Facilitator Tab

Facilitator R&R Form: Technical Reminders

- Items repeated on multiple forms were grayed out on your forms but appear multiple times on ours
  - Always in adjacent rows
  - You will know they are repeat by
    - observing the same item number (same _1 or _2 as well)
    - Observing that they have the same Order of Difficulty
    - Seeing the Consensus Alignment as “Not Aligned,”
      - Which could also be a tie, if not a repeat. There may be no ties.
  - To skip repeated items, set R&R Discussion Status to “Discussed.”

CREATIVE MEASUREMENT SOLUTIONS LLC

Facilitator R&R Form: Focus Reminders

- Focus on Columns W & X. This will guide discussion.
- Focus on correct panelist row
- Focus on score point and correct rubric

CREATIVE MEASUREMENT SOLUTIONS LLC

Questions?

CREATIVE MEASUREMENT SOLUTIONS LLC

Appendix B: edCount Alignment Evaluation for the Maine Through Year Assessment Training Slides

Alignment Evaluation for the Maine Through Year Assessment

July 18 - 20, 2023
Reading

edCount LLC because all students count
CREATIVE MEASUREMENT SOLUTIONS

Welcome & Introductions

- Maine DOE team
- edCount team
- Creative Measurement Solutions team
- Maine educator panels

Friendly Virtual Reminders:
- Please keep your computer camera on at all times.
- Throughout the sessions, please keep your computer microphone muted when not talking.
- Save your work often.

High-level Agenda

- Day 1
  - Alignment & Item Rating Training
  - Independent Item Ratings & Panel Discussions
- Day 2
  - Complete Independent Item Ratings & Panel Discussions
  - Complete Days 1 & 2 Evaluation
- Day 3
  - Inconsistent Item Review Training
  - Inconsistent Item Reviews & Panel Discussions
  - Complete Day 3 Evaluation

Housekeeping

- Should you encounter any difficulties during the study, please contact one of the following edCount staff members for assistance:
  - Laura Heslop: lheslop@edcount.com / 202-819-0821
  - Kevin Konort: kkonort@edcount.com / 202-302-0508

Reading Panel Facilitators

Panel	Facilitator	Email	Cell
Grade 3	Bethany Ferron	bethanyferron@gmail.com	952-994-7710
Grade 4	Jean Clayton	jclayton@edcount.com	202-257-2284
Grade 5	Kila Posey	kposey@edcount.com	678-458-5208
Grade 6	Erin Buchanan	ebuchanan@edcount.com	202-631-0111
Grade 7	Michelle Shipman	mshipman@edcount.com	202-590-7439
Grade 8	Charlene Turner	cturner@edcount.com	202-322-9419
High School	Alice Garcia	agarcia@edcount.com	202-302-4959

OVERVIEW OF THE ALIGNMENT EVALUATION PROCESS

Purpose of Training

- Explain the purpose of our work
- Provide an overview of the rating process
- Explain each task and review any rating rubrics
- Prepare for the hands-on work in the breakout sessions

Logistics and Materials

- Your Non-disclosure Form was collected prior to this meeting.
- View items in a secure environment/private room.
  - Items may not be viewed or discussed with others around you.
- Keep your computer camera on at all times.
- Complete the evaluation surveys at the end of Days 2 and 3.

Test Security Reminders

- Your signed Non-disclosure Form indicates your understanding that:
  - ✓ The test materials and items are **secure** materials.
  - ✓ You may **not** take pictures/screenshots, print, or save copies of the test items in any form.
- During the meeting we ask that you keep closed all programs/applications other than the ones you will use—Microsoft Teams, Box (and the associated files within Box), and Content Review Tool (CRT), which is the application in which you will view the secure test items.
  - Other programs/applications should not be opened or used while we are actively working.

Test Security Reminders

- You are working with secure content. Sharing information about the content of these items outside of this study would undermine the validity of the test and negate the investment the State of Maine has made in the tests.
- You may talk freely about the methods used in this study, but you may not share anything about item content.
- Please notify your facilitator if you need to step away to use your phone/check email.

Meeting Materials

- **You received an individual folder in Box, labeled with your unique ID, containing several folders:**
  - Agenda & Training Materials
  - Activity 1 Materials: Standards and ALD Alignment
    - Common Core State Standards (CCSS) for ELA
    - Achievement Level Descriptors (ALDs) - Reading
    - Domain Concurrence & Differentiation Rubrics
    - Domain Concurrence & Differentiation Rating Sheet

Meeting Materials

- Activities 2 and 3: Item Rating Materials
  - Guide to Accessing Resources and Data Entry
  - Panelist Rating Sheet
  - Common Core State Standards – ELA
  - Depth of Knowledge (DOK) Rubric
  - Achievement Level Descriptors - Reading
  - Item Scoring Rubrics

Meeting Materials

- End of Workshop
  - Evaluation Survey Link for Day 2
  - Evaluation Survey Link for Day 3
- After the workshop, NWEA will send your stipend and any paperwork related to that.

Workshop Agenda – Day 1

Time	Activity
8:45 to 9:00	Log into general training session
9:00 to 10:30	Full group training
10:30 to 10:45	Break/move to individual panel breakout sessions
10:45 to 12:00	Log into Box Form 1: Review/Discuss CCSS and Maine Through Year Assessment ALDs
12:00 to 12:45	Lunch
12:45 to 2:15	Log into Content Review Tool Form 1: Item Rating Activities
2:15 to 2:30	Break
2:30 to 4:00	Form 1: Complete Item Rating Activities

ALIGNMENT EVALUATION

Overview of Alignment

In general:

- The purpose is to evaluate the degree to which the scores from a standards-based assessment can be interpreted in relation to the standards.
- Evidence from an independent alignment study should be considered in combination with evidence about the design and development of the assessment.
- This alignment evaluation will provide evidence of the validity of score use and interpretations for Through-Year Assessments.

Alignment Evaluation Purpose

- To investigate alignment of test forms to the intended constructs (i.e., do tests measure what they are intended to measure?)
- To provide independent evaluation for use in federal peer review

Peer Review Background

- As part of the Peer Review process, the United States Department of Education (USDOE) requires all states submit standards and assessments documentation for approval.
  - Reading/language arts
  - Mathematics
  - Science
- Alignment evaluation information from this workshop will be used as evidence for the Through Year Assessment submission.

Maine Through Year Assessment

- The **Through Year Assessment** is the summative accountability measurement designed to measure the state's academic content standards.
- This evaluation focuses on the grades 3–8 and second year of high school assessments for reading and mathematics.

Review of Key Terms

- **Common Core State Standards:** Statements of what a student should know, understand, and be able to do by the end of the year
- **Domains:** Key areas or core ideas, under which standards are clustered
  - *Examples:* Reading Literature, Reading Information, Vocabulary
- **Achievement Level Descriptors:** Statements that describe the knowledge and skills students typically demonstrate at each achievement level
  - The Through Year Assessment has four levels: *Well-Below State Expectations*; *Below State Expectations*; *At State Expectations*; and *Above State Expectations.*

ALIGNMENT ACTIVITIES

What is Alignment?

- **Alignment** is about coherent connections across various aspects within and across an assessment system.
  - Relates not simply to an assessment but to the assessment scores and the resulting interpretations of those scores

(Forte, 2013a, 2013b)

Our Focus

Three main questions guide our work today:

1. How well do the ALDs reflect the key knowledge and skills in the CCSS for reading?
2. How well do the sets of items that contribute to students' scores reflect the CCSS for reading?
3. How well do the test items align with the ALDs?

The ratings you make will help us answer these evaluation questions.

Roles and Responsibilities

Maine DOE

- Observe the process, but do not participate
- Answer any assessment or policy-related questions

Facilitators

- Provide training and materials
- Answer questions only to facilitate the ratings
- Facilitate the group process and discussion
- Ensure all individual ratings are captured appropriately
- Capture discussions

Roles and Responsibilities

Panelists

- Review items and CCSS for the assigned grade level
- Review all rating rubrics and frameworks
- Ask questions to clarify understanding
- Code each item individually, checking agreement as led by a facilitator
- Participate in all discussions with colleagues
- Maintain test security at all times

Your Activities

- **Activity 1:** Panel review and discussion of the CCSS for the purpose of increasing familiarity with the content standards, identifying key knowledge and skills, and determining alignment of ALDs to the reading content standards
- **Activity 2:** Independent review of items for the purpose of identifying standard alignment, Depth of Knowledge (DOK), and ALD alignment
  - At the end of this activity, each panel will take a few minutes to reflect on the test forms – looking across the three forms (grades 3–8) or the single form (high school) to discuss the standards represented and the range of cognitive challenge present.
- **Activity 3:** Panel review and discussion of items with ALD alignments that appear inconsistent with how students actually performed on the items. A group training will be held on the morning of Day 3 (Thursday) to guide you in this work.

Activity 1 Panelist Tasks – Part 1

- Independently review the CCSS for your grade level.
- Consider the key knowledge and skills associated with your grade level.
  - These are the big ideas from the grade level that, as a teacher, you work for every student to master by the end of the year.

Remember, the Through Year focuses on reading, so some standards may not be measured by design.

Activity 1 Panelist Tasks – Part 2

- Consider how well the ALDs capture the key knowledge and skills included in the grade-level standards and apply the Domain Concurrence Rubric.
- Consider how well the ALDs differentiate as they move from one level to the next, using the ALD Differentiation Rubric.
- ***This is a group activity.***
- ***Your facilitator will capture your panel's ratings.***

Achievement Level Descriptors (ALDs)

- Describe the range of knowledge and skills possessed by students in each of the four levels

ALD Example: Grade 3 Reading

		Well Below State Expectations	Below State Expectations	At State Expectations	Above State Expectations
CCSS Standard and Text		A student at this level:	A student at this level:	A student at this level:	A student at this level:
Reading Standards for Literature					
Key Ideas and Details					
CCSS.ELA-Literacy.RL.3.1	Ask and answer questions to demonstrate understanding of a text, referring explicitly to the text as the basis for the answers.	answers simple questions referring explicitly to the text.	asks basic questions and draws on simple details or basic connections as the basis for answers.	asks and answers explicit questions to demonstrate understanding of a text,referring explicitly to the text as the basis for the answers.	asks and answers inferential questions and uses textual evidence to support answers.
CCSS.ELA-Literacy.RL.3.2	Recount stories, including fables, folktales, and myths from diverse cultures; determine the central message, lesson, or moral and explain how it is conveyed through key details in the text.	provides a basic recounting of a story; identifies an explicit central message, lesson, or moral.	recounts basic ideas/events in stories, including fables, folktales, and myths from diverse cultures; identifies a basic central message, lesson, or moral given simple connections that can be made from the evidence in the text.	recounts stories, including fables, folktales, and myths from diverse cultures; determines the central message, lesson, or moral and explains how it is conveyed through key details in the text.	recounts, in their entirety, stories, including fables, folktales, and myths from diverse cultures; determines the implicit central message, lesson, or moral and analyzes how it is conveyed through key details in the text.

Domain Concurrence Rubric

Rating	Description
Fully Represents	The ALDs fully represent the key knowledge and skills found in the grade-level academic content standards.
Somewhat Represents	The ALDs represent some of the key knowledge and skills found in the grade-level academic content standards, but there are gaps.
Does Not Represent	The ALDs minimally represent or do not represent the key knowledge and skills found in the grade-level academic content standards.

Differentiation of ALDs Rubric

Rating	Description
Adequately Reflects Progression	The progressions from lower to higher levels in the ALDs adequately reflect reasonable progressions from lower levels of sophistication to higher levels of sophistication.
Somewhat Reflects Progression	The progressions from lower to higher levels in the ALDs somewhat reflect reasonable progressions from lower levels of sophistication to higher levels of sophistication.
Does Not Reflect Progression	The progressions from lower to higher levels in the ALDs do not reflect reasonable progressions from lower levels of sophistication to higher levels of sophistication.

Domain Concurrence/Differentiation Panelist Tasks

- If the rating chosen is less than ***fully represents/adequately reflects progression***:
  - Provide additional information regarding what is missing or misrepresented
  - Discuss the rationale for choosing the rating
  - Your facilitator will capture the group discussion/rationale

This activity is completed via panel discussion.

Activity 2: Individual Item Ratings

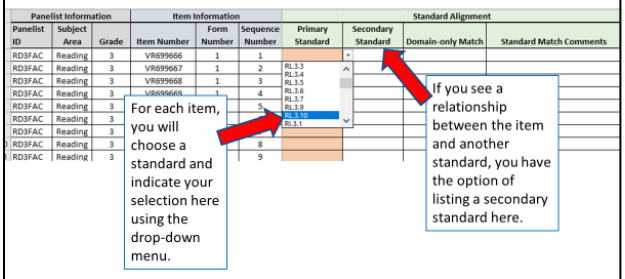
- Calibrate as a group on item-level ratings.
  - Review each item for the purpose of assigning the following ratings:
    - **2a:** Standards match
    - **2b**: DOK level
    - **2c:** Achievement level using the ALDs
    - **2d:** Reflective discussion about cognitive challenge
- After the initial calibration activity, rate items independently, then discuss/share ratings.
  - During discussion, you have the opportunity but are ***not required*** to edit any item ratings or to add secondary ratings.

Friendly reminder: As you work throughout each day, save your rating form to Box often – we don't want to lose your ratings!

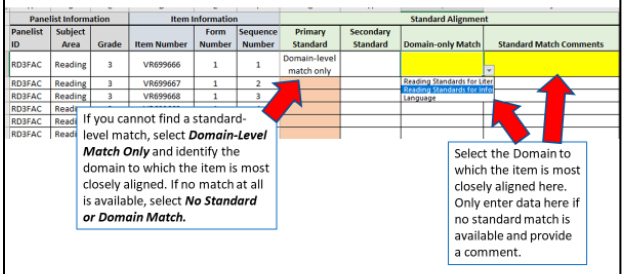
Activity 2a: Item Standard Ratings

- Purpose of this activity:
  - Identify standard(s) assessed by each item.
- Select content match to a standard:
  - Primary
  - Secondary
- If alignment is not found at the standard level, indicate this in the rating sheet and move up to the domain level.

Standard Rating



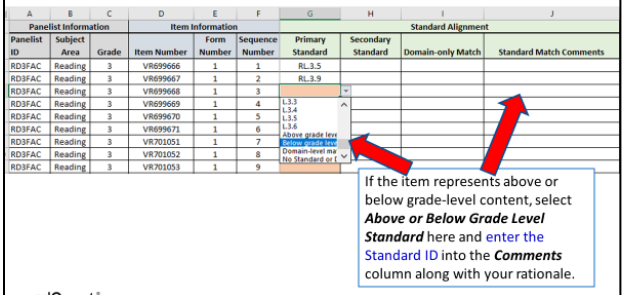
No Standard Match



If you cannot find a standard-level match, select ***Domain-Level Match Only*** and identify the domain to which the item is most closely aligned. If no match at all is available, select ***No Standard or Domain Match.***

Select the Domain to which the item is most closely aligned here. Only enter data here if no standard match is available and provide a comment.

Above or Below Grade-Level Match



If the item represents above or below grade-level content, select ***Above or Below Grade Level Standard*** here and enter the Standard ID into the ***Comments*** column along with your rationale.

Activity 2b: Item DOK Ratings

- Depth of Knowledge: Levels used to describe the degree of complexity of knowledge and skill that an item requires
  - What must the student know, understand, and be aware of in order to answer and explain the answer(s), outcome(s), result(s), and solution(s)?

Webb's DOK Rubric

Level	Depth of Knowledge
4	**Extended Thinking** – Requires an investigation, time to think and process multiple conditions of the problem; most on-demand assessments will not include Level 4 activities
3	**Strategic Thinking** – Requires reasoning, developing a plan or a sequence of steps, some complexity
2	**Skill/Concept** – Use information or conceptual knowledge; requires two or more steps
1	**Recall** – Recall a fact, information, or procedure; process information on a low level; perform one-step calculation

Difficulty

- Different than cognitive complexity
- Based on the percentage of students who respond correctly to the item on the test
  - Difficulty is empirical – it can be quantified
- Determined after the test has been taken

DOK is a judgment about the cognitive demands, considering both content and skill, required to answer an item using a 4-point scale.

Cognitive Complexity and Difficulty

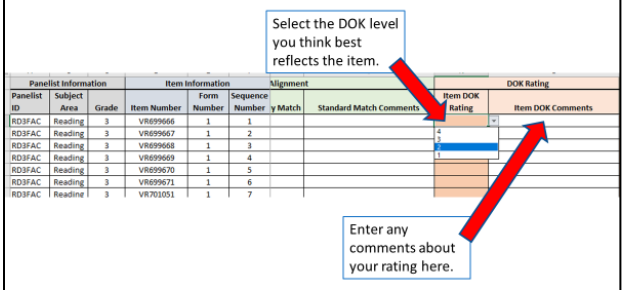
- These two concepts are not synonymous
- *Example item:*
  - Place the following US Vice Presidents in the order in which they served in the role:

1. Elbridge Gerry
2. Daniel Tompkins
3. John Calhoun
4. Martin Van Buren

A. 1 – 2 – 3 – 4
B. 3 – 4 – 1 – 2
C. 4 – 3 – 2 – 1
D. 2 – 4 – 3 – 1

Low Complexity
High Difficulty

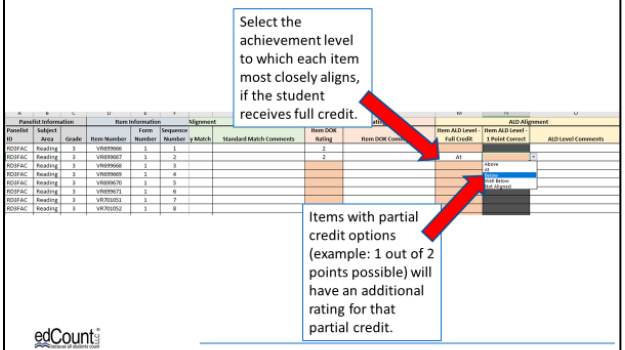
Item DOK Rating



Activity 2c: Item ALD Ratings

- Assign each item to an Achievement Level using the ALDs.
- Evaluate the alignment of the knowledge and skills required to answer the item to the knowledge and skill requirements of the ALD level.
- For items with no content or performance match to any ALD, select *Not Aligned*.
- Add comments as needed to explain your decision.

Item ALD Rating



Activity 2: Discussion of Ratings

- Throughout each day, each panel will discuss ratings at certain intervals to ensure calibration.
- Panelists have the opportunity (but are not required) to make a final rating/selection that differs from the original selection as well as add a secondary standard alignment.

Activity 2d: Discussion of Cognitive Challenge

- After your panel has completed rating all items on the test forms, your facilitator will lead you through a discussion about the cognitive challenge of the collection of items on the test form.
  - Do the items on the test form reflect the range of cognitive challenge present in the content standards and ALDs?
  - Do the forms have a range of items, allowing students across the achievement continuum to show what they know and can do?
  - Are there any gaps?

Conclusion of Day 2

- After all individual and panel activities are complete, panelists will be asked to fill out and return:
  - An evaluation of the alignment workshop and process
- Remember, we will convene as a large group again on Thursday morning at 9:00 am to receive additional training.

Next Steps

- Using the Microsoft Teams link emailed to you, convene in the panel breakout session in 15 minutes.

Friendly reminder: As you work throughout each day, save your rating form to Box often – we don't want to lose your ratings!

edCount Contact Information

Melissa Fincher, Ph.D.
mfincher@edCount.com
Cell: 404-213-5098

Appendix C: Inconsistent Item Review & Resolution Orientation Slides

Maine DOE

Embedded Standard Setting Inconsistent Item Review & Resolution

Orientation

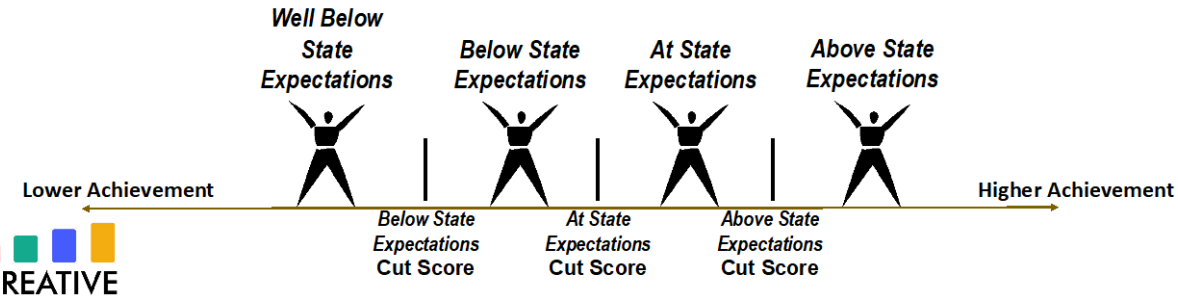
CREATIVE MEASUREMENT SOLUTIONS

Daniel Lewis & Robert Cook
Creative Measurement Solutions LLC

edCount LLC because all students count

What is Standard Setting?

- Standard setting is the process used to recommend cut scores on a test scale to differentiate specified levels of proficiency



CREATIVE MEASUREMENT SOLUTIONS

2

Item Alignment and Standard Setting

- You have been aligning items to standards and achievement levels
- This is the core activity of many standard setting methods including
  - Bookmark method
  - ID Matching method
  - Modified Angoff method
  - ✓ Embedded Standard Setting

CREATIVE MEASUREMENT SOLUTIONS

Embedded Standard Setting (ESS)

ESS is dependent on various elements that are developed throughout the assessment lifecycle:

✓ ALDs

✓ General ALDs

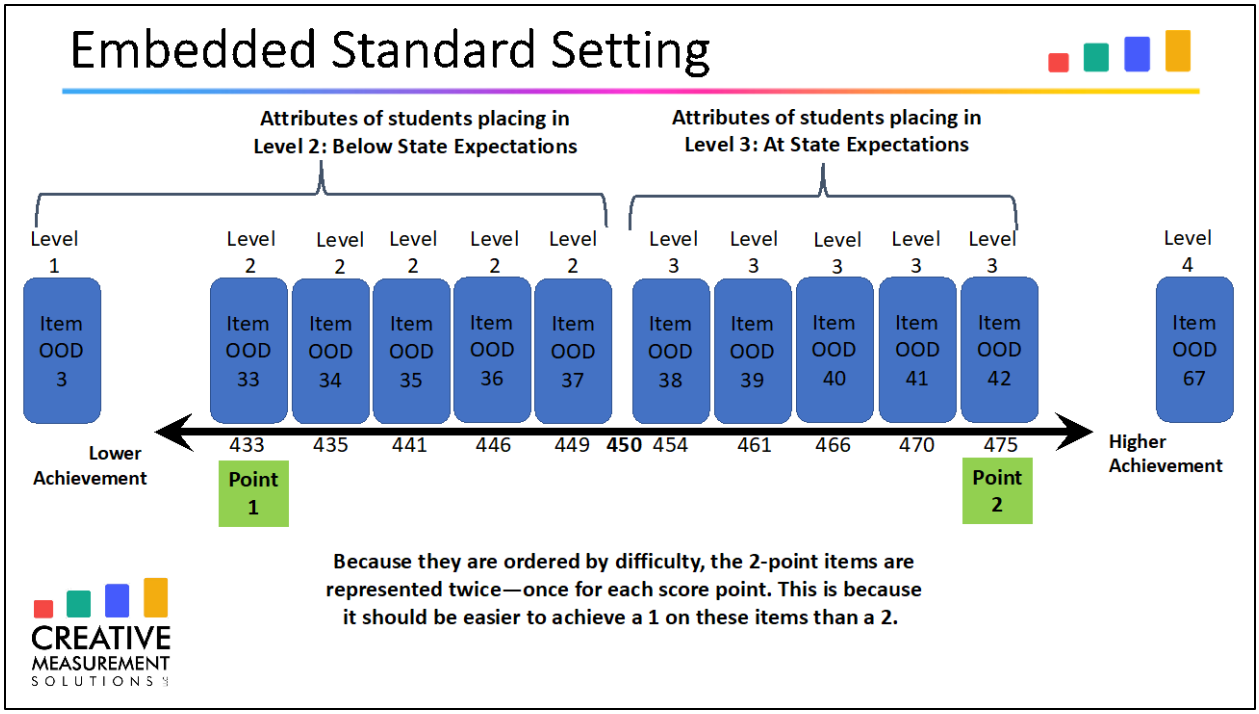
Well-Below State Expectations	Below State Expectations	At State Expectations	Above State Expectations
		On this assessment, students at this achievement level **demonstrate the knowledge and skills** necessary at this grade level, as specified in the Common Core State Standards. The students ***are prepared*** for the next grade level and are on track for college and career readiness.	On this assessment, students at this achievement level **demonstrate advanced understanding of the knowledge and skills** necessary at this grade level, as specified in the Common Core State Standards. The students ***are well prepared*** for the next grade level and are well prepared for college and career readiness.

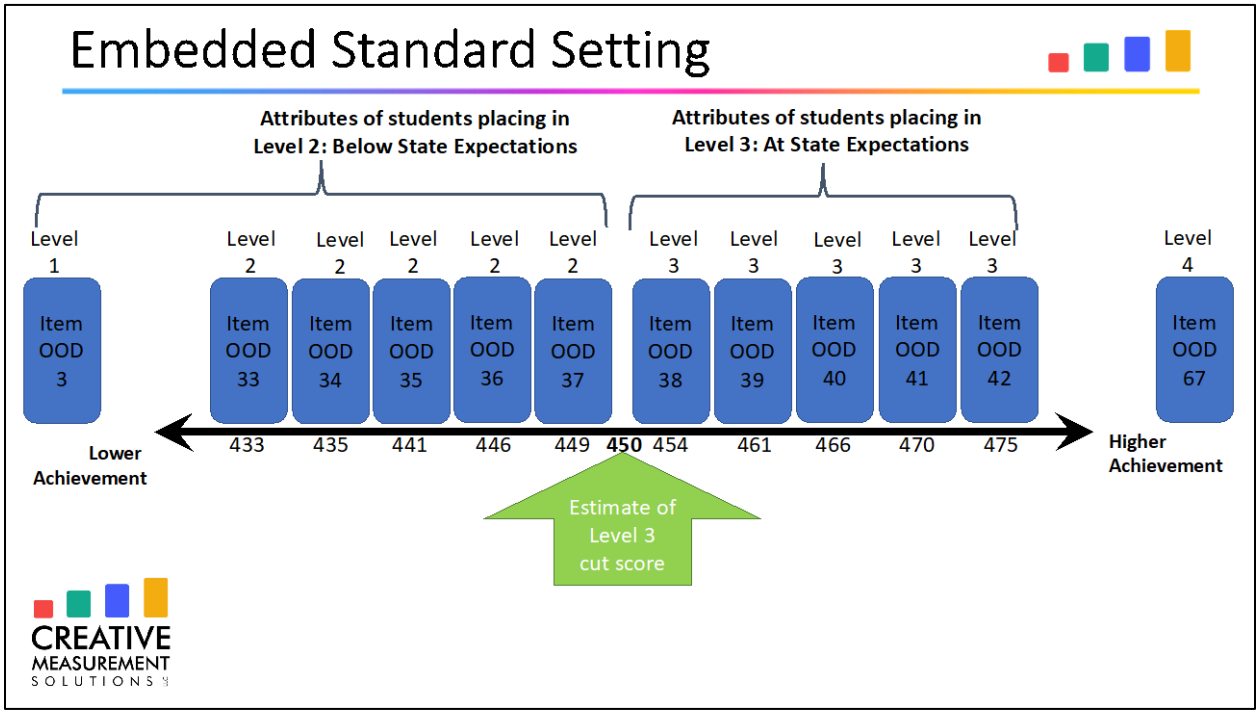
CCSS Reading

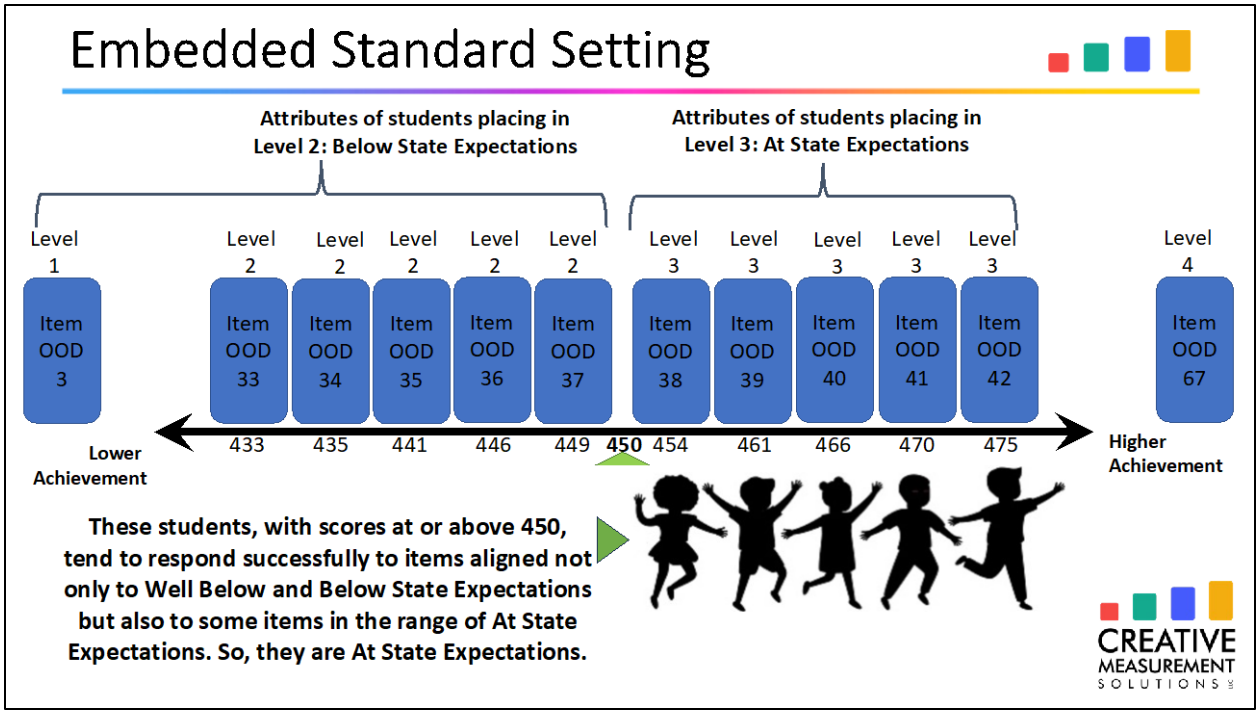
2nd Year of High School Range ALDs

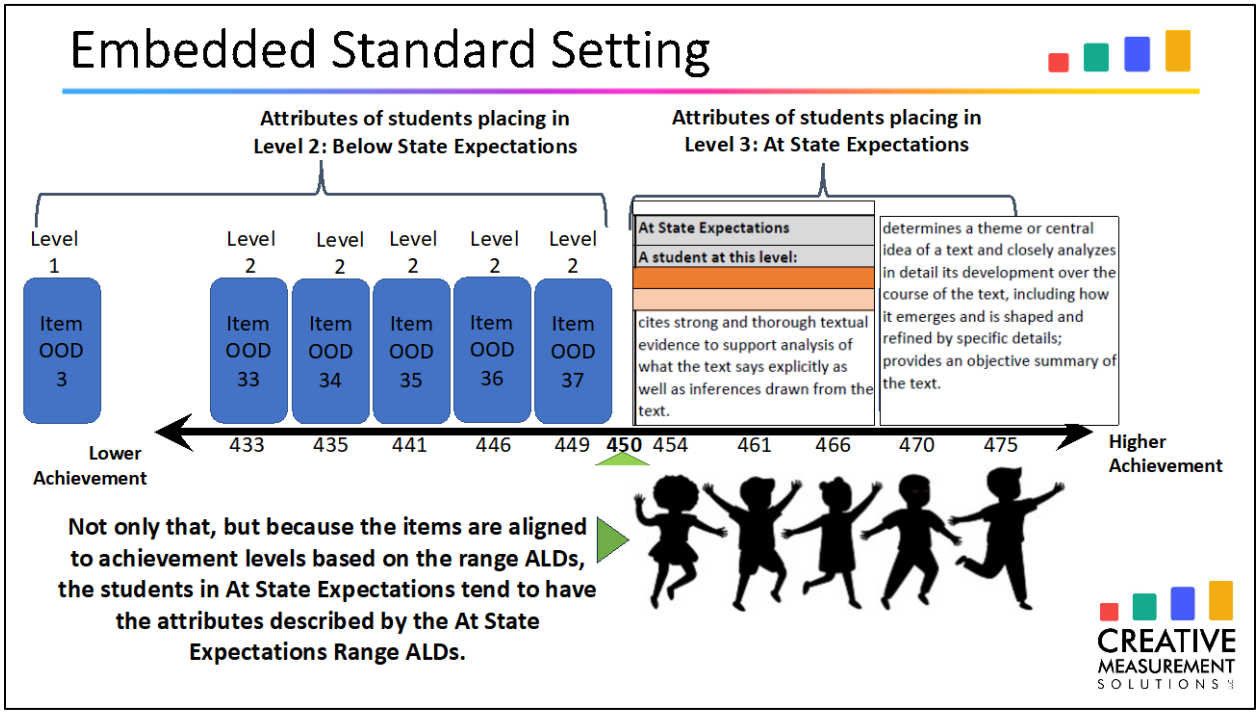
		Well Below State Expectations	Below State Expectations	At State Expectations	Above State Expectations
[illegible]dard and Text		A student at this level:	A student at this level:	A student at this level:	A student at this level:
[illegible]tandards for Literature					
[illegible]and Details					
[illegible]Literacy.RL.9-	Cite strong and thorough textual evidence to support analysis of what the text says explicitly as well as inferences drawn from the text.	cites textual evidence that supports an explanation of what the text says explicitly.	cites textual evidence that supports an explanation of what the text says explicitly as well as basic inferences drawn from the text.	cites strong and thorough textual evidence to support analysis of what the text says explicitly as well as inferences drawn from the text.	cites strong and thorough textual evidence that supports an analysis sophisticated inferences drawn over the course of the text.
[illegible]Literacy.RL.9-	Determine a theme or central idea of a text and analyze in detail its development over the course of the text, including how it emerges and is shaped and refined by specific details; provide an objective summary of the text	identifies a theme or central idea of a text when it is supported by explicit details; identifies details related to the theme or central idea; sequences major events from the text.	identifies a simple theme or central idea of a text; describes how details support development of the theme or central idea over the course of the text; provides a basic summary of the text.	determines a theme or central idea of a text and closely analyzes in detail its development over the course of the text, including how it emerges and is shaped and refined by specific details; provides an objective summary of the text.	analyzes complex theme(s) or central idea(s) and how they develop over the course of the text, including how specific details shape and contribute to the text; provides a concise and objective summary of the text.
[illegible]Literacy[illegible]9-	[illegible]lyze [illegible] com[illegible] characters (e.g., those with [illegible] motivations) develop over the [illegible]ct with other characters, and advance the plot or develop the theme.	identifies how main characters develop over the course of the text and interact with other characters to advance the plot.	identifies how main characters develop over the course of the text (e.g., motivations) and interact with other characters to advance the plot or develop the theme.	analyzes how complex characters (e.g., those with multiple or conflicting motivations) develop over the course of a text, interact with other characters, and advance the plot or develop the theme.	analyzes how effectively complex characters (e.g., those with multiple or conflicting motivations) develop over the course of a text, interact with other characters, and subtly advance the plot or develop complex theme(s).

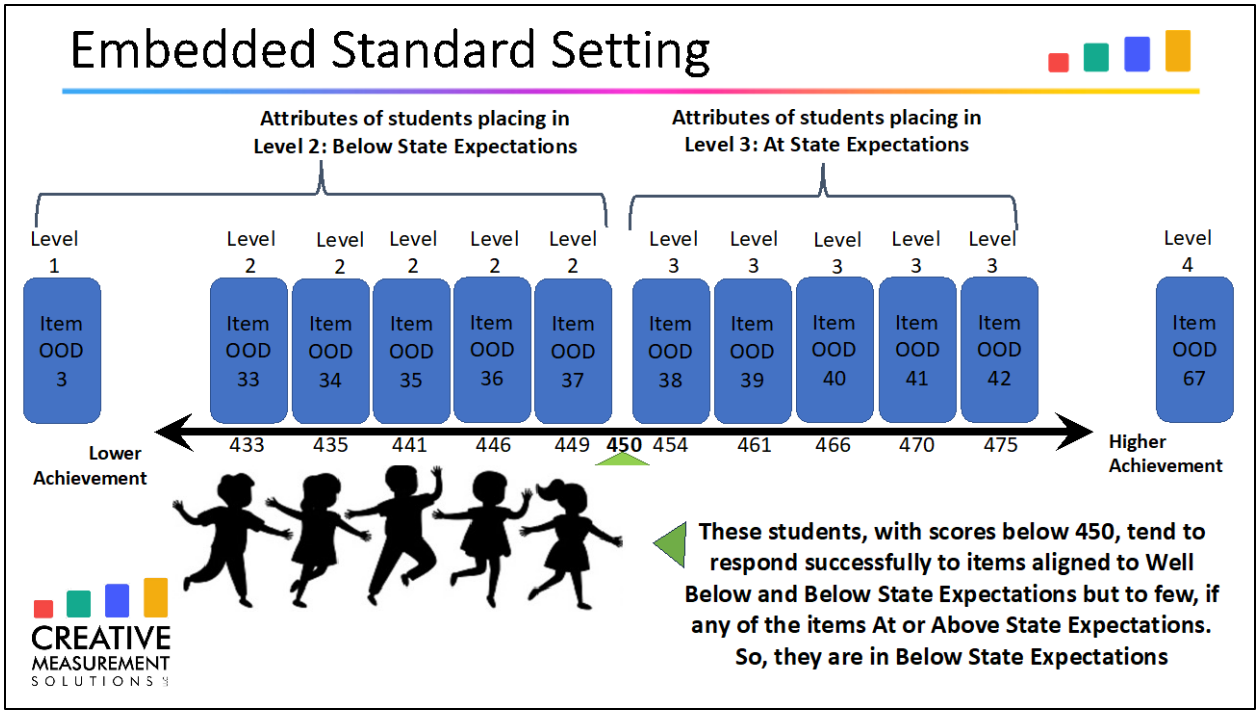
CREATIVE MEASUREMENT SOLUTIONS

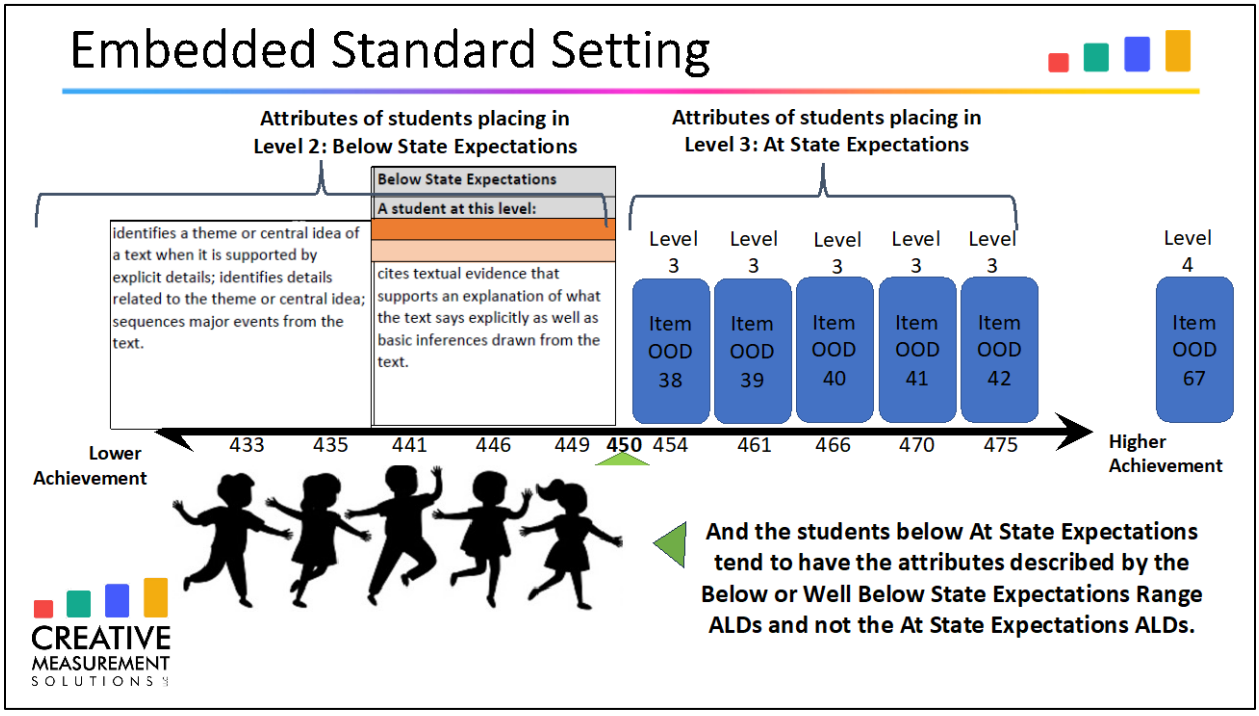

Embedded Standard Setting
Attributes of students placing in Level 2: Below State Expectations
Attributes of students placing in Level 3: At State Expectations
Level 1
Item OOD 3
Level 2
Item OOD 33
Level 2
Item OOD 34
Level 2
Item OOD 35
Level 2
Item OOD 36
Level 2
Item OOD 37
Level 3
Item OOD 38
Level 3
Item OOD 39
Level 3
Item OOD 40
Level 3
Item OOD 41
Level 3
Item OOD 42
Level 4
Item OOD 67
433 435 441 446 449 450 454 461 466 470 475
Lower Achievement
Higher Achievement
Point 1
Point 2
Because they are ordered by difficulty, the 2-point items are represented twice—once for each score point. This is because it should be easier to achieve a 1 on these items than a 2.
CREATIVE MEASUREMENT SOLUTIONS


Embedded Standard Setting
Attributes of students placing in Level 2: Below State Expectations
Attributes of students placing in Level 3: At State Expectations
Level 1
Item OOD 3
Level 2
Item OOD 33
Level 2
Item OOD 34
Level 2
Item OOD 35
Level 2
Item OOD 36
Level 2
Item OOD 37
Level 3
Item OOD 38
Level 3
Item OOD 39
Level 3
Item OOD 40
Level 3
Item OOD 41
Level 3
Item OOD 42
Level 4
Item OOD 67
433 435 441 446 449 450 454 461 466 470 475
Lower Achievement
Higher Achievement
Estimate of Level 3 cut score
CREATIVE MEASUREMENT SOLUTIONS

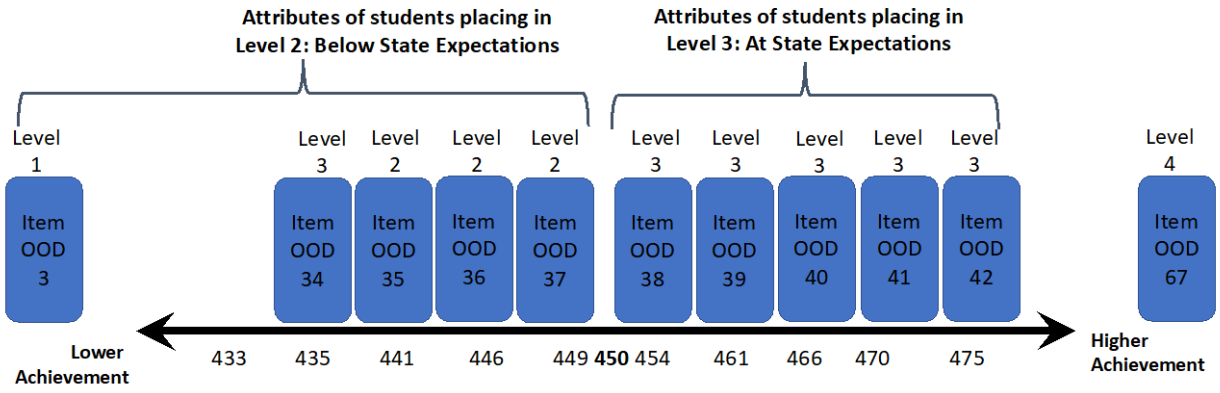

Embedded Standard Setting
Attributes of students placing in Level 2: Below State Expectations
Attributes of students placing in Level 3: At State Expectations
Level 1
Item OOD 3
Level 2
Item OOD 33
Level 2
Item OOD 34
Level 2
Item OOD 35
Level 2
Item OOD 36
Level 2
Item OOD 37
Level 3
Item OOD 38
Level 3
Item OOD 39
Level 3
Item OOD 40
Level 3
Item OOD 41
Level 3
Item OOD 42
Level 4
Item OOD 67
433 435 441 446 449 450 454 461 466 470 475
Lower Achievement
Higher Achievement
These students, with scores at or above 450, tend to respond successfully to items aligned not only to Well Below and Below State Expectations but also to some items in the range of At State Expectations. So, they are At State Expectations.
CREATIVE MEASUREMENT SOLUTIONS


Embedded Standard Setting
Attributes of students placing in Level 2: Below State Expectations
Attributes of students placing in Level 3: At State Expectations
Level 1
Item OOD 3
Level 2
Item OOD 33
Level 2
Item OOD 34
Level 2
Item OOD 35
Level 2
Item OOD 36
Level 2
Item OOD 37
At State Expectations
A student at this level:
cites strong and thorough textual evidence to support analysis of what the text says explicitly as well as inferences drawn from the text.
determines a theme or central idea of a text and closely analyzes in detail its development over the course of the text, including how it emerges and is shaped and refined by specific details; provides an objective summary of the text.
433 435 441 446 449 450 454 461 466 470 475
Lower Achievement
Higher Achievement
Not only that, but because the items are aligned to achievement levels based on the range ALDs, the students in At State Expectations tend to have the attributes described by the At State Expectations Range ALDs.
CREATIVE MEASUREMENT SOLUTIONS


Embedded Standard Setting
Attributes of students placing in Level 2: Below State Expectations
Attributes of students placing in Level 3: At State Expectations
Level 1
Item OOD 3
Level 2
Item OOD 33
Level 2
Item OOD 34
Level 2
Item OOD 35
Level 2
Item OOD 36
Level 2
Item OOD 37
Level 3
Item OOD 38
Level 3
Item OOD 39
Level 3
Item OOD 40
Level 3
Item OOD 41
Level 3
Item OOD 42
Level 4
Item OOD 67
433 435 441 446 449 450 454 461 466 470 475
Lower Achievement
Higher Achievement
These students, with scores below 450, tend to respond successfully to items aligned to Well Below and Below State Expectations but to few, if any of the items At or Above State Expectations. So, they are in Below State Expectations
CREATIVE MEASUREMENT SOLUTIONS


Embedded Standard Setting
Attributes of students placing in Level 2: Below State Expectations
Attributes of students placing in Level 3: At State Expectations
Below State Expectations
A student at this level:
identifies a theme or central idea of a text when it is supported by explicit details; identifies details related to the theme or central idea; sequences major events from the text.
cites textual evidence that supports an explanation of what the text says explicitly as well as basic inferences drawn from the text.
Level 3
Item OOD 38
Level 3
Item OOD 39
Level 3
Item OOD 40
Level 3
Item OOD 41
Level 3
Item OOD 42
Level 4
Item OOD 67
433 435 441 446 449 450 454 461 466 470 475
Lower Achievement
Higher Achievement
And the students below At State Expectations tend to have the attributes described by the Below or Well Below State Expectations Range ALDs and not the At State Expectations ALDs.
CREATIVE MEASUREMENT SOLUTIONS

Embedded Standard Setting



Note

- It is not unusual for item difficulty to be different than anticipated.
- Even experienced item writers develop items with difficulties that are other than they expected.
- By engaging in activities like those you will participate in today, we can better understand the reasons for these items unexpected difficulties and in doing so, support score interpretation and the standard setting process.

CREATIVE MEASUREMENT SOLUTIONS

Item-ALD Alignment: Consistent Items

- You have been aligning items to standards and achievement levels
- For each item (or score point), a "consensus" alignment was established – the level selected by the most panelists.
- When consensus alignment agrees with the empirical level—the level indicated by data—we consider that item to be consistent.
  - These are items with difficulties estimated from Maine student data that are consistent with their expected difficulties, given the consensus alignment.
  - The way these items work empirically is *consistent* with your judgments of how they should work.

CREATIVE MEASUREMENT SOLUTIONS

Item-ALD Alignment: Inconsistent Items

- For some items, consensus alignments do not agree with their empirical levels indicated by data
  - They are not of the difficulty expected given the consensus alignment.
  - How the items work empirically is *inconsistent* with your expectations of how they should be working.
  - Example
    - An item aligned to Above State Expectations should be harder than items At or Below State Expectations.
    - If it is not, it is inconsistent with expectations.

CREATIVE MEASUREMENT SOLUTIONS

Reasons Items May Be Inconsistent

- Items may be inconsistent for several reasons:
  - Upon closer inspection, relevant item attributes may better reflect the attributes of the ALDs associated with the empirically identified achievement level.
  - Language in the ALDs' evidence statements make it challenging to establish a clear best fit
  - Opportunity to Learn (OTL) may make some items easier or harder than expected or some aspect of the item is affecting the difficulty in unintended ways

CREATIVE MEASUREMENT SOLUTIONS

Resolving Inconsistencies

- By discussing and comparing the ALD evidence statements associated with alignments, inconsistencies can be resolved by:
  - identifying items where the best fitting evidence statement is associated with the empirical level, not the aligned level
  - Accepting the empirical alignment because the reasons for the inconsistency reflect the effects of Opportunity to Learn or unintended difficulty factors within the item.
  - refining ALD evidence statements to more precisely reflect attributes that distinguish between achievement levels.
  - You can also maintain your consensus alignment even though it disagrees with the data

CREATIVE MEASUREMENT SOLUTIONS

Up Next

- Your Grade Group Facilitator will guide you, one by one, through each item we identified as inconsistent. For each inconsistent item you will:
  - …examine the ALD evidence statements associated with your panel's consensus alignment(s) and those of the empirical level,
  - discuss which evidence statement is the best match to the content characteristics or measurement attributes of the item, and consider
    - maintaining or adjusting your alignment based on your new perspective following discussion, or
    - recommending refinements to the ALD evidence statements necessary for them to better support the consensus alignments.

CREATIVE MEASUREMENT SOLUTIONS

Review of Easiest and/or Hardest Items

- We observed some grades with few items with Consensus Alignments at the Well Below and/or Above level.
- This is not a problem with anything you have done but we might have expected to see some
  - Well Below State Expectations items among the easiest items
  - Above State Expectations items among the hardest items
- This is because you reviewed three carefully selected forms. We selected:
  - one adaptive form that is targeted to a typical low ability student—one at about the 16th percentile,
  - one adaptive form that is targeted to an average ability student, at about the 50th percentile, and
  - one adaptive form that is targeted to a typical high ability student, at about the 84th percentile
- Because of the systematic selection of items for each of the three forms, we expected to see a diverse range of alignments to the ALDs. We expected to see some items aligning to each of the levels.
- Because we did not find many, if any, Well Below and/or Above items, it would be useful to have you take one last look at some of the easiest and/or hardest items on the forms to confirm or modify your alignments.

CREATIVE MEASUREMENT SOLUTIONS

Review Considerations Based On Order of Difficulty

Panelist Information			Item Information						Standard Alignment	Final ALD Resolution		Consistency Information				
Panelist ID	Subject Area	Grade	Item Number	Panelist Row	Form Number	Sequence Number	Score Point	Max Points	Primary Standard	Final ALD Level - Item/Score Point	Final ALD Comments	Order of Difficulty	Empirical ALD Alignment Level	Consensus ALD Alignment Level	Consistency Status	R&R Discussion Status
RD6FAC	Reading	6	VR699867_1	3	1	1	1	2	RL.6.2			1	Well Below	Below	Inconsistent	
RD6FAC	Reading	6	VR699867_1	57	3	1	1	2	0			1	Well Below	Repeat		
RD6FAC	Reading	6	VR699883_1	8	1	6	1	2	RL.6.4			3	Well Below	At	Inconsistent	
RD6FAC	Reading	6	VR699883_1	62	3	6	1	2	0			3	Well Below	Repeat		
RD6FAC	Reading	6	VR699872_1	4	1	2	1	2	RL.6.3			5	Below	Below	Review	
RD6FAC	Reading	6	VR699878_1	5	1	3	1	2	RL.6.9			6	Below	Below	Review	
RD6FAC	Reading	6	VR703548	25	1	23	1	1	RL.6.1			7	Below	At	Inconsistent	
RD6FAC	Reading	6	VR702320	16	1	14	1	1	L.6.4			8	Below	Below	Consistent	
RD6FAC	Reading	6	VR703547	24	1	22	1	1	RL.6.3			9	Below	Below	Consistent	
RD6FAC	Reading	6	VR703549	26	1	24	1	1	RL.6.3			10	Below	At	Inconsistent	
RD6FAC	Reading	6	VR699881	7	1	5	1	1	RL.6.4			11	Below	Below	Consistent	

CREATIVE MEASUREMENT SOLUTIONS

Review Considerations Based On Order of Difficulty

Panelist Information			Item Information						Standard Alignment	Final ALD Resolution		Consistency Information				
Panelist ID	Subject Area	Grade	Item Number	Panelist Row	Form Number	Sequence Number	Score Point	Max Points	Primary Standard	Final ALD Level - Item/Score Point	Final ALD Comments	Order of Difficulty	Empirical ALD Alignment Level	Consensus ALD Alignment Level	Consistency Status	R&R Discussion Status
RD6FAC	Reading	6	VR699869_2	58	3	2	2	2	RL.6.2			80	At	At	Consistent	
RD6FAC	Reading	6	VR703737	77	3	28	1	1	L.6.4			81	At	At	Consistent	
RD6FAC	Reading	6	VR700289	35	2	6	1	1	RL.6.5			82	At	At	Consistent	
RD6FAC	Reading	6	VR700287	34	2	5	1	1	RL.6.3			83	At	At	Consistent	
RD6FAC	Reading	6	VR700282	30	2	1	1	1	RL.6.6			84	At	At	Consistent	
RD6FAC	Reading	6	VR703742	82	3	40	1	1	RI.6.5			85	At	At	Consistent	
RD6FAC	Reading	6	VR700947	47	2	25	1	1	RI.6.8			86	At	At	Consistent	
RD6FAC	Reading	6	VR703656	75	3	26	1	1	L.6.4			87	At	At	Consistent	
RD6FAC	Reading	6	VR699878_2	5	1	3	2	2	RL.6.9			88	At	At	Consistent	
RD6FAC	Reading	6	VR699875	61	3	5	1	1	RL.6.9			89	At	At	Review	
RD6FAC	Reading	6	VR701286	64	3	8	1	1	RI.6.3			90	At	At	Review	
RD6FAC	Reading	6	VR701283	63	3	7	1	1	RI.6.1			91	At	At	Review	
RD6FAC	Reading	6	VR701288	66	3	10	1	1	RI.6.5			92	At	At	Review	
RD6FAC	Reading	6	VR703739	79	3	37	1	1	RI.6.7			93	At	At	Review	

CREATIVE MEASUREMENT SOLUTIONS

Rating Form: Days 1 & 2

Panelist Information			Item Information			Standard Alignment				DOK Rating		ALD Alignment		
Panelist ID	Subject Area	Grade	Item Number	Form Number	Sequence Number	Primary Standard	Secondary Standard	Domain-only Match	Standard Match Comments	Item DOK Rating	Item DOK Comments	Item ALD Level - Full Credit	Item ALD Level - 1 Point Correct	ALD Level Comments
RD3FAC	Reading	3	VR699667	1	2									
RD3FAC	Reading	3	VR699668	1	3									
RD3FAC	Reading	3	VR699669	1	4									
RD3FAC	Reading	3	VR699670	1	5									
RD3FAC	Reading	3	VR699671	1	6									
RD3FAC	Reading	3	VR701051	1	7									
RD3FAC	Reading	3	VR701052	1	8									
RD3FAC	Reading	3	VR701053	1	9									
RD3FAC	Reading	3	VR701054	1	17									
RD3FAC	Reading	3	VR701055	1	18									
RD3FAC	Reading	3	VR701056	1	19									
RD3FAC	Reading	3	VR702674	1	20									
RD3FAC	Reading	3	VR702675	1	21									
RD3FAC	Reading	3	VR702676	1	22									
RD3FAC	Reading	3	VR702678	1	23									
RD3FAC	Reading	3	VR702679	1	24									
RD3FAC	Reading	3	VR702683	1	25									
RD3FAC	Reading	3	VR702685	1	26									
RD3FAC	Reading	3	VR702686	1	27									
RD3FAC	Reading	3	VR702689	1	28									
RD3FAC	Reading	3	VR702746	1	37									
RD3FAC	Reading	3	VR702747	1	38									
RD3FAC	Reading	3	VR702748	1	39									
RD3FAC	Reading	3	VR702749	1	40									
RD3FAC	Reading	3	VR702750	1	41									
RD3FAC	Reading	3	VR702751	1	42									
RD3FAC	Reading	3	VR699686	2	1									
RD3FAC	Reading	3	VR699688	2	2									
RD3FAC	Reading	3	VR699689	2	3									
RD3FAC	Reading	3	VR699692	2	4									
RD3FAC	Reading	3	VR701532	2	5									
RD3FAC	Reading	3	VR701535	2	6									
RD3FAC	Reading	3	VR701536	2	7									
RD3FAC	Reading	3	VR701537	2	8									
RD3FAC	Reading	3	VR701538	2	9									
RD3FAC	Reading	3	VR701540	2	10									
RD3FAC	Reading	3	VR701586	2	11									

Rating Form: Day 3

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Panelist Information			Item Information			Standard Alignment				DOK Rating		ALD Alignment			Final ALD Resolution		
2	Panelist ID	Subject Area	Grade	Item Number	Form Number	Sequence Number	Primary Standard	Secondary Standard	Domain-only Match	Standard Match Comments	Item DOK Rating	Item DOK Comments	Item ALD Level - Full Credit	Item ALD Level - 1 Point Correct	ALD Level Comments	Final ALD Level - Full Credit	Final ALD Level - 1 Point Correct	Final ALD Comments
3	RD3FAC	Reading	3	VR699666	1	1												
4	RD3FAC	Reading	3	VR699667	1	2												
5	RD3FAC	Reading	3	VR699668	1	3												
6	RD3FAC	Reading	3	VR699669	1	4												
7	RD3FAC	Reading	3	VR699670	1	5												
8	RD3FAC	Reading	3	VR699671	1	6												
9	RD3FAC	Reading	3	VR701051	1	7												
10	RD3FAC	Reading	3	VR701052	1	8												
11	RD3FAC	Reading	3	VR701053	1	9												
12	RD3FAC	Reading	3	VR701054	1	17												
13	RD3FAC	Reading	3	VR701055	1	18												
14	RD3FAC	Reading	3	VR701056	1	19												
15	RD3FAC	Reading	3	VR702674	1	20												
16	RD3FAC	Reading	3	VR702675	1	21												
17	RD3FAC	Reading	3	VR702676	1	22												
18	RD3FAC	Reading	3	VR702678	1	23												
19	RD3FAC	Reading	3	VR702679	1	24												
20	RD3FAC	Reading	3	VR702683	1	25												
21	RD3FAC	Reading	3	VR702685	1	26												
22	RD3FAC	Reading	3	VR702686	1	27												
23	RD3FAC	Reading	3	VR702689	1	28												
24	RD3FAC	Reading	3	VR702746	1	37												
25	RD3FAC	Reading	3	VR702747	1	38												
26	RD3FAC	Reading	3	VR702748	1	39												
27	RD3FAC	Reading	3	VR702749	1	40												
28	RD3FAC	Reading	3	VR702750	1	41												
29	RD3FAC	Reading	3	VR702751	1	42												
30	RD3FAC	Reading	3	VR699686	2	1												
31	RD3FAC	Reading	3	VR699688	2	2												
32	RD3FAC	Reading	3	VR699689	2	3												
33	RD3FAC	Reading	3	VR699692	2	4												
34	RD3FAC	Reading	3	VR701532	2	5												
35	RD3FAC	Reading	3	VR701535	2	6												
36	RD3FAC	Reading	3	VR701536	2	7												
37	RD3FAC	Reading	3	VR701537	2	8												
38	RD3FAC	Reading	3	VR701538	2	9												
39	RD3FAC	Reading	3	VR701540	2	10												
40	RD3FAC	Reading	3	VR701586	2	11												
41	RD3FAC	Reading	3	VR701587	2	12												
42	RD3FAC	Reading	3	VR701588	2	13												
43	RD3FAC	Reading	3	VR701589	2	14												
44	RD3FAC	Reading	3	VR701591	2	23												
45	RD3FAC	Reading	3	VR701592	2	24												
46	RD3FAC	Reading	3	VR701593	2	25												
47	RD3FAC	Reading	3	VR701594	2	26												
48	RD3FAC	Reading	3	VR702671	2	27												

Appendix D: Inconsistent Item Review & Resolution Facilitator Guide and Script

Notes for Facilitator on the Facilitator Review and Resolution (R&R) Rating Form

For the Review and Resolution activity (R&R), facilitators will be working from the Facilitator R&R Rating Form—a slight modification of the rating form that was used for the previous Alignment Evaluation activity. The most notable differences are:

1. Each score point for the 2-point items is in its own row. So rather than one 2-point item in a single row, each score point is in its own row. This makes sense if you think about aligning items to the ALDs because while both score points *could* be aligned to the same level, if is also possible (and in some cases likely) that each score point will be aligned to different levels (with the score of 1 sometimes aligned to a lower level, and never to a higher level, than the score of 2).

We can think of each score point of 2-point items as separate items where if a student gets a 1 on an item, they meet the requirements to get the 1st point correct but not the 2nd point and if they get a 2 on the item, they meet the full credit requirements.

So, each item and score point is on a separate row. And the 2 points of 2-point items are probably not in adjacent rows because of the next notable difference.

2. All items and score points are ordered by difficulty. The ESS analysis establishes cut scores based on the empirical ordering of items and identifying locations for cuts that minimize inconsistency between panelists' consensus alignments and empirical locations of the items. For the Review and Resolution activity (scheduled for Day 3) following the Alignment Evaluation (scheduled for Days 1 and 2), the activity focuses, in turn, on each empirically identified achievement level, specifically on the items found within each whose panelists' consensus Item-ALD alignments are inconsistent with the empirical levels.

3. The form you will use for this R&R activity includes a section that highlights the inconsistency between the educators' Item-ALD alignments and the Empirical alignments. This section will inform the Facilitator as to which items require review and resolution for the R&R activity.

The form is divided into sections in a way similar but not identical to the forms used in the Alignment Evaluation. The following describes each of the sections, noting important differences and how the Facilitator will interact with them.

1. **Panelist Information**

	A	B	C
1	Panelist Information		
2	Panelist ID	Subject Area	Grade
3	RD3FAC	Reading	3
4	RD3FAC	Reading	3
5	RD3FAC	Reading	3
6	RD3FAC	Reading	3
7	RD3FAC	Reading	3
8	RD3FAC	Reading	3
9	RD3FAC	Reading	3
10	RD3FAC	Reading	3
11	RD3FAC	Reading	3

Figure 1. Panelist Information

Panelist information (Figure 1) is an artifact of the previous activity's form (which is still used by panelists in its original form during the R&R) and not relevant to the current R&R activity as we are focusing on group activity and not individual activity. You can ignore this information.
Note that some rows/items are grayed out. The grayed-out rows/items are not subject to R&R as they are items with consistent educator and empirical alignments. This will be explained in the description of the Consistency Information section of the form.

2. **Item Information**

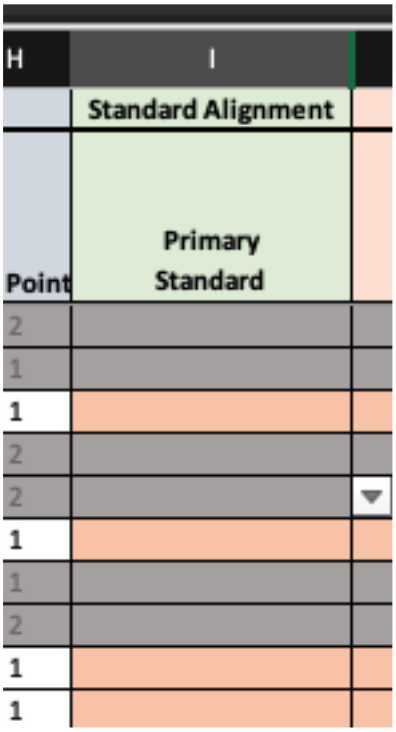
D	E	F	G	H	I
Item Information					
Item Number	Panelist Row	Form Number	Sequence Number	Score Point	Max Points
VR699667_1	75	1	2	1	2
VR701538	82	2	9	1	1
VR701587	32	2	12	1	1
VR702675_1	4	1	21	1	2
VR702675_1	82	3	23	1	2
VR701594	10	2	26	1	1
VR701588	56	2	13	1	1
VR699670_1	11	1	5	1	2
VR701053	71	1	9	1	1
VR701589		2	14	1	1

Figure 2. Item Information

Item information (Figure 2) captures the information required to identify the item (or score point) under review. This section contains the same information presented in the Alignment Evaluation, with additional columns indicating the maximum obtainable points (1 or 2) and which score point is associated with the row. When the Maximum Points is 1, it indicates a dichotomous item (0-1) that appears only once. When the Maximum Points is 2, it indicates a 2-point item (0-2) that will appear twice. The first time the Score Point column will indicate “1” and the second time the Score Point column will indicate “2.”

You may notice that the form number and sequence number are not in the same order as you viewed previously because items are presented by order of difficulty for the R&R activity. This is captured in the Consistency Information section.

3. **Standard Alignment**



H	I
	Standard Alignment
Point	Primary Standard
2	
1	
1	
2	
2	
1	
1	
2	
1	
1	

Figure 3. Standard Alignment

The Standard Alignment column (Figure 3) will be filled in by copying and pasting from Facilitator forms after Day 2.

4. **Final ALD Resolution**

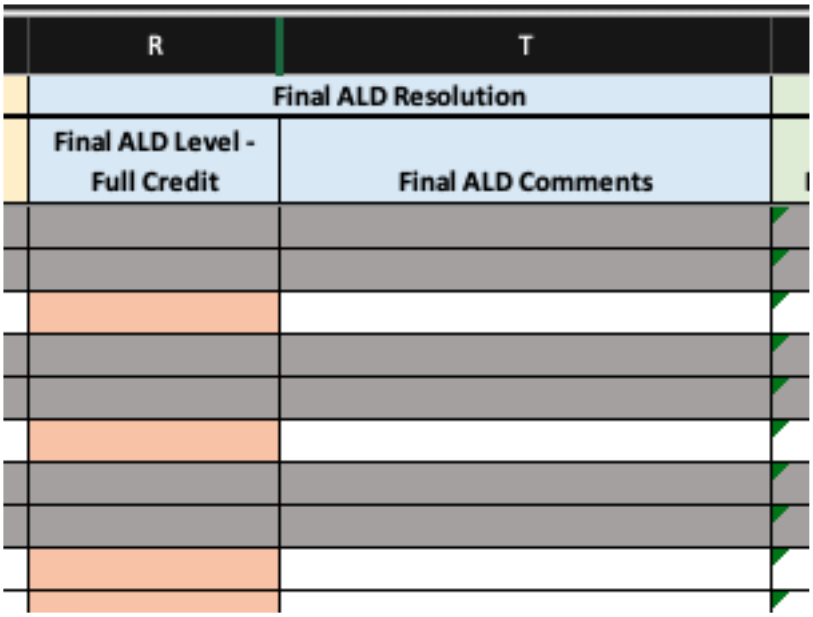


Figure 4. Final ALD Resolution

This section (Figure 4) is included because it is where, **on panelists' forms**, each panelist will make a final alignment judgment based on their updated understanding after receiving new information (consistency status with the Empirical Alignment) and following discussion during the R&R activity. You, the Facilitator, can capture the consensus level informally observed during the R&R if you observe it, but you need not do so as final data will be captured by analyzing panelists' individual alignments on the rating forms after the workshop. Facilitators may make notes in the Comments column if they observe something during discussion that they believe is worth capturing.

5. **Consistency Information**

G	H	I	J
Consistency Information			
Order of Difficulty	Empirical ALD Alignment Level	Consensus ALD Alignment Level	Consistency Status
6	Well Below	Well Below	Consistent
7	Well Below	Well Below	Consistent
9	Well Below	Below	Inconsistent
10	Well Below	Well Below	Consistent
10	Well Below	Well Below	Consistent
12	Well Below	At	Inconsistent
14	Well Below	Well Below	Consistent
15	Well Below	Well Below	Consistent
16	Well Below	At	Inconsistent

Figure 5. Consistency Information

The Consistency Information section (Figure 5) is the essential information for guiding the R&R activity. Columns include:

- **Order of Difficulty.** The position of the item based on ordering by empirical difficulty. This is the order in which the items will be reviewed—easiest inconsistent items first (remember, we are only reviewing items during R&R such that the panelists' consensus Item-ALD alignment ≠ the Empirical Alignment. There may be duplicates as some items appear on multiple forms. Each inconsistent item should be discussed only once.
- **Empirical ALD Alignment Level.** This is the Empirical Alignment. It is based on the cut scores estimated from the panelists' consensus alignments from the Alignment Evaluation activity. The cut scores are estimated by minimizing the number of inconsistent items (items whose empirical alignments disagree with panelists' Item-ALD alignments).
- **Consensus ALD Alignment Level.** This is the "consensus" Item-ALD alignment estimated by identifying the Item-ALD alignment of the majority (or plurality) of panelists during the Alignment Evaluation activity.
- **Consistency Status**: This reflects whether the Empirical Alignment and the panelists' consensus Item-ALD alignment agree (Consistent) or disagree (Inconsistent) for the given item. Items that are "Consistent" are in grayed out rows and are not discussed during R&R. "Inconsistent" items are reviewed in the order they appear on the Facilitator R&R Rating Form.

6. **R&R Discussion Status**

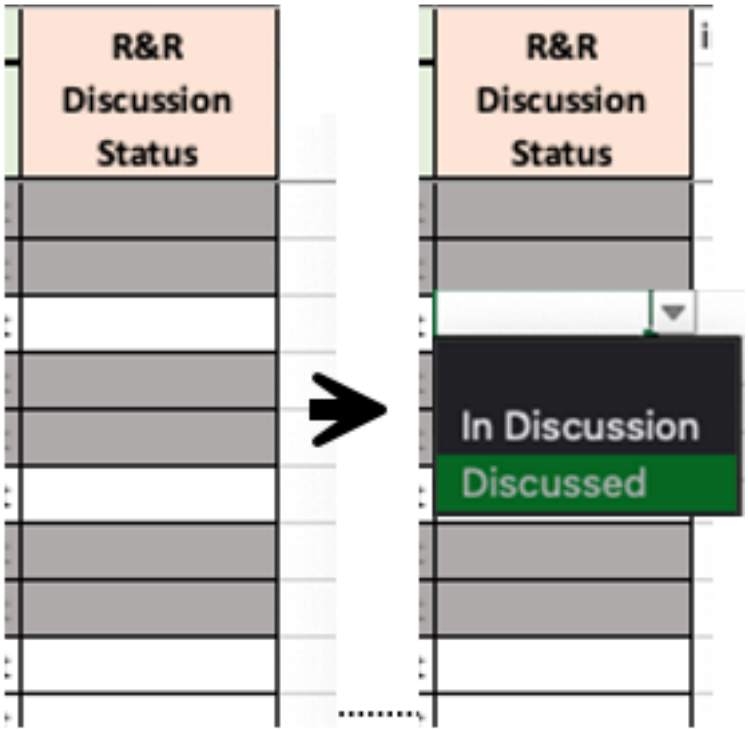


Figure 6. R&R Discussion Status

The R&R Discussion Status column (Figure 6) is included as a convenience for facilitators so they can track which items they have already discussed and track progress during the activity. By setting the status to "Discussed," the row will be grayed out (with a slightly different shading scheme than used for the "Consistent" items) to make it easier for the facilitator to track where they are in the R&R activity.

Introductory Facilitator Script

In the opening session today, we learned that for all items or score points, our colleagues at CMS estimated a "consensus alignment," the alignment selected by a majority or at least a plurality of you.

For many of these items or score points, the consensus alignment agreed with the Empirical Alignment, meaning that these items had difficulties estimated from Maine student data that were consistent with your consensus alignment. These items are considered consistent because the items are working as desired.

CMS also identified items for which your consensus alignments do not agree with the Empirical Alignment. In other words, these items are more (or less) difficult based on Maine student data, than would be expected based on your consensus alignments. These items are Inconsistent

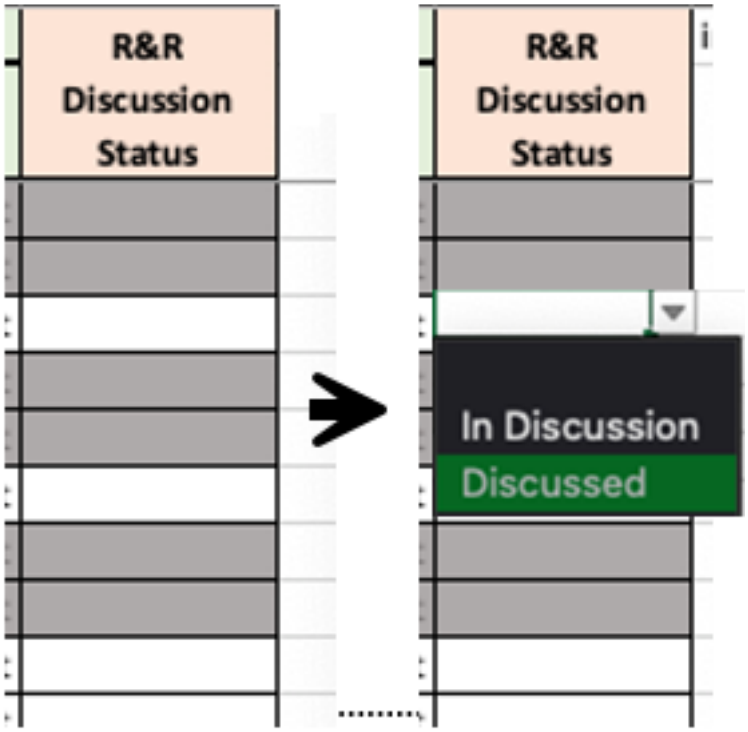
items, and we will be reviewing these items today, and hopefully resolve the reason for the inconsistency as described during the opening session.

We will follow the same basic process for each of the inconsistent items. You have aligned each item to a content standard and achievement level. Each content standard has evidence statements associated with each achievement level. We will look at the evidence statements associated with (a) the panel's consensus alignment and (b) the Empirical Alignment and consider which evidence statement is a better fit to the content characteristics and other attributes of the item.

Depending on your conclusion and the quality of the fit, as a last resort, and better, when multiple items suggest it, additional steps to refine the ALDs may be recommended to achieve the best resolution. There is no specific desired resolution. The best resolution is the one that reflects your judgments and that best captures your expertise.

Facilitator script for each item

1. Identify the first item for review. Mark R&R Discussion Status as "In Discussion."



2. Identify the Item Number, Panelist Row, and Score Point from your Facilitator form and ask panelists to confirm they are working with the correct item and row on their forms.

"The next item is "Item Number." You will find it on your form in "Panelist Row."

We are looking at “Score Point” (1 or 2 when item number is a 2-point item, that is indicated by an underscore attached to the item number and also in the column “Score Point”).

Confirm all panelists are working with the same item and that they understand the score point being considered. Their alignment must be associated with the given score point.

D	E	F	G	H	I
Item Information					
Item Number	Panelist Row	Form Number	Sequence Number	Score Point	Max Points
VR699667_1	75	1	2	1	2
VR701538	82	2	9	1	1
VR701587	**32**	**2**	**12**	**1**	**1**
VR702675_1	4	1	21	1	2
VR702675_1	82	3	23	1	2
VR701594	**10**	**2**	**26**	**1**	**1**
VR701588	56	2	13	1	1
VR699670_1	11	1	5	1	2
VR701053	**71**	**1**	**9**	**1**	**1**
VR701589		**2**	**14**	**1**	**1**

3. Read the item.

4. Identify
    a. the Primary Standard alignment and the Consensus ALD alignment Level from your form
    b. The Empirical Alignment from the Facilitator R&R Rating Form.

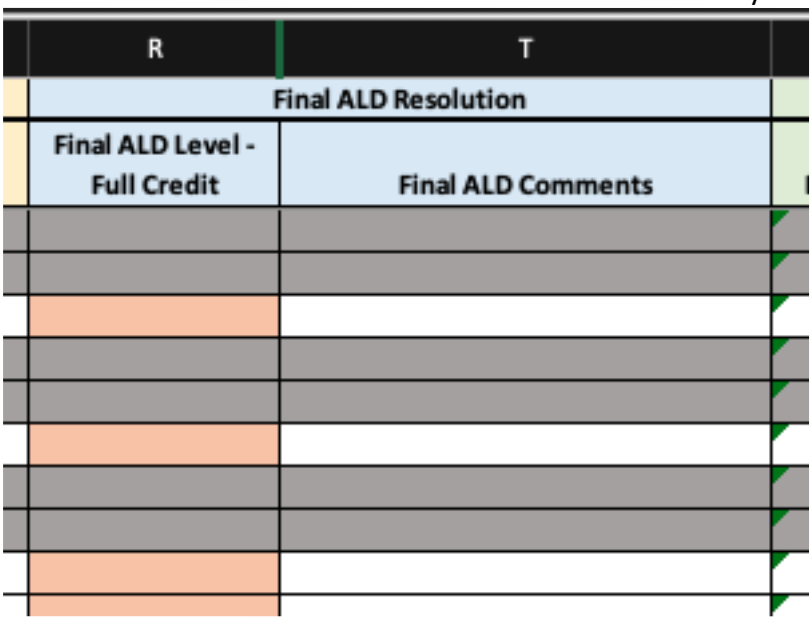
5. Read (or have a panelist read) the evidence statements associated with the empirical and consensus alignments. These are the statements in the ALDs associated with the Primary Standard Alignment and the Consensus and Empirical ALD Alignments.

*If the consensus alignment is a tie (unlikely, but possible), read both tie consensus alignment evidence statements and the empirical alignment evidence statement. Individual panelists may disagree with the consensus alignment and in that case, they may want to read and argue for a different evidence statement.

6. Encourage discussion of the item/score point's characteristics (knowledge, skills, and other content-based item attributes) and how they align with the various identified evidence statements. After discussion:

a. Panelists enter an updated ALD alignment in their rating forms in the relevant Final ALD Resolution Column. For single-point items and for Full Credit on two-point items, this will be the column labeled "Final ALD Level – Full Credit." For the first score point on the 2-point items, this will be the column labeled "Final ALD Level – 1 Point Correct." Panelists may choose to change or maintain their previous alignments **but must enter a rating in the Final ALD Resolution section**.

b. Panelists may, as a group, agree to suggest refinements to the ALD evidence statements to better support their updated alignments. If so, the Facilitator should use tracked changes to edit the ALDs. These recommendations will be considered by the Maine DOE.



7. Proceed to next item and repeat for all Inconsistent Items

Appendix E: ESS Alignment Workshop Day 3 Evaluation Survey Results

Table E1. Evaluation Results: Mathematics - Overall

Overall (N=28)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	1	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	1	4	4
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	1	4	4
It was easy to align items to specific achievement levels using the ALDs.	1	4	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	1	4	4
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	1	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	1	4	4
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	1	4	4
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	1	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	1	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	1	4	3

Table E2.. Evaluation Results: Mathematics – Grade 3

Grade 3 (N=4)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	4	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	4	4	4
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	3	4	4
It was easy to align items to specific achievement levels using the ALDs.	3	4	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	3	4	4
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	3	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	4	4	4
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	4	4	4
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	4	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	4	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	3	4	3.5

Table E3. Evaluation Results: Mathematics – Grade 4

Grade 4 (N=5)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	4	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	3	4	3
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	3	4	4
It was easy to align items to specific achievement levels using the ALDs.	2	3	2
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	3	4	4
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	4	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	2	3	3
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	2	4	3
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	3	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	4	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	2	3	2

Table E4. Evaluation Results: Mathematics – Grade 5

Grade 5 (N=4)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	4	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	4	4	4
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	4	4	4
It was easy to align items to specific achievement levels using the ALDs.	2	4	2
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	3	4	3.5
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	3	4	3.5
The ALDs helped illustrate the standards being tested in my grade and subject.	2	3	2.5
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	3	3	3
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	4	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	4	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	2	4	3

Table E5. Evaluation Results: Mathematics – Grade 6

Grade 6 (N=4)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	4	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	4	4	4
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	4	4	4
It was easy to align items to specific achievement levels using the ALDs.	3	4	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	4	4	4
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	4	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	4	4	4
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	3	4	4
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	4	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	4	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	4	4	4

Table E6. Evaluation Results: Mathematics – Grade 7

Grade 7 (N=3)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	3	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	3	4	4
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	2	4	3
It was easy to align items to specific achievement levels using the ALDs.	2	3	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	3	3	3
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	3	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	3	4	3
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	3	4	4
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	4	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	4	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	2	4	3

Table E7. Evaluation Results: Mathematics – Grade 8

Grade 8 (N=4)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	1	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	1	4	3
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	1	4	3
It was easy to align items to specific achievement levels using the ALDs.	1	3	2.5
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	1	4	4
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	1	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	1	4	3.5
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	1	4	3.5
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	1	4	3
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	1	4	3.5
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	1	4	3

Table E8. Evaluation Results: Mathematics – High School

Grade HS (N=4)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	4	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	3	4	3.5
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	4	4	4
It was easy to align items to specific achievement levels using the ALDs.	3	4	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	3	4	3.5
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	3	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	4	4	4
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	3	4	4
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	4	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	4	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	3	4	4

Table E9. Evaluation Results: Reading - Overall

Overall (N=33)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	1	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	1	4	4
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	1	4	4
It was easy to align items to specific achievement levels using the ALDs.	1	4	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	1	4	4
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	1	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	1	4	3
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	1	4	4
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	1	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	1	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	1	4	3

Table E10. Evaluation Results: Reading – Grade 3

Grade 3 (N=4)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	4	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	3	4	4
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	4	4	4
It was easy to align items to specific achievement levels using the ALDs.	3	4	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	3	4	4
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	3	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	3	4	4
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	3	4	4
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	3	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	4	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	4	4	4

Table E11. Evaluation Results: Reading – Grade 4

Grade 4 (N=5)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	4	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	3	4	3
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	4	4	4
It was easy to align items to specific achievement levels using the ALDs.	2	3	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	3	4	4
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	4	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	2	4	3
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	3	4	3
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	4	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	4	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	3	4	3

Table E12. Evaluation Results: Reading – Grade 5

Grade 5 (N=5)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	1	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	1	4	4
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	1	4	3
It was easy to align items to specific achievement levels using the ALDs.	2	4	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	1	4	3
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	1	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	1	4	3
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	1	4	3
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	1	4	3
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	1	4	3
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	1	4	3

Table E13. Evaluation Results: Reading – Grade 6

Grade 6 (N=5)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	1	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	2	4	3
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	2	4	4
It was easy to align items to specific achievement levels using the ALDs.	2	4	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	2	4	3
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	2	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	2	4	3
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	2	4	4
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	2	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	2	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	2	4	3

Table E14. Evaluation Results: Reading – Grade 7

Grade 7 (N=4)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	3	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	3	4	4
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	3	4	4
It was easy to align items to specific achievement levels using the ALDs.	2	4	3.5
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	3	4	3
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	3	4	3.5
The ALDs helped illustrate the standards being tested in my grade and subject.	3	4	3.5
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	3	4	3.5
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	3	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	4	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	3	4	3.5

Table E15. Evaluation Results: Reading – Grade 8

Grade 8 (N=5)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	1	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	1	4	4
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	1	4	4
It was easy to align items to specific achievement levels using the ALDs.	2	4	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	1	4	4
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	1	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	1	4	3
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	1	4	4
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	1	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	1	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	1	4	4

Table E16. Evaluation Results: Reading – High School

High School (N=5)	Min	Max	Median
I was provided the opportunity to ask questions about the Item Review and Resolution activity.	4	4	4
The training and resources provided for the Inconsistent Item Review & Resolution activity were clear and helped me understand the task.	3	4	4
Discussing and resolving item inconsistencies was useful for improving my item-ALD alignments.	3	4	4
It was easy to align items to specific achievement levels using the ALDs.	1	4	3
I often agreed with my fellow panelists on the most appropriate Item-ALD alignment.	3	4	3
When I disagreed with my fellow panelists on the most appropriate Item-ALD alignment, I felt comfortable expressing my thinking.	4	4	4
The ALDs helped illustrate the standards being tested in my grade and subject.	3	4	3
Aligning the items to the ALDs helped me understand how we measure the standards being tested in my grade and subject.	2	4	3
I understand from the Day 1 opening session that the forms we reviewed during the first two days were examples of the Maine Through Year Assessments that will be used for federal peer review purposes.	3	4	4
I understand that student results will be reported in terms of achievement levels and that the work we did in the workshop will help establish those achievement levels.	3	4	4
The expectations reflected by the ALDs for Maine students were appropriate in terms of content and rigor.	2	4	3

Table E17. Evaluation Results: Mathematics – Day 3 Mean Percentage

Question	Overall Average	Grade						
		3	4	5	6	7	8	HS
What is the lowest percentage of students At or Above State Expectations that you would expect and support?	35	36	28	41	33	20	29	52
What is the highest percentage of students At or Above State Expectations that you would expect and support?	66	85	76	62	55	52	51	75
What is the lowest percentage of students Above State Expectations that you would expect and support?	9	29	5	8	8	3	3	8
What is the highest percentage of students Above Expectations that you would expect and support?	25	76	16	22	16	20	9	18
What is the lowest percentage of students Well Below State Expectations that you would expect and support?	14	5	23	12	14	4	26	6
What is the highest percentage of students Well Below State Expectations that you would expect and support?	36	39	40	28	38	35	46	16

Table E18. Evaluation Results: Reading – Day 3 Mean Percentage

Question	Overall Average	Grade						
		3	4	5	6	7	8	HS
What is the lowest percentage of students At or Above State Expectations that you would expect and support?	43	34	45	53	52	30	46	39
What is the highest percentage of students At or Above State Expectations that you would expect and support?	68	61	64	68	76	67	74	63
What is the lowest percentage of students Above State Expectations that you would expect and support?	12	12	7	12	25	10	10	9
What is the highest percentage of students Above Expectations that you would expect and support?	23	26	19	22	31	22	21	21
What is the lowest percentage of students Well Below State Expectations that you would expect and support?	13	15	12	13	10	20	10	14
What is the highest percentage of students Well Below State Expectations that you would expect and support?	27	37	22	21	20	38	23	30

Appendix F: Detailed ESS Item Maps

Table F1. Detailed ESS Item Maps: Mathematics Grade 3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M3_VR694464	1	-4.4863	Level2	4	12.86	30	87.99	65	251.11	Level1	Level1
M3_VR695822	2	-3.5326	Level2	5	10	29	60.33	64	190.07	Level1	Level1
M3_VR696271	3	-3.0459	Level4	6	9.03	28	46.7	63	159.41	Level1	Level1
M3_VR698039	4	-2.7375	Level2	7	8.72	29	38.37	64	140.29	Level2	Level1
M3_VR694605	5	-2.7358	Level4	8	8.72	28	38.33	63	140.19	Level2	Level1
M3_VR695260	6	-2.6578	Level2	9	8.8	29	36.38	64	135.51	Level2	Level1
M3_VR694338_1	7	-2.533	Level2	10	9.05	28	33.39	63	128.15	Level2	Level1
M3_VR697637	8	-2.4336	Level2	11	9.35	27	31.1	62	122.38	Level2	Level1
M3_VR696189	9	-2.2198	Level1	12	10.2	26	26.4	61	110.19	Level2	Level1
M3_VR698053	10	-2.1875	Level2	11	10.36	25	25.72	60	108.38	Level2	Level1
M3_VR696129_1	11	-2.176	Level2	12	10.43	24	25.49	59	107.75	Level2	Level1
M3_VR697378	12	-2.0358	Level2	13	11.41	23	22.82	58	100.18	Level2	Level1
M3_VR694738	13	-1.9612	Level2	14	12.01	22	21.48	57	96.23	Level2	Level2
M3_VR695335	14	-1.8779	Level2	15	12.76	21	20.06	56	91.9	Level2	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M3_VR696443	15	-1.778	Level1	16	13.76	20	18.47	55	86.8	Level2	Level2
M3_VR696999	16	-1.6336	Level2	15	15.35	19	16.3	54	79.58	Level2	Level2
M3_VR696153	17	-1.5451	Level2	16	16.41	18	15.06	53	75.24	Level2	Level2
M3_VR696615	18	-1.5272	Level2	17	16.64	17	14.83	52	74.39	Level2	Level2
M3_VR695848	19	-1.3942	Level3	18	18.5	16	13.23	51	68.13	Level2	Level2
M3_VR694837_1	20	-1.379	Level2	19	18.73	17	13.07	50	67.44	Level2	Level2
M3_VR696485	21	-1.3168	Level2	20	19.73	16	12.44	49	64.64	Level2	Level2
M3_VR697259	22	-1.2876	Level2	21	20.22	15	12.18	48	63.35	Level2	Level2
M3_VR696469	23	-1.2653	Level2	22	20.63	14	12	47	62.39	Level2	Level2
M3_VR695460_1	24	-1.234	Level4	23	21.22	13	11.78	46	61.08	Level2	Level2
M3_VR696399	25	-1.1358	Level2	24	23.18	14	11.19	47	57.05	Level2	Level2
M3_VR698098	26	-1.1175	Level3	25	23.57	13	11.1	46	56.32	Level2	Level2
M3_VR697700	27	-0.8766	Level3	26	28.87	14	10.14	45	46.92	Level2	Level3
M3_VR698164	28	-0.7301	Level3	27	32.24	15	9.7	44	41.36	Level3	Level3
M3_VR695321	29	-0.7168	Level4	28	32.56	16	9.67	43	40.87	Level3	Level3
M3_VR694747	30	-0.6739	Level3	29	33.63	17	9.63	44	39.32	Level3	Level3
M3_VR695248	31	-0.6664	Level2	30	33.82	18	9.63	43	39.06	Level3	Level3
M3_VR698082_1	32	-0.652	Level3	31	34.21	17	9.64	42	38.57	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M3_VR694837_2	33	-0.584	Level2	32	36.12	18	9.78	41	36.32	Level3	Level3
M3_VR695460_2	34	-0.578	Level4	33	36.29	17	9.8	40	36.13	Level3	Level3
M3_VR696313	35	-0.5658	Level2	34	36.66	18	9.85	41	35.75	Level3	Level3
M3_VR695405	36	-0.5611	Level1	35	36.8	17	9.87	40	35.61	Level3	Level3
M3_VR694732	37	-0.5604	Level3	34	36.83	16	9.87	39	35.59	Level3	Level3
M3_VR696980	38	-0.5227	Level1	35	38.07	17	10.14	38	34.54	Level3	Level3
M3_VR698072_1	39	-0.508	Level3	34	38.57	16	10.26	37	34.14	Level3	Level3
M3_VR698584	40	-0.4951	Level3	35	39.02	17	10.37	36	33.81	Level3	Level3
M3_VR698159	41	-0.342	Level2	36	44.53	18	11.9	35	29.98	Level3	Level3
M3_VR698578	42	-0.2933	Level3	37	46.33	17	12.44	34	28.81	Level3	Level3
M3_VR698437_1	43	-0.263	Level3	38	47.49	18	12.8	33	28.11	Level3	Level3
M3_VR694908	44	-0.2027	Level3	39	49.84	19	13.59	32	26.79	Level3	Level3
M3_VR697278	45	-0.1625	Level4	40	51.45	20	14.15	31	25.94	Level3	Level3
M3_VR695275	46	-0.1333	Level2	41	52.64	21	14.59	32	25.36	Level3	Level3
M3_VR698100	47	-0.087	Level2	42	54.59	20	15.33	31	24.48	Level3	Level3
M3_VR694354	48	-0.0817	Level3	43	54.81	19	15.42	30	24.38	Level3	Level3
M3_VR694599	49	-0.0794	Level3	44	54.92	20	15.46	29	24.34	Level3	Level3
M3_VR696129_2	50	-0.065	Level4	45	55.56	21	15.73	28	24.11	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M3_VR695685_1	51	0.01	Level3	46	59.01	22	17.23	29	22.99	Level3	Level3
M3_VR696682	52	0.0315	Level3	47	60.02	23	17.68	28	22.69	Level3	Level3
M3_VR695278	53	0.0424	Level3	48	60.55	24	17.92	27	22.55	Level3	Level3
M3_VR695953_1	54	0.132	Level4	49	64.94	25	19.99	26	21.47	Level3	Level3
M3_VR694597	55	0.1454	Level3	50	65.61	26	20.31	27	21.32	Level3	Level3
M3_VR698081	56	0.1573	Level3	51	66.22	27	20.6	26	21.2	Level3	Level3
M3_VR695792	57	0.1714	Level4	52	66.95	28	20.97	25	21.08	Level3	Level3
M3_VR694918_1	58	0.197	Level4	53	68.3	29	21.66	26	20.87	Level3	Level3
M3_VR697892	59	0.257	Level4	54	71.55	30	23.34	27	20.45	Level3	Level3
M3_VR694338_2	60	0.32	Level3	55	75.01	31	25.17	28	20.07	Level3	Level3
M3_VR695551	61	0.4183	Level3	56	80.51	32	28.12	27	19.58	Level3	Level3
M3_VR695002	62	0.4482	Level3	57	82.22	33	29.04	26	19.46	Level3	Level3
M3_VR698625	63	0.4931	Level3	58	84.82	34	30.48	25	19.33	Level3	Level3
M3_VR696006	64	0.5205	Level3	59	86.44	35	31.39	24	19.27	Level3	Level3
M3_VR694305	65	0.5625	Level4	60	88.96	36	32.81	23	19.23	Level4	Level3
M3_VR695352	66	0.5671	Level2	61	89.24	37	32.97	24	19.23	Level4	Level3
M3_VR698082_2	67	0.577	Level4	62	89.85	36	33.33	23	19.24	Level4	Level3
M3_VR695370	68	0.6244	Level3	63	92.84	37	35.09	24	19.34	Level4	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M3_VR698072_2	69	0.664	Level3	64	95.37	38	36.59	23	19.45	Level4	Level3
M3_VR694754	70	0.6755	Level3	65	96.12	39	37.04	22	19.5	Level4	Level3
M3_VR695023	71	0.7113	Level3	66	98.49	40	38.47	21	19.68	Level4	Level3
M3_VR697132	72	0.7208	Level3	67	99.12	41	38.86	20	19.74	Level4	Level3
M3_VR697390	73	0.8271	Level3	68	106.35	42	43.32	19	20.48	Level4	Level3
M3_VR695336	74	0.9493	Level3	69	114.78	43	48.58	18	21.46	Level4	Level3
M3_VR695953_2	75	1.006	Level4	70	118.75	44	51.07	17	21.97	Level4	Level3
M3_VR698525	76	1.1289	Level4	71	127.48	45	56.6	18	23.2	Level4	Level4
M3_VR695456	77	1.1668	Level3	72	130.21	46	58.35	19	23.61	Level4	Level4
M3_VR695207	78	1.3223	Level4	73	141.56	47	65.66	18	25.48	Level4	Level4
M3_VR695685_2	79	1.383	Level3	74	146.05	48	68.57	19	26.27	Level4	Level4
M3_VR698605	80	1.5132	Level4	75	155.81	49	74.95	18	28.09	Level4	Level4
M3_VR696803	81	1.5988	Level4	76	162.32	50	79.23	19	29.38	Level4	Level4
M3_VR697044	82	1.65	Level4	77	166.26	51	81.84	20	30.2	Level4	Level4
M3_VR694658	83	1.7892	Level3	78	177.12	52	89.08	21	32.56	Level4	Level4
M3_VR697288	84	1.9347	Level3	79	188.61	53	96.79	20	35.18	Level4	Level4
M3_VR698437_2	85	2.401	Level4	80	225.92	54	121.97	19	44.04	Level4	Level4
M3_VR696639	86	2.567	Level4	81	239.36	55	131.1	20	47.36	Level4	Level4

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M3_VR696976	87	3.1385	Level4	82	286.23	56	163.11	21	59.36	Level4	Level4
M3_VR694918_2	88	3.692	Level4	83	332.17	57	194.65	22	71.54	Level4	Level4

Table F2. Detailed ESS Item Maps: Mathematics Grade 4

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M4_VR698095	1	-3.8689	Level1	3	0.8	19	24.95	79	268.43	Level1	Level1
M4_VR694353_1	2	-3.746	Level2	2	0.55	18	22.73	78	258.85	Level1	Level1
M4_VR697344_1	3	-3.743	Level1	3	0.55	17	22.68	77	258.62	Level1	Level1
M4_VR696271	4	-3.1996	Level1	2	0.55	16	13.99	76	217.32	Level2	Level1
M4_VR695113_1	5	-3.08	Level2	1	0.67	15	12.19	75	208.35	Level2	Level1
M4_VR694888	6	-2.9778	Level2	2	0.87	14	10.76	74	200.79	Level2	Level1
M4_VR697696	7	-2.9582	Level2	3	0.93	13	10.51	73	199.35	Level2	Level1
M4_VR694785	8	-2.5511	Level2	4	2.56	12	5.62	72	170.04	Level2	Level1
M4_VR695427	9	-2.5196	Level2	5	2.72	11	5.28	71	167.81	Level2	Level1
M4_VR694919	10	-2.4999	Level2	6	2.83	10	5.08	70	166.43	Level2	Level1
M4_VR695460_1	11	-2.357	Level2	7	3.83	9	3.79	69	156.57	Level2	Level1
M4_VR694455	12	-2.3009	Level3	8	4.28	8	3.35	68	152.75	Level2	Level1
M4_VR697344_2	13	-2.257	Level2	9	4.68	9	3.04	67	149.81	Level2	Level1
M4_VR695142	14	-2.2498	Level2	10	4.75	8	3	66	149.34	Level2	Level1
M4_VR696407	15	-2.2439	Level2	11	4.81	7	2.97	65	148.95	Level2	Level1
M4_VR697357	16	-2.2226	Level2	12	5.07	6	2.88	64	147.59	Level2	Level1
M4_VR697854	17	-2.1721	Level3	13	5.73	5	2.73	63	144.41	Level2	Level1

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M4_VR697979	18	-2.0655	Level2	14	7.22	6	2.52	62	137.8	Level2	Level1
M4_VR695624	19	-2.0615	Level3	15	7.28	5	2.51	61	137.55	Level2	Level1
M4_VR697120	20	-2.0196	Level2	16	7.95	6	2.51	60	135.04	Level2	Level1
M4_VR695860	21	-1.9454	Level3	17	9.21	5	2.59	59	130.66	Level3	Level1
M4_VR694516	22	-1.7738	Level3	18	12.3	6	2.93	58	120.71	Level3	Level2
M4_VR695460_2	23	-1.658	Level3	19	14.5	7	3.28	57	114.11	Level3	Level2
M4_VR696572	24	-1.557	Level3	20	16.52	8	3.68	56	108.45	Level3	Level2
M4_VR697179	25	-1.5082	Level3	21	17.54	9	3.92	55	105.77	Level3	Level2
M4_VR698234	26	-1.5025	Level3	22	17.67	10	3.96	54	105.46	Level3	Level2
M4_VR695751	27	-1.4268	Level3	23	19.41	11	4.49	53	101.45	Level3	Level2
M4_VR696563	28	-1.3779	Level3	24	20.58	12	4.88	52	98.91	Level3	Level2
M4_VR695248	29	-1.3749	Level3	25	20.66	13	4.91	51	98.75	Level3	Level2
M4_VR696389	30	-1.3167	Level2	26	22.17	14	5.49	50	95.84	Level3	Level2
M4_VR696034	31	-1.2594	Level4	27	23.72	13	6.12	49	93.04	Level3	Level2
M4_VR695089	32	-1.228	Level3	28	24.6	14	6.5	50	91.53	Level3	Level2
M4_VR695953_1	33	-1.226	Level3	29	24.66	15	6.52	49	91.44	Level3	Level2
M4_VR697027_1	34	-1.086	Level3	30	28.86	16	8.48	48	85	Level3	Level2
M4_VR696174	35	-1.0843	Level3	31	28.91	17	8.51	47	84.92	Level3	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M4_VR695194	36	-1.0514	Level4	32	29.96	18	9.03	46	83.47	Level3	Level2
M4_VR697491	37	-0.9243	Level3	33	34.16	19	11.19	47	78.01	Level3	Level2
M4_VR694876	38	-0.7989	Level3	34	38.42	20	13.45	46	72.74	Level3	Level2
M4_VR698164	39	-0.6866	Level2	35	42.35	21	15.59	45	68.13	Level3	Level3
M4_VR695481	40	-0.6852	Level3	36	42.4	20	15.61	44	68.08	Level3	Level3
M4_VR695953_2	41	-0.583	Level3	37	46.18	21	17.76	43	64.09	Level3	Level3
M4_VR696168	42	-0.5241	Level3	38	48.42	22	19.06	42	61.85	Level3	Level3
M4_VR698296_1	43	-0.508	Level3	39	49.05	23	19.43	41	61.26	Level3	Level3
M4_VR696555	44	-0.375	Level3	40	54.37	24	22.62	40	56.47	Level3	Level3
M4_VR697063	45	-0.356	Level3	41	55.15	25	23.09	39	55.81	Level3	Level3
M4_VR696040	46	-0.3474	Level3	42	55.51	26	23.32	38	55.51	Level3	Level3
M4_VR696952	47	-0.2549	Level3	43	59.49	27	25.81	37	52.46	Level3	Level3
M4_VR694378	48	-0.1703	Level3	44	63.21	28	28.18	36	49.75	Level3	Level3
M4_VR695798	49	-0.1328	Level3	45	64.9	29	29.27	35	48.59	Level3	Level3
M4_VR695551	50	-0.1223	Level3	46	65.38	30	29.59	34	48.28	Level3	Level3
M4_VR697027_2	51	-0.051	Level3	47	68.73	31	31.8	33	46.21	Level3	Level3
M4_VR694918_1	52	0.261	Level3	48	83.71	32	41.78	32	37.47	Level3	Level3
M4_VR698191	53	0.3111	Level3	49	86.16	33	43.43	31	36.12	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M4_VR694353_2	54	0.314	Level3	50	86.31	34	43.53	30	36.04	Level3	Level3
M4_VR696871	55	0.3143	Level3	51	86.32	35	43.54	29	36.04	Level3	Level3
M4_VR698267	56	0.3207	Level3	52	86.65	36	43.77	28	35.88	Level3	Level3
M4_VR694387_1	57	0.656	Level4	53	104.43	37	56.18	27	28.17	Level3	Level3
M4_VR695113_2	58	0.683	Level3	54	105.88	38	57.2	28	27.58	Level3	Level3
M4_VR696998	59	0.9191	Level3	55	118.87	39	66.41	27	22.62	Level3	Level3
M4_VR697214	60	0.9401	Level3	56	120.05	40	67.25	26	22.2	Level3	Level3
M4_VR698431	61	0.9455	Level3	57	120.35	41	67.47	25	22.1	Level3	Level3
M4_VR696976	62	0.984	Level3	58	122.59	42	69.09	24	21.4	Level3	Level3
M4_VR696732	63	0.9869	Level4	59	122.76	43	69.22	23	21.35	Level3	Level3
M4_VR694345	64	0.9965	Level3	60	123.33	44	69.64	24	21.2	Level3	Level3
M4_VR696197_1	65	1.034	Level3	61	125.62	45	71.33	23	20.64	Level3	Level3
M4_VR696206	66	1.0895	Level3	62	129.06	46	73.88	22	19.86	Level3	Level3
M4_VR695354	67	1.2119	Level4	63	136.77	47	79.63	21	18.27	Level3	Level3
M4_VR698045	68	1.4116	Level3	64	149.55	48	89.22	22	15.87	Level3	Level3
M4_VR694387_2	69	1.478	Level4	65	153.87	49	92.47	21	15.14	Level3	Level3
M4_VR695898	70	1.5058	Level3	66	155.7	50	93.86	22	14.87	Level3	Level3
M4_VR697888	71	1.5904	Level3	67	161.37	51	98.17	21	14.1	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M4_VR698296_2	72	1.673	Level3	68	166.99	52	102.47	20	13.44	Level3	Level3
M4_VR696739	73	1.7076	Level3	69	169.38	53	104.3	19	13.2	Level3	Level3
M4_VR696197_2	74	1.757	Level3	70	172.83	54	106.97	18	12.9	Level3	Level3
M4_VR698361	75	1.7951	Level3	71	175.54	55	109.07	17	12.71	Level3	Level3
M4_VR696478	76	1.8012	Level3	72	175.98	56	109.41	16	12.69	Level3	Level4
M4_VR698242	77	1.8208	Level3	73	177.41	57	110.53	15	12.63	Level3	Level4
M4_VR698593_1	78	1.85	Level3	74	179.57	58	112.22	14	12.57	Level3	Level4
M4_VR695351	79	2.0492	Level3	75	194.51	59	123.97	13	12.37	Level4	Level4
M4_VR696115	80	2.088	Level3	76	197.46	60	126.3	12	12.37	Level4	Level4
M4_VR694853	81	2.2164	Level3	77	207.35	61	134.13	11	12.5	Level4	Level4
M4_VR698137	82	2.2636	Level3	78	211.03	62	137.06	10	12.6	Level4	Level4
M4_VR695528	83	2.3823	Level3	79	220.41	63	144.54	9	12.95	Level4	Level4
M4_VR698468	84	2.3999	Level4	80	221.81	64	145.66	8	13.02	Level4	Level4
M4_VR697079	85	2.4032	Level4	81	222.08	65	145.88	9	13.04	Level4	Level4
M4_VR694918_2	86	2.539	Level4	82	233.22	66	154.84	10	13.85	Level4	Level4
M4_VR698593_2	87	2.593	Level3	83	237.7	67	158.46	11	14.23	Level4	Level4
M4_VR697706	88	2.7409	Level4	84	250.12	68	168.52	10	15.42	Level4	Level4
M4_VR697125	89	2.8519	Level3	85	259.56	69	176.17	11	16.41	Level4	Level4

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M4_VR695222	90	4.5272	Level4	86	403.63	70	293.45	10	33.17	Level4	Level4
M4_VR694816	91	4.6264	Level4	87	412.26	71	300.49	11	34.26	Level4	Level4

Table F3. Detailed ESS Item Maps: Mathematics Grade 5

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M5_VR697805	1	-3.0963	Level1	12	27.48	45	113.68	82	233.14	Level1	Level1
M5_VR694627_1	2	-2.531	Level2	11	21.26	44	88.8	81	187.35	Level1	Level1
M5_VR694534_1	3	-2.446	Level2	12	20.41	43	85.15	80	180.55	Level1	Level1
M5_VR695408	4	-2.425	Level3	13	20.22	42	84.27	79	178.89	Level1	Level1
M5_VR695891	5	-2.3519	Level3	14	19.64	43	81.27	78	173.19	Level1	Level1
M5_VR695604	6	-2.3022	Level4	15	19.29	44	79.28	77	169.36	Level1	Level1
M5_VR694548	7	-2.2126	Level3	16	18.75	45	75.79	78	162.55	Level1	Level1
M5_VR696065	8	-2.0842	Level1	17	18.11	46	70.91	77	152.92	Level1	Level1
M5_VR697340	9	-2.012	Level2	16	17.82	45	68.24	76	147.58	Level1	Level1
M5_VR694415	10	-1.9817	Level2	17	17.73	44	67.14	75	145.37	Level1	Level1
M5_VR695799_1	11	-1.856	Level1	18	17.48	43	62.75	74	136.32	Level1	Level1
M5_VR694627_2	12	-1.699	Level2	17	17.32	42	57.41	73	125.17	Level2	Level2
M5_VR695159	13	-1.6018	Level1	18	17.32	41	54.2	72	118.37	Level2	Level2
M5_VR696671	14	-1.5354	Level3	17	17.39	40	52.07	71	113.79	Level2	Level2
M5_VR696008	15	-1.4996	Level2	18	17.46	41	50.97	70	111.35	Level2	Level2
M5_VR696423	16	-1.4996	Level3	18	17.46	41	50.97	70	111.35	Level2	Level2
M5_VR696591	17	-1.4256	Level3	20	17.76	41	48.82	68	106.47	Level2	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M5_VR697134	18	-1.383	Level3	21	17.97	42	47.63	67	103.7	Level2	Level2
M5_VR696813	19	-1.3657	Level1	22	18.07	43	47.16	66	102.59	Level2	Level2
M5_VR696555	20	-1.3242	Level2	21	18.36	42	46.08	65	99.98	Level2	Level2
M5_VR694861_1	21	-1.317	Level2	22	18.42	41	45.9	64	99.53	Level2	Level2
M5_VR694534_2	22	-1.299	Level2	23	18.58	40	45.47	63	98.43	Level2	Level2
M5_VR695813	23	-1.2416	Level3	24	19.16	39	44.15	62	94.99	Level2	Level2
M5_VR695566	24	-1.1997	Level1	25	19.62	40	43.23	61	92.52	Level2	Level2
M5_VR696338	25	-1.1427	Level3	24	20.3	39	42.03	60	89.21	Level2	Level2
M5_VR695989	26	-1.1212	Level2	25	20.58	40	41.6	59	87.98	Level2	Level2
M5_VR696178	27	-1.054	Level2	26	21.52	39	40.32	58	84.22	Level2	Level2
M5_VR694710	28	-0.9727	Level2	27	22.74	38	38.86	57	79.75	Level2	Level2
M5_VR696401	29	-0.9463	Level2	28	23.16	37	38.41	56	78.32	Level2	Level2
M5_VR697929	30	-0.8973	Level3	29	24	36	37.63	55	75.73	Level2	Level2
M5_VR694910	31	-0.8766	Level2	30	24.37	37	37.32	54	74.65	Level2	Level2
M5_VR694387_1	32	-0.839	Level2	31	25.08	36	36.79	53	72.73	Level2	Level2
M5_VR698360	33	-0.8363	Level2	32	25.14	35	36.75	52	72.6	Level2	Level2
M5_VR694861_2	34	-0.739	Level3	33	27.18	34	35.59	51	67.83	Level2	Level2
M5_VR695660	35	-0.7137	Level2	34	27.74	35	35.31	50	66.62	Level2	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M5_VR694949	36	-0.3081	Level2	35	37.07	34	31.25	49	47.55	Level2	Level2
M5_VR697925_1	37	-0.3	Level2	36	37.26	33	31.18	48	47.18	Level2	Level2
M5_VR698337	38	-0.2054	Level1	37	39.63	32	30.42	47	42.92	Level2	Level2
M5_VR695799_2	39	-0.187	Level1	36	40.1	31	30.29	46	42.11	Level2	Level3
M5_VR694387_2	40	-0.142	Level2	35	41.32	30	30.02	45	40.18	Level2	Level3
M5_VR697558	41	-0.1204	Level1	36	41.92	29	29.92	44	39.27	Level2	Level3
M5_VR697883	42	-0.0901	Level2	35	42.8	28	29.79	43	38.03	Level2	Level3
M5_VR697910	43	-0.0882	Level2	36	42.86	27	29.79	42	37.95	Level2	Level3
M5_VR694708	44	0.0038	Level3	37	45.71	26	29.6	41	34.37	Level2	Level3
M5_VR694838	45	0.046	Level3	38	47.06	27	29.56	40	32.76	Level2	Level3
M5_VR696511	46	0.1271	Level2	39	49.74	28	29.56	39	29.76	Level3	Level3
M5_VR694701	47	0.1752	Level2	40	51.37	27	29.61	38	28.03	Level3	Level3
M5_VR697925_2	48	0.2	Level2	41	52.24	26	29.66	37	27.16	Level3	Level3
M5_VR694390	49	0.217	Level1	42	52.85	25	29.71	36	26.58	Level3	Level3
M5_VR694461	50	0.2506	Level3	41	54.1	24	29.85	35	25.47	Level3	Level3
M5_VR697614	51	0.3092	Level3	42	56.32	25	30.14	34	23.6	Level3	Level3
M5_VR696125	52	0.3565	Level3	43	58.17	26	30.42	33	22.13	Level3	Level3
M5_VR697181_1	53	0.373	Level2	44	58.83	27	30.54	32	21.64	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M5_VR696672	54	0.4095	Level3	45	60.32	26	30.83	31	20.58	Level3	Level3
M5_VR696958	55	0.4489	Level2	46	61.98	27	31.18	30	19.48	Level3	Level3
M5_VR695597	56	0.4904	Level3	47	63.76	26	31.6	29	18.36	Level3	Level3
M5_VR698355_1	57	0.523	Level3	48	65.2	27	31.96	28	17.51	Level3	Level3
M5_VR697388	58	0.5492	Level2	49	66.38	28	32.27	27	16.85	Level3	Level3
M5_VR696984	59	0.6042	Level1	50	68.91	27	32.99	26	15.53	Level3	Level3
M5_VR698358	60	0.6066	Level3	49	69.02	26	33.02	25	15.48	Level3	Level3
M5_VR695933	61	0.7365	Level3	50	75.26	27	34.97	24	12.62	Level3	Level3
M5_VR697769	62	0.7439	Level3	51	75.62	28	35.09	23	12.46	Level3	Level3
M5_VR694953	63	0.7771	Level3	52	77.28	29	35.65	22	11.8	Level3	Level3
M5_VR696998	64	0.7805	Level3	53	77.45	30	35.71	21	11.74	Level3	Level3
M5_VR695442	65	0.7922	Level3	54	78.06	31	35.94	20	11.53	Level3	Level3
M5_VR696446	66	0.8132	Level3	55	79.17	32	36.36	19	11.17	Level3	Level3
M5_VR697521	67	0.8862	Level3	56	83.11	33	37.89	18	10	Level3	Level3
M5_VR697707	68	0.9115	Level4	57	84.51	34	38.45	17	9.62	Level3	Level3
M5_VR697536	69	0.9194	Level2	58	84.95	35	38.63	18	9.51	Level3	Level3
M5_VR695834	70	1.0193	Level3	59	90.64	34	41.03	17	8.21	Level3	Level3
M5_VR698305	71	1.0537	Level3	60	92.64	35	41.89	16	7.8	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M5_VR697133	72	1.076	Level2	61	93.95	36	42.46	15	7.55	Level3	Level3
M5_VR697181_2	73	1.143	Level3	62	97.97	35	44.27	14	6.88	Level3	Level3
M5_VR696500	74	1.1431	Level2	63	97.98	36	44.28	13	6.88	Level3	Level3
M5_VR698593_1	75	1.149	Level3	64	98.35	35	44.45	12	6.84	Level3	Level3
M5_VR698468	76	1.2034	Level2	65	101.77	36	46.08	11	6.45	Level3	Level3
M5_VR697376	77	1.2049	Level3	66	101.87	35	46.13	10	6.45	Level3	Level3
M5_VR696215	78	1.2184	Level1	67	102.75	36	46.56	9	6.38	Level3	Level3
M5_VR696108	79	1.2399	Level3	66	104.17	35	47.27	8	6.29	Level3	Level3
M5_VR698356	80	1.3605	Level3	67	112.25	36	51.37	7	5.93	Level3	Level3
M5_VR694964	81	1.4905	Level3	68	121.09	37	55.92	6	5.67	Level3	Level3
M5_VR695528	82	1.7186	Level3	69	136.82	38	64.13	5	5.44	Level4	Level3
M5_VR698593_2	83	1.85	Level3	70	146.02	39	68.99	4	5.44	Level4	Level3
M5_VR698355_2	84	1.943	Level4	71	152.63	40	72.53	3	5.54	Level4	Level3
M5_VR696088	85	2.2018	Level2	72	171.26	41	82.62	4	6.05	Level4	Level4
M5_VR695879_1	86	3.03	Level4	73	231.72	40	115.75	3	8.54	Level4	Level4
M5_VR695879_2	87	3.216	Level4	74	245.48	41	123.37	4	9.28	Level4	Level4

Table F4. Detailed ESS Item Maps: Mathematics Grade 6

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M6_VR694788	1	-2.3332	Level1	16	20.74	62	109.45	85	195.53	Level1	Level1
M6_VR694913_1	2	-2.295	Level1	15	20.16	61	107.12	84	192.32	Level1	Level1
M6_VR696307	3	-2.2663	Level1	14	19.76	60	105.4	83	189.94	Level1	Level1
M6_VR697274	4	-2.2576	Level1	13	19.65	59	104.88	82	189.22	Level1	Level1
M6_VR696297	5	-2.0381	Level1	12	17.01	58	92.15	81	171.44	Level1	Level1
M6_VR695945	6	-1.9525	Level3	11	16.07	57	87.27	80	164.6	Level1	Level1
M6_VR698188	7	-1.9209	Level2	12	15.76	58	85.5	79	162.1	Level1	Level1
M6_VR695647	8	-1.9046	Level2	13	15.61	57	84.61	78	160.83	Level1	Level1
M6_VR695856	9	-1.8963	Level2	14	15.54	56	84.16	77	160.19	Level1	Level1
M6_VR695774	10	-1.8561	Level2	15	15.26	55	82.03	76	157.13	Level1	Level1
M6_VR694861_1	11	-1.773	Level1	16	14.76	54	77.71	75	150.9	Level1	Level2
M6_VR697111	12	-1.7248	Level2	15	14.52	53	75.25	74	147.34	Level1	Level2
M6_VR695920	13	-1.7052	Level2	16	14.44	52	74.27	73	145.9	Level1	Level2
M6_VR694584	14	-1.633	Level2	17	14.23	51	70.73	72	140.71	Level1	Level2
M6_VR696888	15	-1.6001	Level2	18	14.16	50	69.15	71	138.37	Level1	Level2
M6_VR695754	16	-1.5949	Level2	19	14.16	49	68.91	70	138.01	Level1	Level2
M6_VR694872_1	17	-1.511	Level2	20	14.16	48	65.05	69	132.22	Level1	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M6_VR697411	18	-1.3712	Level1	21	14.3	47	58.76	68	122.71	Level2	Level2
M6_VR696219	19	-1.356	Level2	20	14.33	46	58.09	67	121.69	Level2	Level2
M6_VR695114	20	-1.3106	Level2	21	14.46	45	56.14	66	118.7	Level2	Level2
M6_VR694861_2	21	-1.265	Level1	22	14.64	44	54.22	65	115.73	Level2	Level2
M6_VR695086	22	-1.2028	Level2	21	14.96	43	51.67	64	111.75	Level2	Level2
M6_VR696007	23	-1.196	Level2	22	15	42	51.4	63	111.32	Level2	Level2
M6_VR695700	24	-1.194	Level1	23	15.01	41	51.32	62	111.2	Level2	Level2
M6_VR696157	25	-1.1369	Level2	22	15.47	40	49.15	61	107.72	Level2	Level2
M6_VR694387_1	26	-1.135	Level1	23	15.48	39	49.08	60	107.6	Level2	Level2
M6_VR695221	27	-1.1342	Level2	22	15.49	38	49.05	59	107.55	Level2	Level2
M6_VR694696	28	-1.0872	Level2	23	16.01	37	47.41	58	104.83	Level2	Level2
M6_VR695076	29	-1.0509	Level2	24	16.45	36	46.17	57	102.76	Level2	Level2
M6_VR695004	30	-0.9758	Level2	25	17.42	35	43.69	56	98.55	Level2	Level2
M6_VR694913_2	31	-0.944	Level2	26	17.87	34	42.68	55	96.8	Level2	Level2
M6_VR695549	32	-0.9228	Level1	27	18.18	33	42.02	54	95.66	Level2	Level2
M6_VR695868	33	-0.8041	Level2	26	20.08	32	38.46	53	89.37	Level2	Level2
M6_VR696909	34	-0.7336	Level2	27	21.28	31	36.41	52	85.7	Level2	Level2
M6_VR697925_1	35	-0.733	Level2	28	21.29	30	36.4	51	85.67	Level2	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M6_VR694604	36	-0.703	Level2	29	21.86	29	35.59	50	84.17	Level2	Level2
M6_VR695372	37	-0.633	Level1	30	23.26	28	33.77	49	80.74	Level2	Level2
M6_VR694603	38	-0.5766	Level3	29	24.45	27	32.36	48	78.04	Level2	Level2
M6_VR695538	39	-0.501	Level2	30	26.11	28	30.54	47	74.48	Level2	Level2
M6_VR694387_2	40	-0.374	Level1	31	29.03	27	27.62	46	68.64	Level2	Level2
M6_VR698498_1	41	-0.363	Level2	30	29.3	26	27.38	45	68.15	Level2	Level2
M6_VR695353	42	-0.3403	Level2	31	29.86	25	26.9	44	67.15	Level2	Level2
M6_VR694473_1	43	-0.241	Level2	32	32.44	24	24.92	43	62.88	Level2	Level2
M6_VR697265	44	-0.237	Level3	33	32.55	23	24.84	42	62.71	Level2	Level2
M6_VR696521	45	-0.1931	Level3	34	33.78	24	24.05	41	60.91	Level2	Level2
M6_VR697925_2	46	-0.181	Level2	35	34.13	25	23.84	40	60.42	Level2	Level2
M6_VR695083	47	-0.1107	Level2	36	36.24	24	22.72	39	57.68	Level2	Level2
M6_VR696772	48	-0.0621	Level2	37	37.75	23	21.99	38	55.84	Level2	Level3
M6_VR697784	49	-0.0299	Level2	38	38.78	22	21.54	37	54.64	Level2	Level3
M6_VR695555	50	-0.0262	Level2	39	38.9	21	21.49	36	54.51	Level2	Level3
M6_VR695961	51	0.0147	Level2	40	40.29	20	21	35	53.08	Level2	Level3
M6_VR694552	52	0.0335	Level2	41	40.95	19	20.79	34	52.44	Level2	Level3
M6_VR697181_1	53	0.13	Level2	42	44.42	18	19.83	33	49.26	Level2	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M6_VR695877	54	0.1347	Level3	43	44.6	17	19.79	32	49.11	Level2	Level3
M6_VR698498_2	55	0.144	Level3	44	44.95	18	19.71	31	48.82	Level2	Level3
M6_VR694300	56	0.1881	Level1	45	46.67	19	19.4	30	47.49	Level2	Level3
M6_VR697050	57	0.4018	Level4	44	55.22	18	18.12	29	41.3	Level2	Level3
M6_VR694726	58	0.5317	Level2	45	60.54	19	17.47	30	37.66	Level2	Level3
M6_VR695879_1	59	0.689	Level2	46	67.15	18	16.84	29	33.41	Level2	Level3
M6_VR694473_2	60	0.774	Level2	47	70.81	17	16.59	28	31.2	Level2	Level3
M6_VR695879_2	61	0.988	Level2	48	80.22	16	16.16	27	25.85	Level2	Level3
M6_VR697181_2	62	1.071	Level2	49	83.96	15	16.08	26	23.86	Level2	Level3
M6_VR698433	63	1.2531	Level1	50	92.33	14	16.08	25	19.67	Level3	Level3
M6_VR698320	64	1.2854	Level3	49	93.85	13	16.11	24	18.96	Level3	Level3
M6_VR697957_1	65	1.341	Level2	50	96.52	14	16.22	23	17.79	Level3	Level3
M6_VR695265	66	1.3505	Level2	51	96.99	13	16.25	22	17.6	Level3	Level3
M6_VR695646	67	1.3981	Level2	52	99.37	12	16.44	21	16.7	Level3	Level3
M6_VR698250	68	1.4048	Level4	53	99.71	11	16.47	20	16.58	Level3	Level3
M6_VR697818	69	1.4678	Level3	54	102.98	12	16.85	21	15.51	Level3	Level3
M6_VR696247_1	70	1.531	Level3	55	106.33	13	17.29	20	14.5	Level3	Level3
M6_VR696262	71	1.5978	Level3	56	109.94	14	17.83	19	13.5	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M6_VR697957_2	72	1.717	Level3	57	116.5	15	18.9	18	11.83	Level3	Level3
M6_VR697940	73	1.7576	Level3	58	118.77	16	19.31	17	11.3	Level3	Level3
M6_VR695239	74	1.7741	Level4	59	119.71	17	19.49	16	11.1	Level3	Level3
M6_VR696784	75	1.8214	Level1	60	122.45	18	20.06	17	10.58	Level3	Level3
M6_VR696298_1	76	1.883	Level3	59	126.09	17	20.86	16	9.96	Level3	Level3
M6_VR695763	77	1.9317	Level3	60	129.01	18	21.54	15	9.53	Level3	Level3
M6_VR697653	78	2.0169	Level3	61	134.21	19	22.82	14	8.84	Level3	Level3
M6_VR696408	79	2.1495	Level3	62	142.43	20	24.94	13	7.92	Level3	Level4
M6_VR696247_2	80	2.183	Level3	63	144.54	21	25.51	12	7.72	Level3	Level4
M6_VR696286	81	2.2325	Level3	64	147.71	22	26.4	11	7.47	Level3	Level4
M6_VR694925	82	2.3292	Level3	65	153.99	23	28.24	10	7.08	Level3	Level4
M6_VR696298_2	83	2.378	Level4	66	157.21	24	29.21	9	6.93	Level3	Level4
M6_VR695097	84	2.4243	Level3	67	160.32	25	30.18	10	6.84	Level3	Level4
M6_VR697995	85	2.4268	Level3	68	160.49	26	30.24	9	6.84	Level3	Level4
M6_VR694882	86	2.5963	Level2	69	172.18	27	34.14	8	6.84	Level3	Level4
M6_VR696237	87	2.8889	Level3	70	192.66	26	41.16	7	7.13	Level4	Level4
M6_VR694872_2	88	3.275	Level3	71	220.08	27	50.81	6	7.9	Level4	Level4
M6_VR696121	89	4.0378	Level2	72	275	28	70.65	5	10.19	Level4	Level4

Table F5. Detailed ESS Item Maps: Mathematics Grade 7

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M7_VR696641_1	1	-3.396	Level1	5	10.61	52	134.87	86	259.73	Level1	Level1
M7_VR694513	2	-3.1474	Level2	4	9.62	51	122.19	85	238.6	Level1	Level1
M7_VR694742	3	-3.0856	Level2	5	9.43	50	119.1	84	233.41	Level1	Level1
M7_VR696297	4	-3.0199	Level2	6	9.3	49	115.88	83	227.96	Level1	Level1
M7_VR697431	5	-2.886	Level3	7	9.16	48	109.45	82	216.98	Level1	Level1
M7_VR694913_1	6	-2.781	Level2	8	9.16	49	104.52	81	208.47	Level1	Level1
M7_VR696527	7	-2.7087	Level2	9	9.24	48	101.19	80	202.69	Level2	Level1
M7_VR695314	8	-2.3535	Level2	10	9.95	47	85.21	79	174.63	Level2	Level1
M7_VR695187	9	-2.2998	Level3	11	10.11	46	82.84	78	170.44	Level2	Level1
M7_VR694872_1	10	-2.263	Level2	12	10.26	47	81.26	77	167.61	Level2	Level1
M7_VR697135	11	-2.2395	Level3	13	10.37	46	80.27	76	165.82	Level2	Level1
M7_VR697061	12	-2.2361	Level2	14	10.39	47	80.13	75	165.56	Level2	Level1
M7_VR695279_1	13	-2.232	Level1	15	10.42	46	79.97	74	165.26	Level2	Level1
M7_VR696135	14	-2.1727	Level3	14	10.9	45	77.66	73	160.93	Level2	Level2
M7_VR696011	15	-2.1217	Level2	15	11.36	46	75.72	72	157.26	Level2	Level2
M7_VR695791	16	-2.1203	Level2	16	11.37	45	75.67	71	157.16	Level2	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M7_VR695109	17	-2.118	Level2	17	11.4	44	75.59	70	157	Level2	Level2
M7_VR695622	18	-2.0441	Level2	18	12.28	43	73	69	151.9	Level2	Level2
M7_VR697039	19	-2.0432	Level2	19	12.29	42	72.97	68	151.84	Level2	Level2
M7_VR694290	20	-1.6666	Level1	20	17.57	41	60.54	67	126.61	Level2	Level2
M7_VR694526	21	-1.5937	Level2	19	18.66	40	58.21	66	121.8	Level2	Level2
M7_VR694752	22	-1.5867	Level2	20	18.77	39	57.99	65	121.34	Level2	Level2
M7_VR694678	23	-1.5635	Level2	21	19.17	38	57.29	64	119.86	Level2	Level2
M7_VR696641_2	24	-1.533	Level2	22	19.72	37	56.41	63	117.93	Level2	Level2
M7_VR694496	25	-1.5119	Level2	23	20.12	36	55.82	62	116.63	Level2	Level2
M7_VR697353	26	-1.4704	Level2	24	20.95	35	54.7	61	114.09	Level2	Level2
M7_VR697736	27	-1.4692	Level2	25	20.97	34	54.67	60	114.02	Level2	Level2
M7_VR694913_2	28	-1.46	Level2	26	21.17	33	54.44	59	113.48	Level2	Level2
M7_VR695909	29	-1.3974	Level3	27	22.61	32	52.94	58	109.85	Level2	Level2
M7_VR695345	30	-1.3695	Level3	28	23.28	33	52.29	57	108.26	Level2	Level2
M7_VR696644	31	-1.3317	Level2	29	24.23	34	51.46	56	106.14	Level2	Level2
M7_VR695990_1	32	-1.309	Level1	30	24.82	33	50.99	55	104.89	Level2	Level2
M7_VR696735	33	-1.1677	Level3	29	28.63	32	48.16	54	97.26	Level2	Level2
M7_VR694294	34	-0.9741	Level3	30	34.05	33	44.48	53	87	Level2	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M7_VR697945	35	-0.7759	Level3	31	39.8	34	40.91	52	76.7	Level2	Level2
M7_VR695832	36	-0.7069	Level2	32	41.87	35	39.74	51	73.18	Level2	Level2
M7_VR695524	37	-0.6611	Level3	33	43.29	34	39.01	50	70.89	Level2	Level2
M7_VR694319	38	-0.6137	Level2	34	44.81	35	38.3	49	68.56	Level2	Level2
M7_VR697606	39	-0.5938	Level2	35	45.47	34	38.02	48	67.61	Level2	Level2
M7_VR697781	40	-0.5772	Level2	36	46.03	33	37.8	47	66.83	Level2	Level2
M7_VR695279_2	41	-0.394	Level2	37	52.44	32	35.6	46	58.4	Level2	Level3
M7_VR697452	42	-0.223	Level2	38	58.6	31	33.72	45	50.71	Level2	Level3
M7_VR697154	43	-0.1527	Level3	39	61.2	30	33.02	44	47.61	Level2	Level3
M7_VR695990_2	44	-0.089	Level2	40	63.62	31	32.45	43	44.87	Level2	Level3
M7_VR695165	45	-0.0807	Level2	41	63.94	30	32.38	42	44.53	Level2	Level3
M7_VR695811	46	-0.0366	Level3	42	65.71	29	32.07	41	42.72	Level2	Level3
M7_VR697977	47	0.0376	Level3	43	68.75	30	31.63	40	39.75	Level2	Level3
M7_VR695154	48	0.1124	Level2	44	71.89	31	31.25	39	36.83	Level2	Level3
M7_VR695879_1	49	0.128	Level2	45	72.56	30	31.19	38	36.24	Level2	Level3
M7_VR696840_1	50	0.151	Level4	46	73.57	29	31.12	37	35.39	Level2	Level3
M7_VR697191	51	0.2077	Level2	47	76.13	30	31.01	38	33.35	Level2	Level3
M7_VR694902	52	0.2471	Level2	48	77.94	29	30.97	37	31.97	Level2	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M7_VR695784_1	53	0.341	Level2	49	82.35	28	30.97	36	28.78	Level3	Level3
M7_VR697271	54	0.3791	Level2	50	84.18	27	31.01	35	27.52	Level3	Level3
M7_VR695879_2	55	0.389	Level2	51	84.66	26	31.03	34	27.2	Level3	Level3
M7_VR698111	56	0.4613	Level3	52	88.28	25	31.24	33	24.96	Level3	Level3
M7_VR698408	57	0.5048	Level2	53	90.5	26	31.42	32	23.66	Level3	Level3
M7_VR696360	58	0.5504	Level3	54	92.87	25	31.64	31	22.33	Level3	Level3
M7_VR696247_1	59	0.554	Level2	55	93.06	26	31.67	30	22.23	Level3	Level3
M7_VR697408	60	0.5937	Level3	56	95.2	25	31.94	29	21.16	Level3	Level3
M7_VR696296_1	61	0.629	Level2	57	97.15	26	32.23	28	20.24	Level3	Level3
M7_VR695504	62	0.6777	Level3	58	99.87	25	32.67	27	19.02	Level3	Level3
M7_VR695850	63	0.764	Level3	59	104.79	26	33.53	26	16.95	Level3	Level3
M7_VR696442	64	0.85	Level2	60	109.78	27	34.47	25	14.98	Level3	Level3
M7_VR698194	65	0.8806	Level2	61	111.59	26	34.84	24	14.3	Level3	Level3
M7_VR695768	66	0.9235	Level3	62	114.16	25	35.4	23	13.4	Level3	Level3
M7_VR697488	67	1.013	Level2	63	119.62	26	36.65	22	11.61	Level3	Level3
M7_VR696840_2	68	1.098	Level4	64	124.89	25	37.93	21	10	Level3	Level3
M7_VR696298_1	69	1.099	Level3	65	124.95	26	37.94	22	9.98	Level3	Level3
M7_VR694730	70	1.1793	Level3	66	130.09	27	39.31	21	8.61	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M7_VR694360	71	1.236	Level3	67	133.78	28	40.33	20	7.71	Level3	Level3
M7_VR696296_2	72	1.26	Level2	68	135.36	29	40.78	19	7.35	Level3	Level3
M7_VR695384	73	1.2639	Level3	69	135.62	28	40.86	18	7.29	Level3	Level3
M7_VR696480	74	1.3085	Level3	70	138.65	29	41.8	17	6.71	Level3	Level3
M7_VR696247_2	75	1.312	Level3	71	138.9	30	41.88	16	6.67	Level3	Level3
M7_VR695630	76	1.34	Level3	72	140.86	31	42.52	15	6.36	Level3	Level3
M7_VR696510	77	1.3596	Level3	73	142.25	32	42.99	14	6.17	Level3	Level3
M7_VR694568	78	1.3957	Level3	74	144.85	33	43.89	13	5.84	Level3	Level3
M7_VR695784_2	79	1.409	Level2	75	145.82	34	44.24	12	5.73	Level3	Level3
M7_VR695777	80	1.5449	Level3	76	155.87	33	47.91	11	4.78	Level3	Level3
M7_VR695506	81	1.5856	Level3	77	158.93	34	49.05	10	4.54	Level3	Level3
M7_VR696298_2	82	1.618	Level3	78	161.39	35	49.99	9	4.38	Level3	Level3
M7_VR697336	83	1.6356	Level2	79	162.74	36	50.52	8	4.31	Level3	Level3
M7_VR695307_1	84	1.693	Level3	80	167.22	35	52.29	7	4.13	Level3	Level3
M7_VR697150	85	1.7428	Level3	81	171.16	36	53.89	6	4.04	Level3	Level3
M7_VR696869	86	1.8481	Level3	82	179.58	37	57.36	5	3.93	Level3	Level4
M7_VR698425	87	2.2337	Level1	83	210.81	38	70.47	4	3.93	Level4	Level4
M7_VR695307_2	88	2.399	Level4	82	224.37	37	76.26	3	4.09	Level4	Level4

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M7_VR694872_2	89	2.945	Level2	83	269.69	38	95.92	4	5.19	Level4	Level4

Table F6. Detailed ESS Item Maps: Mathematics Grade 8

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M8_VR698235_1	1	-2.749	Level1	46	71.17	71	135.3	88	195.58	Level1	Level1
M8_VR694519_1	2	-2.482	Level1	45	59.15	70	116.61	87	172.35	Level1	Level1
M8_VR698066	3	-2.4811	Level2	44	59.11	69	116.55	86	172.27	Level1	Level1
M8_VR695679	4	-2.2171	Level1	45	47.76	68	98.6	85	149.83	Level1	Level1
M8_VR696785	5	-2.1797	Level1	44	46.19	67	96.09	84	146.69	Level1	Level1
M8_VR694764	6	-2.1459	Level1	43	44.8	66	93.86	83	143.88	Level1	Level1
M8_VR694875	7	-2.1062	Level1	42	43.21	65	91.28	82	140.63	Level1	Level1
M8_VR695015	8	-2.0697	Level1	41	41.79	64	88.94	81	137.67	Level1	Level1
M8_VR694509	9	-2.0043	Level1	40	39.31	63	84.82	80	132.44	Level1	Level1
M8_VR697576_1	10	-1.96	Level1	39	37.67	62	82.08	79	128.94	Level1	Level1
M8_VR694379	11	-1.9332	Level1	38	36.7	61	80.44	78	126.85	Level1	Level1
M8_VR695584	12	-1.9025	Level2	37	35.63	60	78.6	77	124.49	Level1	Level1
M8_VR694929	13	-1.8779	Level1	38	34.79	59	77.15	76	122.62	Level1	Level2
M8_VR694725	14	-1.8627	Level1	37	34.29	58	76.27	75	121.48	Level1	Level2
M8_VR694814_1	15	-1.838	Level1	36	33.5	57	74.86	74	119.65	Level1	Level2
M8_VR697244	16	-1.7796	Level1	35	31.69	56	71.59	73	115.39	Level1	Level2
M8_VR694643_1	17	-1.763	Level1	34	31.19	55	70.67	72	114.19	Level1	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M8_VR694624	18	-1.7596	Level4	33	31.09	54	70.49	71	113.95	Level1	Level2
M8_VR694659	19	-1.756	Level1	34	30.99	55	70.3	72	113.7	Level1	Level2
M8_VR698235_2	20	-1.695	Level1	33	29.34	54	67.13	71	109.49	Level1	Level2
M8_VR697147	21	-1.666	Level1	32	28.59	53	65.65	70	107.52	Level1	Level2
M8_VR697175	22	-1.6472	Level1	31	28.12	52	64.71	69	106.26	Level1	Level2
M8_VR694545	23	-1.6029	Level1	30	27.06	51	62.54	68	103.33	Level1	Level2
M8_VR695741	24	-1.596	Level1	29	26.9	50	62.21	67	102.88	Level1	Level2
M8_VR695315	25	-1.572	Level1	28	26.37	49	61.08	66	101.35	Level1	Level2
M8_VR697804	26	-1.5684	Level1	27	26.29	48	60.91	65	101.12	Level1	Level2
M8_VR697180	27	-1.557	Level2	26	26.07	47	60.4	64	100.42	Level1	Level2
M8_VR698395_1	28	-1.544	Level1	27	25.82	46	59.83	63	99.62	Level1	Level2
M8_VR694473_1	29	-1.51	Level1	26	25.21	45	58.37	62	97.58	Level1	Level2
M8_VR697014	30	-1.4773	Level1	25	24.65	44	56.99	61	95.65	Level1	Level2
M8_VR696305_1	31	-1.45	Level2	24	24.21	43	55.87	60	94.07	Level1	Level2
M8_VR694582	32	-1.4336	Level1	25	23.97	42	55.22	59	93.13	Level1	Level2
M8_VR697164	33	-1.4045	Level1	24	23.56	41	54.08	58	91.51	Level1	Level2
M8_VR695064	34	-1.3923	Level2	23	23.4	40	53.62	57	90.83	Level1	Level2
M8_VR694790	35	-1.3312	Level1	24	22.67	39	51.36	56	87.53	Level1	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M8_VR695987	36	-1.2739	Level3	23	22.04	38	49.3	55	84.5	Level1	Level2
M8_VR695719	37	-1.227	Level1	24	21.57	39	47.65	54	82.06	Level1	Level2
M8_VR694957	38	-1.201	Level1	23	21.34	38	46.77	53	80.73	Level1	Level2
M8_VR698445	39	-1.1856	Level1	22	21.21	37	46.26	52	79.96	Level1	Level2
M8_VR696257	40	-1.127	Level1	21	20.8	36	44.39	51	77.09	Level1	Level2
M8_VR694527	41	-1.0906	Level1	20	20.58	35	43.26	50	75.34	Level1	Level2
M8_VR695510	42	-1.0831	Level1	19	20.55	34	43.03	49	74.99	Level1	Level2
M8_VR694519_2	43	-0.936	Level2	18	19.96	33	38.77	48	68.23	Level1	Level2
M8_VR695038	44	-0.861	Level3	19	19.73	32	36.67	47	64.85	Level1	Level2
M8_VR696277	45	-0.8066	Level1	20	19.62	33	35.2	46	62.46	Level1	Level2
M8_VR696949	46	-0.8008	Level1	19	19.62	32	35.05	45	62.21	Level1	Level2
M8_VR695828	47	-0.7594	Level3	18	19.62	31	34.01	44	60.47	Level2	Level2
M8_VR696344	48	-0.7307	Level2	19	19.65	32	33.32	43	59.29	Level2	Level2
M8_VR694473_2	49	-0.708	Level2	20	19.69	31	32.8	42	58.38	Level2	Level2
M8_VR695926	50	-0.6078	Level2	21	19.99	30	30.6	41	54.48	Level2	Level2
M8_VR695181	51	-0.4359	Level3	22	20.68	29	26.99	40	47.94	Level2	Level2
M8_VR694814_2	52	-0.297	Level1	23	21.38	30	24.21	39	42.8	Level2	Level3
M8_VR694575	53	-0.2949	Level2	22	21.39	29	24.17	38	42.73	Level2	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M8_VR698395_2	54	-0.23	Level1	23	21.84	28	23	37	40.46	Level2	Level3
M8_VR694975	55	-0.193	Level1	22	22.14	27	22.37	36	39.2	Level2	Level3
M8_VR696296_1	56	-0.117	Level2	21	22.82	26	21.16	35	36.69	Level2	Level3
M8_VR695056	57	-0.0984	Level2	22	23.01	25	20.88	34	36.1	Level2	Level3
M8_VR696567_1	58	-0.076	Level2	23	23.26	24	20.56	33	35.4	Level2	Level3
M8_VR696923_1	59	0.014	Level2	24	24.34	23	19.39	32	32.7	Level2	Level3
M8_VR698466	60	0.1869	Level3	25	26.58	22	17.32	31	27.69	Level2	Level3
M8_VR696305_2	61	0.284	Level2	26	27.94	23	16.25	30	24.97	Level2	Level3
M8_VR695307_1	62	0.378	Level2	27	29.35	22	15.31	29	22.43	Level2	Level3
M8_VR696333	63	0.3868	Level1	28	29.49	21	15.23	28	22.2	Level2	Level3
M8_VR696754	64	0.4302	Level2	27	30.23	20	14.88	27	21.12	Level2	Level3
M8_VR695107	65	0.4616	Level2	28	30.8	19	14.66	26	20.36	Level2	Level3
M8_VR696296_2	66	0.476	Level2	29	31.07	18	14.58	25	20.03	Level2	Level3
M8_VR696243	67	0.5184	Level1	30	31.92	17	14.37	24	19.1	Level2	Level3
M8_VR696538	68	0.6708	Level3	29	35.12	16	13.76	23	15.9	Level2	Level3
M8_VR695341	69	0.6787	Level2	30	35.29	17	13.73	22	15.74	Level2	Level3
M8_VR698307	70	0.7247	Level3	31	36.35	16	13.64	21	14.87	Level2	Level3
M8_VR698469	71	0.7327	Level3	32	36.54	17	13.63	20	14.72	Level2	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
M8_VR696158	72	0.7505	Level2	33	36.99	18	13.63	19	14.42	Level2	Level3
M8_VR696393	73	0.8571	Level2	34	39.76	17	13.74	18	12.71	Level3	Level3
M8_VR696669	74	0.8821	Level3	35	40.43	16	13.79	17	12.34	Level3	Level3
M8_VR696597	75	0.8902	Level3	36	40.66	17	13.81	16	12.23	Level3	Level3
M8_VR694643_2	76	0.926	Level2	37	41.7	18	13.96	15	11.76	Level3	Level3
M8_VR696567_2	77	1.123	Level2	38	47.61	17	14.94	14	9.4	Level3	Level3
M8_VR698481	78	1.1586	Level1	39	48.71	16	15.16	13	9.01	Level3	Level3
M8_VR696795	79	1.3456	Level3	38	54.7	15	16.46	12	7.14	Level3	Level3
M8_VR697576_2	80	1.379	Level1	39	55.8	16	16.73	11	6.83	Level3	Level3
M8_VR697165	81	1.3813	Level2	38	55.88	15	16.75	10	6.82	Level3	Level3
M8_VR695307_2	82	1.422	Level3	39	57.3	14	17.16	9	6.53	Level3	Level3
M8_VR694906	83	1.4843	Level3	40	59.54	15	17.84	8	6.16	Level3	Level3
M8_VR696923_2	84	1.565	Level3	41	62.53	16	18.81	7	5.75	Level3	Level3
M8_VR697908	85	1.6231	Level1	42	64.74	17	19.57	6	5.52	Level3	Level3
M8_VR696502	86	1.6597	Level1	41	66.16	16	20.08	5	5.41	Level3	Level3
M8_VR695892	87	1.987	Level3	40	79.26	15	24.99	4	4.76	Level3	Level4
M8_VR696880	88	2.3156	Level3	41	92.73	16	30.25	3	4.43	Level4	Level4
M8_VR696703	89	2.6693	Level3	42	107.58	17	36.26	2	4.43	Level4	Level4

Table F7. Detailed ESS Item Maps: Mathematics High School

				Level 2		Level 3		Level 4			
ID	**OOD**	**LOC**	**Consensus Level**	**Count**	**Weight**	**Count**	**Weight**	**Count**	**Weight**	**Initial Empirical Level**	**Final Empirical Level**
MHS_VR688105_1	1	-2.363	Level2	3	4.6	16	30.57	28	58.12	Level1	Level1
MHS_VR688119_1	2	-1.604	Level3	4	3.08	15	19.18	27	37.62	Level1	Level2
MHS_VR688107	3	-1.3896	Level3	5	2.87	16	16.18	26	32.05	Level1	Level2
MHS_VR688116_1	4	-1.211	Level2	6	2.87	17	13.86	25	27.58	Level2	Level3
MHS_VR688080	5	-1.1924	Level2	7	2.89	16	13.64	24	27.14	Level2	Level3
MHS_VR688074	6	-1.1347	Level1	8	3	15	13	23	25.81	Level2	Level3
MHS_VR688089	7	-1.1172	Level3	7	3.06	14	12.83	22	25.43	Level2	Level3
MHS_VR710121	8	-1.0643	Level4	8	3.27	15	12.35	21	24.32	Level2	Level3
MHS_VR710122	9	-0.9518	Level4	9	3.83	16	11.45	22	22.07	Level2	Level3
MHS_VR688088	10	-0.9167	Level2	10	4.04	17	11.21	23	21.4	Level2	Level3
MHS_VR710124	11	-0.857	Level1	11	4.46	16	10.85	22	20.32	Level2	Level3
MHS_VR688105_2	12	-0.689	Level3	10	5.8	15	10.01	21	17.47	Level2	Level3
MHS_VR710125	13	-0.6167	Level3	11	6.45	16	9.72	20	16.31	Level2	Level3
MHS_VR710123	14	-0.5632	Level4	12	6.99	17	9.56	19	15.51	Level2	Level3
MHS_VR688070	15	-0.5522	Level2	13	7.11	18	9.54	20	15.35	Level2	Level3
MHS_VR688122	16	-0.5184	Level2	14	7.52	17	9.5	19	14.92	Level2	Level3
MHS_VR688102_1	17	-0.495	Level1	15	7.82	16	9.5	18	14.63	Level2	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
MHS_VR688111	18	-0.3906	Level2	14	9.28	15	9.61	17	13.49	Level3	Level3
MHS_VR688118	19	-0.3822	Level2	15	9.41	14	9.62	16	13.4	Level3	Level3
MHS_VR688068	20	-0.1503	Level3	16	13.12	13	10.32	15	11.32	Level3	Level3
MHS_VR688124	21	-0.0932	Level2	17	14.09	14	10.55	14	10.86	Level3	Level3
MHS_VR688091	22	0.002	Level2	18	15.8	13	11.02	13	10.19	Level3	Level3
MHS_VR688117	23	0.1799	Level4	19	19.18	12	12.09	12	9.12	Level3	Level4
MHS_VR688069	24	0.246	Level3	20	20.5	13	12.55	13	8.79	Level3	Level4
MHS_VR710128	25	0.2573	Level4	21	20.74	14	12.64	12	8.75	Level3	Level4
MHS_VR688083	26	0.3296	Level2	22	22.33	15	13.29	13	8.53	Level3	Level4
MHS_VR688114	27	0.5019	Level3	23	26.29	14	15.02	12	8.19	Level3	Level4
MHS_VR688119_2	28	0.554	Level3	24	27.55	15	15.59	11	8.14	Level3	Level4
MHS_VR688081	29	0.555	Level3	25	27.57	16	15.6	10	8.14	Level3	Level4
MHS_VR688116_2	30	0.632	Level2	26	29.57	17	16.6	9	8.21	Level4	Level4
MHS_VR688104	31	1.3067	Level3	27	47.79	16	26.05	8	9.56	Level4	Level4
MHS_VR688071	32	1.5953	Level3	28	55.87	17	30.38	7	10.43	Level4	Level4
MHS_VR688102_2	33	1.904	Level2	29	64.82	18	35.32	6	11.66	Level4	Level4
MHS_VR688079	34	2.166	Level4	30	72.68	17	39.77	5	12.97	Level4	Level4

Table F8. Detailed ESS Item Maps: Reading Grade 3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R3_VR702675_1	1	-3.429	Level2	1	2.6	43	99.63	74	241.47	Level1	Level1
R3_VR699667_1	2	-2.498	Level2	2	2.6	42	60.53	73	173.51	Level2	Level1
R3_VR701589	3	-2.4693	Level2	3	2.63	41	59.35	72	171.44	Level2	Level1
R3_VR699670_1	4	-2.268	Level2	4	3.03	40	51.3	71	157.15	Level2	Level1
R3_VR701587	5	-2.2385	Level2	5	3.12	39	50.15	70	155.09	Level2	Level1
R3_VR701540	6	-1.8689	Level2	6	4.6	38	36.11	69	129.58	Level2	Level2
R3_VR701594	7	-1.8511	Level2	7	4.69	37	35.45	68	128.37	Level2	Level2
R3_VR701538	8	-1.8357	Level2	8	4.78	36	34.89	67	127.34	Level2	Level2
R3_VR701536	9	-1.7855	Level2	9	5.13	35	33.14	66	124.03	Level2	Level2
R3_VR701591	10	-1.7787	Level2	10	5.19	34	32.9	65	123.59	Level2	Level2
R3_VR701588	11	-1.752	Level2	11	5.43	33	32.02	64	121.88	Level2	Level2
R3_VR701535	12	-1.6824	Level2	12	6.12	32	29.8	63	117.49	Level2	Level2
R3_VR701586	13	-1.6324	Level2	13	6.67	31	28.25	62	114.39	Level2	Level2
R3_VR701592	14	-1.4992	Level2	14	8.27	30	24.25	61	106.27	Level2	Level2
R3_VR702677	15	-1.4653	Level2	15	8.71	29	23.27	60	104.23	Level2	Level2
R3_VR701053	16	-1.3739	Level2	16	9.99	28	20.71	59	98.84	Level2	Level2
R3_VR701593	17	-1.2508	Level2	17	11.84	27	17.38	58	91.7	Level2	Level2
R3_VR702674	18	-1.172	Level2	18	13.1	26	15.34	57	87.21	Level2	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R3_VR699668	19	-1.1322	Level2	19	13.78	25	14.34	56	84.98	Level2	Level2
R3_VR701056	20	-1.1219	Level2	20	13.96	24	14.09	55	84.41	Level2	Level2
R3_VR701537	21	-0.9409	Level2	21	17.4	23	9.93	54	74.64	Level2	Level2
R3_VR701051	22	-0.9253	Level2	22	17.71	22	9.59	53	73.81	Level2	Level2
R3_VR702671_1	23	-0.865	Level2	23	18.98	21	8.32	52	70.68	Level2	Level2
R3_VR702678	24	-0.8271	Level1	24	19.81	20	7.56	51	68.75	Level2	Level2
R3_VR702679	25	-0.8087	Level2	23	20.24	19	7.21	50	67.83	Level2	Level2
R3_VR699686_1	26	-0.802	Level2	24	20.4	18	7.09	49	67.5	Level2	Level2
R3_VR702689	27	-0.7083	Level2	25	22.74	17	5.5	48	63	Level2	Level2
R3_VR699666	28	-0.6805	Level2	26	23.46	16	5.06	47	61.69	Level2	Level2
R3_VR702751	29	-0.6452	Level2	27	24.42	15	4.53	46	60.07	Level2	Level2
R3_VR700971_1	30	-0.613	Level2	28	25.32	14	4.07	45	58.62	Level2	Level2
R3_VR699689_1	31	-0.593	Level2	29	25.9	13	3.81	44	57.74	Level2	Level3
R3_VR699669	32	-0.5358	Level2	30	27.61	12	3.13	43	55.28	Level2	Level3
R3_VR701532	33	-0.5059	Level2	31	28.54	11	2.8	42	54.02	Level2	Level3
R3_VR699688	34	-0.442	Level2	32	30.58	10	2.16	41	51.41	Level2	Level3
R3_VR699671	35	-0.4214	Level2	33	31.26	9	1.98	40	50.58	Level2	Level3
R3_VR701052	36	-0.4138	Level2	34	31.52	8	1.91	39	50.28	Level2	Level3
R3_VR701054	37	-0.3967	Level2	35	32.12	7	1.79	38	49.63	Level2	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R3_VR702746	38	-0.2758	Level2	36	36.47	6	1.07	37	45.16	Level2	Level3
R3_VR702686	39	-0.1861	Level2	37	39.79	5	0.62	36	41.93	Level2	Level3
R3_VR702748	40	-0.1551	Level2	38	40.97	4	0.5	35	40.85	Level2	Level3
R3_VR702673_1	41	-0.146	Level2	39	41.33	3	0.47	34	40.54	Level2	Level3
R3_VR699692	42	-0.1364	Level3	40	41.71	2	0.45	33	40.22	Level2	Level3
R3_VR702749	43	-0.0197	Level3	41	46.49	3	0.33	32	36.49	Level3	Level3
R3_VR699670_2	44	-0.009	Level3	42	46.94	4	0.33	31	36.16	Level3	Level3
R3_VR702676_1	45	-0.008	Level2	43	46.99	5	0.33	30	36.13	Level3	Level3
R3_VR702747	46	-0.0056	Level3	44	47.09	4	0.34	29	36.06	Level3	Level3
R3_VR699686_2	47	0.078	Level3	45	50.85	5	0.59	28	33.71	Level3	Level3
R3_VR702750	48	0.1854	Level2	46	55.79	6	1.02	27	30.81	Level3	Level3
R3_VR702675_2	49	0.226	Level3	47	57.7	5	1.22	26	29.76	Level3	Level3
R3_VR699689_2	50	0.292	Level3	48	60.87	6	1.62	25	28.11	Level3	Level3
R3_VR702671_2	51	0.332	Level3	49	62.83	7	1.9	24	27.15	Level3	Level3
R3_VR702683	52	0.3652	Level3	50	64.49	8	2.16	23	26.39	Level3	Level3
R3_VR699667_2	53	0.715	Level3	51	82.33	9	5.31	22	18.69	Level3	Level3
R3_VR702685	54	0.8656	Level3	52	90.16	10	6.82	21	15.53	Level3	Level3
R3_VR702676_2	55	0.875	Level3	53	90.66	11	6.92	20	15.34	Level3	Level3
R3_VR701838	56	0.8912	Level3	54	91.53	12	7.12	19	15.03	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R3_VR701055	57	0.8936	Level3	55	91.67	13	7.15	18	14.99	Level3	Level3
R3_VR701845	58	1.0343	Level3	56	99.55	14	9.12	17	12.6	Level3	Level3
R3_VR701843_1	59	1.068	Level3	57	101.47	15	9.62	16	12.06	Level3	Level3
R3_VR701841	60	1.2055	Level3	58	109.44	16	11.82	15	9.99	Level3	Level3
R3_VR701843_2	61	1.46	Level3	59	124.46	17	16.15	14	6.43	Level3	Level3
R3_VR700967	62	1.5227	Level3	60	128.22	18	17.28	13	5.62	Level3	Level4
R3_VR700960	63	1.5455	Level3	61	129.61	19	17.71	12	5.34	Level3	Level4
R3_VR700956	64	1.5502	Level3	62	129.9	20	17.81	11	5.29	Level3	Level4
R3_VR701839	65	1.69	Level3	63	138.71	21	20.74	10	3.89	Level3	Level4
R3_VR701840	66	1.833	Level3	64	147.86	22	23.89	9	2.61	Level3	Level4
R3_VR701844	67	1.8887	Level3	65	151.48	23	25.17	8	2.16	Level3	Level4
R3_VR700971_2	68	1.935	Level3	66	154.54	24	26.28	7	1.84	Level3	Level4
R3_VR700957	69	1.9706	Level3	67	156.92	25	27.17	6	1.62	Level3	Level4
R3_VR700955	70	2.1039	Level3	68	165.99	26	30.64	5	0.96	Level3	Level4
R3_VR700964	71	2.151	Level3	69	169.24	27	31.91	4	0.77	Level3	Level4
R3_VR700963	72	2.2696	Level3	70	177.54	28	35.23	3	0.41	Level3	Level4
R3_VR700973	73	2.4346	Level4	71	189.25	29	40.01	2	0.08	Level3	Level4
R3_VR702673_2	74	2.445	Level3	72	190	30	40.32	3	0.07	Level3	Level4
R3_VR700969	75	2.5067	Level3	73	194.51	31	42.24	2	0.07	Level3	Level4

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R3_VR700961	76	2.6377	Level4	74	204.2	32	46.43	1	0.2	Level4	Level4
R3_VR700966	77	2.8337	Level4	75	218.9	33	52.9	2	0.6	Level4	Level4
R3_VR700972	78	3.9511	Level4	76	303.82	34	90.89	3	3.95	Level4	Level4

Table F9. Detailed ESS Item Maps: Reading Grade 4

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R4_VR700688_1	1	-3.202	Level2	0	0	9	24.46	66	218.02	Level2	Level1
R4_VR700703	2	-2.0035	Level3	1	1.2	8	14.87	65	140.12	Level2	Level1
R4_VR702966	3	-1.7849	Level3	2	1.64	9	13.34	64	126.13	Level2	Level1
R4_VR702224_1	4	-1.527	Level3	3	2.41	10	11.79	63	109.88	Level2	Level1
R4_VR703417	5	-1.4755	Level3	4	2.62	11	11.53	62	106.68	Level2	Level1
R4_VR702962	6	-1.4099	Level3	5	2.94	12	11.27	61	102.68	Level2	Level1
R4_VR701363	7	-1.2359	Level2	6	3.99	13	10.75	60	92.24	Level2	Level2
R4_VR700124	8	-0.9916	Level4	7	5.7	12	10.26	59	77.83	Level2	Level2
R4_VR701360	9	-0.9728	Level2	8	5.85	13	10.24	60	76.74	Level2	Level2
R4_VR700128	10	-0.9175	Level4	9	6.35	12	10.24	59	73.59	Level3	Level2
R4_VR703419	11	-0.8942	Level2	10	6.58	13	10.26	60	72.28	Level3	Level2
R4_VR703420	12	-0.7871	Level3	11	7.76	12	10.48	59	66.39	Level3	Level2
R4_VR700688_2	13	-0.708	Level3	12	8.71	13	10.72	58	62.12	Level3	Level2
R4_VR702957	14	-0.6719	Level3	13	9.18	14	10.86	57	60.21	Level3	Level2
R4_VR702963	15	-0.5762	Level4	14	10.52	15	11.34	56	55.23	Level3	Level2
R4_VR702958	16	-0.5524	Level3	15	10.87	16	11.48	57	54.02	Level3	Level2
R4_VR700120_1	17	-0.439	Level2	16	12.69	17	12.28	56	48.35	Level3	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R4_VR700693	18	-0.4285	Level3	17	12.86	16	12.36	55	47.83	Level3	Level2
R4_VR700126	19	-0.4095	Level3	18	13.21	17	12.53	54	46.92	Level3	Level2
R4_VR700686	20	-0.36	Level3	19	14.15	18	13.03	53	44.59	Level3	Level2
R4_VR700701	21	-0.3256	Level3	20	14.84	19	13.4	52	43.01	Level3	Level2
R4_VR702973	22	-0.3187	Level3	21	14.98	20	13.49	51	42.7	Level3	Level2
R4_VR700687	23	-0.2457	Level3	22	16.59	21	14.44	50	39.49	Level3	Level3
R4_VR703422	24	-0.239	Level3	23	16.74	22	14.53	49	39.2	Level3	Level3
R4_VR702959	25	-0.2191	Level3	24	17.22	23	14.83	48	38.37	Level3	Level3
R4_VR700122	26	-0.2188	Level3	25	17.23	24	14.83	47	38.35	Level3	Level3
R4_VR700696	27	-0.1977	Level2	26	17.77	25	15.19	46	37.51	Level3	Level3
R4_VR700120_2	28	-0.185	Level3	27	18.12	24	15.42	45	37.01	Level3	Level3
R4_VR702975_1	29	-0.161	Level3	28	18.79	25	15.88	44	36.1	Level3	Level3
R4_VR700691	30	-0.154	Level3	29	18.99	26	16.02	43	35.84	Level3	Level3
R4_VR700697	31	-0.1339	Level3	30	19.6	27	16.44	42	35.12	Level3	Level3
R4_VR701842_1	32	-0.044	Level3	31	22.38	28	18.42	41	31.97	Level3	Level3
R4_VR700127	33	-0.0332	Level3	32	22.73	29	18.66	40	31.61	Level3	Level3
R4_VR701364	34	0.0005	Level3	33	23.84	30	19.47	39	30.49	Level3	Level3
R4_VR702222	35	0.0317	Level3	34	24.9	31	20.25	38	29.49	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R4_VR700702	36	0.1581	Level3	35	29.32	32	23.54	37	25.58	Level3	Level3
R4_VR701838	37	0.1784	Level3	36	30.06	33	24.09	36	24.97	Level3	Level3
R4_VR702224_2	38	0.251	Level3	37	32.74	34	26.12	35	22.86	Level3	Level3
R4_VR702970	39	0.2635	Level3	38	33.22	35	26.48	34	22.51	Level3	Level3
R4_VR701841	40	0.2752	Level2	39	33.67	36	26.83	33	22.2	Level3	Level3
R4_VR700690	41	0.3036	Level3	40	34.81	35	27.71	32	21.46	Level3	Level3
R4_VR703421	42	0.3081	Level3	41	34.99	36	27.86	31	21.35	Level3	Level3
R4_VR701361	43	0.3716	Level3	42	37.66	37	29.95	30	19.82	Level3	Level3
R4_VR700137	44	0.3849	Level3	43	38.23	38	30.41	29	19.52	Level3	Level3
R4_VR702974	45	0.3862	Level3	44	38.29	39	30.45	28	19.49	Level3	Level3
R4_VR700689_1	46	0.461	Level3	45	41.66	40	33.14	27	17.92	Level3	Level3
R4_VR702975_2	47	0.521	Level3	46	44.42	41	35.36	26	16.72	Level3	Level3
R4_VR700692	48	0.5615	Level3	47	46.32	42	36.9	25	15.95	Level3	Level3
R4_VR701843_1	49	0.607	Level3	48	48.5	43	38.68	24	15.13	Level3	Level3
R4_VR700689_2	50	0.61	Level3	49	48.65	44	38.8	23	15.08	Level3	Level3
R4_VR701845	51	0.6429	Level3	50	50.29	45	40.15	22	14.55	Level3	Level3
R4_VR700140	52	0.6518	Level3	51	50.75	46	40.52	21	14.42	Level3	Level3
R4_VR700121	53	0.6749	Level3	52	51.95	47	41.51	20	14.09	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R4_VR700138	54	0.8344	Level3	53	60.4	48	48.53	19	12.02	Level3	Level3
R4_VR702221	55	0.9386	Level3	54	66.03	49	53.22	18	10.77	Level3	Level3
R4_VR701843_2	56	0.995	Level3	55	69.13	50	55.81	17	10.15	Level3	Level3
R4_VR702226	57	1.0598	Level3	56	72.76	51	58.86	16	9.5	Level3	Level3
R4_VR700131	58	1.0837	Level3	57	74.12	52	60.01	15	9.29	Level3	Level3
R4_VR700139	59	1.1096	Level4	58	75.63	53	61.28	14	9.08	Level3	Level3
R4_VR700133_1	60	1.131	Level2	59	76.89	54	62.35	15	8.93	Level3	Level3
R4_VR700684	61	1.1322	Level3	60	76.96	53	62.41	14	8.92	Level3	Level3
R4_VR702223	62	1.1731	Level2	61	79.46	54	64.53	13	8.72	Level3	Level3
R4_VR702225_1	63	1.204	Level3	62	81.37	53	66.17	12	8.59	Level3	Level3
R4_VR701840	64	1.2406	Level3	63	83.68	54	68.15	11	8.48	Level3	Level3
R4_VR700698	65	1.306	Level3	64	87.86	55	71.75	10	8.35	Level3	Level3
R4_VR702225_2	66	1.333	Level3	65	89.62	56	73.26	9	8.33	Level3	Level3
R4_VR701842_2	67	1.449	Level3	66	97.27	57	79.87	8	8.33	Level4	Level3
R4_VR701844	68	1.5199	Level3	67	102.02	58	83.98	7	8.4	Level4	Level3
R4_VR702968	69	1.6058	Level4	68	107.86	59	89.05	6	8.57	Level4	Level3
R4_VR700133_2	70	1.943	Level3	69	131.13	60	109.28	7	9.58	Level4	Level4
R4_VR702967	71	2.0381	Level3	70	137.79	61	115.08	6	9.96	Level4	Level4

Table F10. Detailed ESS Item Maps: Reading Grade 5

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R5_VR702939_1	1	-3.155	Level3	2	3.07	9	17.26	68	216.7	Level1	Level1
R5_VR702952	2	-2.7549	Level2	3	2.67	10	14.06	67	189.89	Level1	Level1
R5_VR702941_1	3	-1.865	Level1	4	2.67	9	7.83	66	131.16	Level2	Level1
R5_VR702472	4	-1.5667	Level2	3	2.97	8	6.04	65	111.77	Level2	Level1
R5_VR702940_1	5	-1.37	Level1	4	3.37	7	5.06	64	99.18	Level2	Level2
R5_VR702941_2	6	-1.129	Level2	3	4.09	6	4.1	63	83.99	Level2	Level2
R5_VR702473	7	-0.971	Level3	4	4.72	5	3.62	62	74.2	Level3	Level2
R5_VR702940_2	8	-0.968	Level2	5	4.74	6	3.62	61	74.02	Level3	Level2
R5_VR699771	9	-0.9661	Level3	6	4.75	5	3.61	60	73.9	Level3	Level2
R5_VR699770	10	-0.9563	Level3	7	4.82	6	3.61	59	73.32	Level3	Level2
R5_VR699768	11	-0.9422	Level3	8	4.93	7	3.63	58	72.51	Level3	Level2
R5_VR702469	12	-0.9041	Level3	9	5.27	8	3.7	57	70.33	Level3	Level2
R5_VR701037	13	-0.8243	Level3	10	6.07	9	3.94	56	65.86	Level3	Level2
R5_VR699764	14	-0.8051	Level3	11	6.28	10	4.02	55	64.81	Level3	Level2
R5_VR702474	15	-0.7822	Level3	12	6.56	11	4.14	54	63.57	Level3	Level2
R5_VR702944	16	-0.7422	Level3	13	7.08	12	4.38	53	61.45	Level3	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R5_VR701039	17	-0.7414	Level3	14	7.09	13	4.38	52	61.41	Level3	Level2
R5_VR702470	18	-0.7189	Level3	15	7.43	14	4.56	51	60.26	Level3	Level2
R5_VR701041	19	-0.6838	Level2	16	7.99	15	4.88	50	58.51	Level3	Level2
R5_VR702471	20	-0.5824	Level3	17	9.71	14	5.89	49	53.54	Level3	Level2
R5_VR702942_1	21	-0.577	Level3	18	9.81	15	5.95	48	53.28	Level3	Level2
R5_VR703738	22	-0.5328	Level3	19	10.65	16	6.48	47	51.2	Level3	Level2
R5_VR699769	23	-0.4878	Level3	20	11.55	17	7.07	46	49.13	Level3	Level2
R5_VR699772	24	-0.4261	Level2	21	12.84	18	7.93	45	46.36	Level3	Level2
R5_VR701043	25	-0.3676	Level2	22	14.13	17	8.81	44	43.78	Level3	Level3
R5_VR701038	26	-0.2757	Level3	23	16.24	16	10.28	43	39.83	Level3	Level3
R5_VR702934	27	-0.2203	Level3	24	17.57	17	11.22	42	37.5	Level3	Level3
R5_VR702114	28	-0.219	Level3	25	17.61	18	11.24	41	37.45	Level3	Level3
R5_VR702951	29	-0.2068	Level3	26	17.92	19	11.47	40	36.96	Level3	Level3
R5_VR702113	30	-0.1982	Level3	27	18.16	20	11.65	39	36.63	Level3	Level3
R5_VR699763	31	-0.147	Level3	28	19.59	21	12.72	38	34.68	Level3	Level3
R5_VR701559	32	-0.1212	Level3	29	20.34	22	13.29	37	33.73	Level3	Level3
R5_VR702475	33	-0.1052	Level3	30	20.82	23	13.66	36	33.15	Level3	Level3
R5_VR702937	34	-0.1051	Level3	31	20.82	24	13.66	35	33.15	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R5_VR702953	35	-0.0098	Level3	32	23.87	25	16.04	34	29.91	Level3	Level3
R5_VR701560	36	0.0087	Level3	33	24.48	26	16.52	33	29.3	Level3	Level3
R5_VR702939_2	37	0.113	Level4	34	28.03	27	19.34	32	25.96	Level3	Level3
R5_VR702115	38	0.202	Level3	35	31.14	28	21.83	33	23.2	Level3	Level3
R5_VR700357	39	0.2129	Level3	36	31.53	29	22.15	32	22.87	Level3	Level3
R5_VR702116	40	0.2566	Level3	37	33.15	30	23.46	31	21.61	Level3	Level3
R5_VR701556	41	0.303	Level3	38	34.91	31	24.9	30	20.31	Level3	Level3
R5_VR702221	42	0.3799	Level3	39	37.91	32	27.36	29	18.23	Level3	Level3
R5_VR700352	43	0.5058	Level3	40	42.95	33	31.51	28	14.96	Level3	Level3
R5_VR701040	44	0.5217	Level3	41	43.6	34	32.05	27	14.56	Level3	Level3
R5_VR702112	45	0.5584	Level3	42	45.14	35	33.34	26	13.68	Level3	Level3
R5_VR702946_1	46	0.608	Level3	43	47.28	36	35.12	25	12.54	Level3	Level3
R5_VR702225_1	47	0.672	Level3	44	50.09	37	37.49	24	11.13	Level3	Level3
R5_VR702223	48	0.7333	Level3	45	52.85	38	39.82	23	9.84	Level3	Level3
R5_VR702935	49	0.7497	Level3	46	53.6	39	40.46	22	9.51	Level3	Level3
R5_VR702942_2	50	0.797	Level4	47	55.83	40	42.35	21	8.62	Level3	Level3
R5_VR702226	51	0.8148	Level3	48	56.68	41	43.08	22	8.3	Level3	Level3
R5_VR702225_2	52	0.82	Level3	49	56.94	42	43.3	21	8.21	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R5_VR702946_2	53	0.858	Level3	50	58.84	43	44.93	20	7.6	Level3	Level3
R5_VR702117	54	0.8696	Level3	51	59.43	44	45.44	19	7.43	Level3	Level3
R5_VR702949	55	0.878	Level3	52	59.86	45	45.82	18	7.31	Level3	Level3
R5_VR702943	56	0.9084	Level3	53	61.48	46	47.22	17	6.91	Level3	Level3
R5_VR701555	57	1.0054	Level3	54	66.71	47	51.78	16	5.75	Level3	Level3
R5_VR701557	58	1.0267	Level3	55	67.89	48	52.8	15	5.51	Level3	Level3
R5_VR700349	59	1.046	Level3	56	68.97	49	53.75	14	5.32	Level3	Level3
R5_VR702938	60	1.0629	Level3	57	69.93	50	54.59	13	5.17	Level3	Level3
R5_VR700354	61	1.0735	Level3	58	70.54	51	55.13	12	5.08	Level3	Level3
R5_VR702111	62	1.1012	Level3	59	72.18	52	56.57	11	4.89	Level3	Level3
R5_VR700350	63	1.1278	Level3	60	73.77	53	57.98	10	4.73	Level3	Level3
R5_VR699884	64	1.1546	Level3	61	75.41	54	59.43	9	4.6	Level3	Level3
R5_VR702932	65	1.3097	Level3	62	85.03	55	67.96	8	3.98	Level3	Level3
R5_VR703737	66	1.4849	Level3	63	96.06	56	77.77	7	3.45	Level3	Level3
R5_VR699882	67	1.4901	Level4	64	96.4	57	78.07	6	3.44	Level3	Level3
R5_VR700358	68	1.5782	Level3	65	102.12	58	83.18	7	3.35	Level4	Level4
R5_VR702933	69	1.692	Level3	66	109.63	59	89.89	6	3.35	Level4	Level4
R5_VR703743	70	1.704	Level4	67	110.44	60	90.61	5	3.36	Level4	Level4

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R5_VR703742	71	1.7591	Level3	68	114.18	61	93.97	6	3.47	Level4	Level4
R5_VR700351	72	1.9065	Level4	69	124.35	62	103.11	5	3.92	Level4	Level4
R5_VR699865	73	2.0696	Level4	70	135.77	63	113.39	6	4.57	Level4	Level4
R5_VR699880_1	74	2.301	Level3	71	152.2	64	128.2	7	5.73	Level4	Level4
R5_VR699880_2	75	2.324	Level4	72	153.86	65	129.69	6	5.86	Level4	Level4
R5_VR703739	76	2.4445	Level4	73	162.65	66	137.65	7	6.71	Level4	Level4

Table F11. Detailed ESS Item Maps: Reading Grade 6

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R6_VR699867_1	1	-2.693	Level2	5	5.97	27	40.58	85	225.49	Level1	Level1
R6_VR699883_1	2	-2.642	Level1	6	5.76	26	39.26	84	221.2	Level1	Level1
R6_VR699872_1	3	-2.581	Level2	5	5.58	25	37.73	83	216.14	Level1	Level1
R6_VR703548	4	-1.9616	Level3	6	4.34	24	22.87	82	165.35	Level1	Level1
R6_VR699878_1	5	-1.955	Level2	7	4.34	25	22.71	81	164.82	Level1	Level1
R6_VR702320	6	-1.919	Level2	8	4.34	24	21.92	80	161.94	Level2	Level1
R6_VR703547	7	-1.7738	Level2	9	4.48	23	18.87	79	150.46	Level2	Level1
R6_VR703549	8	-1.5918	Level2	10	4.85	22	15.23	78	136.27	Level2	Level1
R6_VR699881	9	-1.5902	Level2	11	4.85	21	15.2	77	136.15	Level2	Level1
R6_VR703546	10	-1.5589	Level2	12	4.98	20	14.64	76	133.77	Level2	Level1
R6_VR703544	11	-1.5534	Level3	13	5	19	14.55	75	133.35	Level2	Level1
R6_VR702318	12	-1.5141	Level2	14	5.24	20	13.92	74	130.45	Level2	Level1
R6_VR703550	13	-1.4503	Level3	15	5.69	19	12.96	73	125.79	Level2	Level2
R6_VR702319	14	-1.3803	Level1	16	6.25	20	11.98	72	120.75	Level2	Level2
R6_VR703738	15	-1.3299	Level2	15	6.7	19	11.32	71	117.17	Level2	Level2
R6_VR701266	16	-1.2936	Level1	16	7.06	18	10.89	70	114.63	Level2	Level2
R6_VR703552	17	-1.2666	Level1	15	7.36	17	10.59	69	112.77	Level2	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R6_VR701268	18	-1.1191	Level2	14	9.13	16	9.12	68	102.74	Level2	Level2
R6_VR701265	19	-1.107	Level2	15	9.29	15	9.01	67	101.93	Level2	Level2
R6_VR702321	20	-1.082	Level2	16	9.64	14	8.81	66	100.28	Level2	Level2
R6_VR702326	21	-1.0713	Level2	17	9.8	13	8.73	65	99.58	Level2	Level2
R6_VR703551	22	-0.943	Level3	18	11.85	12	7.96	64	91.37	Level2	Level2
R6_VR701263	23	-0.9353	Level2	19	11.98	13	7.93	63	90.88	Level2	Level2
R6_VR701267	24	-0.9136	Level1	20	12.37	12	7.84	62	89.54	Level2	Level2
R6_VR699871_1	25	-0.684	Level2	19	16.73	11	7.15	61	75.53	Level2	Level2
R6_VR702325	26	-0.6772	Level2	20	16.87	10	7.14	60	75.12	Level2	Level2
R6_VR703740	27	-0.6425	Level3	21	17.6	9	7.1	59	73.08	Level2	Level2
R6_VR699879	28	-0.6186	Level3	22	18.12	10	7.1	58	71.69	Level3	Level2
R6_VR702323	29	-0.5555	Level3	23	19.58	11	7.16	57	68.09	Level3	Level2
R6_VR703741	30	-0.4648	Level3	24	21.75	12	7.35	56	63.01	Level3	Level2
R6_VR702322	31	-0.4476	Level2	25	22.18	13	7.4	55	62.07	Level3	Level2
R6_VR703506_1	32	-0.445	Level2	26	22.25	12	7.41	54	61.93	Level3	Level2
R6_VR703501	33	-0.4024	Level3	27	23.4	11	7.62	53	59.67	Level3	Level2
R6_VR703653	34	-0.3525	Level3	28	24.8	12	7.92	52	57.08	Level3	Level3
R6_VR700291	35	-0.3425	Level3	29	25.09	13	7.99	51	56.57	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R6_VR703654	36	-0.2186	Level3	30	28.8	14	8.98	50	50.37	Level3	Level3
R6_VR701293	37	-0.1535	Level3	31	30.82	15	9.57	49	47.18	Level3	Level3
R6_VR700939	38	-0.1523	Level3	32	30.86	16	9.58	48	47.12	Level3	Level3
R6_VR703655	39	-0.1454	Level3	33	31.09	17	9.66	47	46.8	Level3	Level3
R6_VR703651	40	-0.124	Level3	34	31.82	18	9.91	46	45.81	Level3	Level3
R6_VR703652	41	-0.0924	Level3	35	32.92	19	10.32	45	44.39	Level3	Level3
R6_VR700942	42	0.062	Level3	36	38.48	20	12.48	44	37.6	Level3	Level3
R6_VR700944	43	0.0741	Level3	37	38.93	21	12.67	43	37.08	Level3	Level3
R6_VR700950	44	0.1541	Level3	38	41.97	22	13.95	42	33.72	Level3	Level3
R6_VR699883_2	45	0.169	Level3	39	42.55	23	14.2	41	33.11	Level3	Level3
R6_VR699874	46	0.1697	Level3	40	42.58	24	14.21	40	33.08	Level3	Level3
R6_VR700940_1	47	0.183	Level2	41	43.12	25	14.46	39	32.56	Level3	Level3
R6_VR699867_2	48	0.198	Level3	42	43.75	24	14.76	38	31.99	Level3	Level3
R6_VR700946_1	49	0.283	Level2	43	47.41	25	16.55	37	28.85	Level3	Level3
R6_VR703506_2	50	0.317	Level3	44	48.9	24	17.3	36	27.62	Level3	Level3
R6_VR700940_2	51	0.323	Level3	45	49.17	25	17.44	35	27.41	Level3	Level3
R6_VR700934	52	0.345	Level3	46	50.19	26	17.96	34	26.66	Level3	Level3
R6_VR701289	53	0.3571	Level3	47	50.75	27	18.27	33	26.26	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R6_VR703502	54	0.4011	Level3	48	52.87	28	19.41	32	24.86	Level3	Level3
R6_VR700941	55	0.4674	Level3	49	56.12	29	21.2	31	22.8	Level3	Level3
R6_VR703504	56	0.4702	Level3	50	56.26	30	21.28	30	22.72	Level3	Level3
R6_VR700938	57	0.4741	Level3	51	56.45	31	21.39	29	22.6	Level3	Level3
R6_VR703736	58	0.476	Level3	52	56.55	32	21.45	28	22.55	Level3	Level3
R6_VR700946_2	59	0.569	Level3	53	61.48	33	24.33	27	20.04	Level3	Level3
R6_VR700285	60	0.5815	Level3	54	62.16	34	24.73	26	19.72	Level3	Level3
R6_VR700936	61	0.6271	Level3	55	64.66	35	26.24	25	18.58	Level3	Level3
R6_VR700286	62	0.6548	Level3	56	66.22	36	27.18	24	17.91	Level3	Level3
R6_VR700284	63	0.7574	Level3	57	72.06	37	30.77	23	15.55	Level3	Level3
R6_VR701287	64	0.7588	Level3	58	72.15	38	30.82	22	15.52	Level3	Level3
R6_VR701290	65	0.8374	Level3	59	76.78	39	33.73	21	13.87	Level3	Level3
R6_VR699869_1	66	0.977	Level2	60	85.16	40	39.03	20	11.08	Level3	Level3
R6_VR699872_2	67	1.003	Level3	61	86.74	39	40.05	19	10.58	Level3	Level3
R6_VR699871_2	68	1.033	Level3	62	88.6	40	41.25	18	10.04	Level3	Level3
R6_VR700945_1	69	1.034	Level3	63	88.67	41	41.29	17	10.03	Level3	Level3
R6_VR703743	70	1.0612	Level3	64	90.41	42	42.43	16	9.59	Level3	Level3
R6_VR703737	71	1.1588	Level3	65	96.75	43	46.63	15	8.13	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R6_VR700289	72	1.1842	Level3	66	98.43	44	47.74	14	7.77	Level3	Level3
R6_VR700945_2	73	1.226	Level3	67	101.23	45	49.63	13	7.23	Level3	Level3
R6_VR699869_2	74	1.357	Level3	68	110.14	46	55.65	12	5.66	Level3	Level3
R6_VR700287	75	1.4416	Level3	69	115.97	47	59.63	11	4.73	Level3	Level3
R6_VR700282	76	1.5114	Level3	70	120.86	48	62.98	10	4.03	Level3	Level4
R6_VR703742	77	1.5941	Level3	71	126.73	49	67.03	9	3.28	Level3	Level4
R6_VR700947	78	1.7239	Level3	72	136.08	50	73.52	8	2.24	Level3	Level4
R6_VR703656	79	1.7538	Level3	73	138.26	51	75.05	7	2.04	Level3	Level4
R6_VR699878_2	80	1.835	Level3	74	144.27	52	79.27	6	1.55	Level3	Level4
R6_VR699875	81	1.8683	Level3	75	146.77	53	81.03	5	1.38	Level3	Level4
R6_VR701286	82	2.0269	Level3	76	158.82	54	89.6	4	0.75	Level3	Level4
R6_VR701283	83	2.1167	Level3	77	165.74	55	94.54	3	0.48	Level3	Level4
R6_VR701288	84	2.2846	Level3	78	178.83	56	103.94	2	0.14	Level3	Level4
R6_VR703739	85	2.4268	Level3	79	190.07	57	112.04	1	0	Level4	Level4

Table F12. Detailed ESS Item Maps: Reading Grade 7

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R7_VR700444_1	1	-2.937	Level1	8	18.34	60	150.82	91	273.21	Level1	Level1
R7_VR701897	2	-2.7531	Level2	7	17.05	59	139.97	90	256.66	Level1	Level1
R7_VR700529_1	3	-2.03	Level2	8	12.71	58	98.03	89	192.3	Level1	Level1
R7_VR700447_1	4	-1.967	Level2	9	12.4	57	94.44	88	186.76	Level1	Level1
R7_VR702192	5	-1.7523	Level2	10	11.54	56	82.41	87	168.08	Level1	Level1
R7_VR701642	6	-1.559	Level2	11	10.96	55	71.78	86	151.46	Level1	Level1
R7_VR700545	7	-1.4025	Level2	12	10.65	54	63.33	85	138.15	Level1	Level2
R7_VR702191	8	-1.3068	Level2	13	10.55	53	58.26	84	130.11	Level2	Level2
R7_VR701607	9	-1.3043	Level2	14	10.55	52	58.13	83	129.91	Level2	Level2
R7_VR700518_1	10	-1.269	Level2	15	10.58	51	56.33	82	127.01	Level2	Level2
R7_VR702187	11	-1.2387	Level1	16	10.65	50	54.81	81	124.56	Level2	Level2
R7_VR701640	12	-1.0628	Level2	15	11.17	49	46.19	80	110.49	Level2	Level2
R7_VR700544	13	-1.0612	Level3	16	11.18	48	46.12	79	110.36	Level2	Level2
R7_VR703654	14	-1.0115	Level2	17	11.43	49	43.78	78	106.48	Level2	Level2
R7_VR702190	15	-0.9744	Level1	18	11.65	48	42.08	77	103.63	Level2	Level2
R7_VR700548	16	-0.9622	Level2	17	11.74	47	41.53	76	102.7	Level2	Level2
R7_VR701605	17	-0.9496	Level2	18	11.84	46	40.97	75	101.75	Level2	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R7_VR700540	18	-0.9146	Level2	19	12.15	45	39.47	74	99.16	Level2	Level2
R7_VR700547	19	-0.9141	Level1	20	12.16	44	39.45	73	99.13	Level2	Level2
R7_VR700451	20	-0.8842	Level2	19	12.49	43	38.22	72	96.97	Level2	Level2
R7_VR701184	21	-0.8354	Level2	20	13.07	42	36.27	71	93.51	Level2	Level2
R7_VR700443_1	22	-0.825	Level2	21	13.21	41	35.86	70	92.78	Level2	Level2
R7_VR701183	23	-0.781	Level1	22	13.82	40	34.19	69	89.75	Level2	Level2
R7_VR700449	24	-0.7769	Level2	21	13.88	39	34.04	68	89.47	Level2	Level2
R7_VR700549	25	-0.7478	Level2	22	14.35	38	32.99	67	87.52	Level2	Level2
R7_VR702193	26	-0.7457	Level2	23	14.39	37	32.92	66	87.38	Level2	Level2
R7_VR700551	27	-0.693	Level2	24	15.33	36	31.13	65	83.95	Level2	Level2
R7_VR700539	28	-0.6463	Level2	25	16.22	35	29.59	64	80.96	Level2	Level2
R7_VR700546	29	-0.5847	Level3	26	17.45	34	27.61	63	77.08	Level2	Level2
R7_VR701887	30	-0.5668	Level3	27	17.83	35	27.06	62	75.97	Level2	Level2
R7_VR701634	31	-0.5383	Level2	28	18.46	36	26.2	61	74.24	Level2	Level2
R7_VR703655	32	-0.4995	Level2	29	19.35	35	25.08	60	71.91	Level2	Level2
R7_VR703652	33	-0.4046	Level2	30	21.63	34	22.42	59	66.31	Level2	Level2
R7_VR703653	34	-0.3675	Level2	31	22.55	33	21.42	58	64.16	Level2	Level2
R7_VR701182	35	-0.2946	Level2	32	24.45	32	19.52	57	60	Level2	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R7_VR701638	36	-0.2635	Level1	33	25.29	31	18.75	56	58.26	Level2	Level2
R7_VR700443_2	37	-0.241	Level2	32	25.92	30	18.21	55	57.02	Level2	Level2
R7_VR701886	38	-0.2013	Level2	33	27.07	29	17.29	54	54.88	Level2	Level2
R7_VR701894	39	-0.1981	Level2	34	27.17	28	17.22	53	54.71	Level2	Level3
R7_VR701615_1	40	-0.193	Level2	35	27.32	27	17.12	52	54.44	Level2	Level3
R7_VR700781	41	-0.16	Level2	36	28.38	26	16.46	51	52.76	Level2	Level3
R7_VR700444_2	42	-0.131	Level3	37	29.34	25	15.91	50	51.31	Level2	Level3
R7_VR702789	43	-0.067	Level2	38	31.51	26	14.75	49	48.17	Level2	Level3
R7_VR700447_2	44	-0.051	Level2	39	32.07	25	14.48	48	47.41	Level2	Level3
R7_VR703651	45	-0.0363	Level2	40	32.6	24	14.25	47	46.72	Level2	Level3
R7_VR700768	46	0.1828	Level2	41	40.71	23	10.96	46	36.64	Level2	Level3
R7_VR701181	47	0.2143	Level2	42	41.91	22	10.52	45	35.22	Level2	Level3
R7_VR700519	48	0.2256	Level2	43	42.35	21	10.37	44	34.72	Level2	Level3
R7_VR700770	49	0.3539	Level2	44	47.48	20	8.83	43	29.21	Level2	Level3
R7_VR701890	50	0.3655	Level3	45	47.95	19	8.7	42	28.72	Level2	Level3
R7_VR700553	51	0.4407	Level2	46	51.11	20	7.95	41	25.64	Level2	Level3
R7_VR700535	52	0.4551	Level2	47	51.73	19	7.82	40	25.06	Level2	Level3
R7_VR700558_1	53	0.497	Level2	48	53.58	18	7.49	39	23.43	Level2	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R7_VR702796	54	0.546	Level2	49	55.78	17	7.14	38	21.56	Level2	Level3
R7_VR700779	55	0.5471	Level1	50	55.83	16	7.14	37	21.52	Level2	Level3
R7_VR700782	56	0.5515	Level2	49	56.04	15	7.12	36	21.36	Level2	Level3
R7_VR701888	57	0.5623	Level3	50	56.56	14	7.07	35	20.99	Level2	Level3
R7_VR701606_1	58	0.597	Level2	51	58.26	15	6.97	34	19.81	Level2	Level3
R7_VR702788_1	59	0.608	Level2	52	58.81	14	6.95	33	19.44	Level2	Level3
R7_VR701608	60	0.6281	Level2	53	59.83	13	6.93	32	18.8	Level2	Level3
R7_VR700537	61	0.6302	Level2	54	59.94	12	6.93	31	18.73	Level2	Level3
R7_VR700778	62	0.6319	Level2	55	60.03	11	6.93	30	18.68	Level2	Level3
R7_VR701651	63	0.6675	Level2	56	61.95	10	7	29	17.65	Level3	Level3
R7_VR701891	64	0.7097	Level3	57	64.27	9	7.13	28	16.47	Level3	Level3
R7_VR700554	65	0.7223	Level3	58	64.98	10	7.18	27	16.13	Level3	Level3
R7_VR700532	66	0.7227	Level3	59	65	11	7.18	26	16.12	Level3	Level3
R7_VR700775	67	0.7607	Level2	60	67.21	12	7.41	25	15.17	Level3	Level3
R7_VR701641	68	0.7618	Level3	61	67.27	11	7.41	24	15.14	Level3	Level3
R7_VR700558_2	69	0.797	Level3	62	69.38	12	7.7	23	14.33	Level3	Level3
R7_VR701615_2	70	0.856	Level3	63	72.98	13	8.23	22	13.04	Level3	Level3
R7_VR702790	71	0.862	Level3	64	73.35	14	8.29	21	12.91	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R7_VR700772_1	72	0.91	Level3	65	76.38	15	8.81	20	11.95	Level3	Level3
R7_VR700780	73	0.9303	Level3	66	77.68	16	9.06	19	11.56	Level3	Level3
R7_VR700527	74	1.2507	Level3	67	98.5	17	13.22	18	5.8	Level3	Level3
R7_VR700536	75	1.2659	Level3	68	99.51	18	13.44	17	5.54	Level3	Level3
R7_VR700529_2	76	1.287	Level3	69	100.92	19	13.75	16	5.2	Level3	Level3
R7_VR702787_1	77	1.346	Level3	70	104.93	20	14.7	15	4.32	Level3	Level3
R7_VR702787_2	78	1.403	Level3	71	108.87	21	15.67	14	3.52	Level3	Level3
R7_VR701614	79	1.403	Level3	72	108.87	22	15.67	13	3.52	Level3	Level3
R7_VR701613	80	1.4035	Level1	73	108.9	23	15.68	12	3.51	Level3	Level3
R7_VR703656	81	1.409	Level3	72	109.3	22	15.79	11	3.45	Level3	Level3
R7_VR702795	82	1.4171	Level2	73	109.89	23	15.96	10	3.37	Level3	Level3
R7_VR700783	83	1.4709	Level3	74	113.87	22	17.14	9	2.89	Level3	Level3
R7_VR701650	84	1.5284	Level3	75	118.18	23	18.46	8	2.43	Level3	Level3
R7_VR702788_2	85	1.533	Level3	76	118.53	24	18.57	7	2.39	Level3	Level3
R7_VR701656_1	86	1.559	Level3	77	120.53	25	19.22	6	2.24	Level3	Level3
R7_VR701606_2	87	1.614	Level3	78	124.82	26	20.65	5	1.96	Level3	Level3
R7_VR700518_2	88	1.653	Level4	79	127.9	27	21.71	4	1.81	Level3	Level3
R7_VR701656_2	89	1.706	Level3	80	132.14	28	23.19	5	1.65	Level3	Level4

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R7_VR701655	90	1.7218	Level3	81	133.42	29	23.65	4	1.62	Level3	Level4
R7_VR700772_2	91	1.894	Level4	82	147.55	30	28.81	3	1.44	Level3	Level4
R7_VR700533	92	2.4453	Level3	83	193.3	31	45.9	4	1.44	Level4	Level4
R7_VR700526	93	2.5454	Level3	84	201.71	32	49.11	3	1.54	Level4	Level4

Table F13. Detailed ESS Item Maps: Reading Grade 8

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R8_VR700815	1	-1.9199	Level2	2	1.81	17	17.28	89	206.37	Level1	Level1
R8_VR701703_1	2	-1.554	Level2	3	1.44	16	11.42	88	174.18	Level1	Level1
R8_VR702185	3	-1.523	Level2	4	1.44	15	10.96	87	171.48	Level1	Level1
R8_VR701735	4	-1.2768	Level2	5	1.69	14	7.51	86	150.31	Level2	Level2
R8_VR701734	5	-1.2678	Level3	6	1.7	13	7.39	85	149.54	Level2	Level2
R8_VR700817	6	-1.1955	Level3	7	1.92	14	6.53	84	143.47	Level2	Level2
R8_VR701733	7	-1.0497	Level1	8	2.5	15	4.92	83	131.37	Level2	Level2
R8_VR700846_1	8	-1.043	Level2	7	2.54	14	4.86	82	130.82	Level2	Level2
R8_VR702182	9	-1.0217	Level2	8	2.67	13	4.66	81	129.09	Level2	Level2
R8_VR701736	10	-0.9845	Level1	9	2.93	12	4.37	80	126.12	Level2	Level2
R8_VR701698	11	-0.9186	Level2	8	3.45	11	3.91	79	120.91	Level2	Level2
R8_VR701704_1	12	-0.901	Level2	9	3.61	10	3.8	78	119.54	Level2	Level2
R8_VR700848	13	-0.897	Level2	10	3.65	9	3.78	77	119.23	Level2	Level2
R8_VR701731	14	-0.8924	Level2	11	3.7	8	3.76	76	118.88	Level2	Level2
R8_VR700821	15	-0.8517	Level2	12	4.19	7	3.64	75	115.83	Level2	Level2
R8_VR701738	16	-0.8376	Level2	13	4.37	6	3.61	74	114.78	Level2	Level2
R8_VR700834	17	-0.8231	Level3	14	4.58	5	3.6	73	113.72	Level2	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R8_VR702177	18	-0.7658	Level3	15	5.44	6	3.6	72	109.6	Level3	Level2
R8_VR701737	19	-0.7266	Level2	16	6.06	7	3.64	71	106.82	Level3	Level2
R8_VR702183	20	-0.6445	Level3	17	7.46	6	3.8	70	101.07	Level3	Level2
R8_VR700837	21	-0.4513	Level3	18	10.94	7	4.38	69	87.74	Level3	Level2
R8_VR700819	22	-0.4329	Level2	19	11.29	8	4.45	68	86.49	Level3	Level2
R8_VR701704_2	23	-0.311	Level3	20	13.72	7	5.06	67	78.32	Level3	Level2
R8_VR701699	24	-0.1385	Level3	21	17.35	8	6.1	66	66.93	Level3	Level2
R8_VR700823	25	0.0073	Level3	22	20.55	9	7.12	65	57.46	Level3	Level3
R8_VR700825_1	26	0.025	Level3	23	20.96	10	7.26	64	56.32	Level3	Level3
R8_VR703205	27	0.0511	Level3	24	21.59	11	7.5	63	54.68	Level3	Level3
R8_VR701732	28	0.0902	Level3	25	22.57	12	7.89	62	52.26	Level3	Level3
R8_VR701529	29	0.0958	Level3	26	22.71	13	7.95	61	51.91	Level3	Level3
R8_VR702179	30	0.1058	Level3	27	22.98	14	8.07	60	51.31	Level3	Level3
R8_VR701652	31	0.1572	Level3	28	24.42	15	8.74	59	48.28	Level3	Level3
R8_VR701702	32	0.1765	Level3	29	24.98	16	9.01	58	47.16	Level3	Level3
R8_VR700831	33	0.2425	Level3	30	26.96	17	10	57	43.4	Level3	Level3
R8_VR703209	34	0.2984	Level3	31	28.69	18	10.89	56	40.27	Level3	Level3
R8_VR701701	35	0.3168	Level3	32	29.28	19	11.2	55	39.26	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R8_VR702788_1	36	0.321	Level3	33	29.42	20	11.28	54	39.03	Level3	Level3
R8_VR700860	37	0.3348	Level3	34	29.89	21	11.54	53	38.3	Level3	Level3
R8_VR701700	38	0.3396	Level3	35	30.06	22	11.64	52	38.05	Level3	Level3
R8_VR700827_1	39	0.341	Level3	36	30.11	23	11.67	51	37.98	Level3	Level3
R8_VR700850	40	0.3433	Level3	37	30.19	24	11.72	50	37.86	Level3	Level3
R8_VR701530_1	41	0.348	Level3	38	30.37	25	11.83	49	37.63	Level3	Level3
R8_VR702796	42	0.378	Level3	39	31.54	26	12.55	48	36.19	Level3	Level3
R8_VR701653	43	0.412	Level3	40	32.9	27	13.4	47	34.6	Level3	Level3
R8_VR700822	44	0.4663	Level3	41	35.13	28	14.81	46	32.1	Level3	Level3
R8_VR700862	45	0.5575	Level3	42	38.96	29	17.27	45	27.99	Level3	Level3
R8_VR701521_1	46	0.601	Level3	43	40.83	30	18.49	44	26.08	Level3	Level3
R8_VR701651	47	0.6144	Level3	44	41.42	31	18.88	43	25.5	Level3	Level3
R8_VR702790	48	0.6163	Level3	45	41.5	32	18.93	42	25.42	Level3	Level3
R8_VR700816	49	0.6875	Level3	46	44.78	33	21.14	41	22.5	Level3	Level3
R8_VR700863	50	0.7637	Level3	47	48.36	34	23.58	40	19.46	Level3	Level3
R8_VR700840	51	0.7904	Level3	48	49.64	35	24.46	39	18.41	Level3	Level3
R8_VR703280	52	0.795	Level3	49	49.87	36	24.62	38	18.24	Level3	Level3
R8_VR703212	53	0.8003	Level3	50	50.13	37	24.8	37	18.04	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R8_VR700825_2	54	0.802	Level3	51	50.22	38	24.86	36	17.98	Level3	Level3
R8_VR703210	55	0.8514	Level3	52	52.79	39	26.69	35	16.25	Level3	Level3
R8_VR703208	56	0.9215	Level3	53	56.5	40	29.35	34	13.87	Level3	Level3
R8_VR700838	57	0.9252	Level3	54	56.7	41	29.5	33	13.75	Level3	Level3
R8_VR701530_2	58	0.932	Level3	55	57.08	42	29.77	32	13.53	Level3	Level3
R8_VR700851	59	0.9427	Level3	56	57.68	43	30.21	31	13.2	Level3	Level3
R8_VR700810	60	1.0508	Level3	57	63.84	44	34.75	30	9.96	Level3	Level3
R8_VR702787_1	61	1.069	Level3	58	64.89	45	35.53	29	9.43	Level3	Level3
R8_VR703207	62	1.0702	Level3	59	64.96	46	35.59	28	9.39	Level3	Level3
R8_VR700861	63	1.1144	Level3	60	67.62	47	37.57	27	8.2	Level3	Level3
R8_VR700842	64	1.1483	Level3	61	69.68	48	39.13	26	7.32	Level3	Level3
R8_VR702787_2	65	1.149	Level3	62	69.73	49	39.17	25	7.3	Level3	Level3
R8_VR701521_2	66	1.175	Level3	63	71.37	50	40.41	24	6.68	Level3	Level3
R8_VR702788_2	67	1.204	Level3	64	73.22	51	41.84	23	6.01	Level3	Level3
R8_VR700818	68	1.2042	Level3	65	73.23	52	41.85	22	6.01	Level3	Level3
R8_VR703206	69	1.229	Level3	66	74.87	53	43.11	21	5.49	Level3	Level3
R8_VR701656_1	70	1.235	Level3	67	75.27	54	43.42	20	5.37	Level3	Level3
R8_VR703279	71	1.2356	Level3	68	75.31	55	43.45	19	5.35	Level3	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R8_VR700864	72	1.2826	Level3	69	78.56	56	45.99	18	4.51	Level3	Level3
R8_VR703278	73	1.3002	Level3	70	79.79	57	46.96	17	4.21	Level3	Level3
R8_VR701524	74	1.3441	Level3	71	82.91	58	49.42	16	3.51	Level3	Level3
R8_VR703277	75	1.3678	Level3	72	84.61	59	50.77	15	3.15	Level3	Level3
R8_VR700847	76	1.3749	Level3	73	85.13	60	51.18	14	3.05	Level3	Level3
R8_VR700857	77	1.4173	Level3	74	88.27	61	53.68	13	2.5	Level3	Level3
R8_VR700813	78	1.4588	Level3	75	91.38	62	56.17	12	2	Level3	Level3
R8_VR701656_2	79	1.462	Level3	76	91.62	63	56.37	11	1.97	Level3	Level3
R8_VR701528_1	80	1.47	Level2	77	92.24	64	56.86	10	1.89	Level3	Level3
R8_VR701655	81	1.4875	Level3	78	93.61	63	57.97	9	1.73	Level3	Level3
R8_VR701650	82	1.4983	Level3	79	94.46	64	58.66	8	1.64	Level3	Level3
R8_VR701703_2	83	1.512	Level3	80	95.55	65	59.55	7	1.55	Level3	Level3
R8_VR700827_2	84	1.556	Level3	81	99.12	66	62.45	6	1.28	Level3	Level3
R8_VR700846_2	85	1.558	Level3	82	99.28	67	62.59	5	1.27	Level3	Level3
R8_VR703281_1	86	1.722	Level3	83	112.89	68	73.74	4	0.62	Level3	Level3
R8_VR703283	87	1.8423	Level3	84	123	69	82.04	3	0.26	Level3	Level3
R8_VR701528_2	88	1.9	Level3	85	127.9	70	86.08	2	0.14	Level3	Level3
R8_VR700845	89	2.0418	Level3	86	140.1	71	96.15	1	0	Level3	Level4

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
R8_VR700867	90	2.2822	Level4	87	161.01	72	113.45	0	0	Level4	Level4
R8_VR700868	91	2.6051	Level4	88	189.43	73	137.03	1	0.32	Level4	Level4
R8_VR703282	92	2.6673	Level4	89	194.96	74	141.63	2	0.45	Level4	Level4
R8_VR703281_2	93	2.863	Level4	90	212.58	75	156.31	3	1.03	Level4	Level4

Table F14. Detailed ESS Item Maps: Reading High School

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
RHS_VR703955	1	-2.0727	Level2	0	0	14	16.27	35	62.37	Level2	Level1
RHS_VR703974_1	2	-1.928	Level2	1	0.14	13	14.39	34	57.45	Level2	Level1
RHS_VR703954_1	3	-1.128	Level2	2	1.74	12	4.79	33	31.05	Level2	Level1
RHS_VR703970	4	-1.1217	Level3	3	1.76	11	4.72	32	30.85	Level2	Level1
RHS_VR703936_1	5	-1.088	Level2	4	1.9	12	4.39	31	29.8	Level2	Level1
RHS_VR703935	6	-1.057	Level2	5	2.05	11	4.11	30	28.87	Level2	Level1
RHS_VR703988_1	7	-1.027	Level2	6	2.23	10	3.87	29	28	Level2	Level1
RHS_VR703941_1	8	-0.991	Level2	7	2.49	9	3.62	28	26.99	Level2	Level1
RHS_VR703939	9	-0.9286	Level2	8	2.98	8	3.24	27	25.31	Level2	Level1
RHS_VR703943_1	10	-0.69	Level2	9	5.13	7	2.05	26	19.1	Level2	Level2
RHS_VR703954_2	11	-0.66	Level3	10	5.43	6	1.93	25	18.35	Level2	Level2
RHS_VR703985_1	12	-0.646	Level2	11	5.59	7	1.89	24	18.02	Level2	Level2
RHS_VR703975_1	13	-0.644	Level2	12	5.61	6	1.88	23	17.97	Level2	Level2
RHS_VR703953_1	14	-0.639	Level3	13	5.67	5	1.88	22	17.86	Level2	Level2
RHS_VR703984	15	-0.5859	Level2	14	6.42	6	1.88	21	16.75	Level2	Level2
RHS_VR703968	16	-0.4464	Level3	15	8.51	5	2.02	20	13.96	Level3	Level2
RHS_VR703969	17	-0.4078	Level3	16	9.13	6	2.09	19	13.22	Level3	Level2

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
RHS_VR703981	18	-0.3833	Level3	17	9.55	7	2.17	18	12.78	Level3	Level2
RHS_VR703980	19	-0.3807	Level3	18	9.59	8	2.18	17	12.74	Level3	Level2
RHS_VR703985_2	20	-0.054	Level3	19	15.8	9	3.81	16	7.51	Level3	Level3
RHS_VR703971	21	-0.0184	Level3	20	16.51	10	4.03	15	6.98	Level3	Level3
RHS_VR703937	22	-0.0082	Level2	21	16.73	11	4.1	14	6.83	Level3	Level3
RHS_VR703948_1	23	0.051	Level2	22	18.03	10	4.57	13	6.06	Level3	Level3
RHS_VR703982	24	0.1387	Level3	23	20.04	9	5.36	12	5.01	Level3	Level3
RHS_VR703975_2	25	0.167	Level3	24	20.72	10	5.64	11	4.7	Level3	Level3
RHS_VR703972	26	0.2504	Level3	25	22.81	11	6.56	10	3.87	Level3	Level3
RHS_VR703977	27	0.3107	Level3	26	24.38	12	7.28	9	3.32	Level3	Level3
RHS_VR703948_2	28	0.426	Level3	27	27.49	13	8.78	8	2.4	Level3	Level3
RHS_VR703987	29	0.4759	Level3	28	28.89	14	9.48	7	2.05	Level3	Level3
RHS_VR703943_2	30	0.527	Level3	29	30.37	15	10.25	6	1.75	Level3	Level3
RHS_VR703976_1	31	0.574	Level3	30	31.78	16	11	5	1.51	Level3	Level3
RHS_VR703944	32	0.6827	Level3	31	35.15	17	12.85	4	1.08	Level3	Level3
RHS_VR703951	33	0.7205	Level3	32	36.36	18	13.53	3	0.96	Level3	Level3
RHS_VR703988_2	34	0.722	Level4	33	36.41	19	13.56	2	0.96	Level3	Level3
RHS_VR703976_2	35	0.976	Level4	34	45.04	20	18.64	3	0.71	Level4	Level3

				Level 2		Level 3		Level 4			
ID	OOD	LOC	Consensus Level	Count	Weight	Count	Weight	Count	Weight	Initial Empirical Level	Final Empirical Level
RHS_VR703974_2	36	0.984	Level4	35	45.32	21	18.8	4	0.71	Level4	Level3
RHS_VR703941_2	37	1.055	Level3	36	47.88	22	20.37	5	0.78	Level4	Level3
RHS_VR703936_2	38	1.341	Level4	37	58.46	23	26.94	4	1.35	Level4	Level3
RHS_VR703950	39	1.3489	Level3	38	58.76	24	27.13	5	1.37	Level4	Level3
RHS_VR703953_2	40	1.43	Level4	39	61.92	25	29.16	4	1.7	Level4	Level3
RHS_VR703973	41	2.0422	Level4	40	86.41	26	45.08	5	4.76	Level4	Level4

Appendix G: Rosters of Inconsistent and Essentially Consistent Items

Table G1. Roster of Inconsistent Items: Mathematics Grade 3

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
M3	M3_VR694464	1	Level2	Level1	1	-1.6863	1.686
M3	M3_VR695822	2	Level2	Level1	1	-0.7326	0.733
M3	M3_VR696271	3	Level4	Level1	3	-3.5759	3.576
M3	M3_VR694605	5	Level4	Level2	2	-3.2658	3.266
M3	M3_VR696189	9	Level1	Level2	-1	1.5802	1.58
M3	M3_VR696443	15	Level1	Level2	-1	2.022	2.022
M3	M3_VR695848	19	Level3	Level2	1	-0.6442	0.644
M3	M3_VR695460_1	24	Level4	Level2	2	-1.764	1.764
M3	M3_VR695321	29	Level4	Level3	1	-1.2468	1.247
M3	M3_VR695248	31	Level2	Level3	-1	1.0836	1.084
M3	M3_VR694837_2	33	Level2	Level3	-1	1.166	1.166
M3	M3_VR695460_2	34	Level4	Level3	1	-1.108	1.108
M3	M3_VR696313	35	Level2	Level3	-1	1.1842	1.184
M3	M3_VR695405	36	Level1	Level3	-2	3.2389	3.239
M3	M3_VR696980	38	Level1	Level3	-2	3.2773	3.277
M3	M3_VR698159	41	Level2	Level3	-1	1.408	1.408
M3	M3_VR697278	45	Level4	Level3	1	-0.6925	0.693
M3	M3_VR695275	46	Level2	Level3	-1	1.6167	1.617
M3	M3_VR698100	47	Level2	Level3	-1	1.663	1.663
M3	M3_VR696129_2	50	Level4	Level3	1	-0.595	0.595
M3	M3_VR695352	66	Level2	Level4	-2	2.3171	2.317
M3	M3_VR695370	68	Level3	Level4	-1	1.0944	1.094
M3	M3_VR698072_2	69	Level3	Level4	-1	1.134	1.134
M3	M3_VR694754	70	Level3	Level4	-1	1.1455	1.146
M3	M3_VR695023	71	Level3	Level4	-1	1.1813	1.181
M3	M3_VR697132	72	Level3	Level4	-1	1.1908	1.191
M3	M3_VR697390	73	Level3	Level4	-1	1.2971	1.297
M3	M3_VR695336	74	Level3	Level4	-1	1.4193	1.419
M3	M3_VR695456	77	Level3	Level4	-1	1.6368	1.637
M3	M3_VR695685_2	79	Level3	Level4	-1	1.853	1.853
M3	M3_VR694658	83	Level3	Level4	-1	2.2592	2.259

Table G2. Roster of Essentially Consistent Items: Mathematics Grade 3

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
M3	M3_VR698098	26	Level3	Level2	1	-0.3675	0.367
M3	M3_VR697700	27	Level3	Level2	1	-0.1266	0.127
M3	M3_VR695953_1	54	Level4	Level3	1	-0.398	0.398
M3	M3_VR695792	57	Level4	Level3	1	-0.3586	0.359
M3	M3_VR694918_1	58	Level4	Level3	1	-0.333	0.333
M3	M3_VR697892	59	Level4	Level3	1	-0.273	0.273

Table G3. Roster of Inconsistent Items: Mathematics Grade 4

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
M4	M4_VR696271	4	Level1	Level2	-1	1.2004	1.2
M4	M4_VR696389	30	Level2	Level3	-1	1.6333	1.633
M4	M4_VR696034	31	Level4	Level3	1	-3.2794	3.279
M4	M4_VR695194	36	Level4	Level3	1	-3.0714	3.071
M4	M4_VR698164	39	Level2	Level3	-1	2.2634	2.263
M4	M4_VR694387_1	57	Level4	Level3	1	-1.364	1.364
M4	M4_VR696732	63	Level4	Level3	1	-1.0331	1.033
M4	M4_VR695354	67	Level4	Level3	1	-0.8081	0.808
M4	M4_VR694387_2	69	Level4	Level3	1	-0.542	0.542
M4	M4_VR695351	79	Level3	Level4	-1	1.0292	1.029
M4	M4_VR696115	80	Level3	Level4	-1	1.068	1.068
M4	M4_VR694853	81	Level3	Level4	-1	1.1964	1.196
M4	M4_VR698137	82	Level3	Level4	-1	1.2436	1.244
M4	M4_VR695528	83	Level3	Level4	-1	1.3623	1.362
M4	M4_VR698593_2	87	Level3	Level4	-1	1.573	1.573
M4	M4_VR697125	89	Level3	Level4	-1	1.8319	1.832

Table G4. Roster of Essentially Consistent Items: Mathematics Grade 4

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
M4	M4_VR694353_1	2	Level2	Level1	1	-0.346	0.346
M4	M4_VR694455	12	Level3	Level2	1	-0.3509	0.351
M4	M4_VR697854	17	Level3	Level2	1	-0.2221	0.222
M4	M4_VR695624	19	Level3	Level2	1	-0.1115	0.112

Table G5. Roster of Inconsistent Items: Mathematics Grade 5

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
M5	M5_VR694627_1	2	Level2	Level1	1	-0.801	0.801
M5	M5_VR694534_1	3	Level2	Level1	1	-0.716	0.716
M5	M5_VR695408	4	Level3	Level1	2	-2.535	2.535
M5	M5_VR695891	5	Level3	Level1	2	-2.4619	2.462
M5	M5_VR695604	6	Level4	Level1	3	-3.9322	3.932
M5	M5_VR694548	7	Level3	Level1	2	-2.3226	2.323
M5	M5_VR695159	13	Level1	Level2	-1	1.1282	1.128
M5	M5_VR696671	14	Level3	Level2	1	-1.6454	1.645
M5	M5_VR696423	16	Level3	Level2	1	-1.6096	1.61
M5	M5_VR696591	17	Level3	Level2	1	-1.5356	1.536
M5	M5_VR697134	18	Level3	Level2	1	-1.493	1.493
M5	M5_VR696813	19	Level1	Level2	-1	1.3643	1.364
M5	M5_VR695813	23	Level3	Level2	1	-1.3516	1.352
M5	M5_VR695566	24	Level1	Level2	-1	1.5303	1.53
M5	M5_VR696338	25	Level3	Level2	1	-1.2527	1.253
M5	M5_VR697929	30	Level3	Level2	1	-1.0073	1.007
M5	M5_VR694861_2	34	Level3	Level2	1	-0.849	0.849
M5	M5_VR698337	38	Level1	Level2	-1	2.5246	2.525
M5	M5_VR695799_2	39	Level1	Level2	-1	2.543	2.543
M5	M5_VR697558	41	Level1	Level2	-1	2.6096	2.61
M5	M5_VR696511	46	Level2	Level3	-1	1.0171	1.017
M5	M5_VR694701	47	Level2	Level3	-1	1.0652	1.065
M5	M5_VR697925_2	48	Level2	Level3	-1	1.09	1.09
M5	M5_VR694390	49	Level1	Level3	-2	2.947	2.947
M5	M5_VR697181_1	53	Level2	Level3	-1	1.263	1.263
M5	M5_VR696958	55	Level2	Level3	-1	1.3389	1.339
M5	M5_VR697388	58	Level2	Level3	-1	1.4392	1.439
M5	M5_VR696984	59	Level1	Level3	-2	3.3342	3.334
M5	M5_VR697707	68	Level4	Level3	1	-0.7185	0.718
M5	M5_VR697536	69	Level2	Level3	-1	1.8094	1.809
M5	M5_VR697133	72	Level2	Level3	-1	1.966	1.966
M5	M5_VR696500	74	Level2	Level3	-1	2.0331	2.033
M5	M5_VR698468	76	Level2	Level3	-1	2.0934	2.093
M5	M5_VR696215	78	Level1	Level3	-2	3.9484	3.948
M5	M5_VR695528	82	Level3	Level4	-1	1.0886	1.089
M5	M5_VR698593_2	83	Level3	Level4	-1	1.22	1.22
M5	M5_VR696088	85	Level2	Level4	-2	3.0918	3.092

Table G6. Roster of Essentially Consistent Items: Mathematics Grade 5

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
M5	M5_VR694415	10	Level2	Level1	1	-0.2517	0.252
M5	M5_VR694708	44	Level3	Level2	1	-0.1062	0.106
M5	M5_VR694838	45	Level3	Level2	1	-0.064	0.064

Table G7. Roster of Inconsistent Items: Mathematics Grade 6

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
M6	M6_VR695945	6	Level3	Level1	2	-3.0425	3.042
M6	M6_VR698188	7	Level2	Level1	1	-0.4209	0.421
M6	M6_VR695647	8	Level2	Level1	1	-0.4046	0.405
M6	M6_VR697411	18	Level1	Level2	-1	1.1288	1.129
M6	M6_VR694861_2	21	Level1	Level2	-1	1.235	1.235
M6	M6_VR695700	24	Level1	Level2	-1	1.306	1.306
M6	M6_VR694387_1	26	Level1	Level2	-1	1.365	1.365
M6	M6_VR695549	32	Level1	Level2	-1	1.5772	1.577
M6	M6_VR695372	37	Level1	Level2	-1	1.867	1.867
M6	M6_VR694603	38	Level3	Level2	1	-1.6666	1.667
M6	M6_VR694387_2	40	Level1	Level2	-1	2.126	2.126
M6	M6_VR697265	44	Level3	Level2	1	-1.327	1.327
M6	M6_VR696521	45	Level3	Level2	1	-1.2831	1.283
M6	M6_VR695877	54	Level3	Level2	1	-0.9553	0.955
M6	M6_VR698498_2	55	Level3	Level2	1	-0.946	0.946
M6	M6_VR694300	56	Level1	Level2	-1	2.6881	2.688
M6	M6_VR697050	57	Level4	Level2	2	-2.2582	2.258
M6	M6_VR698433	63	Level1	Level3	-2	3.7531	3.753
M6	M6_VR697957_1	65	Level2	Level3	-1	1.251	1.251
M6	M6_VR695265	66	Level2	Level3	-1	1.2605	1.26
M6	M6_VR695646	67	Level2	Level3	-1	1.3081	1.308
M6	M6_VR698250	68	Level4	Level3	1	-1.2552	1.255
M6	M6_VR695239	74	Level4	Level3	1	-0.8859	0.886
M6	M6_VR696784	75	Level1	Level3	-2	4.3214	4.321
M6	M6_VR694882	86	Level2	Level3	-1	2.5063	2.506
M6	M6_VR696237	87	Level3	Level4	-1	1.2289	1.229
M6	M6_VR694872_2	88	Level3	Level4	-1	1.615	1.615
M6	M6_VR696121	89	Level2	Level4	-2	3.9478	3.948

Table G8. Roster of Essentially Consistent Items: Mathematics Grade 6

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
M6	M6_VR695856	9	Level2	Level1	1	-0.3963	0.396
M6	M6_VR695774	10	Level2	Level1	1	-0.3561	0.356
M6	M6_VR697111	12	Level2	Level1	1	-0.2248	0.225
M6	M6_VR695920	13	Level2	Level1	1	-0.2052	0.205
M6	M6_VR694584	14	Level2	Level1	1	-0.133	0.133
M6	M6_VR696888	15	Level2	Level1	1	-0.1001	0.1
M6	M6_VR695754	16	Level2	Level1	1	-0.0949	0.095
M6	M6_VR694872_1	17	Level2	Level1	1	-0.011	0.011
M6	M6_VR696298_2	83	Level4	Level3	1	-0.282	0.282

Table G9. Roster of Inconsistent Items: Mathematics Grade 7

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
M7	M7_VR694513	2	Level2	Level1	1	-0.4174	0.417
M7	M7_VR697431	5	Level3	Level1	2	-3.156	3.156
M7	M7_VR695187	9	Level3	Level2	1	-2.5698	2.57
M7	M7_VR697135	11	Level3	Level2	1	-2.5095	2.51
M7	M7_VR695279_1	13	Level1	Level2	-1	1.498	1.498
M7	M7_VR696135	14	Level3	Level2	1	-2.4427	2.443
M7	M7_VR694290	20	Level1	Level2	-1	2.0634	2.063
M7	M7_VR695909	29	Level3	Level2	1	-1.6674	1.667
M7	M7_VR695345	30	Level3	Level2	1	-1.6395	1.64
M7	M7_VR695990_1	32	Level1	Level2	-1	2.421	2.421
M7	M7_VR696735	33	Level3	Level2	1	-1.4377	1.438
M7	M7_VR694294	34	Level3	Level2	1	-1.2441	1.244
M7	M7_VR697945	35	Level3	Level2	1	-1.0459	1.046
M7	M7_VR695524	37	Level3	Level2	1	-0.9311	0.931
M7	M7_VR697154	43	Level3	Level2	1	-0.4227	0.423
M7	M7_VR696840_1	50	Level4	Level2	2	-1.809	1.809
M7	M7_VR695784_1	53	Level2	Level3	-1	1.071	1.071
M7	M7_VR697271	54	Level2	Level3	-1	1.1091	1.109
M7	M7_VR695879_2	55	Level2	Level3	-1	1.119	1.119
M7	M7_VR698408	57	Level2	Level3	-1	1.2348	1.235
M7	M7_VR696247_1	59	Level2	Level3	-1	1.284	1.284
M7	M7_VR696296_1	61	Level2	Level3	-1	1.359	1.359
M7	M7_VR696442	64	Level2	Level3	-1	1.58	1.58
M7	M7_VR698194	65	Level2	Level3	-1	1.6106	1.611
M7	M7_VR697488	67	Level2	Level3	-1	1.743	1.743
M7	M7_VR696840_2	68	Level4	Level3	1	-0.862	0.862
M7	M7_VR696296_2	72	Level2	Level3	-1	1.99	1.99
M7	M7_VR695784_2	79	Level2	Level3	-1	2.139	2.139
M7	M7_VR697336	83	Level2	Level3	-1	2.3656	2.366
M7	M7_VR698425	87	Level1	Level4	-3	5.9637	5.964
M7	M7_VR694872_2	89	Level2	Level4	-2	3.675	3.675

Table G10. Roster of Essentially Consistent Items: Mathematics Grade 7

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
M7	M7_VR694742	3	Level2	Level1	1	-0.3556	0.356
M7	M7_VR696297	4	Level2	Level1	1	-0.2899	0.29
M7	M7_VR694913_1	6	Level2	Level1	1	-0.051	0.051
M7	M7_VR695811	46	Level3	Level2	1	-0.3066	0.307
M7	M7_VR697977	47	Level3	Level2	1	-0.2324	0.232

Table G11. Roster of Inconsistent Items: Mathematics Grade 8

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
M8	M8_VR698066	3	Level2	Level1	1	-1.7211	1.721
M8	M8_VR695584	12	Level2	Level1	1	-1.1425	1.143
M8	M8_VR694624	18	Level4	Level1	3	-4.0496	4.05
M8	M8_VR697180	27	Level2	Level1	1	-0.797	0.797
M8	M8_VR696305_1	31	Level2	Level1	1	-0.69	0.69
M8	M8_VR695064	34	Level2	Level1	1	-0.6323	0.632
M8	M8_VR695987	36	Level3	Level1	2	-2.0639	2.064
M8	M8_VR695038	44	Level3	Level1	2	-1.651	1.651
M8	M8_VR695828	47	Level3	Level2	1	-1.5494	1.549
M8	M8_VR695181	51	Level3	Level2	1	-1.2259	1.226
M8	M8_VR694814_2	52	Level1	Level2	-1	1.463	1.463
M8	M8_VR698395_2	54	Level1	Level2	-1	1.53	1.53
M8	M8_VR694975	55	Level1	Level2	-1	1.567	1.567
M8	M8_VR698466	60	Level3	Level2	1	-0.6031	0.603
M8	M8_VR696333	63	Level1	Level2	-1	2.1468	2.147
M8	M8_VR696243	67	Level1	Level2	-1	2.2784	2.278
M8	M8_VR696393	73	Level2	Level3	-1	1.0671	1.067
M8	M8_VR694643_2	76	Level2	Level3	-1	1.136	1.136
M8	M8_VR696567_2	77	Level2	Level3	-1	1.333	1.333
M8	M8_VR698481	78	Level1	Level3	-2	2.9186	2.919
M8	M8_VR697576_2	80	Level1	Level3	-2	3.139	3.139
M8	M8_VR697165	81	Level2	Level3	-1	1.5913	1.591
M8	M8_VR697908	85	Level1	Level3	-2	3.3831	3.383
M8	M8_VR696502	86	Level1	Level3	-2	3.4197	3.42
M8	M8_VR696880	88	Level3	Level4	-1	1.0256	1.026
M8	M8_VR696703	89	Level3	Level4	-1	1.3793	1.379

Table G12. Roster of Essentially Consistent Items: Mathematics Grade 8

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
M8	M8_VR694519_2	43	Level2	Level1	1	-0.176	0.176
M8	M8_VR696538	68	Level3	Level2	1	-0.1192	0.119
M8	M8_VR698307	70	Level3	Level2	1	-0.0653	0.065
M8	M8_VR698469	71	Level3	Level2	1	-0.0573	0.057

Table G13. Roster of Inconsistent Items: Mathematics High School

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
MHS	MHS_VR688105_1	1	Level2	Level1	1	-1.003	1.003
MHS	MHS_VR688119_1	2	Level3	Level1	2	-1.134	1.134
MHS	MHS_VR688107	3	Level3	Level1	2	-0.9196	0.92
MHS	MHS_VR688074	6	Level1	Level2	-1	1.2253	1.225
MHS	MHS_VR688089	7	Level3	Level2	1	-0.6472	0.647
MHS	MHS_VR710121	8	Level4	Level2	2	-1.6843	1.684
MHS	MHS_VR710122	9	Level4	Level2	2	-1.5718	1.572
MHS	MHS_VR710124	11	Level1	Level2	-1	1.503	1.503
MHS	MHS_VR710123	14	Level4	Level2	2	-1.1832	1.183
MHS	MHS_VR688102_1	17	Level1	Level2	-1	1.865	1.865
MHS	MHS_VR688111	18	Level2	Level3	-1	1.0794	1.079
MHS	MHS_VR688118	19	Level2	Level3	-1	1.0878	1.088
MHS	MHS_VR688124	21	Level2	Level3	-1	1.3768	1.377
MHS	MHS_VR688091	22	Level2	Level3	-1	1.472	1.472
MHS	MHS_VR688117	23	Level4	Level3	1	-0.4401	0.44
MHS	MHS_VR688083	26	Level2	Level3	-1	1.7996	1.8
MHS	MHS_VR688116_2	30	Level2	Level4	-2	2.102	2.102
MHS	MHS_VR688104	31	Level3	Level4	-1	1.6867	1.687
MHS	MHS_VR688071	32	Level3	Level4	-1	1.9753	1.975
MHS	MHS_VR688102_2	33	Level2	Level4	-2	3.374	3.374

Table G14. Roster of Essentially Consistent Items: Mathematics High School

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
MHS	MHS_VR688105_2	12	Level3	Level2	1	-0.219	0.219
MHS	MHS_VR710125	13	Level3	Level2	1	-0.1467	0.147
MHS	MHS_VR710128	25	Level4	Level3	1	-0.3627	0.363

Table G15. Roster of Inconsistent Items: Reading Grade 3

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
R3	R3_VR702675_1	1	Level2	Level1	1	-0.559	0.559
R3	R3_VR702678	24	Level1	Level2	-1	3.0429	3.043
R3	R3_VR702676_1	45	Level2	Level3	-1	1.042	1.042
R3	R3_VR702750	48	Level2	Level3	-1	1.2354	1.235

Table G16. Roster of Essentially Consistent Items: Reading Grade 3

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
R3	R3_VR699692	42	Level3	Level2	1	-0.0864	0.086
R3	R3_VR700973	73	Level4	Level3	1	-0.1854	0.185

Table G17. Roster of Inconsistent Items: Reading Grade 4

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
R4	R4_VR700703	2	Level3	Level2	1	-1.0435	1.043
R4	R4_VR702966	3	Level3	Level2	1	-0.8249	0.825
R4	R4_VR702224_1	4	Level3	Level2	1	-0.567	0.567
R4	R4_VR703417	5	Level3	Level2	1	-0.5155	0.516
R4	R4_VR702962	6	Level3	Level2	1	-0.4499	0.45
R4	R4_VR700124	8	Level4	Level2	2	-2.4216	2.422
R4	R4_VR700128	10	Level4	Level3	1	-2.3475	2.348
R4	R4_VR703419	11	Level2	Level3	-1	1.0658	1.066
R4	R4_VR702963	15	Level4	Level3	1	-2.0062	2.006
R4	R4_VR700120_1	17	Level2	Level3	-1	1.521	1.521
R4	R4_VR700696	27	Level2	Level3	-1	1.7623	1.762
R4	R4_VR701841	40	Level2	Level3	-1	2.2352	2.235
R4	R4_VR700133_1	60	Level2	Level3	-1	3.091	3.091
R4	R4_VR702223	62	Level2	Level3	-1	3.1331	3.133
R4	R4_VR701842_2	67	Level3	Level4	-1	1.019	1.019
R4	R4_VR701844	68	Level3	Level4	-1	1.0899	1.09
R4	R4_VR700133_2	70	Level3	Level4	-1	1.513	1.513
R4	R4_VR702967	71	Level3	Level4	-1	1.6081	1.608

Table G18. Roster of Essentially Consistent Items: Reading Grade 4

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
R4	R4_VR700139	59	Level4	Level3	1	-0.3204	0.32

Table G19. Roster of Inconsistent Items: Reading Grade 5

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
R5	R5_VR702939_1	1	Level3	Level1	2	-2.165	2.165
R5	R5_VR702952	2	Level2	Level1	1	-0.5549	0.555
R5	R5_VR702941_1	3	Level1	Level2	-1	1.335	1.335
R5	R5_VR702940_1	5	Level1	Level2	-1	1.83	1.83
R5	R5_VR702940_2	8	Level2	Level3	-1	1.022	1.022
R5	R5_VR701041	19	Level2	Level3	-1	1.3062	1.306
R5	R5_VR699772	24	Level2	Level3	-1	1.5639	1.564
R5	R5_VR701043	25	Level2	Level3	-1	1.6224	1.622
R5	R5_VR702939_2	37	Level4	Level3	1	-1.407	1.407
R5	R5_VR702942_2	50	Level4	Level3	1	-0.723	0.723
R5	R5_VR700358	68	Level3	Level4	-1	1.0582	1.058
R5	R5_VR702933	69	Level3	Level4	-1	1.172	1.172
R5	R5_VR703742	71	Level3	Level4	-1	1.2391	1.239
R5	R5_VR699880_1	74	Level3	Level4	-1	1.781	1.781

Table G20. Roster of Essentially Consistent Items: Reading Grade 5

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
R5	R5_VR699882	67	Level4	Level3	1	-0.0299	0.03

Table G21. Roster of Inconsistent Items: Reading Grade 6

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
R6	R6_VR699867_1	1	Level2	Level1	1	-0.773	0.773
R6	R6_VR699872_1	3	Level2	Level1	1	-0.661	0.661
R6	R6_VR703548	4	Level3	Level1	2	-1.3316	1.332
R6	R6_VR703544	11	Level3	Level2	1	-0.9234	0.923
R6	R6_VR703550	13	Level3	Level2	1	-0.8203	0.82
R6	R6_VR702319	14	Level1	Level2	-1	1.5397	1.54
R6	R6_VR701266	16	Level1	Level2	-1	1.6264	1.626
R6	R6_VR703552	17	Level1	Level2	-1	1.6534	1.653
R6	R6_VR701267	24	Level1	Level2	-1	2.0064	2.006
R6	R6_VR702322	31	Level2	Level3	-1	1.1824	1.182
R6	R6_VR703506_1	32	Level2	Level3	-1	1.185	1.185
R6	R6_VR700940_1	47	Level2	Level3	-1	1.813	1.813
R6	R6_VR700946_1	49	Level2	Level3	-1	1.913	1.913
R6	R6_VR699869_1	66	Level2	Level3	-1	2.607	2.607
R6	R6_VR703739	85	Level3	Level4	-1	0.9968	0.997

Table G22. Roster of Essentially Consistent Items: Reading Grade 6

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
R6	R6_VR699878_1	5	Level2	Level1	1	-0.035	0.035
R6	R6_VR703551	22	Level3	Level2	1	-0.313	0.313
R6	R6_VR703740	27	Level3	Level2	1	-0.0125	0.012

Table G23. Roster of Inconsistent Items: Reading Grade 7

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
R7	R7_VR701897	2	Level2	Level1	1	-1.4131	1.413
R7	R7_VR700529_1	3	Level2	Level1	1	-0.69	0.69
R7	R7_VR700447_1	4	Level2	Level1	1	-0.627	0.627
R7	R7_VR702192	5	Level2	Level1	1	-0.4123	0.412
R7	R7_VR702187	11	Level1	Level2	-1	1.1013	1.101
R7	R7_VR700544	13	Level3	Level2	1	-1.7012	1.701
R7	R7_VR702190	15	Level1	Level2	-1	1.3656	1.366
R7	R7_VR700547	19	Level1	Level2	-1	1.4259	1.426
R7	R7_VR701183	23	Level1	Level2	-1	1.559	1.559
R7	R7_VR700546	29	Level3	Level2	1	-1.2247	1.225
R7	R7_VR701887	30	Level3	Level2	1	-1.2068	1.207
R7	R7_VR701638	36	Level1	Level2	-1	2.0765	2.076
R7	R7_VR700444_2	42	Level3	Level2	1	-0.771	0.771
R7	R7_VR700779	55	Level1	Level2	-1	2.8871	2.887
R7	R7_VR701651	63	Level2	Level3	-1	1.0275	1.027
R7	R7_VR700775	67	Level2	Level3	-1	1.1207	1.121
R7	R7_VR701613	80	Level1	Level3	-2	3.7435	3.744
R7	R7_VR702795	82	Level2	Level3	-1	1.7771	1.777
R7	R7_VR700533	92	Level3	Level4	-1	1.3953	1.395
R7	R7_VR700526	93	Level3	Level4	-1	1.4954	1.495

Table G24. Roster of Essentially Consistent Items: Reading Grade 7

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
R7	R7_VR701642	6	Level2	Level1	1	-0.219	0.219
R7	R7_VR700545	7	Level2	Level1	1	-0.0625	0.062
R7	R7_VR701890	50	Level3	Level2	1	-0.2745	0.274
R7	R7_VR701888	57	Level3	Level2	1	-0.0777	0.078
R7	R7_VR700518_2	88	Level4	Level3	1	-0.397	0.397
R7	R7_VR700772_2	91	Level4	Level3	1	-0.156	0.156

Table G25. Roster of Inconsistent Items: Reading Grade 8

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
R8	R8_VR700815	1	Level2	Level1	1	-0.4999	0.5
R8	R8_VR701734	5	Level3	Level2	1	-0.4978	0.498
R8	R8_VR700817	6	Level3	Level2	1	-0.4255	0.426
R8	R8_VR701733	7	Level1	Level2	-1	1.3703	1.37
R8	R8_VR701736	10	Level1	Level2	-1	1.4355	1.435
R8	R8_VR701737	19	Level2	Level3	-1	1.0434	1.043
R8	R8_VR700819	22	Level2	Level3	-1	1.3371	1.337
R8	R8_VR701528_1	80	Level2	Level3	-1	3.24	3.24

Table G26. Roster of Essentially Consistent Items: Reading Grade 8

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
R8	R8_VR701703_1	2	Level2	Level1	1	-0.134	0.134
R8	R8_VR702185	3	Level2	Level1	1	-0.103	0.103
R8	R8_VR700834	17	Level3	Level2	1	-0.0531	0.053

Table G27. Roster of Inconsistent Items: Reading High School

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
RHS	RHS_VR703970	4	Level3	Level2	1	-0.5617	0.562
RHS	RHS_VR703937	22	Level2	Level3	-1	1.5518	1.552
RHS	RHS_VR703948_1	23	Level2	Level3	-1	1.611	1.611
RHS	RHS_VR703941_2	37	Level3	Level4	-1	1.105	1.105
RHS	RHS_VR703950	39	Level3	Level4	-1	1.3989	1.399

Table G28. Roster of Essentially Consistent Items: Reading High School

GCA	Item ID	OOD	Consensus Level	Empirical Level	Level Difference	Distance	Absolute Distance
RHS	RHS_VR703954_2	11	Level3	Level2	1	-0.1	0.1
RHS	RHS_VR703953_1	14	Level3	Level2	1	-0.079	0.079
RHS	RHS_VR703988_2	34	Level4	Level3	1	-0.228	0.228

Appendix H: Considerations in the Use of Standard Errors for the Adjustment of Cut Scores

Introduction

Research indicates that different standard setting methods may produce different results (Green, Trimble, & Lewis, 2003) and different panels of standard setting experts may recommend different cut scores (Lewis, Green, Mitzel, Baum, & Patz, 1998). Therefore, when adopting cut scores resulting from a standard setting study, it is reasonable to consider the stability of the recommended cut scores. Embedded Standard Setting is an item-based standard setting method; thus, the stability of ESS cut scores should be considered relative to the specific sample of items upon which SMEs' Item-ALD Alignments are made. Decision-making with regard to the adoption of cut scores may also consider the stability of student scores based on the standard error of measurement (SEM_{Test}) of the test.

The Standard Error of Measurement

First, we can consider the role that the standard error of measurement (SEM_{Test}) of the test plays in the misclassification of students. The standard error of measurement of the test is a measure of the precision of a student's obtained test score or ability estimate. When a student's obtained score is close to the cut score relative to the standard error, two types of errors could occur. When a student's test score meets or exceeds a cut score because of measurement error and not because of true ability it is called a false-positive error. When a student's test score is less than a cut score because of measurement error and not because of true ability it is called a false-negative error.

One reason that a recommended cut score would be adjusted following a standard setting would be to decrease the likelihood of a false-positive or false-negative error. Raising the recommended cut point would decrease the likelihood of false-positive errors and lowering the recommended cut point would decrease the likelihood of a false-negative error. Thus, a decision to raise or lower the recommended cut score might be based on whether the greater concern was for passing students who should have failed based on their true ability or failing students who should have passed based on their true ability.

The Standard Error of the ESS-estimated Cut Score (SE_{ESS}) Based on the Sampling of Items

Embedded Standard Setting (Lewis & Cook, 2020) cut score estimates may vary depending on the selection of items sampled from the item pool. The item pool itself may be considered as a sample of items from the universe of items from the domain. It is assumed that the items are a sample selected from the pool or universe, and if a different sample of items had been selected, it is likely that a somewhat different cut score would result. The standard error of the ESS cut score (SE_{ESS}) is an estimate of the stability of an ESS-estimated cut score across different item samples of the same size, randomly selected from the pool of available items. The

variability of the ESS cut score estimates from multiple samples of items can be used to estimate SE_{ESS}.

The ESS-estimated cut score that would be estimated if the entire universe of items were considered in the ESS analyses has a .68 likelihood of being in the interval defined by the ESS-estimated cut score plus or minus 1 SE_{ESS} and has a.95 likelihood of being in the interval defined by the ESS-estimated cut score plus or minus 1.96 SE_{ESS}.

Lewis, Lee, & Choi (2021, May) describe the Bootstrap analyses employed by EmStanS to estimate SE_{ESS}.

A Standard Error Reflecting both the Standard Error of Measurement and the Standard Error of the Cut Score

When both types of error (SEM_{Test} and SE_{ESS}) are of concern, a single standard error can be calculated ($SE_{Test+ESS}$), as reported by Jaeger (1989). If errors due to the unreliability of the test (SEM_{Test}) and errors due to ESS sampling of items (SE_{ESS}) are considered to be independent (a plausible assumption), a standard error that reflects both sources is given by

$$SE_{Test+ESS} = \sqrt{(SEM_{Test}^2 + SE_{ESS}^2)} .$$

This combined standard error will tend to be larger than either the SEM_{Test} or SE_{ESS}.

Precedent and Rationale for the Adjustment of Cut Scores by Sponsoring Agencies

There are reasons why adjustments should be made to cut scores prior to adoption by a sponsoring agency including (a) the need to reduce the likelihood of false-positive or false-negative errors and (b) the need for well-articulated cut scores within and across grades. There is precedent for adjustments to recommended cut scores such as the 1992 NAEP Mathematics recommended cut scores for grades 4, 8, and 12, which were each adjusted downward by one standard error (American College Testing [ACT], 1993).

> To set the final cut scores in mathematics, NAGB took the average Round 3 rating for each level, subtracted one standard error of measurement (based on split-sample variation in judges' ratings), and adopted these values as the final cut scores (see Shepard, Glaser, Linn, & Bohrnstedt, 1993, p.32).

In his review of standard setting guidelines, Cizek (1996) noted that adjustments to recommended cut scores often consist of raising or lowering the passing score by a fraction or multiple of the SEM. He noted, however, that the rationale for such adjustments must be well documented. If adjustments to recommended cut scores are being considered, Mehrens (1986) argued that the relative effects of incorrect decisions (e.g., false-positives and false-negatives) be considered and that the values underlying those considerations be made explicit.

Cizek (1996) noted the following:

> If adjustments to the passing score are to be made, evidence should be presented that supports the reasonableness (i.e., validity) of such an adjustment in terms of other indicators of achievement, cost-benefit analyses, or other relevant considerations. . . . If it is decided that adjustments to participants' individual judgments or passing scores are necessary, a detailed explication of the rationale, method, and effect of the adjustment is clearly warranted in a report on the standard-setting procedure.

Geisinger (1991) concurs with Cizek, noting that "we should explicitly decide whether or not to modify our passing score on the basis of established techniques," and "we must be clear as to the rationale for such adjustments." Cizek (1996) also noted that "it may be tempting to justify these adjustments on the basis that all standard setting involves the synthesis of judgment—adjustments merely reflect the incorporation of additional information or judgment."

In sum, the literature suggests that under the proper conditions and with strong rationales, adjustments to participants' recommended cut scores can and should be made. Such adjustments should be done with only an explicit rationale, and all information contributing to such a decision should be well documented.

References

American College Testing (1993). Setting achievement levels on the 1992 National Assessment of Educational Progress in Mathematics, Reading, and Writing: A Technical Report on Reliability and Validity. ACT.

Cizek, G. J. (1996). Standard Setting Guidelines. *Educational Measurement: Issues and Practice, 15*(1), pp. 13–21.

Geisinger, K. F. (1991). Using Standard Setting Data to Establish Cutoff Score. *Educational Measurement: Issues and Practice, 10*(2), pp. 17–22.

Green, D. R., Trimble, S., & Lewis, D. M. (2003). Interpreting the results of three different standards setting procedures. *Educational Measurement: Issues and Practices, 22*(1), 22–32.

Jaeger, R. M. (1989). Selection of Judges for Standard Setting: What Kinds? How Many? Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA. MEHRENS, W. A. (1986). Measurement Specialists: Motive to Achieve or Motive to Avoid Failure? *Educational Measurement: Issues and Practice, 5*(4), pp. 5–10.

Lewis, D. & Cook, R. (2020). Embedded Standard Setting: Aligning standard setting methodology with contemporary assessment design principles. *Educational Measurement: Issues and Practices, 39*(1), 8–21.

Lewis, D. M., Green, D. R., Mitzel, H. C., Baum, K., & Patz, R. J. (1998, April). The Bookmark Standard Setting Procedure: Methodology and recent implementations. Paper presented at the 1998 National Council for Measurement in Education annual meeting, San Diego.

Lewis, D., Lee, S., & Choi, S. (2021, May). An Investigation of Two Embedded Standard Setting Cut Score Estimation Algorithms: ESS-Count & ESS-Weight. In D. Lewis (Organizer), Embedded Standard Setting: Research & Advances. Symposium at the 2021 annual meeting of the National Council on Measurement in Education, virtual.

Mehrens, W. A. (1986). Measurement Specialists Motive to Achieve or Motive to Avoid Failure? *Educational Measurement: Issues and Practices, 5*(4), 5–10.

Shepard, L. Glaser, R., Linn, R., & Bohrnstedt, G. (1993). Setting Achievement Standards for Student Achievement. Stanford: National Academy of Education.

Appendix I: Moderation and Smoothing of Panelists' Cut Scores

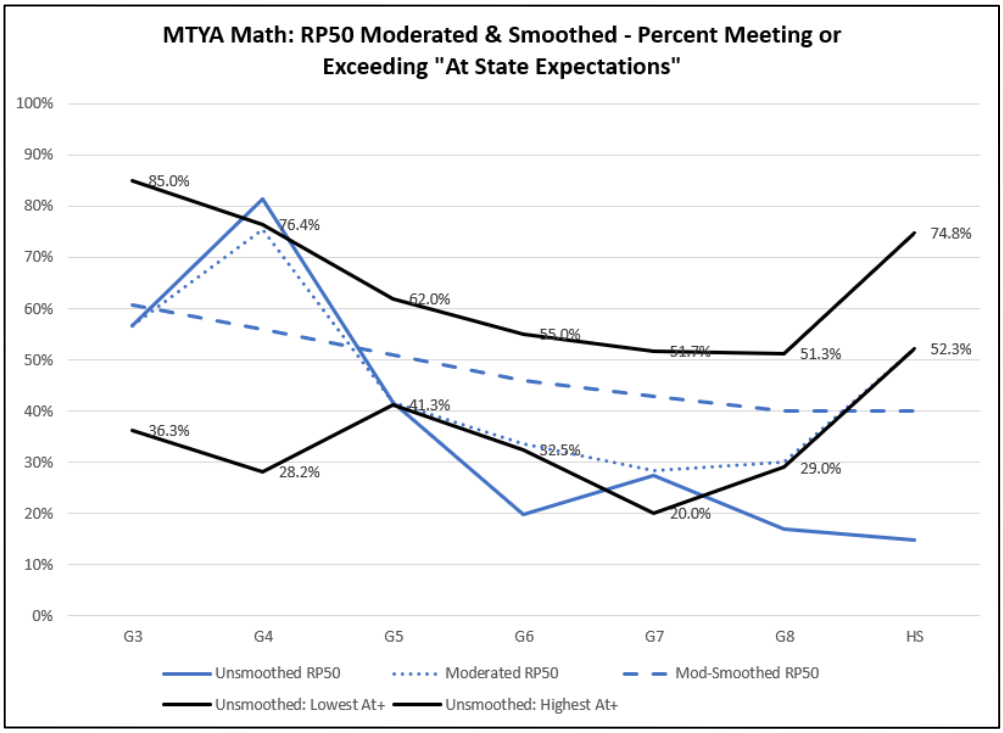


Figure 1. Math: Moderation and Smoothing of the At State Expectations Cut Score

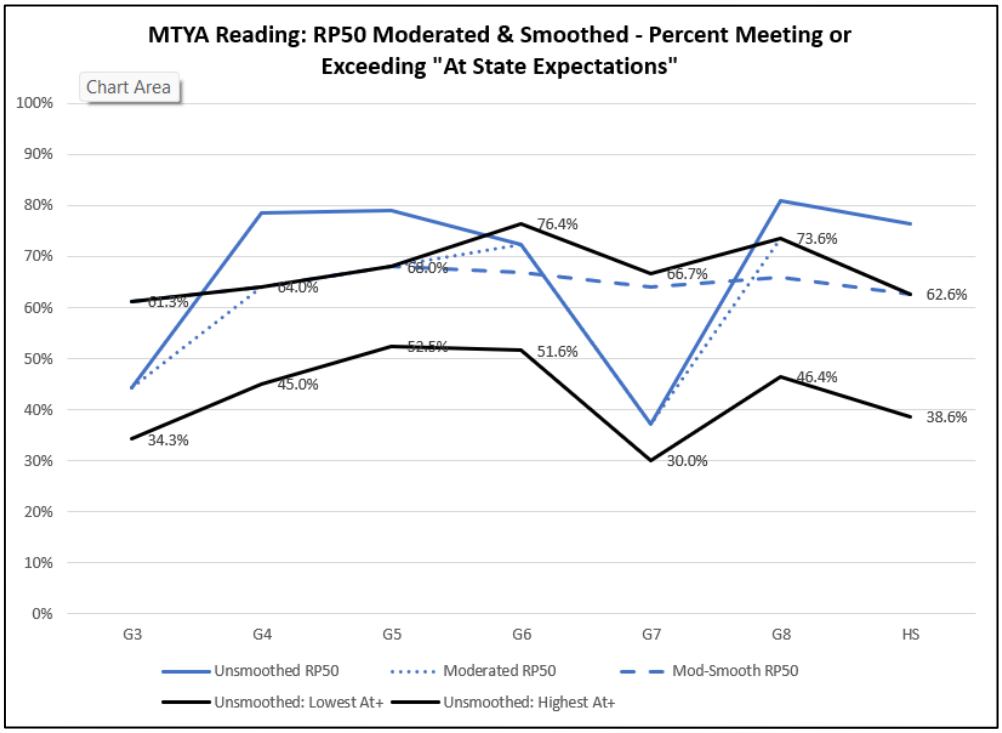


Figure 2. Reading: Moderation and Smoothing of the At State Expectations Cut Score

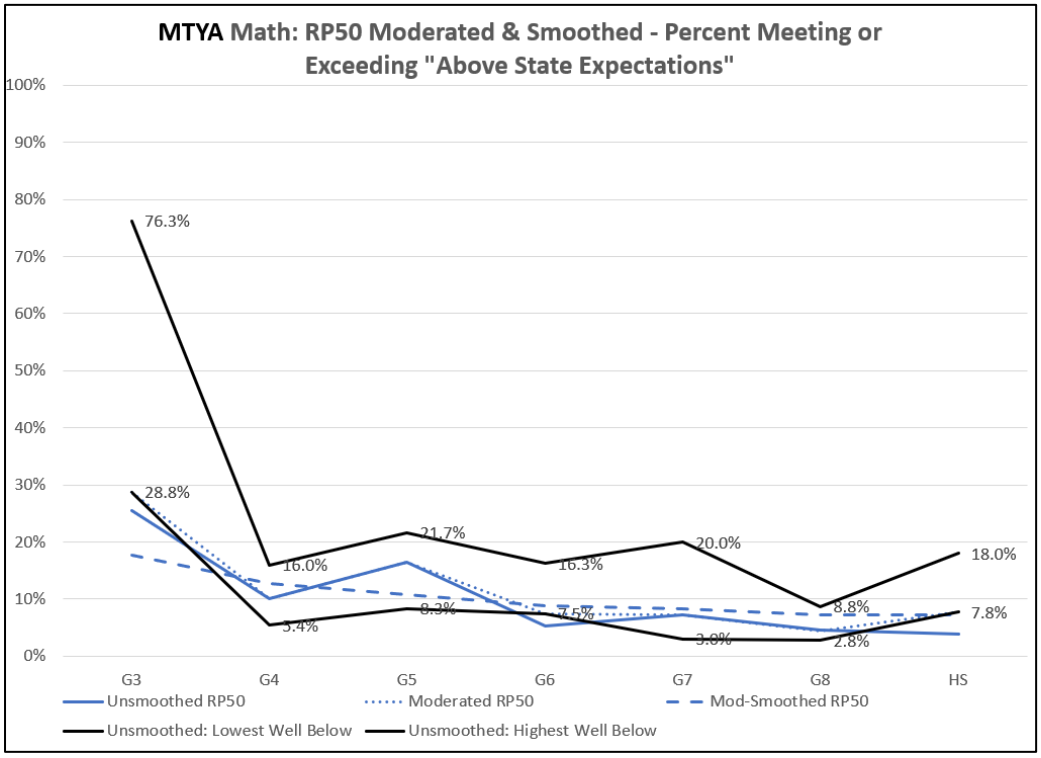


Figure 3. Math: Moderation and Smoothing of the Above State Expectations Cut Score

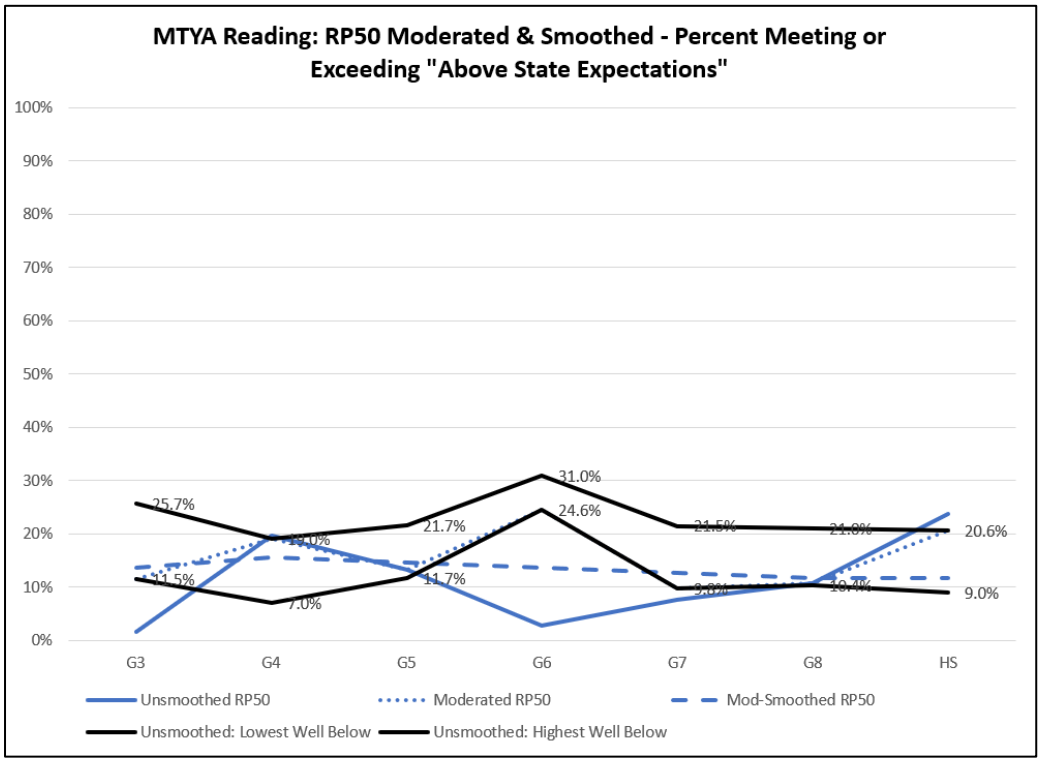


Figure 4. Reading: Moderation and Smoothing of the Above State Expectations Cut Score

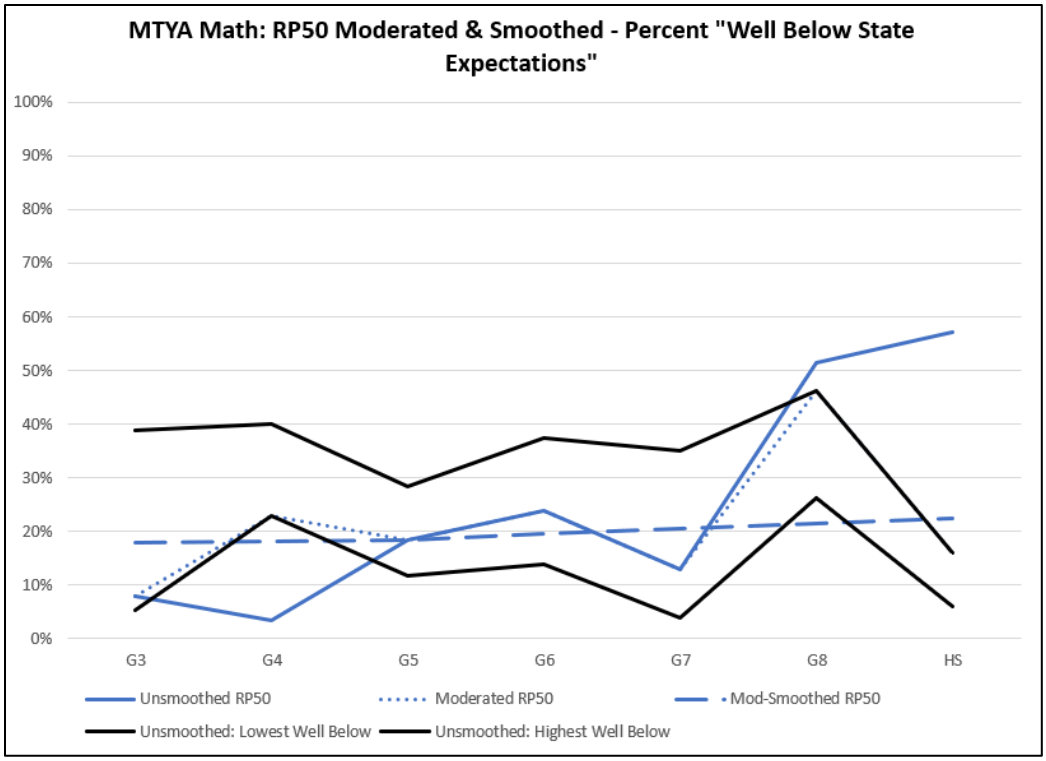


Figure 5. Math: Moderation and Smoothing of the Below State Expectations Cut Score

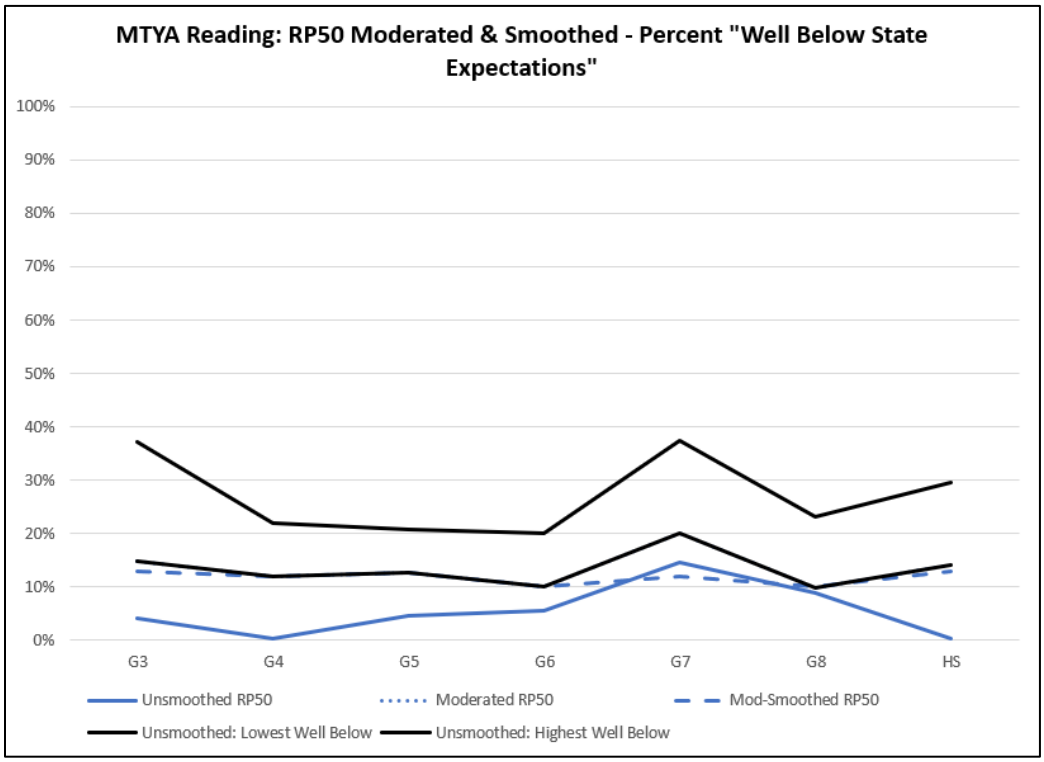


Figure 6. Reading: Moderation and Smoothing of the Below State Expectations Cut Score