LEAP 2025 Social Studies 3–8 Technical Report: 2021–2022

Prepared by DRC, Pearson, and WestEd

EXECUTIVE SUMMARY

The Louisiana Educational Assessment Program 2025 (LEAP 2025) is composed of tests that are carefully constructed to fairly assess the achievement of Louisiana students. This technical report provides information on the operational test administrations, scoring activities, analyses, and results of the spring 2022 administration of the LEAP 2025 Social Studies test, which included both operational and field test items.

While this technical report and its associated materials have been produced in a way that can help educators understand the technical characteristics of the assessment used to measure student achievement, the information is primarily intended for use by those who evaluate tests, interpret scores, or use test results in making educational decisions. It is assumed that the reader has technical knowledge of test construction and measurement procedures, as stated in *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014).

The chapters of this technical report outline general information about the administration and scoring activities of the LEAP 2025 assessments, CTT (Classical Test Theory) and IRT (Item Response Theory) analysis results, 2022 test results, demographic characteristics of students, reliability and validity, and the interpretation of the scores on the tests. Additionally, because of conditions related to COVID-19, please use caution when making any inferences from the statistical results of the spring 2022 administration.

Table of Contents

1. Introduction

The Louisiana Department of Education (LDOE) has a long and distinguished history in the development and administration of assessments that support its state accountability system and are aligned to its state content standards. Per state law, the LDOE is to administer statewide Social Studies assessments in grades 3–8 and high school annually. Fulfilling the directive of the Louisiana State Board of Elementary and Secondary Education (BESE), the LDOE must deliver high-quality, Louisiana-specific standards-based assessments. Further, the LDOE and the BESE are committed to the development of rigorous assessments as one component of their comprehensive plan—Louisiana Believes—designed to ensure that every Louisiana student is on track to be successful in postsecondary education and the workforce.

The purpose of this technical report is to describe the process for the operational administration of the statewide summative Social Studies assessments for grades 3–8. This report outlines the testing procedures, including forms construction, administration, scoring and analyses, and reporting of scores.

Summary of the 2019–2022 Activities

WestEd and Pearson, in partnership with the LDOE and Data Recognition Corporation (DRC), the administration vendor, developed a timeline to capture the major activities necessary to produce the spring 2022 grades 3–8 operational forms with embedded field test (EFT) items. Table 1.1 summarizes the key activities along with the months during which the activities were completed.

Table 1.1
Key Activities from August 2019 to May 2022

Date	Activity
August 2018–May 2019	• The LDOE and WestEd planned item development and determined item sets and standalone items for revise and re-field test and new development • The LDOE and WestEd worked to revise sources and items
August–December 2019	• Data review of Spring 2019 items • WestEd and the LDOE constructed operational tests for spring 2020 • Technical Advisory Committee convened
January–May 2020	• Source Review Committees convened • Technical Advisory Committee convened • The LDOE staff conducted Source Review Committees
	• Spring 2020 administration of the LEAP 2025 assessments due to COVID-19 pandemic
June–July 2020	• Item Content and Bias Review Committees convened • Reconciliation meetings held between the LDOE and WestEd staff
August–December 2020	• The LDOE and WestEd prepared operational test forms for spring 2021 administration • Technical Advisory Committee convened • BESE authorized the review and revision of the Louisiana Student Standards for Social Studies (LSSSS)
January–May 2021	• Style guide updated • WestEd updated 2020–2021 Assessment Framework • Technical Advisory Committee convened • Spring 2021 test administration

September–November 2021	• The LDOE, WestEd, and DRC prepared operational test forms and field test forms • Technical Advisory Committee convened
March–April 2022	• Technical Advisory Committee convened • BESE approved new LSSSS
April–May 2022	• Spring 2022 tests administered, including field test items

2. Assessment Frameworks

The initial assessment frameworks developed at the start of the project included:

- proposed test designs;
- test blueprints;
- the range of standards and Grade-Level Expectations (GLEs) to be covered;
- reporting categories;
- percentages of assessment items and score points by reporting category;
- projected testing times; and
- the numbers of forms to be administered.

Before the spring 2022 operational test forms were constructed, the Assessment Frameworks were updated to reflect changes to the design and field test plan, as well as to clarify the criteria used to guide item and form selection.

3. Overview of the Development Process

This section describes the processes used to develop field test item sets, tasks, and standalone items to embed within the LEAP 2025 Social Studies assessments.

Item Development Plan

WestEd's proposed item development plans may include item sets, tasks, and standalone items. WestEd's item development plans for 2019 and 2020 focused on new and revise and re-field test items for grades 3–8. The proposed item development plans included item sets and standalone items. In grade 4, WestEd proposed to develop TE items as part of item sets. Table 3.1 shows the 2019–2020 revise and re-field test item development plan for grades 3–8. Table 3.2 shows the 2020–2021 development plan for grades 3, 4, and 8. No new item development was proposed for grades 5, 6, and 7.

Table 3.1

Item Development Revise and Re-field Test Plan: Grades 3–8 (2019–2020)

Grades 3–8 Item Development Plan		Total Tasks	Total Item Sets	Total Items per Set	MC/MS	CR	TE	ER	Total Items
2020	Tasks	–	–	–	–	–	–	–	–
	Item sets	–	24	7 to 15	221	30	30	–	281
	Standalone items (MC/MS/TE)	–	–	–	34	–	–	–	34
	TOTALS	–	24	–	255	30	30	–	315

Key

MC: multiple choice MS: multiple select CR: constructed response
TE: technology enhanced ER: extended response

Table 3.2

Item Development Plan: Grades 3, 4, and 8 (2020–2021)

Grades 3, 4, and 8 Item Development Plan		Total Tasks	Total Item Sets	Total Items per Set	MC/MS	CR	TE	ER	Total Items
2021	Tasks	–	–	–	–	–	–	–	–
	Item sets	–	12	12 to 14	117	19	16	–	152
	Standalone items (MC/MS/TE)	–	–	–	35	0	0	–	35
	TOTALS	–	12	–	152	19	16	–	187

Key

MC: multiple choice MS: multiple select CR: constructed response
TE: technology enhanced ER: extended response

Proposal and Review of Topics and Sources

Determining Topics

The WestEd content leads reviewed the existing item bank, the LDOE instructional materials, and the Louisiana Student Standards for Social Studies to help determine the content eligible for assessment and what was needed to support the development of the operational assessment. After studying these resources, the content leads made recommendations for which new and revise and re-field test item sets and standalone items should be developed.

When identifying possible topics, WestEd content leads consider the following:

- Which topics have already been developed and which topics need development
- What content is eligible according to the companion documents and scope and sequence documents
- Whether proposed topics will support the required item types and number of items, including overage

- How GLEs will be combined to provide meaningful assessment of content and concepts
- How a topic reflects the LDOE's goal of assessing larger ideas rather than discrete facts

Topics are chosen to represent the breadth of assessable Social Studies content while complementing the balance of topics in the existing pool. The process of choosing assessable GLEs for each topic is iterative and includes the identification of potential GLEs that could be assessed together. It also requires an understanding of the need to create an item pool with the broadest possible content coverage.

Tasks and Item Sets. Tasks and item sets contain multiple related sources that provide the context from which students answer groups of questions. Sets allow students to delve deeply into a topic and may include items aligned to GLEs across reporting categories—allowing a set to highlight the interrelated nature of history, geography, civics, and economics—or from a subset of those categories.

Standalone Items. Standalone items assess content that may or may not be connected to a source. A goal in standalone item development is to have a source for 80% of the standalone items to best support students in answering the questions. With the exception of grade 4, which has standalone TE items, all standalone items are selected response (SR) items (multiple choice, multiple select). Standalone items are included in the test design to provide greater coverage of the assessable content and GLEs and to provide flexibility in meeting the blueprints and test characteristic curve targets across test administrations. Content leads select topics for standalone items based on content and GLEs that may not be sufficiently covered across the sets, with the goal of providing maximum flexibility during test construction. Consequently, the standalone items are typically developed last.

GLE Coverage

Grade 3. By the end of the 2018–2019 development cycle, WestEd had developed at least 1 item aligned to each of the 40 assessable GLEs in grade 3. (GLEs 3.1.3, 3.1.4, 3.1.5, 3.3.5,

and 3.3.6 have been identified as those that are unlikely to be addressed directly in assessment items.)

Grade 4. By the end of the 2018–2019 development cycle, WestEd had developed at least 1 item aligned to each of the 39 assessable GLEs in grade 4. (GLEs 4.1.3, 4.1.4, 4.1.7, and 4.4.7 have been identified as those that are unlikely to be addressed directly in assessment items.)

Grade 5. By the end of the 2018–2019 development cycle, WestEd had developed at least 1 item aligned to each of the 24 assessable GLEs in grade 5. (GLEs 5.1.2, 5.1.4, and 5.4.2 have been identified as those that are unlikely to be addressed directly in assessment items.)

Grade 6. By the end of the 2018–2019 development cycle, WestEd had developed at least 1 item aligned to each of the 23 assessable GLEs in grade 6. (GLEs 6.1.1, 6.1.2, 6.1.3, and 6.1.4 have been identified as those that are unlikely to be addressed directly in assessment items.)

Grade 7. By the end of the 2018–2019 development cycle, WestEd had developed at least 1 item aligned to each of the 39 assessable GLEs in grade 7. (GLEs 7.1.1, 7.1.2, 7.1.4, 7.1.5, 7.5.2, and 7.9.1 have been identified as those that are unlikely to be addressed directly in assessment items.)

Grade 8. By the end of the 2018–2019 development cycle, WestEd had developed at least 1 item aligned to each of the 30 assessable GLEs in grade 8. (GLEs 8.1.1, 8.1.2, 8.2.10, 8.3.3, and 8.10.5 have been identified as those that are unlikely to be addressed directly in assessment items.)

While some GLEs at each grade level may have only 1 item aligned to them, others have several. This variation is a result of differences in the importance and scope of content covered by individual GLEs.

Obtaining LDOE Approval for Topics

For tasks and item sets, WestEd submits lists of proposed topics at each grade level to the LDOE for review prior to item development. These lists describe the topics and possible related sources so that the LDOE can review and approve them simultaneously. The proposed topic lists also include the GLEs and reporting categories that might be assessed by the tasks and item sets. Once the LDOE approves the topics to be developed for the development cycle, source-searching and development of the task and item set overviews begin.

For standalone items, there has been no separate approval phase for the topics or sources. However, WestEd and the LDOE have a process to identify the appropriate alignment of the standalone items.

For revised and re-field-tested item sets and standalone items, WestEd submits lists of previously developed and field tested item sets and standalone items to the LDOE for review. Working with WestEd and reviewing the field test data, the LDOE determines which sets should be revised, including their sources, and re-field tested.

Identifying Sources

The LEAP 2025 Social Studies assessments focus on the use of authentic historical and contemporary documents, including maps, letters, journal entries, speeches, photographs, paintings, reports, and other primary source documents. The assessments also include secondary source documents, such as authentic newspaper articles and book excerpts. These documents are supplemented by timelines, tables, charts, and graphic organizers created by WestEd's Design Team. On rare occasions, a source, such as a scenario, is written by WestEd editors to meet a specific assessment purpose.

Both internal and external editors locate appropriate sources for tasks, item sets, and standalone items. Before the source searchers begin, WestEd trains them on the search process, on the LDOE's objectives, and on best practices, including bias and sensitivity

training. For an outline of the training, see LEAP 2025 Social Studies Grades 3–8 Source Search Training Agenda in Appendix A.

All sources are submitted to WestEd for evaluation for alignment and appropriateness for the approved topics. Based on this evaluation, the WestEd content leads select the final sources to propose to the LDOE.

Public Domain versus Permissioned Work. WestEd endeavors to maintain a ratio of 80% royalty-free sources from the public domain or created internally to a maximum of 20% permissioned work. The actual percentages vary from year to year and grade to grade, depending on the needs of the content in development. Across all grades, the total percentage of permissioned work is not less than 20%. Before administration of the assessment, WestEd's permissions coordinator obtains permissions from the rights holders for five years of use of any work that was not in the public domain or created internally.

Evaluating the Readability of Sources. WestEd performs both a Lexile analysis and an ATOS analysis on each passage in the tasks and item sets to obtain a quantitative measure of the readability of the texts. The Lexile Analyzer, developed by MetaMetrics, analyzes the semantic and syntactic features of a text and assigns it a Lexile measure. MetaMetrics also provides grade-level ranges corresponding to Lexile ranges. It should be noted that the grade-level ranges include overlap across grade levels. The ATOS readability tool, developed by Renaissance, also analyzes the reading level of passages. It focuses on elements of text complexity, such as average sentence length, average word length, and word difficulty. Using the Lexile and ATOS measurements provides important statistical information to determine if the passages are grade-level appropriate. Besides the Lexile and ATOS measurements, the *Children's Writer's Word Book* (Mogilner, 2006) and *EDL Core Vocabularies* (Taylor, Frackenpohl, White, Nieroroda, Browning, & Birsner, 1989) are used as additional measures of grade-level appropriateness. WestEd and the LDOE also draw on the professional experience of educators during content review to verify that sources are accessible to students and make changes based on their feedback.

Most of the sources chosen as part of the development cycles for 2019 and beyond were found to be below or at grade level; however, some of the authentic historical documents were evaluated as being above grade level. In those cases, footnotes were added for words that were above grade level and for words or phrases that were thought to be a potential source of confusion for students. When an authentic historical document required many footnotes or if the language was considered too arcane or incomprehensible for students at the given grade level, the document was adapted to improve readability and accessibility. These modifications were made evident by the use of the phrase "Adapted from" in the title of the document. After modification, the sources were re-evaluated to ensure that the changes resulted in the desired outcomes.

Obtaining LDOE Approval for Tasks, Item Sets, and Sources

As sources for tasks and item sets are reviewed and approved for submission to the LDOE, WestEd content leads finalize set overviews, which outline the content of the sets, identify the number and types of items to be assessed in the sets, identify the GLEs and sources associated with each item, and provide rough drafts of the item stems. WestEd then submits the set overviews and sources to the LDOE for another round of approval before beginning item writing.

For standalone items, WestEd submits the items along with their corresponding sources.

Item Writing and Review Process

WestEd employs item writers and editors for grades 3–8. Some of the WestEd writers have been part of item development since the first development cycle in 2015–2016. WestEd secures the required approval from the LDOE for each writer during their first development cycle. Writers and editors receive training from WestEd that outlines lessons learned from previous development cycles, the LDOE's expectations, and best practices for item development, including consideration of bias and sensitivity. For an outline of the information covered at the 2020–2021 training, see Appendix A for the LEAP 2025 Social Studies Grades 3–8 Item Writer and Editor Training Agenda.

After the training, item writers are provided with approved set overviews, which identify the set topics, list the GLEs to be addressed, specify the number and type of items to be written, and offer specific guidance to the item writer about how the content for each item within a set should be assessed. The use of the set overviews allows WestEd to control the quality of the task and item sets.

Once written, items go through two rounds of content editing, one round of proofreading, and a final round of review before being submitted to the LDOE for their first round of review. The LDOE has two rounds of review prior to content and bias review committee meetings. WestEd revises items based on feedback provided by the LDOE assessment staff.

Item Development Platform. Items are developed in Assessment Banking and Building solutions for Interoperable assessment (ABBI), Pearson's proprietary item development platform. In addition to the items and sources, the platform captures item metadata and allows viewers to preview items using Pearson's format viewer (TestNav 8). In this view, items appear together with their associated sources. The ability to examine the items and sources together is critical in the item review and in the evaluation of the content and cognitive demands on students.

Style Guidelines. The *LEAP Social Studies and Science Content Style Guide* is updated immediately following test construction to reflect final formatting decisions made by the LDOE. Throughout the development and review process, when questions of style arise that are unanswered by existing documentation, WestEd consults the LDOE, and approved changes are added to the Style Guide.

LDOE Content Review. As writing and editing for batches of tasks, item sets, and standalone items are completed, the batches are sent to the LDOE for content lead review. Feedback from the LDOE review is implemented before educator committees convene for content and bias review.

Content and Bias Review Committees. After the completion of item development and the initial rounds of the LDOE review, virtual content and bias review meetings are held. The LDOE recruits educators from different parts of Louisiana, who represent all Louisiana students, to serve on the committees. The meetings are led jointly by facilitators from the LDOE and WestEd. Items developed as part of the 2019–2020 development cycle were revised from previously developed items; therefore, content and bias review committees were not held for those items. Content and bias review committees were held in June 2020 for the new items developed as part of the 2020–2021 development cycle. Table 3.3 provides information about the representation of educators who participated in the content and bias reviews in June 2020.

Table 3.3

Representation of Educators Participating in the June 2020 Content and Bias Reviews

Grade	Number of Committee Participants	Classroom Teacher	Special Education	Instructional Lead or Supervisor	Visually Impaired Teacher	EL Teacher/ Supervisor	Other
3	9	5	1	—	1	1	1 Representative of Native American people
4	9	5	—	1	1	2	—
8	9	5	—	1	1	1	1 Representative of Native American people

Training and Security for Virtual Content and Bias Review. The virtual format of content and bias review allows participants to access the item development platform and vote on sources and items individually before coming together in an online meeting format to discuss the items and sources as a group. Prior to accessing the platform, WestEd provides training to explain the content and bias review process and to review the security protocols associated with the virtual pre-review and review. To orient educators

to the process, WestEd describes the criteria for evaluating items for content and bias considerations, explains how to use ABBI for item review, and shows educators how to individually review the items and record their recommendation to accept, accept with edits, or reject an item.

Committee members are provided with a pre-review day during which they access the items using ABBI and vote on the items. Comments are compiled and shared with the LDOE and WestEd facilitators prior to the joint virtual committee review. When the committee convenes as a group, the committee members revisit and discuss items and sources. A WestEd recorder takes detailed notes about discussions and records the final committee recommendations. These notes are compiled for reconciliation with the LDOE and post-review implementation. Access to the items is tightly controlled by WestEd, with password access shutting off immediately following the close of each pre-review and review section. At the close of each session, committee members are instructed to clear their internet browser history. In addition, all participants complete a nondisclosure agreement prior to accessing any items.

Results of Content and Bias Review. The results of the reviewers' recommendations are captured in ABBI. Table 3.4 provides the results based on the participants' votes following their initial review of the sources and the items. Table 3.5 shows the results of the group votes after discussing and reaching a consensus on the disposition of the sources and the items.

Table 3.4
Vote Totals Based on Individual Votes Following Initial Review of Sources and Items (June 2020)

Grade	Number of Items	Accept	Accept with Edits*	No Vote	Reject	Grand Total
3	58	450	68	2	2	522
4	49	394	40	6	3	443
8	73	562	79	8	4	653

* Votes cast as "Accept with Reconciliation" were counted as "Accept with Edits" since this vote was not used during this round of review.

Table 3.5

Vote Totals for Items Based on Group Consensus for Sources and Items (June 2020)

Grade	Number of Items and Sources	Accept	Accept with Edits	No Vote	Reject
3	58	44	14	0	0
4	49	32	17	0	0
8	73	45	28	0	0

Post Committee Finalization. At the conclusion of the content and bias reviews, WestEd content leads consult the LDOE to reconcile any unresolved committee feedback. Following the implementation of the committee's feedback, the LDOE and WestEd content leads meet virtually for final item reconciliation. WestEd provides records of all the implemented changes to the LDOE prior to the virtual reconciliation meetings. During the reconciliation meetings, the leads review the items to ensure that they were correctly edited. Once content considerations are resolved, all items and sources go through a final formal fact-checking round and two additional rounds of proofreading. Any changes resulting from these reviews are submitted to the LDOE for approval.

4. Construction of Test Forms

Initial Construction

The purpose of forms construction activities is to create an operational (OP) form for each tested grade and to embed field test items for potential use in future operational assessments. This section describes the process used to create the operational and field test forms.

Spring 2022 Operational Forms

The operational forms for grades 3–8 were the operational tests originally selected in fall 2019 for spring 2020 administration. They were not used because of the suspension of that administration due to the COVID-19 pandemic. Items approved during the data review from spring 2019 that embedded field tests were available for use on the spring 2022 operational assessments. (See the 2018–2019 Technical Report for results from the data review and reconciliation of the spring 2019 field test items.) Items approved during data review from previous years were also available for use on the spring 2022 operational assessments.

WestEd and the LDOE content staff worked together to complete item selection for one operational form per grade for consideration. WestEd submitted the form to Pearson psychometricians for consideration before formal submission to the LDOE. The operational forms were designed to adhere to the blueprint for the tested grade and exhibit the broadest possible balance of content and breadth of GLE and content coverage. The task was selected first for grades 5–8, followed by item sets with CRs, other item sets, and standalone items. For grades 3–4, item sets with CRs were selected first, followed by other item sets and standalone items. Test form developers worked to avoid cueing and clanging between items. Cueing occurs when the content in one item provides clues to the answer of another item. Clanging refers to the overlap or similarity of content. Because the content was purposely distributed across sessions, cueing and clanging were intended to have been avoided; however, developers also conducted a separate review of

the forms to check for inadvertent cueing and clanging. During item selection, test maps were created to capture details of the forms, including each item's unique identification number (UIN), test session, item sequence, item descriptions, and associated item metadata. Tables 4.1–4.6 provide the operational test composition for the grades 3–8 spring 2022 forms.

Table 4.1

Grade 3 Social Studies Operational Test Composition for 2022

Item Sets/Item Types	Total Sets	Total Points per Set	SR	CR	Total Items	Total Points
6-Item Set	1	6	6	–	6	6
6-Item Set with CR	2	7	10	2	12	14
5-Item Set	2	5	10	–	10	10
Standalone Items	–	–	15	–	15	15
Total Items	5	–	41	2	43	45

Table 4.2

Grade 4 Social Studies Operational Test Composition for 2022

Item Sets/Item Types	Total Sets	Total Points per Set	SR	CR	Total Items	Total Points
6-Item Set	3	6	18	–	18	18
5-Item Set with CR	2	6	8	2	10	12
Standalone Items	–	–	15	–	15	15
Total Items	5	–	41	2	43	45

Table 4.3

Grade 5 Social Studies Operational Test Composition for 2022

Item Sets/Item Types	Total Sets	Total Points per Set	SR	CR	TE	ER	Total Items	Total Points
6-Item Set with TE	2	7	10	–	2	–	12	14
5-Item Set with CR	2	6	8	2	–	–	10	12
5-Item Set with TE	1	6	4	–	1	–	5	6
Standalone Items	–	–	14	–	–	–	14	14
5-Item Task	1	12	4	–	–	1	5	12
Total Items	6	–	40	2	3	1	46	58

Table 4.4

Grade 6 Social Studies Operational Test Composition for 2022

Sets and	Total	Total Points	SR	CR	TE	ER	Total	Total
6 Item Set SR Only	1	6	6	–	–	–	6	6
6-Item Set with TE	2	7	10	–	2	–	12	14
6-Item Set with CR	2	7	10	2	–	–	12	14
5-Item Set with TE	1	6	4	–	1	–	5	6
Standalone Items	–	–	14	–	–	–	14	14
5-Item Task	1	–	4	–	–	1	5	12
Total Items	7	–	48	2	3	1	54	66

Table 4.5

Grade 7 Social Studies Operational Test Composition for 2022

Sets and	Total	Total Points	SR	CR	TE	ER	Total	Total
6-Item Set with TE	2	7	10	–	2	–	12	14
6-Item Set with CR	1	7	5	1	–	–	6	7
5-Item Set with TE	2	6	8	2	–	–	10	12
5-Item Set with CR	1	6	4	1	–	–	5	6
Standalone Items	–	–	15	–	–	–	15	15
5-Item Task	1	12	4	–	–	1	5	12
Total Items	–	–	46	4	2	1	53	66

Table 4.6

Grade 8 Social Studies Operational Test Composition for 2022

Sets and	Total	Total Points	SR	CR	TE	ER	Total	Total
6-Item Set with SR	1	6	6	–	–	–	6	6
6-Item Set with TE	2	7	10	–	2	–	12	14
6-Item Set with CR	2	7	10	2	–	–	12	14
5-Item Set with TE	1	6	4	–	1	–	5	6
Standalone Items	–	–	14	–	–	–	14	14
5-Item Task	1	12	4	–	–	1	5	12
Total Items	7	–	48	2	3	1	54	66

Spring 2022 Field Test Forms

For grades 3–8, item sets and standalone items that informed the 2019–2020 development cycle were made available for field testing. Items that informed the 2020–2021 development cycle were field-tested in grade 3 only. Sets and standalone items were placed on multiple field test forms, with a different combination of items on each form, to ensure field testing of the maximum number of items in each set. Six field test forms were administered for grades 3–4. Eight field test forms were administered for grade 5. Nine field test forms were administered for grades 6–8. Test maps captured details about the field test forms.

Field test items were embedded in Session 2 for grades 3 and 4 and part of a fourth session for grades 5–8.

Field test forms for the spring 2022 administration were organized as follows.

Grade 3

- Four forms had one 7-item set with CR (in Session 2)
- One form had one 6-item set with CR and one standalone item (in Session 2)
- One form had one 5-item set with CR and two standalone items (in Session 2)

Grade 4

- Two forms had one 6-item set with CR and one standalone item (in Session 2)
- Two forms had one 5-item set with CR and two standalone item (in Session 2)
- One form had one 5-item set with SR items only and two standalone SR items (in Session 2)
- One form had one 4-item set with CR and three standalone items (in Session 2)

Grade 5

- Two forms had one 7-item set with TE and one standalone item (in Session 4)
- Four forms had one 6-item set with CR and two standalone items (in Session 4)
- Two forms had one 6-item set with TE and two standalone items (in Session 4)

Grade 6

- Seven forms had one 5-item set with CR and one 5-item set with TE (in Session 4)
- One form had one 5-item set with TE and five standalone items (in Session 4)
- One form had one 5-item set with CR and five standalone items (in Session 4)

Grade 7

- Three forms had two 5-item sets with TE (in Session 4)
- Three forms had one 5-item set with CR and one 5-item set with TE (in Session 4)
- Two forms had one 5-item set with CR and five standalone items (in Session 4)
- One form had one 5-item set with TE and five standalone items (in Session 4)

Grade 8

- Four forms had two 5-item sets with TE (in Session 4)
- Four forms had one 5-item set with CR and five standalone items (in Session 4)
- One form had one 5-item set with TE and five standalone items (in Session 4)

Items were repeated on field test forms as necessary to fill all available positions.

Revision and Review

Psychometric Approval of Operational Forms

Prior to submitting the forms to the LDOE staff for review, Pearson psychometricians and WestEd content specialists participate in an iterative process of reviewing and revising the forms. The psychometric review consists of comparisons of the expected representation and the actual representation of reporting categories (History, Geography, Civics, Economics) and item types—selected response (SR), constructed response (CR), technology enhanced (TE), and extended response (ER)—on the operational forms.

The answer keys for multiple-choice (MC) items also are examined, to determine whether any forms have significantly non-uniform distributions of correct responses (A, B, C, and D). Spreadsheets are used to generate frequency tables of reporting categories, item types, and MC answer keys for each form. They are also used to compare the operational forms from previous years for each grade. Deviations from the blueprint are identified and addressed. Test characteristic curves (TCC) based on item response theoretic models are applied to the data, and conditional standard errors of measurement are computed for each iteration during the test construction process to evaluate how well a proposed test form matches psychometric targets. Psychometric approval from Pearson is provided for all the forms prior to submission to the LDOE for their review. Please refer to the following table for criteria to flag items based on scoring points.

Table 4.7

Summary of Flagging Criteria to Select/Flag Items: Classical Analysis and IRT

Point	*P*-value		P-B	DIF	IRT		
	Low Bound	Upper Bound	Lower Bound	Exclude	A	b	C
1	0.25	0.90	0.20	C	0.35 – 3.50	-3.00 – 3.00	< 0.35
2 and higher	0.25	0.90	0.20		0.35 – 3.50	-3.00 – 3.00	N/A

Note: Detailed information can be found in the 2018–2019 Framework and Test Construction Document. It should be noted that these values are psychometric recommendations. Actual item decision occurs by content staff based on these recommendation criteria.

LDOE Review

Following the psychometric reviews, the test maps and constructed sets for each grade are delivered to the LDOE for approval. Forms are reviewed by both the LDOE content and psychometric staff. Based on the LDOE review, sets or items are replaced and resequenced as requested. After these changes, the overall balance of answer choices and key runs is re-evaluated, and final adjustments are made to achieve the appropriate balance. Pearson also updates the TCC and SEM curves to be sure that the selection meets psychometric targets.

Finalized test maps are used to create PDF versions, or constructed sets, of the forms, which are reviewed by WestEd's proofreaders before the items are transferred from ABBI to DRC.

Online and Paper Versions

For grade 3, forms are mainly delivered on paper, and at least one of the forms is identified for delivery online. For grades 4–8, all forms are delivered online. One form in each grade is designated by the LDOE as the accommodated version, to be used with

students who have accommodations. For grades 4–8, the accommodated version is available in print form to students who require paper testing. The accommodated version is also rendered in braille. To support students with low or no vision, additional text (alternate text) is provided to describe the graphic components of the assessments. Content specialists evaluate the graphics and draft the alternate text. Table 4.8 shows the number of online and paper forms for each grade in spring 2022.

Table 4.8
Numbers of LEAP 2025 Forms for Spring 2022 Operational and Embedded Field Test

Grade	Paper Forms	Online Forms
3	6	1*
4	0	6**
5	0	8
6	0	9
7	0	9
8	0	9

* Same form as one of the paper forms.
** A form 1A was created as an online form with TE companion forms as the regular and online accommodated form.

Whenever possible, the comparability between the primary test mode at a grade level (paper for grade 3; online for grades 4–8) and the secondary mode (online for grade 3; accommodated forms for grades 4–8) is evaluated empirically. For grade 3, for example, assessment results are separately analyzed by paper and by online versions. For grades 4–8, a historical limitation for the same types of analysis exists because of the lack of examinees who take an accommodated version of the test. Comparability between online and accommodated versions of tests in grades 4–8 is defined by comprehensive content evaluations.

5. Test Administration

This chapter describes processes and activities implemented and information disseminated to help ensure standardized test administration procedures and, thus, uniform test administration conditions for students. According to the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (2014) *Standards for Educational and Psychological Testing* (hereafter the *Standards*), "The usefulness and interpretability of test scores require that a test be administered and scored according to the developer's instructions" (111). This chapter examines how test administration procedures implemented for the Louisiana Education Assessment Program 2025 (LEAP 2025) strengthen and support the intended score interpretations and reduce construct-irrelevant variance that could threaten the validity of score interpretations.

Training of School Systems

To ensure that the LEAP 2025 assessments are administered and scored in accordance with the department's policies, the LDOE takes a primary role in communicating with and training school system personnel. The LDOE provides train-the-trainer opportunities for the district test coordinators, who in turn convey test administration training to schools within their school systems. The LDOE conducts quality-assurance visits during testing to ensure adherence to the standardized administration of the tests.

The district test coordinators are responsible for the schools within their systems. They disseminate information to each school, offer assistance with test administration, and serve as liaisons between the LDOE and their school systems. The LDOE also provides assistance with and interpretation of assessment data and test results.

Ancillary Materials

Ancillary materials for LEAP 2025 test administration contribute to the body of evidence of the validity of score interpretation. This section examines how the test materials address the *Standards* related to test administration procedures.

For the spring test administration, Data Recognition Corporation (DRC) produces two administration manuals: the *LEAP 2025 Grade 3 Paper-Based Test Administration Manual* (TAM) and the *LEAP 2025 Grades 3–8 Computer-Based Test Administration Manual* (TAM). The TAMs provide detailed instructions for administering the LEAP assessments. The manuals include information on test security, test administrator responsibilities, test preparation, administration of tests (computer-based or paper-based), and post-test procedures.

Table of Contents for *LEAP 2025 Paper-Based Test Administration Manual* (TAM)

- Spring 2022 Notes and Reminders
- Test Administrator Pre-Administration Oath of Security and Confidentiality Statement
- Test Administrator Post-Administration Oath of Security and Confidentiality Statement
- Overview
- Test Security
    - Secure Test Materials
    - Testing Irregularities and Security Breaches
    - Testing Environment
    - Violations of Test Security
    - Answer Change Analysis
    - Voiding Student Tests
- Test Administrator Responsibilities
- Test Administration Checklists
    - Before Testing

    - During Testing
    - After Testing (Daily)
    - After Testing (Last Day)
- Test Administrators' Frequently Asked Questions
- Test Materials
    - Receipt of Test Materials
- Testing Guidelines
    - Testing Eligibility
    - Test Schedule
    - Extended Time for Testing
- Testing Times
    - Makeup Testing
    - Testing Conditions
- Special Populations and Accommodations
    - IDEA Special Education Students
    - Students with One or More Disabilities According to Section 504
    - Gifted and Talented Special Education Students
    - Test Accommodations for Special Education and Section 504 Students
    - Special Considerations for Deaf and Hard-of-Hearing Students
    - English Learners (ELs)
- Hand-Coded Consumable Test Booklets
- Students Absent from Testing
- Consumable Test Booklet Coding
    - Coding the Demographic Section
- Sample Grade 3 English Language Arts Consumable Test Booklet
- General Instructions for LEAP 2025
    - Student Marking/Erasing on Consumable Test Booklet

    - Reading Directions to Students
    - Special Instructions
- Directions for Administering LEAP 2025 Tests
- Post-Test Procedures
    - Test Administrator Oath of Security and Confidentiality Statement
    - Used and Unused Consumable Test Booklets (Defined)
    - Transferring Student Responses
    - Returning Test Materials to the School Test Coordinator
- Index

Table of Contents for *LEAP 2025 Computer-Based Test Administration Manual* (TAM)

- Spring 2022 Notes and Reminders
- Test Administrator Pre-Administration Oath of Security and Confidentiality Statement
- Test Administrator Post-Administration Oath of Security and Confidentiality Statement
- Overview
- Test Security
    - Secure Test Materials
    - Testing Irregularities and Security Breaches
    - Testing Environment
    - Violations of Test Security
    - Voiding Student Tests
- Test Administrator Responsibilities
    - Software Tools and Features for Test Administrators
- Test Administration Checklists
    - Before Testing

    - During Testing
    - After Testing (Daily)
    - After Testing (Last Day)
- Test Administrators' Frequently Asked Questions
- Test Materials
    - Receipt of Test Materials
- Testing Guidelines
    - Testing Eligibility
    - Testing Schedule
    - Extended Time for Testing
- Testing Times for Grades 3 through 8
    - Makeup Testing
    - Testing Conditions
- Online Tools Training
- Student Tutorials
    - Student Tutorials
- Special Populations and Accommodations
    - IDEA Special Education Students
    - Students with One or More Disabilities According to Section 504
    - Gifted and Talented Special Education Students
    - Test Accommodations for Special Education and Section 504 Students
    - Special Considerations for Deaf and Hard-of-Hearing Students
    - English Learners (ELs)
- General Instructions
    - Reading Directions to Students
- LEAP 2025: Grades 3–8 English Language Arts (All Sessions)
- LEAP 2025: Grades 3–8 Mathematics (All Sessions)

- LEAP 2025: Grades 3–8 Science (Sessions 1–2)
- LEAP 2025: Grades 5–8 Science Session 3 Select Schools Only
- LEAP 2025: Grades 3–8 Social Studies (Grades 3–4 Sessions 1–2, Grades 5–8 Sessions 1–3)
- LEAP 2025: Grades 5–8 Social Studies Session 4 Select Schools Only
- Post-Test Procedures
    - Test Administrator Post-Administration Oath of Security and Confidentiality Statement
    - Returning Test Materials to the School Test Coordinator
- Index

DRC also produces test coordinator manuals for paper- and computer-based test administrations. The TCMs provide detailed instructions for district and school test coordinators' responsibilities for distributing, collecting, and returning test materials to DRC for scoring.

Table of Contents for *LEAP 2025 Paper-Based Testing Test Coordinators Manual* (TCM)

- Key Dates
- Spring 2022 Alerts
- Pre-Administration Oath of Security and Confidentiality Statement
- Post-Administration Oath of Security and Confidentiality Statement
- General Information
- Test Security
    - Key Definitions
    - Violations of Test Security
    - Answer Change Analysis
    - Voiding Student Tests
- Testing Guidelines
    - Testing Eligibility

    - Testing Conditions
    - Test Schedule
    - Extended Time for Testing
    - Extended Breaks
    - Makeup Testing
    - Test Administration Resources
- Testing Times for Grade 3
- District Test Coordinator
    - Conduct Training Session
    - Receive Test Materials
    - Large-Print and Braille Test Materials and Communication Assistance Scripts (CAS)
    - Accommodated Materials
    - Verify and Distribute Test Materials to School Test Coordinators
    - Request Additional Test Materials and Bar-Code Labels
    - Collect Materials from Schools After Testing
    - Used and Unused Consumable Test Booklets (Defined)
    - Unscorable Documents and Unscorable Document Labels
- Directions for Returning Test Materials to DRC in May
    - Pickup 1: ELA and Mathematics Scorable Test Materials
    - Pickup 2: Science and Social Studies Scorable Test Materials
    - Pickup 3: Nonscorable Test Materials
    - Final Checklist for Returning Test Materials to DRC
- School Test Coordinator
    - Receive and Verify Test Materials
    - Conduct Test Administration and Security Training Session

    - Supervise Application of Bar-Code Labels and Coding of Consumable Test Booklets
    - Soiled, Damaged, and Other Unscorable Consumable Test Booklets
    - Verify and Distribute Materials to Test Administrators
    - Supervise Test Administration
    - Collect Test Materials
    - Used and Unused Consumable Test Booklets (Defined)
    - Coding Responsibilities of Principals—Before Testing
    - Coding Responsibilities of Principals—Before or After Testing
    - Coding Responsibilities of Principals—After Testing
- Directions for Returning Test Materials to District Test Coordinator
    - Pickup 1: ELA and Mathematics Scorable Test Materials
    - Pickup 2: Science and Social Studies Scorable Test Materials
    - Pickup 3: Nonscorable Test Materials
    - Final Checklist for Returning Test Materials to DTC
- Void Notification—Spring 2022
- Index

Table of Contents for *LEAP 2025 Computer-Based Testing Test Coordinators Manual* (TCM)

- Key Dates Spring 2022
- Resources Available in DRC INSIGHT Portal (eDIRECT) Spring 2022
- Spring 2022 Alerts
- Pre-Administration Oath of Security and Confidentiality Statement
- Post-Administration Oath of Security and Confidentiality Statement
- General Information
    - DRC INSIGHT Portal (eDIRECT) and INSIGHT
- Test Security

    - Key Definitions
    - Violations of Test Security
- Testing Guidelines
    - Testing Eligibility
    - Testing Conditions
    - Testing Schedule
    - Extended Time for Testing
    - Extended Breaks
    - Accommodations
    - Makeup Testing
    - Test Administration Resources
- Testing Times for Grades 3 through 8
- Roles and Responsibilities
    - District Test Coordinator
    - School Test Coordinator
    - Technology Coordinator
- Managing Test Tickets
    - Student Transfers
    - Locked Test Tickets
    - Technical Issues
    - Invalidating Test Tickets
- Resources for Online Testing
    - Test Administration Manuals
    - *DRC INSIGHT Portal (eDIRECT) User Guide*
    - *LEAP 2025 Accommodations and Accessibility Features User Guide*
    - *INSIGHT Technology User Guide*
    - Online Tools Training (OTT)

    - Student Tutorials
- Void Notification—Spring 2022

The LDOE assessment staff review, provide feedback, and give final approval for these manuals. The manuals are inclusive of grades 3–8 English Language Arts (ELA), Mathematics, Social Studies, and Science.

The *Standards* contain multiple references relevant to test administration. Information in the TAMs addresses these in the following manner.

Standard 4.15.

> The directions for test administration should be presented with sufficient clarity so that it is possible for others to replicate the administration conditions under which the data on reliability, validity, and (where appropriate) norms were obtained. Allowable variations in administration procedures should be clearly described. The process for reviewing requests for additional testing variations should also be documented. (90)

The TAMs provide instructions for activities that happen before, during, and after testing with sufficient detail and clarity to support reliable test administrations by qualified test administrators. To ensure uniform administration conditions throughout the state, instructions in the TAMs describe the following: general rules of paper and online testing; assessment duration, timing, and sequencing information; and the materials required for testing.

Standard 6.1. Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer and any instructions from the test user. (114)

To ensure the usefulness and interpretability of test scores and to minimize sources of construct-irrelevant variance, it was essential that the LEAP 2025 tests were administered according to the prescribed TAMs. It should be noted that adhering to the test schedule is

also a critical component. The TCMs included instructions for scheduling the test within the state testing window. The TAMs and TCMs also contained the schedule for timing each test session.

Standard 6.3. Changes or disruptions to standardized test administration procedures or scoring should be documented and reported to the test user. (115)

Department staff release annual test security reports that describe a wide range of improper activities that may occur during testing, including the following: copying and reviewing test questions with students; cueing students during testing, verbally or with written materials on the classroom walls; cueing students nonverbally, such as by tapping or nodding the head; allowing students to correct or complete answers after tests have been submitted; splitting sessions into two parts; ignoring the standardized directions in the online assessment; paraphrasing parts of the test to students; changing or completing (or allowing other school personnel to change or complete) student answers; allowing accommodations that are not written in the Individualized Education Program (IEP), Individual Accommodation Plan/504 Plan (IAP), or English Learner Plan (EL plan); allowing accommodations for students who do not have an IEP, IAP, or EL plan; or defining terms on the test.

Standard 6.4. The testing environment should furnish reasonable comfort with minimal distractions to avoid construct-irrelevant variance. (116)

The TAMs outline the steps that teachers should take to prepare the classroom testing environment for administering the LEAP 2025 test. These include the following:

- Determine the layout of the classroom environment.
- Plan seating arrangements. Allow enough space between students to prevent the sharing of answers.
- Eliminate distractions such as bells or telephones.
- Use a Do Not Disturb sign on the door of the testing room.
- Make sure classroom maps, charts, and any other materials that relate to the content and processes of the test are covered or removed or are out of the students' view.

Standard 6.6. Reasonable efforts should be made to ensure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent or deceptive means. (116)

The TAMs present instructions for post-test activities to ensure that online tests are submitted and printed test materials are handled properly to maintain the integrity of student information and test scores. Detailed instructions guide test examiners in submitting all online test records. For students who were administered a large-print or braille version of the LEAP 2025 assessment, examiners are instructed to transcribe students' responses from the large-print or braille test book into the online testing system (INSIGHT) exactly as they responded in the large-print or braille test book.

Standard 6.7. Test users have the responsibility of protecting the security of test materials at all times. (117)

Throughout the manuals, test coordinators and examiners are reminded of test security requirements and procedures to maintain test security. Specific actions that are direct violations of test security are noted. Detailed information about test security procedures is presented under "Test Security" in the manuals.

Return Material Forms and Guidelines

The paper-based TCM instructs test coordinators regarding procedures for organizing and packing materials and returning them to DRC for secure inventory purposes. The LDOE assessment staff have opportunities to review, provide feedback, and give final approval of the guidelines. The purpose of the instructions is to ensure that secure test materials are properly accounted for and organized appropriately for the return shipment.

Security Checklists

As soon as printed test materials are received by a school system, the district test coordinator ensures that the first and last security barcodes on the tests match the packing list they received. The district test coordinator then packages the tests to be sent to schools. Upon returning test books to DRC, school and district test coordinators are required to complete and submit an accountability form that details the number of test books or printed test forms returned. This form also requires that systems/schools document nonstandard situations, including lost, damaged, destroyed, extra, or missing test books.

Interpretive Guides

Essential to making valid interpretations of test scores is an understanding of what the test scores mean and how to interpret score reports. The Interpretive Guide is written for Louisiana teachers and administrators who receive the LEAP 2025 score reports. https://www.louisianabelieves.com/resources/library/assessment-guidance

Time

Each session of each content area test is timed to provide sufficient time for students to attempt all items. Only students with extended time accommodation were permitted to exceed the established time limits of any given session. The manuals provide examiners with timing guidelines for the assessments.

Online Forms Administration, Grades 3–8

The online forms are administered via DRC's INSIGHT online assessment system. School system and school personnel set up test sessions via DRC's INSIGHT portal (eDIRECT) and print test tickets. Students enter their ticket information to access the test in INSIGHT. In addition, students have access to the Online Tools Training (OTT) before the testing window, which allows them to practice using tools and features within INSIGHT. Tutorials with online video clips that demonstrate features of the system are also available to students before testing.

Paper-Based Forms Administration, Grade 3

Schools with testers at grade 3 had the option to participate in either paper-based or computer-based testing for the spring 2022 test. DRC prints and ships paper materials to the sites that choose paper-based testing. These materials are returned to DRC after testing for processing and scoring with the online tests.

Accessibility and Accommodations

Accessibility features and accommodations include Access for All, Accessibility Features, and Accommodations.

- Access for All features are available to all students taking an assessment.
- Accessibility Features are available to students when deemed appropriate by a team of educators.
- Accommodations must appear in a student's IEP/IAP/EL plan.

Accommodations may be used with students who qualify under the Individuals with Disabilities Education Act (IDEA) and have an IEP or Section 504 of the Americans with Disabilities Act and have an IAP, or who are identified as English Learners (ELs) and have an EL plan.

Accommodations must be specified in the qualifying student's individual plan and must be

consistent with accommodations used during daily classroom instruction and testing. The use of any accommodation must be indicated on the student information sheet at the time of test administration. AERA, APA, and NCME Standard 6.2 states:

> When formal procedures have been established for requesting and receiving accommodations, test takers should be informed of these procedures in advance of testing. (115)

In compliance with this standard, the TAM contains the list of Universal Tools, Designated Supports, and Accommodations permissible for the LEAP assessments. The following accommodations were provided by DRC for this administration:

- Braille
- Text-to-Speech
- Directions in Native Language

The following additional access and accommodation features were also available:

- Answers Recorded
- Extended Time
- Transferred Answers
- Individual/Small Group Administration
- Tests Read Aloud
- English/Native Language Word-to-Word Dictionary
- Directions Read Aloud/Clarified in Native Language
- Text-to-Speech for online testers
- Human Read Aloud
- Directions in Native Language

For more details about these accommodations, please refer to the *LEAP 2025 Accessibility and Accommodations Manual*.

Testing Windows

The computer-based testing window was available from April 25 through May 25, 2022. Paper-based testing occurred from April 27 through May 3, 2022.

Test Security Procedures

Maintaining the security of all test materials is crucial to preventing the possibility of random or systematic errors, such as unauthorized exposure of test items that would affect the valid interpretation of test scores. Several test security measures are implemented for the LEAP 2025 assessments. Test security procedures are discussed throughout the TCMs and TAMs.

Test coordinators and administrators are instructed to keep all test materials in locked storage, except during actual test administration, and access to secure materials must be restricted to authorized individuals only (e.g., test administrators and the school test coordinator). During the testing sessions, test administrators are directly responsible for the security of the LEAP 2025 assessment and must account for all test materials and supervise the test administrators at all times.

Data Forensic Analyses

Due to the importance of the LEAP 2025 assessment, it is prudent to ensure that the results from the assessments are based on effective instruction and true student achievement. To help ensure that scores are related to actual learning and that results are valid, data forensic analyses take place to assist in separating meaningful gains from spurious gains. It is important to note that although the results of the analyses may be used to identify potential problems within a school, the identification of a problem is not an accusation of misconduct.

Multiple methods are incorporated into the forensic analysis. The following methods are applied:

- Response Change Analysis

- Score Fluctuation Analysis
- Web Monitoring
- Plagiarism Detection
- Alerts for Disturbing Content

Response Change Analysis. Students make changes to answer choices when taking the LEAP 2025 assessments, and this behavior is expected. Unfortunately, changes to student answers are sometimes influenced by school personnel who want to improve performance. Therefore, the response change analysis is conducted to identify school- and test administrator-level response change patterns that are statistically improbable when compared to the expected pattern at the state level.

Score Fluctuation Analysis. It is anticipated that performance on the LEAP 2025 assessments will improve over time for legitimate reasons such as changes in the curriculum and improvement in instruction. However, large and unexpected score changes may be a sign of testing impropriety. The LDOE applied an approach where the state's level of change in performance from one year to the next is compared to schools' and test administrators' change in performance during the same time frame. Schools and test administrators are identified when the level of change is statistically unexpected.

Web Monitoring. The content of the LEAP 2025 assessments should not appear outside the boundaries of the forms administered. To protect Louisiana test content, the internet is monitored for postings that contain, or appear to contain, potentially exposed and/or copied test content. When test content is verified, steps are taken to quickly remove the infringing content.

Plagiarism Detection. The LDOE monitors for two different plagiarism situations: copying from student to student and copying from an outside source, such as Wikipedia or another internet source. Instances of plagiarism are identified by human scorers and artificial intelligence. Alerts are set to identify responses that may indicate the possibility of teacher interference or plagiarism. Alerted responses are given additional review so that the appropriate response can be taken.

Alerts for Disturbing Content. Scorers for the LEAP 2025 assessments also have the ability to apply an alert flag to student responses that may indicate disturbing content (e.g., possible physical or emotional abuse, suicidal ideation, threats of harm to themselves or others). All alerted responses are automatically routed to the scoring director, who reviews and forwards appropriate responses to senior project staff for review. If it is concluded that a response warrants an alert, project management will contact the LDOE to take the necessary action. At no point during this process do scorers or staff have access to demographic information for any students participating in the assessment.

6. Scoring Activities

Directory of Test Specification (DOTS) Process. DRC creates a DOTS file, based on the approved test selection. The DOTS is a document containing information about each item on a test form, such as item identifier, item sequence, answer key, score points, subtype, session, alignment, and prior use of item. WestEd reviews and confirms the contents of the DOTS file as part of test review rounds. The DOTS file is then provided to the LDOE for review and final approval. Once approved, the information contained in the DOTS is used in scoring the test and in reporting.

Selected-Response (SR) Item Keycheck. SR items for Social Studies include multiple-choice (MC) and multiple-select (MS) questions. Pearson calculates MC and MS item statistics and flags items if item statistics fall outside expected ranges. For example, items are flagged if few students select the correct response (*p*-value less than 0.15), if the item does not discriminate well between students of lower and higher ability (point-biserial correlation less than 0.20), or if many students (more than 40%) select a certain incorrect response. Lists of flagged MC and MS items, with the reasons for flagging, are provided to the LDOE and WestEd content staff for key verification. The staff reviews the list of flagged MC and MS items to confirm that the answer keys are accurate. The scoring of MC and MS items is also evaluated at data review.

Scoring of Technology-Enhanced (TE) Items. All TE items are processed through DRC's autoscoring engine and scored according to the assigned scoring rules established during content creation by WestEd in conjunction with the LDOE. DRC ensures that all rubrics and scoring rules are verified for accuracy before scoring any TE items. DRC has an established adjudication process for TE items to verify that correct answers are identified. DRC's TE scoring process includes the following procedures:

- A scoring rubric is created for each TE item. The rubric describes the one and only correct answer for dichotomously scored items (i.e., items scored as either right or wrong). If partial credit is possible, the rubric describes in detail the type of response that could receive credit for each score point.

- The information from each scoring rubric is entered into the scoring system within the item banking system so that the truth resides in one place along with the item image and other metadata. This scoring information designates specific information that varies by item type. For example, for a drag-and-drop item, the information includes which objects are to be placed in each drop region to receive credit.
- The information is then verified by another autoscoring expert.
- After testing starts, reports are generated that show every response, how many students gave that response, and the score the scoring system provided for that response.
- The scoring is then checked against the scoring rubric using two levels of verification.
- If any discrepancies are found, the scoring information is modified and verified again. The scoring process is then rerun. This checking and modification process continues until no other issues are found.
- As a final check, a final report is generated that shows all student responses, their frequencies, and their received scores.

In the case of braille and accommodated print test forms, student responses to TE items are transcribed into the online system by a test administrator.

Adjudication. TE items and other eligible items identified in the test map are automatically scored as tests are processed. TE items are scored according to scoring rules in the DOTS, which includes scoring information for all item types.

The adjudication process focuses on detecting possible errors in scoring TE and MS items. DRC provides a report listing the frequency distributions of TE item responses and MS items. Members of the LDOE and WestEd content staff examine the TE and MS response distributions and the auto-frequency reports to evaluate whether the items are scored appropriately. In the event that scoring issues are identified, WestEd content staff and the LDOE recommend changes to the scoring algorithm. Any changes to the scoring algorithm are based on the LDOE's decisions. DRC, in turn, applies the approved scoring changes to any affected items.

Constructed-Response and Extended-Response Scoring

Constructed-response items are scored by human raters trained by DRC. Extended-response items are scored by Project Essay Grade (PEG), an Artificial Intelligence (AI) scoring engine. Ten percent of the responses are scored twice to monitor and maintain inter-rater reliability. Scoring supervisors also conduct read-behinds and review all nonscores and alerts. Handscoring processing rules are detailed in the LEAP 2025 Spring 2022 Handscoring/AI Documentation document.

Selection of Scoring Evaluators. Standard 4.20 states the following:

> The process for selecting, training, qualifying, and monitoring scorers should be specified by the test developer. The training materials, such as the scoring rubrics and examples of test takers' responses that illustrate the levels on the rubric score scale, and the procedures for training scorers should result in a degree of accuracy and agreement among scorers that allows the scores to be interpreted as originally intended by the test developer. Specifications should also describe processes for assessing scorer consistency and potential drift over time in raters' scoring. (92)

The following sections explain how scorers are selected and trained for the LEAP 2025 handscoring process and describe how the scorers are monitored throughout the handscoring process.

Recruitment and Interview Process. DRC strives to develop a highly qualified, experienced core of evaluators to appropriately maintain the integrity of all projects. All readers hired by DRC to score 2021–2022 LEAP 2025 test responses had at least a four-year college degree.

DRC has a human resources director dedicated solely to recruiting and retaining the handscoring staff. Applications for reader positions are screened by the handscoring project manager, the human resources director, or recruiting staff to create a large pool of potential readers. In the screening process, preference is given to candidates with previous experience scoring large-scale assessments and with degrees emphasizing the

appropriate content areas. At the personal interview, reader candidates are asked to demonstrate their proficiency in writing by responding to a DRC writing topic and their proficiency in mathematics by solving word problems with correct work shown. These steps result in a highly qualified and diverse workforce. DRC personnel files for readers and team leaders include evaluations for each project completed. DRC uses these evaluations to place individuals on projects that best fit their professional backgrounds, their college degrees, and their performances on similar projects at DRC. Once placed, all readers go through rigorous training and qualifying procedures specific to the project on which they are placed. Any scorer who does not complete this training and does not demonstrate the ability to apply the scoring criteria by qualifying at the end of the process is not allowed to score live student responses.

Security. Whether training and scoring are conducted within a DRC facility or done remotely, security is essential to the handscoring process. When users log into DRC's secure, web-based scoring application, ScoreBoard, they are required to read and accept the security policy before they are allowed to access any project. For each project, scorers are also required to read and sign non-disclosure agreements, and during training emphasis is always given to what security means, the importance of maintaining security, and how this is accomplished.

Readers only have access to student responses they are qualified to score. Each scorer is assigned a unique username and password to access DRC's imaging system and must qualify before viewing any live student responses. DRC maintains full control of who may access the system and which item each scorer may score. No demographic data is available to scorers at any time.

Each DRC scoring center is a secure facility. Access to scoring centers is limited to badge-wearing staff and to visitors accompanied by authorized staff. All readers are made aware that no scoring materials may leave the scoring center. To prevent the unauthorized duplication of secure materials, cell phone/camera use within the scoring rooms is strictly forbidden. Readers only have access to student responses they are qualified to score.

In a remote environment, security reminders are given on a daily basis. Similar to the work that occurs within DRC scoring sites, in a remote environment, education about security expectations is the best way to maintain security of any project materials. DRC requires scorers working remotely to work in a private environment away from other people (including family members). Restrictions are in place that define the hours during the day scorers are able to log into the system. If any type of security breach were to occur, immediate action would be taken to secure materials, and the employee would be terminated. DRC has the same policy within the scoring centers.

Handscoring Training Process. Standard 6.9 specifies:

> Those responsible for test scoring should establish and document quality control processes and criteria. Adequate training should be provided. The quality of scoring should be monitored and documented. Any systematic source of scoring errors should be documented and corrected. (118)

Training Material Development. DRC scoring supervisors train scorers using the LDOE-approved training materials. These materials are developed by DRC and LDOE staff from a selection scored by Louisiana educators at rangefinding and include the following:

- Prompts and associated sources
- Rubrics
- Anchor sets
- Practice sets
- Qualifying sets

Training and Qualifying Procedures. Handscoring involves training and qualifying team leaders and evaluators, monitoring scoring accuracy and production, and ensuring security of both the test materials and the scoring facilities. The LDOE reviews training materials and oversees the training process.

Qualifying Standards. Scorers demonstrate their ability to apply the scoring criteria by qualifying (i.e., scoring with acceptable agreement with true scores on qualifying sets).

After each qualifying set is scored, the DRC scoring director responsible for training leads the scorers in a discussion of the set.

Any scorer who does not qualify by the end of the qualifying process for an item is not allowed to score live student responses.

Monitoring the Scoring Process. Standard 6.8 states:

> Those responsible for test scoring should establish scoring protocols. Test scoring that involves human judgment should include rubrics, procedures, and criteria for scoring. When scoring of complex responses is done by computer, the accuracy of the algorithm and processes should be documented. (118)

The following section explains the monitoring procedures that DRC uses to ensure that handscoring evaluators follow established scoring criteria while items are being scored. Detailed scoring rubrics, which specify the criteria for scoring, are available for all constructed- and extended-response items.

Reader Monitoring Procedures. Throughout the handscoring process, DRC project managers, scoring directors, and team leaders review the statistics that are generated daily. DRC uses one team leader for every 10 to 12 readers. If scoring concerns are apparent among individual scorers or if a scorer needs clarification on the scoring rules, team leaders address those issues on an individual basis. DRC supervisors typically monitor one out of five of the scorer's readings, making adjustments to that ratio as needed. If a supervisor disagrees with a reader's scores during monitoring, the supervisor provides retraining in the form of direct feedback to the reader, using rubric language and applicable training responses.

Validity Sets and Inter-Rater Reliability. In addition to the feedback that supervisors provide to readers during regular read-behinds and the continuous monitoring of inter-rater reliability and score point distributions, DRC also conducts validity scoring using the LDOE-approved validity responses identified by the DRC scoring supervisors during live scoring for newly operational items. Validity responses are inserted among the live student responses.

The validity responses are added to DRC's image handscoring system prior to the beginning of scoring. Validity reports compare readers' scores to predetermined scores and are used to help detect potential room drift as well as individual scorer drift. This data is used to make decisions regarding the retraining and/or release of scorers, as well as the rescoring of responses.

Approximately 10% of all live student responses are scored by a second reader to establish inter-rater reliability statistics for all constructed- and extended-response items. This procedure is called a "double-blind read" because the second reader does not know the first reader's score. DRC monitors inter-rater reliability based on the responses that are scored by two readers. If a scorer falls below the expected rate of agreement, the team leader or scoring director retrains the scorer. If a scorer fails to improve after retraining and feedback, DRC removes the scorer from the project. In this situation, DRC removes all scores assigned by the scorer in question. The responses are then reassigned and rescored.

To monitor inter-rater reliability, DRC produces scoring summary reports daily. DRC's scoring summary reports display exact, adjacent, and nonadjacent agreement rates for each reader. These rates are calculated based on responses that are scored by two readers, and their definitions are included below.

- Percentage Exact (%EX)—total number of responses by reader where scores are the same, divided by the number of responses that were scored twice
- Percentage Adjacent (%AD)—total number of responses by reader where scores are one point apart, divided by the number of responses that were scored twice
- Percentage Nonadjacent (%NA)—total number of responses by reader where scores are more than one point apart, divided by the number of responses that were scored twice

Each reader is required to maintain a level of exact agreement on validity responses and on inter-rater reliability. Additionally, readers are required to maintain a low rate of nonadjacent agreement.

Calibration Sets. DRC pulls calibration responses for items. DRC uses these sets to perform calibration across the entire scorer population for an item if trends are detected (e.g., low agreement between certain score points if a certain type of response is missing from initial training). These calibrations are designed to help refocus scorers on how to properly use the scoring guidelines. They are selected to help illustrate particular points and familiarize scorers with the types of responses commonly seen during operational scoring. After readers score a calibration set, the scoring director reviews it from the front of the room, using rubric language and scoring concepts exemplified by the anchor responses to explain the reasoning behind each response's score.

Reports and Reader Feedback. Reader performance and intervention information are recorded in reader feedback logs. These logs track information about actions taken with individual readers to ensure scoring consistency in regard to reliability, score point distribution, and validity performance. In addition to the reader feedback logs, DRC provides the LDOE with handscoring quality control reports for review throughout the scoring window.

Inter-Rater Reliability. A minimum of 10% of the responses for constructed- and extended-response items are scored independently by a second reader. This is the case regardless of whether the first reader is a human rater or AI. The statistics for inter-rater reliability are calculated for all items at all grades. To determine the reliability of scoring, the percentage of perfect agreement and adjacent agreement between the first and second scores is examined.

Tables 6.1–6.4 provide the inter-rater reliability and score point distributions by grade level for the constructed-response and extended-response items administered in the spring 2022 forms.

Table 6.1

Inter-Rater Reliability for Operational Constructed-Response Items

Grade	Item	Inter-Rater Reliability*			
		2x	Exact Agreement (%)	Adjacent Agreement (%)	Nonadjacent (%)
3**	Item 1	≥15,350	89	11	0
	Item 2	≥15,240	92	8	0
4	Item 1	≥17,070	87	13	0
	Item 2	≥19,250	92	8	0
5	Item 1	≥16,880	88	12	0
	Item 2	≥20,930	91	9	0
6	Item 1	≥16,330	89	11	0
	Item 2	≥20,680	90	10	0
7	Item 1	≥18,880	89	11	0
	Item 2	≥24,460	91	9	0
8	Item 1	≥14,600	85	15	0
	Item 2	≥16,230	85	15	0

* The percent may not add up to 100% due to rounding.

** This grade report combines both online and paper forms.

Table 6.2

Score Point Distributions for Operational Constructed-Response Items

Grade	Item	Score Point Distribution*					
		Total	"0" Rating (%)	"1" Rating (%)	"2" Rating (%)	Blank (%)	Nonscore Codes (%)**
3***	Item 1	≥58,940	53	25	8	6	8
	Item 2	≥58,950	45	26	14	7	8
4	Item 1	≥56,910	34	42	10	0	13
	Item 2	≥57,960	57	21	5	0	17
5	Item 1	≥56,740	55	26	5	0	13
	Item 2	≥58,900	37	34	9	0	20
6	Item 1	≥57,010	42	34	11	0	12
	Item 2	≥59,200	54	22	4	0	21
7	Item 1	≥59,820	27	40	16	0	16
	Item 2	≥62,680	30	33	10	0	25
8	Item 1	≥57,270	26	36	30	0	9
	Item 2	≥58,240	33	42	13	0	11

* The percent may not add up to 100% due to rounding.

** Nonscore codes include Foreign language (F), Insufficient (I), Don't Understand (N), Refusal (R), Off Topic (T), and Unintelligible (U). Responses that cannot be assigned a score based on the rubric are assigned a nonscore code and count as zero points toward student scores.

*** This grade report combines both online and paper forms.

Table 6.3

Inter-Rater Reliability for Operational Extended-Response Items

Grade		Inter-Rater Reliability*			
	2x	Dimension	Exact Agreement (%)	Adjacent Agreement (%)	Nonadjacent (%)
5	≥46,480	Content	94	6	0
		Claim	95	5	0
6	≥47,770	Content	94	6	0
		Claim	94	6	0
7	≥45,350	Content	96	4	0
		Claim	96	4	0
8	≥62,720	Content	96	4	0
		Claim	96	4	0

* The percent may not add up to 100% due to rounding.

Table 6.4

Score Point Distributions for Operational Extended-Response Items

Grade	Score Point Distribution*								
	Total	Dimension	"0" (%)	"1" (%)	"2" (%)	"3" (%)	"4" (%)	Blank (%)	Nonscore Codes (%)**
5	≥71,660	Content	44	26	7	1	0	0	21
		Claim	46	24	7	1	0	0	21
6	≥72,740	Content	35	29	8	2	0	0	26
		Claim	30	34	7	1	0	0	26
7	≥73,120	Content	31	37	10	3	0	0	19
		Claim	37	30	10	4	0	0	19
8	≥81,430	Content	20	34	23	8	1	0	13
		Claim	19	38	21	7	2	0	13

* The percent may not add up to 100% due to rounding.

** Nonscore codes include Foreign language (F), Insufficient (I), Don't Understand (N), Refusal (R), Off Topic (T), and Unintelligible (U). Responses that cannot be assigned a score based on the rubric are assigned a nonscore code and count as zero points toward student scores.

7. Data Analysis

Classical Item Statistics

This section describes the classical item analysis for data obtained from the operational LEAP 2025 Social Studies tests. The classical analysis includes statistical analysis based on the following types of items: multiple-choice/multiple-select items, rule-based machine-scored items such as technology-enhanced items, and handscored items such as constructed- and extended-response items. For each operational item, the statistical analysis produces item difficulty (*p*-value) and item discrimination (point-biserial).

Tables and figures that provide additional information on classical item statistics for the spring 2022 test can be found in Appendix C: Item Analysis Summary Report. Tables C.1–C.5 show the summaries of classical item statistics. As a measure of item difficulty, p (or "the *p*-value") indicates the average proportion of total points earned on an item. For example, if p = 0.50 on an MC item, then half of the examinees earned a score of 1. If p = 0.50 on a CR item, then examinees earned half of the possible points on average (e.g., 1 out of 2 possible points). A measure of point-biserial correlation indicates the correlation between an item score and the total test score. Items with higher item-total correlations provide better information about how well items discriminate between lower- and higher-performing students. It should be also noted that a corrected point-biserial correlation indicates the correlation between an item score and the total test score, where the item score is not included in the total score. The results can be found in Tables C.1-C4. By the way, the statistical analysis results for field test (FT) items are stored in Pearson's Assessment Banking and Building solutions for Interoperable assessment (ABBI) system.

Differential Item Functioning

Differential item functioning (DIF) analyses are intended to statistically signal potential item bias. DIF is defined as a difference between similar-ability groups' (e.g., males or females that attain the same total test score) probability of getting an item correct.

Because test scores can reflect many sources of variation, the test developers' task is to create assessments that measure the intended knowledge and skills without introducing construct-irrelevant variance. When tests measure something other than what they are intended to measure, test scores may reflect those extraneous elements in addition to what the test is purported to measure. If this occurs, these tests can be called biased (Angoff, 1993; Camilli & Shepard, 1994; Green, 1975; Zumbo, 1999). Different cultural and socioeconomic experiences are among some factors that can confound test scores intended to reflect the measured construct.

One DIF methodology applied to dichotomous items was the Mantel–Haenszel (*MH*) DIF statistic (Holland & Thayer, 1988; Mantel & Haenszel, 1959). The MH method is a frequently used method that offers efficient statistical power (Clauser & Mazor, 1998). The *MH* chi-square statistic is

$$MH_{\chi^2} = \frac{(\sum_k F_k - \sum_k E(F_k))^2}{\sum_k Var(F_k)},$$

where F_k is the sum of scores for the focal group at the *k*th level of the matching variable (Zwick, Donoghue, & Grima, 1993). Note that the *MH* statistic is sensitive to *N* such that larger sample sizes increase the value of the chi-square.

In addition to the *MH* chi-square statistic, the *MH* delta statistic (*ΔMH*), first developed by the Educational Testing Service (ETS), was computed. To compute the *ΔMH DIF*, the *MH* alpha (the odds ratio) is calculated:

$$\alpha_{MH} = \frac{\sum_{k=1}^{K} N_{r1k} N_{f0k} / N_k}{\sum_{k=1}^{K} N_{f1k} N_{r0k} / N_k},$$

where N_{r1k} is the number of correct responses in the reference group at ability level *k*, N_{f0k} is the number of incorrect responses in the focal group at ability level *k*, N_k is the total number of responses, N_{f1k} is the number of correct responses in the focal group at

ability level *k*, and N_{r0k} is the number of incorrect responses in the reference group at ability level *k*. The *MH DIF* statistic is based on a 2×2×*M* (2 groups × 2 item scores × *M* strata) frequency table, in which students in the reference (male or white) and focal (female or black) groups are matched on their total raw scores.

The *ΔMH DIF* is then computed as

$$\Delta MH\ DIF = -2.35 \ln(\alpha_{MH}).$$

Positive values of *ΔMH DIF* indicate items that favor the focal group (i.e., positive DIF items are differentially easier for the focal group); negative values of *ΔMH DIF* indicate items that favor the reference group (i.e., negative DIF items are differentially easier for the reference group). Ninety-five percent confidence intervals for *ΔMH DIF* are used to conduct statistical tests.

The *MH* chi-square statistic and the *ΔMH DIF* were used in combination to identify operational test items exhibiting strong, weak, or no DIF (Zieky, 1993). Table 7.1 defines the DIF categories for dichotomous items.

Table 7.1
DIF Categories for Dichotomous Items

DIF Category	Criteria
A (negligible)	| *ΔMH DIF* | is not significantly different from 0.0 or is less than 1.0.
B (slight to moderate)	1. | *ΔMH DIF* | is significantly different from 0.0 but not from 1.0, and is at least 1.0; OR 2. | *ΔMH DIF* | is significantly different from 1.0 but is less than 1.5. Positive values are classified as "B+" and negative values as "B–."
C (moderate to large)	| *ΔMH DIF* | is significantly different than 1.0 and is at least 1.5. Positive values are classified as "C+" and negative values as "C–."

For polytomous items, the standardized mean difference (*SMD*) (Dorans & Schmitt, 1991; Zwick, Thayer, & Mazzeo, 1997) and the Mantel χ^2 statistic (Mantel, 1963) are used to identify items with DIF. *SMD* estimates the average difference in performance between the reference group and the focal group while controlling for student ability. To calculate the *SMD*, let *M* represent the matching variable (total test score). For all $M = m$, identify the students with raw score m and calculate the expected item score for the reference group (E_{rm}) and the focal group (E_{fm}). DIF is defined as $D_m = E_{fm} - E_{rm}$, and *SMD* is a weighted average of D_m using the weights $w_m = N_{fm}$ (the number of students in the focal group with raw score m), which gives the greatest weight at score levels most frequently attained by students in the focal group.

$$\text{SMD} = \frac{\sum_m w_m (E_{fm} - E_{rm})}{\sum_m w_m} = \frac{\sum_m w_m D_m}{\sum_m w_m}$$

The *SMD* is converted to an effect-size metric by dividing it by the standard deviation of item scores for the total group. A negative *SMD* value indicates an item on which the focal group has a lower mean than the reference group, conditioned on the matching variable. On the other hand, a positive *SMD* value indicates an item on which the reference group has a lower mean than the focal group, conditioned on the matching variable.

The *MH DIF* statistic is based on a $2\times(T+1)\times M$ (2 groups × T+1 item scores × M strata) frequency table, where students in the reference and focal groups are matched on their total raw scores (T = maximum score for the item). The Mantel χ^2 statistic is defined by the following equation:

$$\text{Mantel } \chi^2 = \frac{\left(\sum_m \sum_t N_{rtm} Y_t - \sum_m \frac{N_{r+m}}{N_{++m}} \sum_t N_{+tm} Y_t\right)^2}{\sum_m Var\left(\sum_t N_{rtm} Y_t\right)}.$$

The *p*-value associated with the Mantel χ^2 statistic and the *SMD* (on an effect-size metric) are used to determine DIF classifications. Table 7.2 defines the DIF categories for polytomous items.

Table 7.2

DIF Categories for Polytomous Items

DIF Category	Criteria
A (negligible)	Mantel χ^2 p-value > 0.05 or $\lvert SMD/SD \rvert \leq 0.17$
B (slight to moderate)	Mantel χ^2 p-value < 0.05 and $0.17 < \lvert SMD/SD \rvert < 0.25$
C (moderate to large)	Mantel χ^2 p-value < 0.05 and $\lvert SMD/SD \rvert \geq 0.25$

Three DIF analyses were conducted for the operational test items only: female/male, black/white, and Hispanic/white. That is, item score data were used to detect items on which female or male students performed unexpectedly well or unexpectedly poorly, given their performance on the full assessment. The same methods were used to detect items on which both black/white and Hispanic/white students performed unexpectedly well or unexpectedly poorly, given their performance on the full assessment. The last two columns of Tables 7.3.1–7.3.3 provide the number of items flagged for DIF. Items flagged with A-DIF show negligible DIF, items flagged with B-DIF are said to exhibit slight to moderate DIF, and items with C-DIF are said to exhibit moderate to large DIF. Very few operational test items were flagged for C-DIF by either analysis.

Note that DIF flags for dichotomous items are based on the *MH* statistics while DIF flags for polytomous items are based on the combination of Mantel χ^2 p-value and *SMD* statistics. Because the spring 2022 test was administered during the COVID-19 pandemic, great caution should be applied when any statistical inference is drawn.

Table 7.3.1

Summary of Female/Male DIF Flags by Grade

Grade	A	[B+],[B-]	[C+],[C-]
3	43	[0],[0]	[0],[0]
4	43	[0],[0]	[0],[0]
5	45	[1],[0]	[0],[0]
6	54	[0],[0]	[0],[0]
7	51	[1],[1]	[0],[0]
8	51	[0],[2]	[1],[0]

Table 7.3.2

Summary of African American/White DIF Flags by Grade

Grade	A	[B+],[B-]	[C+],[C-]
3	42	[0],[1]	[0],[0]
4	43	[0],[0]	[0],[0]
5	45	[0],[1]	[0],[0]
6	53	[0],[1]	[0],[0]
7	52	[0],[1]	[0],[0]
8	53	[0],[1]	[0],[0]

Table 7.3.3

Summary of Hispanic/White DIF Flags by Grade

Grade	A	[B+],[B-]	[C+],[C-]
3	43	[0],[0]	[0],[0]
4	43	[0],[0]	[0],[0]
5	46	[0],[0]	[0],[0]
6	54	[0],[0]	[0],[0]
7	53	[0],[0]	[0],[0]
8	54	[0],[0]	[0],[0]

Measurement Models

IRTPRO, a software application for item calibration and test scoring, was used to estimate IRT parameters from LEAP 2025 test data. MC, MS, and some TE items (i.e., one-point) were scored dichotomously (0/1), so the three-parameter logistic model (3PL) was applied to those data:

$$p_i(\theta_j) = c_i + \frac{1-c_i}{1+e^{-Da_i(\theta_j - b_i)}}.$$

In that model, $p_i(\theta_j)$ is the probability that student *j* would earn a score of 1 on item *i*, b_i is the difficulty parameter for item *i*, a_i is the slope (or discrimination) parameter for item *i*, c_i is the pseudo-chance (or guessing) parameter for item *i*, and *D* is the constant 1.7. Since the Social Studies tests also included polytomous items scored higher than 1 point, the generalized partial credit model (GPCM) (Muraki, 1992) was used to estimate the parameters of these items:

$$p_{im}(\theta_j) = \frac{\exp[\sum_{k=0}^{m} Da_i(\theta_j - b_i + d_{ik})]}{\sum_{v=0}^{M_i-1} \exp[Da_i(\theta_j - b_i + d_{iv})]},$$

where $a_i(\theta_j - b_i + d_{i0}) \equiv 0$, $p_{im}(\theta_j)$ is the probability of an examinee with θ_j getting score *m* on item *i*, and *Mi* is the number of score categories of item *i* with possible item scores as consecutive integers from 0 to *Mi* – 1. In the GPCM, the *d* parameters define the "category intersections" (i.e., the θ value at which examinees have the same probability of scoring 0 and 1, 1 and 2, etc.).

Calibration and Linking

LEAP 2025 Social Studies assessments are standards-based assessments that have been constructed to align to the LSSS, as defined by the LDOE and Louisiana educators. For each course, the content standards specify the subject matter students should know and the skills they should be able to perform. In addition, performance standards specify how much of the content standards students need to master in order to achieve proficiency. Constructing tests to content standards enables the tests to assess the same constructs from one year to the next.

Item Response Theory (IRT) models were used in the item calibration for the LEAP 2025 Social Studies tests. All calibration activities were independently replicated by Pearson staff as an added quality-control check.

The most common and straightforward way to score a test is to simply use the sum of points a student earned on the test, namely, the raw score. Although the raw score is conceptually simple, it can be interpreted only in terms of a particular set of items. When new test forms are administered in subsequent administrations, other types of derived scores must be used to compensate for any differences in the difficulty of the items and to allow direct comparisons of student performance between administrations.

Thus, the primary purpose of form equating is to establish score equivalency between two (or more) forms. Equivalency is established by first building the forms to be equated according to content specifications. Then the form scores are placed on the same scale (by equating), such that students performing on two scaled assessments at the same level of underlying achievement should receive the same scale score on both forms, although they may not receive the same number-correct score (or raw score). The LDOE and Pearson strive to maintain equivalent samples or use near-census samples over the years, minimizing the potential differences caused by the different samples.

It should be noted that spring 2017 is the first operational administration for the LEAP Social Studies tests; in the spring of 2021, the forms used were intact, and when originally administered in 2019, they were post-equated and linked to the LEAP 2025 scale.

Tables 7.4.1–7.4.6 provide scale scores at selected percentiles that can be used to compare the distributional characteristics of the spring 2022 test form to previous administrations. Although these scale scores are rounded values, there were differences

in the scale score values for a given percentile across the forms. These variations could arise for several reasons: (1) differences in the proficiency (i.e., achievement) of the students in the samples or growth in student achievement across years; (2) unevenness in the respective distributions that combine with the number-correct-to-scale-score scoring method, leaving "gaps" in the scale; or (3) other sources of equating error. In general, however, the test characteristic function equating techniques will "level" the equated forms through the raw-to-scale-score adjustment.

Table 7.4.1

Comparisons of Scale Scores at Selected Percentiles: Grade 3 Operational Forms

Percentile	2017 Spring Form A	2018 Spring Form B	2019 Spring Form C	2021 Spring Form C	2022 Spring Form D
99	787	801	810	810	807
95	771	778	786	781	779
90	761	770	772	768	767
85	753	762	764	756	760
80	750	754	756	748	753
75	743	751	752	744	746
70	740	743	744	736	739
65	732	739	740	732	735
60	729	735	736	724	728
55	725	731	728	719	723
50	721	727	724	714	719
45	713	719	719	708	715
40	709	714	714	703	710
35	705	709	708	696	704
30	700	705	703	688	698
25	691	699	696	688	691
20	685	694	688	679	683
15	680	681	679	668	674
10	666	673	668	653	660
5	650	653	650	650	650
1	650	650	650	650	650

Table 7.4.2

Comparisons of Scale Scores at Selected Percentiles: Grade 4 Operational Forms

Percentile	2017 Spring Form A	2018 Spring Form B	2019 Spring Form C	2021 Spring Form C	2022 Spring Form D
99	796	801	810	808	808
95	775	781	784	777	788
90	768	771	774	768	775
85	757	762	766	757	768
80	753	755	758	753	760
75	747	752	754	746	753
70	743	746	747	738	749
65	736	743	743	734	741
60	732	736	739	730	737
55	729	733	735	722	732
50	721	730	727	717	727
45	717	722	723	712	722
40	713	718	719	707	711
35	709	715	714	702	704
30	705	710	709	696	697
25	696	706	704	689	697
20	691	696	698	681	688
15	680	691	685	662	666
10	673	679	676	650	650
5	650	662	650	650	650
1	650	650	650	650	650

Table 7.4.3

Comparisons of Scale Scores at Selected Percentiles: Grade 5 Operational Forms

Percentile	2017 Spring Form A	2018 Spring Form B	2019 Spring Form C	2021 Spring Form C	2022 Spring Form D
99	794	798	803	803	802
95	774	778	783	780	780
90	762	768	772	769	769
85	755	760	764	762	761
80	748	753	759	754	754
75	742	748	754	749	748
70	737	743	749	740	743
65	731	738	743	738	737
60	725	733	738	731	734
55	722	730	735	725	727
50	718	727	728	721	723
45	714	721	725	717	715
40	710	717	721	713	710
35	705	713	717	709	705
30	700	709	713	703	698
25	694	705	709	697	691
20	686	699	703	690	682
15	676	693	697	680	671
10	662	677	680	667	656
5	650	665	667	650	650
1	650	650	650	650	650

Table 7.4.4

Comparisons of Scale Scores at Selected Percentiles: Grade 6 Operational Forms

Percentile	2017 Spring Form A	2018 Spring Form B	2019 Spring Form C	2021 Spring Form C	2022 Spring Form D
99	799	800	807	803	792
95	776	779	783	779	772
90	765	767	772	769	761
85	757	760	763	760	754
80	750	756	757	752	747
75	745	750	752	746	740
70	740	745	746	741	735
65	735	741	741	735	731
60	730	739	738	730	723
55	725	734	733	725	718
50	719	729	727	719	713
45	716	727	725	713	707
40	713	722	719	709	704
35	706	716	713	702	698
30	702	713	709	694	690
25	698	706	702	690	685
20	687	703	694	685	674
15	681	694	690	673	668
10	665	683	679	666	659
5	654	668	658	650	650
1	650	650	650	650	650

Table 7.4.5

Comparisons of Scale Scores at Selected Percentiles: Grade 7 Operational Forms

Percentile	2017 Spring Form A	2018 Spring Form B	2019 Spring Form C	2021 Spring Form C	2022 Spring Form D
99	807	820	816	816	814
95	786	792	795	792	792
90	775	781	783	781	779
85	768	773	775	770	771
80	759	766	768	763	764
75	755	762	760	755	757
70	748	755	755	750	750
65	743	750	750	743	745
60	737	746	745	737	740
55	734	741	740	731	732
50	728	738	734	725	726
45	721	733	728	722	720
40	718	727	725	715	714
35	714	721	718	711	710
30	704	718	715	702	701
25	698	710	706	697	696
20	692	702	702	691	691
15	684	697	691	684	685
10	661	685	677	667	677
5	650	667	656	650	655
1	650	650	650	650	650

Table 7.4.6

Comparisons of Scale Scores at Selected Percentiles: Grade 8 Operational Forms

Percentile	2017 Spring Form A	2018 Spring Form B	2019 Spring Form C	2021 Spring Form C	2022 Spring Form D
99	815	822	821	815	816
95	789	797	796	793	791
90	780	783	786	779	779
85	771	775	779	774	771
80	766	770	771	765	765
75	759	765	765	760	758
70	755	760	760	755	753
65	748	755	758	750	746
60	743	750	753	743	741
55	739	743	748	738	737
50	734	739	743	733	729
45	730	734	738	728	724
40	723	729	733	723	718
35	717	724	728	715	712
30	712	719	723	709	704
25	706	713	715	703	696
20	700	707	709	695	686
15	688	697	699	687	679
10	678	689	691	676	663
5	655	673	676	662	650
1	650	650	650	650	650

Operational Item Parameters

The distributions of item parameters are summarized in Appendix C. Appendix C also provides graphical displays of the distributions of IRT parameter estimates for each grade. TE, CR, and ER items have no *c* parameters because they are polytomous items and are therefore modeled using the GPCM. The number of item parameters associated with the ER items reflect item parameter estimates associated with particular "part scores" that comprise the total ER item. By the way, it should be noted that statistical results of FT items can be found at Pearson ABBI.

Item Fit

IRT scaling algorithms attempt to find item parameters (numerical characteristics) that create a match between observed patterns of item responses and theoretical response patterns defined by the selected IRT models. The Q_1 statistic (Yen, 1981) is used as an index for how well theoretical item curves match observed item responses. Q_1 is computed by first conducting an IRT item parameter estimation, then estimating students' achievement using the estimated item parameters, and, finally, using students' achievement scores in combination with estimated item parameters to compute expected performance on each item. Differences between expected item performance and observed item performance are then compared at 10 selected equal intervals across the range of student achievement. Q_1 is computed as a ratio involving expected and observed item performance. Q_1 is interpretable as a chi-square (χ^2) statistic, which is a statistical test that determines whether the data (observed item performance) fit the hypothesis (the expected item performance). Q_1 for each item type has varying degrees of freedom because the different item types have different numbers of IRT parameters. Therefore, Q_1 is not directly comparable across item types. An adjustment or linear transformation (translation to a *Z*-score, Z_{Q_1}) is made for different numbers of item parameters and sample size to create a more comparable statistic.

It should be noted that Yen's Q_1 statistic (Yen, 1981) was calculated to evaluate item fit for both operational and field test items by comparing observed and expected item

performance. MAP (maximum *a posteriori*) estimates from IRTPRO were used as student ability estimates. For dichotomous items, Q_1 is computed as

$$Q_{1i} = \sum_{j=1}^{j} \frac{N_{ij}(O_{ij}-E_{ij})^2}{E_{ij}(1-E_{ij})},$$

where N_{ij} is the number of examinees in interval (or group) *j* for item *i*, O_{ij} is the observed proportion of the examinees in the same interval, and E_{ij} is the expected proportion of the examinees for that interval. The expected proportion is computed as

$$E_{ij} = \frac{1}{N_{ij}} \sum_{a \in j}^{N_{ij}} P_i(\hat{\theta}_a),$$

where $P_i(\hat{\theta}_a)$ is the item characteristic function for item *i* and examinee *a*. The summation is taken over examinees in interval *j*.

The generalization of Q_1 for items with multiple response categories is

$$Gen\ Q_{1i} = \sum_{j=1}^{10} \sum_{k=1}^{m_i} \frac{N_{ij}(O_{ikj}-E_{ikj})^2}{E_{ikj}},$$

where

$$E_{ikj} = \frac{1}{N_{ij}} \sum_{a \in j}^{N_{ij}} P_{ik}\ (\hat{\theta}_a).$$

Both Q_1 and generalized Q_1 results are transformed to ZQ_1 and are compared to a criterion $ZQ_{1,\text{crit}}$ to determine whether fit is acceptable. The conversion formulas are

$$ZQ_1 = \frac{Q_1 - df}{\sqrt{2df}}$$

and

$$ZQ_{1,crit} = \frac{N}{1500} * 4,$$

where *df* is the degrees of freedom (the number of intervals minus the number of independent item parameters). Items are categorized as exhibiting either fit or misfit.

A summary of IRT item parameter statistics and item fit for operational items is displayed in Appendix D: Dimensionality.

Dimensionality and Local Item Independence

By fitting all items simultaneously to the same achievement scale, IRT is operating under the assumption that there is a single predominant construct that underlies the performance of all items. Under this assumption, item performance should be related to achievement and, additionally, any relationship of performance between pairs of items should be explained or accounted for by variance in students' levels of achievement. This is the "local item independence" assumption of unidimensional IRT and is associated with a test for unidimensionality called the Q_3 statistic (Yen, 1984).

Computation of the Q_3 statistic starts with expected student performance on each item, which is calculated using item parameters and estimated achievement scores. Then, for each student and each item, the difference between expected and observed item performance is calculated. The difference is the remainder in performance after accounting for underlying achievement. If performance on an item is driven by a predominant achievement construct, then the residual will be small (as tested by the Q_1 statistic), and the correlation between residuals of the item pairs will also be small. These correlations are analogous to partial correlations or the relationship between two variables (items) after accounting for the effects of a third variable (underlying achievement). The correlation among IRT residuals is the Q_3 statistic.

When calculating the level of local item dependence for two items (*i* and *j*), the Q_3 statistic is

$$Q_3 = r_{d_i d_j}.$$

The correlation between d_i and d_j values is the correlation of the residuals—that is, the difference between expected and observed scores for each item. For test taker *k*,

$$d_{ik} = u_{ik} - P_i(\theta_k),$$

where u_{ik} is the score of the kth test taker on item i and $P_i(\theta_k)$ represents the probability of test taker k responding correctly to item i.

With n items, there are $n(n - 1)/2$ Q_3 statistics. If an assessment consists of 48 items, for example, there are 1,128 Q_3 values. The Q_3 values should all be small. Summaries of the distributions of Q_3 are provided in Appendix D: Dimensionality. Specifically, Q_3 data are summarized by minimum, 5th percentile, median, 95th percentile, and maximum values for LEAP 2025 Social Studies grades 3 through 8. To add perspective to the meaning of Q_3 distributions, the average zero-order correlation (simple intercorrelation) among item responses is also shown. If the achievement construct accounts for the relationships between items, Q_3 values should be much smaller than the zero-order correlations. The Q_3 summary tables in the dimensionality reports in Appendix D show for all grades and subjects that at least 90% (between the 5th and 95th percentiles) of the items are expectedly small. These data, coupled with the Q_1 data, indicate that the unidimensional IRT model provides a reasonable solution to capture the essence of student social studies achievement defined by the selected set of items for each grade level.

Scaling

Based on the panelist recommendations and LDOE approval, the scale is set using two cut scores, Basic and Mastery, with fixed scale score points of 725 and 750, respectively. The scale scores for Approaching Basic and Advanced vary by grade level. The highest obtainable scale score (HOSS) and lowest obtainable scale score (LOSS) for the scale determined by the LDOE are 650 and 850.

IRT ability estimates (θs) are transformed to the reporting scale with a linear transformation equation of the form

$$SS = A\theta + B,$$

where *SS* is scale score, θ is IRT ability, *A* is a slope coefficient, and *B* is an intercept. The slope can be calculated as

$$A = \frac{SS_{Mastery} - SS_{Basic}}{\theta_{Mastery} - \theta_{Basic}},$$

where $\theta_{Mastery}$ is the Mastery cut score on the theta scale, and θ_{Basic} is the Basic cut score on the theta scale. $SS_{Mastery}$ and SS_{Basic} are the Mastery and Basic scale score cuts, respectively. With *A* calculated, *B* are derived from the equation

$$SS_{Mastery} = A\theta_{Mastery} + B,$$

which are rearranged as

$$B = SS_{Mastery} - A\theta_{Mastery} \text{ or } B = SS_{Mastery} - \frac{SS_{Mastery} - SS_{Basic}}{\theta_{Mastery} - \theta_{Basic}}\theta_{Mastery}.$$

Thus, the general equation for converting θs to scale scores is

$$SS = \left(\frac{SS_{Mastery} - SS_{Basic}}{\theta_{Mastery} - \theta_{Basic}}\right)\theta + \left(SS_{Mastery} - \frac{SS_{Mastery} - SS_{Basic}}{\theta_{Mastery} - \theta_{Basic}}\theta_{Mastery}\right).$$

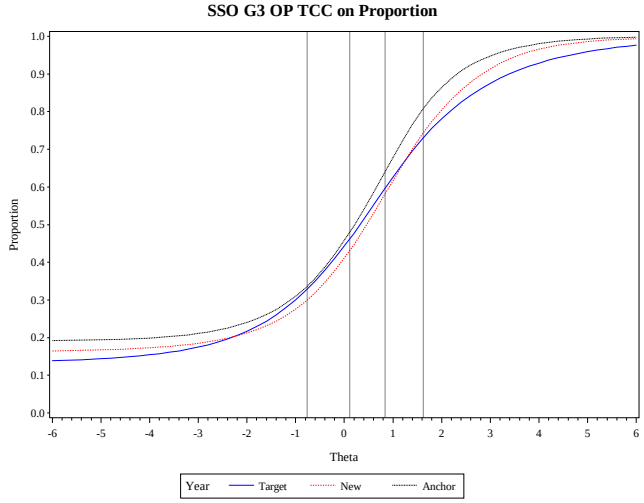
The scaling constants *A* and *B* are calculated, and the Advanced cut score and the Approaching Basic cut score on the θ scale are transformed to the reporting scale, rounded to the nearest integer. At this point, the score ranges associated with the five achievement levels are determined. The same scaling constants *A* and *B* are used to convert student ability estimates to the reporting scale until new achievement level standards are set. Descriptive Statistics and Frequency Distribution of LEAP 2025 Social Studies Scale Scores can be found in Appendix E: Scale Distribution and Statistical Report.

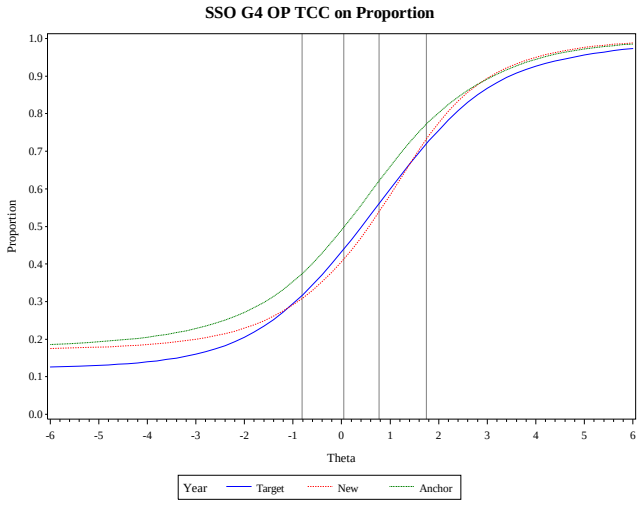
Test Characteristic Curve (TCC)

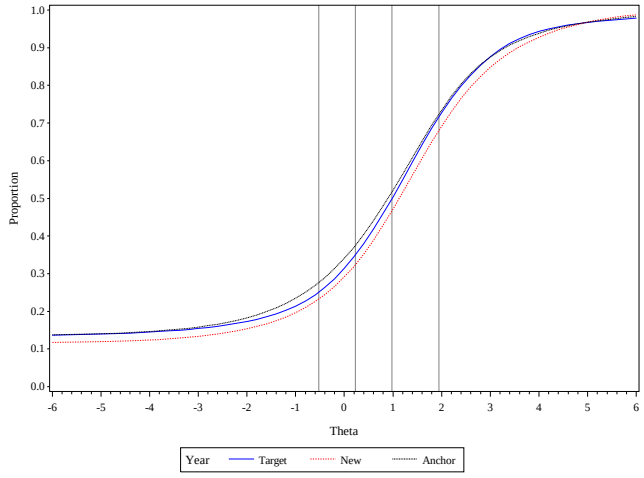
Additional evidence of comparability can be found by reviewing the test characteristic curves (TCCs) across administrations (i.e., 2022 operational form and base year form which was administered in 2017) of the LEAP 2025 Social Studies assessments, as can be seen in the following figure. As seen from Plot 7.1, the TCCs between two years were similar across ability ranges. By the way, Plot 9.1 also indicates that the SEMs between two years are similar across ability ranges, especially in the middle ability ranges; each theta cut matches the scale score of each performance-level cut (e.g., 696, 725, 750, and 783 for Grade 4).

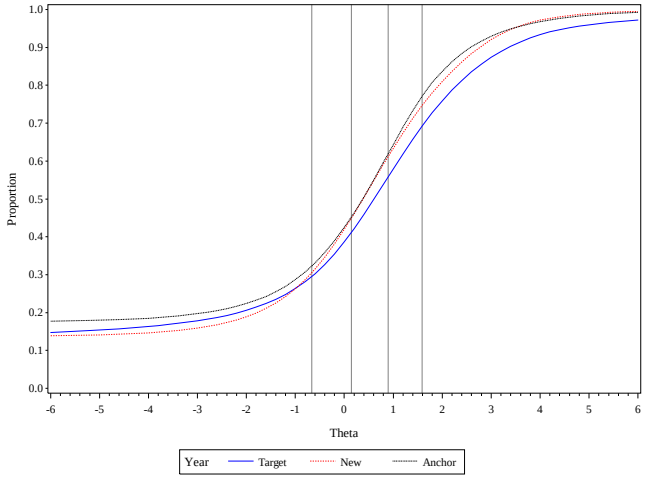
Plot 7.1

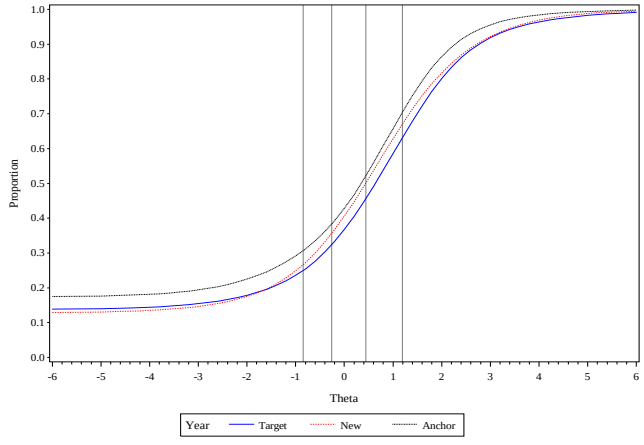
Test Characteristic Curve: Operational Social Studies Gr3–8

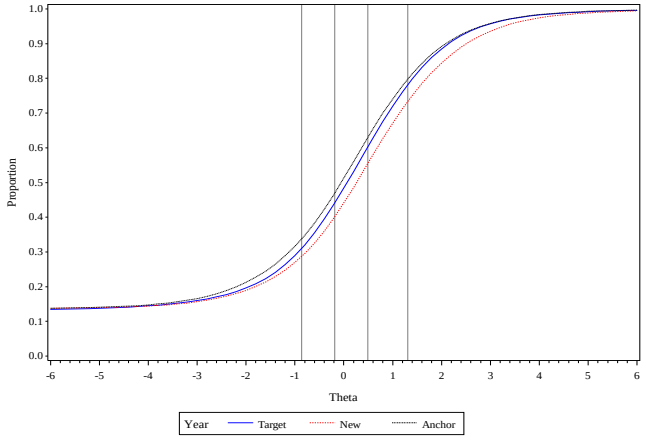












Test Information Curve, Score Distribution, and IRT Difficulty Distribution

In this section, students' Social Studies test score distribution, IRT item difficulty (i.e., b-parameter) distribution, and item information curve are presented. Compared to the base year (i.e., 2017 test), the 2022 test generally followed the shape of the base year's test information and provided more test information around the middle range of theta than other ranges regardless of grades, as can be observed from Tables 7.5.1–7.5.6 and Plot 7.2.

Table 7.5.1

Spring 2022 Students' Score and IRT B-Parameter Distribution: Grade 3

Percent of Students' Theta	Theta Range	Number of Items of IRT-B
3.51	theta < -3.5	0
2.57	-3.5 <= theta < -3.0	0
0.00	-3.0 <= theta < -2.5	0
3.34	-2.5 <= theta < -2.0	0
4.31	-2.0 <= theta < -1.5	0
10.13	-1.5 <= theta < -1.0	1
10.44	-1.0 <= theta < -0.5	3
18.34	-0.5 <= theta < 0.0	4
14.68	0.0 <= theta < 0.5	4
14.59	0.5 <= theta < 1.0	14
10.40	1.0 <= theta < 1.5	10
5.11	1.5 <= theta < 2.0	3
1.88	2.0 <= theta < 2.5	1
0.54	2.5 <= theta < 3.0	3
0.10	3.0 <= theta < 3.5	0
0.05	3.5 <= theta	0
-6.00	**Minimum**	-1.32
4.47	**Maximum**	2.82
-0.30	**Mean**	0.82
1.61	**SD**	0.92
≥49,310	**Total**	43

Table 7.5.2

Spring 2022 Students' Score and IRT B-Parameter Distribution: Grade 4

Percent of Students' Theta	Theta Range	Number of Items of IRT-B
4.55	theta < -3.5	0
0.00	-3.5 <= theta < -3.0	0
2.84	-3.0 <= theta < -2.5	0
3.52	-2.5 <= theta < -2.0	0
4.19	-2.0 <= theta < -1.5	0
9.74	-1.5 <= theta < -1.0	1
10.24	-1.0 <= theta < -0.5	2
14.39	-0.5 <= theta < 0.0	3
12.75	0.0 <= theta < 0.5	8
16.84	0.5 <= theta < 1.0	6
10.01	1.0 <= theta < 1.5	13
6.66	1.5 <= theta < 2.0	4
3.28	2.0 <= theta < 2.5	3
0.65	2.5 <= theta < 3.0	3
0.20	3.0 <= theta < 3.5	0
0.14	3.5 <= theta	0
-6.00	**Minimum**	-1.12
5.09	**Maximum**	2.70
-0.21	**Mean**	0.94
1.69	**SD**	0.89
≥48,880	**Total**	43

Table 7.5.3

Spring 2022 Students' Score and IRT B-Parameter Distribution: Grade 5

Percent of Students' Theta	Theta Range	Number of Items of IRT-B
3.83	theta < -3.5	0
0.00	-3.5 <= theta < -3.0	0
2.89	-3.0 <= theta < -2.5	0
0.00	-2.5 <= theta < -2.0	0
3.95	-2.0 <= theta < -1.5	0
9.66	-1.5 <= theta < -1.0	0
10.74	-1.0 <= theta < -0.5	1
14.08	-0.5 <= theta < 0.0	4
18.00	0.0 <= theta < 0.5	6
13.00	0.5 <= theta < 1.0	3
12.70	1.0 <= theta < 1.5	12
7.07	1.5 <= theta < 2.0	14
2.97	2.0 <= theta < 2.5	6
0.88	2.5 <= theta < 3.0	1
0.19	3.0 <= theta < 3.5	0
0.03	3.5 <= theta	0
-6.00	**Minimum**	-0.63
5.66	**Maximum**	2.71
-0.05	**Mean**	1.23
1.52	**SD**	0.81
≥48,880	**Total**	47

Table 7.5.4

Spring 2022 Students' Score and IRT B-Parameter Distribution: Grade 6

Percent of Students' Theta	Theta Range	Number of Items of IRT-B
2.51	theta < -3.5	0
1.84	-3.5 <= theta < -3.0	0
2.49	-3.0 <= theta < -2.5	0
2.93	-2.5 <= theta < -2.0	0
6.87	-2.0 <= theta < -1.5	0
10.82	-1.5 <= theta < -1.0	1
12.52	-1.0 <= theta < -0.5	3
15.40	-0.5 <= theta < 0.0	9
16.13	0.0 <= theta < 0.5	14
13.48	0.5 <= theta < 1.0	12
9.40	1.0 <= theta < 1.5	11
3.79	1.5 <= theta < 2.0	5
1.37	2.0 <= theta < 2.5	0
0.33	2.5 <= theta < 3.0	0
0.08	3.0 <= theta < 3.5	0
0.03	3.5 <= theta	0
-6.00	**Minimum**	-1.34
4.64	**Maximum**	2.00
-0.37	**Mean**	0.54
1.48	**SD**	0.73
≥49,270	**Total**	55

Table 7.5.5

Spring 2022 Students' Score and IRT B-Parameter Distribution: Grade 7

Percent of Students' Theta	Theta Range	Number of Items of IRT-B
2.30	theta < -3.5	0
0.00	-3.5 <= theta < -3.0	0
1.77	-3.0 <= theta < -2.5	0
2.58	-2.5 <= theta < -2.0	0
7.17	-2.0 <= theta < -1.5	0
12.37	-1.5 <= theta < -1.0	0
13.98	-1.0 <= theta < -0.5	2
15.34	-0.5 <= theta < 0.0	10
14.74	0.0 <= theta < 0.5	11
14.19	0.5 <= theta < 1.0	13
9.08	1.0 <= theta < 1.5	9
4.15	1.5 <= theta < 2.0	8
1.87	2.0 <= theta < 2.5	1
0.29	2.5 <= theta < 3.0	0
0.13	3.0 <= theta < 3.5	0
0.05	3.5 <= theta	0
-6.00	**Minimum**	-0.67
4.75	**Maximum**	2.06
-0.27	**Mean**	0.67
1.32	**SD**	0.72
≥50,930	**Total**	54

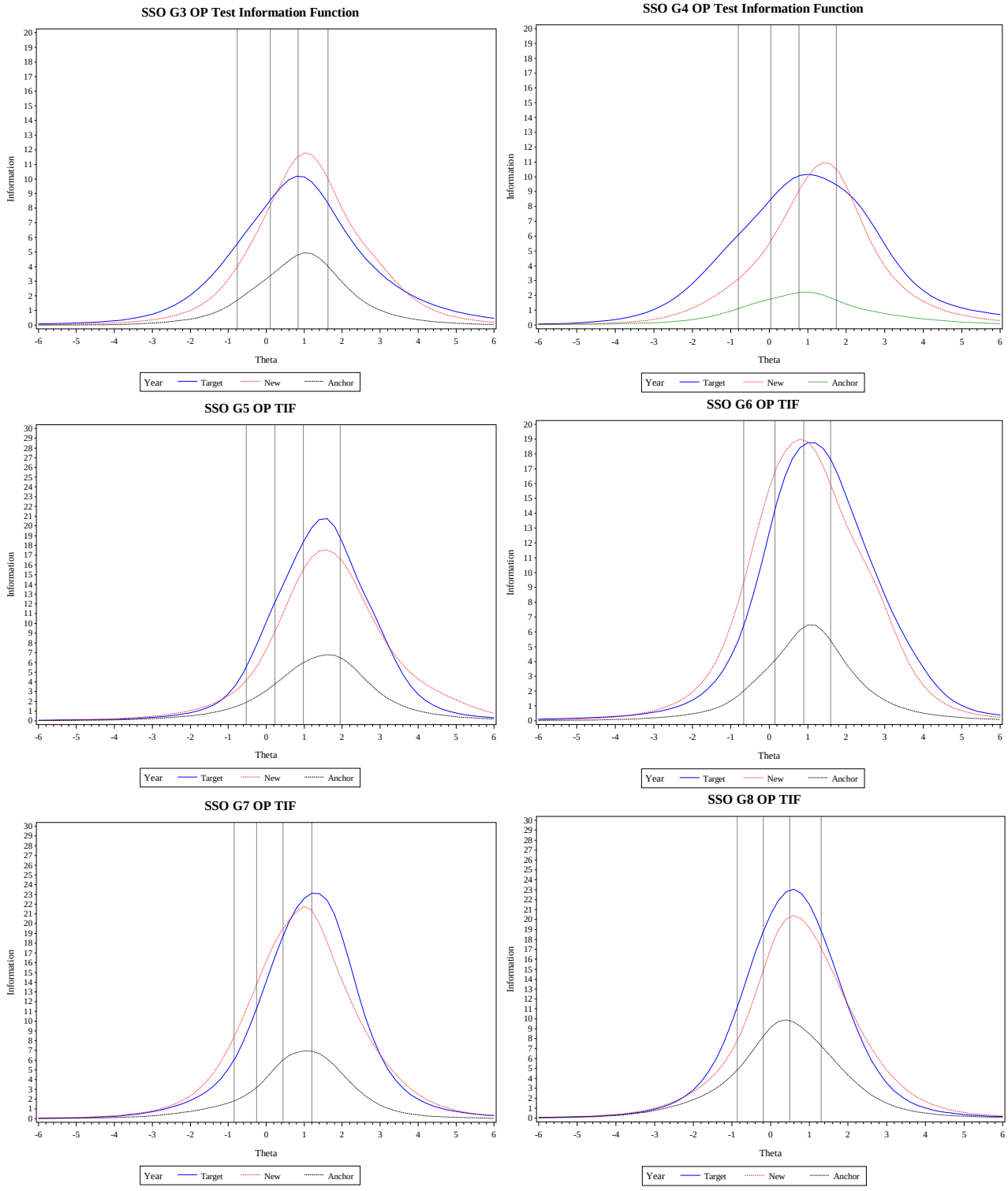
Table 7.5.6

Spring 2022 Students' Score and IRT B-Parameter Distribution: Grade 8

Percent of Students' Theta	Theta Range	Number of Items of IRT-B
2.48	theta < -3.5	0
1.54	-3.5 <= theta < -3.0	0
2.00	-3.0 <= theta < -2.5	0
2.49	-2.5 <= theta < -2.0	0
5.81	-2.0 <= theta < -1.5	0
9.08	-1.5 <= theta < -1.0	1
13.18	-1.0 <= theta < -0.5	0
15.66	-0.5 <= theta < 0.0	9
15.39	0.0 <= theta < 0.5	21
15.83	0.5 <= theta < 1.0	10
9.58	1.0 <= theta < 1.5	10
4.66	1.5 <= theta < 2.0	3
1.76	2.0 <= theta < 2.5	0
0.36	2.5 <= theta < 3.0	0
0.14	3.0 <= theta < 3.5	1
0.04	3.5 <= theta	0
-6.00	**Minimum**	-1.35
6.00	**Maximum**	3.42
-0.25	**Mean**	0.56
1.48	**SD**	0.71
≥50,660	**Total**	55

Plot 7.2

Test Information Curve; Spring 2022 Social Studies G3-8



Field Test Data Review

The process used to complete the field test item equating is an anchored item equating process. In this process, the item parameters from the operational items from the 2022 administration were fixed as constant (i.e., to calculate Stocking-Lord equating constant) and the item parameters for the field test items were freely calibrated, placing the item parameters for the field test items on the same scale as the operational items.

As mentioned previously, field test items are reviewed at the data review meeting for all the same criteria as outlined previously. The data review meeting began with a refresher presentation to data review. The presentation included a review of item statistics (difficulty, discrimination, DIF, score distributions) based on CTT and IRT, appropriate interpretations and inferences, what would be considered reasonable values, and how the values might differ across item types. The result of such reviews is to determine if items are eligible to be placed in the item bank for future test construction or if items need to be updated and field-tested again. It should be noted that all the results of spring 2022 data review are saved in Pearson's ABBI. It should also be noted that the training presentation agenda for data evaluation is included in Appendix A: Training Agendas.

8. Test Results and Score Reports

This chapter provides information on the results of the spring LEAP 2025 Social Studies tests. The scale score results and achievement level information are also presented here. Presenting the results by achievement level translates the quantitative scale provided through scale scores into a qualitative description of student achievement. The levels are Advanced, Mastery, Basic, Approaching Basic, and Unsatisfactory. The results in the following tables are presented as evidence of the reliability and validity of the scores from the LEAP 2025 Social Studies G3–8 tests.

Demographic Characteristics of Students

The operational Social Studies tests were administered to all eligible students in the appropriate grade level during spring 2022. Spring 2022 operational score results were reviewed based on the following student characteristics:

- Gender: Female and Male
- Race and Ethnicity: Hispanic/Latino, American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, and Two or More Races
- Education Classification
- Economic Status
- English Learner (EL)
- Migrant Status
- Homeless Status
- Military Affiliation
- Foster Care Status

Test Results

For the spring 2022 Social Studies tests, the lowest obtainable scale score (LOSS) on the tests is 650 and the highest obtainable scale score (HOSS) is 850. Scale score means and standard deviations as well as the percentages of students in each performance level are reported for the state and disaggregated into various demographic groups. In addition to the descriptive statistics presented in the following tables, scale score frequency distributions are presented in Appendix E: Scale Distribution and Statistical Report. Finally, because the spring 2022 tests were administered during the COVID-19 pandemic, great caution should be applied when any statistical inference is drawn.

Table 8.1.1

*Spring 2022 LEAP 2025 State Test Results: Grade 3***

	Scale Score			% at Performance Level				
	N	Mean	SD	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
TOTAL	≥49,310	717.26	39.51	29	27	23	15	6
Gender								
Female	≥24,090	717.49	38.61	28	28	24	14	5
Male	≥25,220	717.03	40.34	30	27	22	15	7
Ethnicity								
African American	≥20,380	703.82	35.95	41	30	19	8	2
American Indian or Alaska Native	≥270	721.92	35.84	22	29	25	17	6
Asian	≥830	741.25	39.97	13	19	23	26	18
Hispanic/Latino	≥4,970	709.17	38.33	36	29	21	10	4
Two or more races	≥1,850	722.25	37.93	22	30	26	16	7
Native Hawaiian or Other Pacific Islander	≥30	721.34	37.81	24	32	16	21	8
White	≥20,930	730.82	38.16	17	24	27	21	10
Economically Disadvantaged*								
No	≥13,960	738.16	38.37	14	20	26	26	14
Yes	≥35,190	709.06	36.79	35	30	21	10	3

Table 8.1.1 (continued)

	Scale Score			% at Performance Level				
	N	Mean	SD	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
English Learner								
No	≥46,460	718.63	39.50	28	27	23	15	6
Yes	≥2,840	694.85	32.20	49	32	15	3	NR
Education Classification								
Regular	≥43,000	719.41	39.44	27	27	24	16	7
Special	≥6,300	702.56	36.70	43	30	16	7	3
Section 504								
No	≥45,770	717.94	39.69	29	27	23	15	6
Yes	≥3,540	708.41	35.95	34	33	20	10	3
Migrant								
No	≥49,220	717.29	39.50	29	27	23	15	6
Yes	≥90	697.91	39.96	49	26	16	5	3
Homeless Status								
No	≥47,890	717.69	39.51	29	27	23	15	6
Yes	≥1,410	702.67	36.35	41	32	17	8	2
Military Affiliation								
No	≥48,360	716.92	39.44	29	27	23	14	6
Yes	≥950	734.21	39.10	15	24	25	24	12
Foster Care Status								
No	≥49,150	717.29	39.51	29	27	23	15	6
Yes	≥150	707.70	36.03	34	35	21	8	3

* Economic status was not available for all students.

** The results combine both online and paper forms.

Table 8.1.2

Spring 2022 LEAP 2025 State Test Results: Grade 4

	Scale Score			% at Performance Level				
	N	Mean	SD	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
TOTAL	≥48,880	722.49	41.97	25	25	24	20	7
Gender								
Female	≥23,890	721.77	40.43	24	26	25	19	6
Male	≥24,990	723.18	43.37	26	23	23	20	8
Ethnicity								
African American	≥20,480	707.58	38.42	35	30	21	11	2
American Indian or Alaska Native	≥260	727.95	40.80	21	20	26	25	8
Asian	≥800	750.82	42.78	11	13	20	33	23
Hispanic/Latino	≥5,080	716.17	41.42	29	27	22	17	5
Two or more races	≥1,680	729.60	39.34	18	23	28	23	8
Native Hawaiian or Other Pacific Islander	≥30	727.41	32.56	13	33	33	15	5
White	≥20,510	737.19	39.89	15	20	26	28	12
Economically Disadvantaged*								
No	≥13,980	744.80	38.91	10	16	26	32	16
Yes	≥34,610	713.67	39.74	31	28	23	15	4

Table 8.1.2 (continued)

	Scale Score			% at Performance Level				
	N	Mean	SD	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
English Learner								
No	≥46,420	723.87	41.85	24	24	24	21	7
Yes	≥2,460	696.44	35.15	45	32	17	6	1
Education Classification								
Regular	≥42,910	725.63	41.30	22	24	25	21	8
Special	≥5,960	699.92	39.80	45	28	16	8	3
Section 504								
No	≥44,740	723.39	42.07	24	24	24	20	7
Yes	≥4,140	712.71	39.60	31	30	22	14	4
Migrant								
No	≥48,820	722.50	41.97	25	25	24	20	7
Yes	≥60	711.10	41.46	29	32	23	11	5
Homeless Status								
No	≥47,570	722.92	41.95	25	25	24	20	7
Yes	≥1,300	706.71	39.55	38	29	19	12	3
Military Affiliation								
No	≥47,930	722.06	41.92	25	25	24	20	7
Yes	≥950	744.14	38.59	11	18	27	29	16
Foster Care Status								
No	≥48,750	722.51	41.98	25	25	24	20	7
Yes	≥130	712.62	37.28	30	32	21	15	2

* Economic status was not available for all students.

Table 8.1.3

Spring 2022 LEAP 2025 State Test Results: Grade 5

	Scale Score			% at Performance Level**				
	N	Mean	SD	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
TOTAL	≥48,880	720.28	38.57	31	22	23	19	5
Gender								
Female	≥23,810	720.43	37.44	30	23	24	18	4
Male	≥25,060	720.14	39.62	32	21	22	19	5
Ethnicity								
African American	≥20,650	706.37	35.41	44	25	20	10	1
American Indian or Alaska Native	≥250	719.96	35.19	28	24	26	20	2
Asian	≥740	749.54	39.43	12	12	21	35	20
Hispanic/Latino	≥4,790	715.73	38.59	35	22	23	16	4
Two or more races	≥1,640	726.75	36.65	24	21	27	23	5
Native Hawaiian or Other Pacific Islander	≥40	717.75	42.13	30	32	11	23	5
White	≥20,730	733.65	36.36	19	19	27	27	8
Economically Disadvantaged*								
No	≥14,580	740.13	36.32	15	16	26	32	11
Yes	≥33,990	711.99	36.34	38	24	22	14	2

Table 8.1.3 (continued)

	Scale Score			% at Performance Level				
	N	Mean	SD	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
English Learner								
No	≥46,800	721.52	38.41	30	22	24	20	5
Yes	≥2,070	692.56	31.15	61	24	12	3	NR
Education Classification								
Regular	≥42,880	723.52	37.87	27	22	25	21	5
Special	≥6,000	697.19	35.57	57	22	12	7	2
Section 504								
No	≥44,140	721.37	38.64	30	22	24	20	5
Yes	≥4,730	710.14	36.38	41	25	19	13	2
Migrant								
No	≥48,820	720.30	38.58	31	22	23	19	5
Yes	≥50	707.32	34.04	39	27	24	10	NR
Homeless Status								
No	≥47,660	720.65	38.57	31	22	23	19	5
Yes	≥1,210	705.95	35.92	45	24	19	11	1
Military Affiliation								
No	≥47,990	719.91	38.52	31	22	23	19	5
Yes	≥890	740.28	35.86	14	17	26	33	10
Foster Care Status								
No	≥48,770	720.32	38.57	31	22	23	19	5
Yes	≥110	703.40	34.54	43	32	14	9	2

* Economic status was not available for all students.

Table 8.1.4

Spring 2022 LEAP 2025 State Test Results: Grade 6

	Scale Score			% at Performance Level				
	N	Mean	SD	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
TOTAL	≥49,270	712.43	37.51	34	26	23	12	5
Gender								
Female	≥23,930	713.72	36.28	32	28	24	12	4
Male	≥25,340	711.22	38.59	36	25	22	12	5
Ethnicity								
African American	≥20,680	699.59	33.79	47	29	17	6	1
American Indian or Alaska Native	≥280	716.03	34.47	28	29	29	10	4
Asian	≥790	742.60	39.77	14	15	24	25	22
Hispanic/Latino	≥5,060	706.11	37.32	40	26	21	10	3
Two or more races	≥1,620	718.31	36.36	27	28	24	16	5
Native Hawaiian or Other Pacific Islander	≥20	729.00	34.03	14	38	24	10	14
White	≥20,770	725.11	36.12	21	24	29	18	8
Economically Disadvantaged*								
No	≥14,460	732.72	35.01	15	22	30	22	11
Yes	≥34,530	704.12	35.21	42	28	20	8	2

Table 8.1.4 (continued)

	Scale Score			% at Performance Level**				
	N	Mean	SD	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
English Learner								
No	≥47,210	713.79	37.26	33	26	24	13	5
Yes	≥2,060	681.43	28.67	69	22	7	1	NR
Education Classification								
Regular	≥43,710	715.87	36.73	30	27	25	13	5
Special	≥5,560	685.43	32.23	66	20	10	3	1
Section 504								
No	≥44,070	713.97	37.56	32	26	24	13	5
Yes	≥5,200	699.42	34.35	48	28	17	6	2
Migrant								
No	≥49,210	712.44	37.50	34	26	23	12	5
Yes	≥60	711.79	39.10	33	26	21	17	3
Homeless Status								
No	≥48,020	712.77	37.54	34	26	23	12	5
Yes	≥1,250	699.52	33.67	48	27	18	6	2
Military Affiliation								
No	≥48,430	712.13	37.46	34	26	23	12	5
Yes	≥840	730.06	35.72	17	22	30	20	10
Foster Care Status								
No	≥49,150	712.47	37.51	34	26	23	12	5
Yes	≥120	698.00	32.95	52	28	13	5	2

* Economic status was not available for all students.

Table 8.1.5

Spring 2022 LEAP 2025 State Test Results: Grade 7

	Scale Score			% at Performance Level				
	N	Mean	SD	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
TOTAL	≥50,930	727.01	40.54	30	18	20	20	11
Gender								
Female	≥25,070	727.95	39.10	28	19	21	21	11
Male	≥25,850	726.10	41.86	32	17	19	20	12
Ethnicity								
African American	≥21,850	713.30	36.71	42	21	19	13	5
American Indian or Alaska Native	≥290	731.20	35.87	23	20	26	23	9
Asian	≥740	762.21	41.91	9	8	17	27	40
Hispanic/Latino	≥4,830	720.89	40.58	36	18	18	19	9
Two or more races	≥1,690	733.66	39.29	24	17	22	25	13
Native Hawaiian or Other Pacific Islander	≥40	743.10	37.95	17	12	19	33	19
White	≥21,440	740.54	39.02	18	15	22	27	18
Economically Disadvantaged*								
No	≥15,130	747.62	38.69	14	13	21	29	23
Yes	≥35,500	718.41	38.08	37	20	20	17	6

Table 8.1.5 (continued)

	Scale Score			% at Performance Level				
	N	Mean	SD	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
English Learner								
No	≥49,100	728.33	40.31	29	18	20	21	12
Yes	≥1,820	691.73	28.91	67	20	9	3	NR
Education Classification								
Regular	≥45,380	730.66	39.81	26	18	21	22	13
Special	≥5,550	697.22	33.50	63	18	11	6	2
Section 504								
No	≥45,550	728.80	40.52	28	18	20	21	12
Yes	≥5,370	711.87	37.40	44	22	17	13	5
Migrant								
No	≥50,870	727.03	40.54	30	18	20	20	11
Yes	≥60	711.80	37.87	39	20	21	16	3
Homeless Status								
No	≥49,770	727.34	40.54	30	18	20	21	12
Yes	≥1,160	712.75	37.87	45	18	18	14	5
Military Affiliation								
No	≥50,070	726.66	40.48	30	18	20	20	11
Yes	≥860	747.57	38.38	12	14	24	27	23
Foster Care Status								
No	≥50,810	727.05	40.54	30	18	20	20	11
Yes	≥120	711.96	34.86	39	28	18	12	2

* Economic status was not available for all students.

Table 8.1.6

Spring 2022 LEAP 2025 State Test Results: Grade 8

	Scale Score			% at Performance Level				
	N	Mean	SD	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
TOTAL	≥50,660	726.97	42.44	26	20	22	23	10
Gender								
Female	≥24,990	729.47	40.19	23	21	24	23	9
Male	≥25,660	724.53	44.38	30	18	20	22	10
Ethnicity								
African American	≥21,520	712.72	39.30	37	23	21	15	4
American Indian or Alaska Native	≥280	735.43	40.01	18	19	23	30	11
Asian	≥810	761.77	44.30	10	9	15	29	38
Hispanic/Latino	≥4,820	720.55	43.84	33	19	20	21	8
Two or more races	≥1,570	733.28	40.77	20	19	25	25	11
Native Hawaiian or Other Pacific Islander	≥40	746.12	43.46	16	12	21	33	19
White	≥21,600	740.68	39.83	15	16	23	30	15
Economically Disadvantaged*								
No	≥16,040	747.93	38.53	11	14	23	33	19
Yes	≥34,310	717.37	40.61	33	22	21	18	5

Table 8.1.6 (continued)

	Scale Score			% at Performance Level				
	N	Mean	SD	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
English Learner								
No	≥48,850	728.46	42.01	25	20	22	23	10
Yes	≥1,810	686.74	33.18	67	20	8	4	1
Education Classification								
Regular	≥45,460	730.88	41.25	22	19	23	24	11
Special	≥5,200	692.74	36.91	60	20	11	7	2
Section 504								
No	≥45,500	728.73	42.37	25	19	22	24	10
Yes	≥5,150	711.38	39.78	39	24	19	14	5
Migrant								
No	≥50,600	726.98	42.44	26	20	22	23	10
Yes	≥50	719.68	41.02	31	22	22	19	7
Homeless Status								
No	≥49,550	727.27	42.41	26	20	22	23	10
Yes	≥1,110	713.47	41.67	38	21	20	16	5
Military Affiliation								
No	≥49,750	726.53	42.38	27	20	22	22	10
Yes	≥910	750.78	38.57	10	14	22	33	21
Foster Care Status								
No	≥50,530	727.03	42.42	26	20	22	23	10
Yes	≥130	702.54	40.50	51	20	15	11	3

* Economic status was not available for all students.

Effect Size

One way to evaluate the magnitude of the standardized mean difference (SMD) is to calculate the ES. Cohen's *d* was used to calculate the ES and is given by the following formula:

$$d = \frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{(n_a + n_b) - 2}}},$$

where is $\bar{x}_a$ is the mean score of group A, $\bar{x}_b$ is the mean score of group B, s_a^2 is the variance of group A, s_b^2 is the variance of group B, n_a is the number of students in group A, and n_b is the number of students in group B.

Cohen's *d*, then, expresses the difference in group means in terms of the standard deviation. Cohen (1988) offered guidelines for interpreting the meaning of the *d* statistic: *d* = 0.20 is a small ES, *d* = 0.50 is a medium ES, and *d* = 0.80 is a large ES. Based on Cohen's (1988) guidelines, certain trends are observable in Tables B.6.1–B.6.6. Although no big difference in Social Studies tests was seen between females and males, mean raw scores and ESs show that Asian and White students tend to outperform other ethnicity groups. There were clear performance differences among regular education, gifted/talented education, and special education students in Education Classification and Non-English Learner and English Learner in EL status. Performance differences were also observed from Economically Disadvantaged status, Homeless status, Foster Care status, and Military Affiliation status.

Uses of Test Scores

To understand whether a test score is being used properly, one must understand the purpose of the test. The intended uses of the LEAP 2025 test scores include the following:

- evaluating students' overall proficiency of the Louisiana Student Standards
- identifying students' strengths and weaknesses

- evaluating programs at the school, school system, and/or state level
- informing stakeholders, including students, teachers, school administrators, school system administrators, LDOE staff members, parents, and the public, of the status of students' progress toward meeting college and career readiness standards.

This technical report refers to the uses of the test-level scores (i.e., scale scores and achievement levels), and reporting category-level scores and achievement level classifications.

Score Reports

Score reports are the primary means of communicating test scores to appropriate school system personnel (e.g., testing coordinators or superintendents), teachers, and parents. Interpretations of test scores from each administration are disseminated in two ways: the individual score report and the LEAP Interpretive Guide. The LDOE and DRC strive to create documents that will be accessible to parents, teachers, and all other stakeholders. The *Individual Student-Level Report (ISR)* is the primary means for sharing student test results with parents. As such, it is a standalone document from which parents can glean information that is relevant to understanding their children's test scores. For more information about the test, parents are provided the *Parent Guide to the LEAP 2025 Student Reports*. In the 2021–2022 administration year, student reports for each school were posted by subject, then downloaded and printed from eDIRECT by the school systems and schools. eDIRECT is DRC's secure online system that provides schools and districts access to student tests and reports.

School Roster Report. A School Roster Report, which provides summary information about student performance on the LEAP 2025 Grades 3–8 Social Studies tests, is available to school systems and schools through eDIRECT. Total test scores and achievement level indicators are shown for the test of interest. Category and subcategory performance ratings are also reported for students. At the school level, the percentage of students at each achievement level and rating by category and subcategory are summarized. More details can be found in the *LEAP 2025 Grades 3–8 Interpretive Guide (iGUIDE) Spring 2022*.

Individual Student-Level Report. The ISR is another type of report available through the eDIRECT system. ISRs may be downloaded and printed by schools to be sent home to parents. At the top of the page, overall student performance is reported by scale score and achievement level. In the middle of the page, category and subcategory performance indicators are reported. When a student does not receive a scale score, their achievement level will be left blank. ISRs for students whose scores were invalidated will display a blank scale score for a given course.

LEAP 2025 Grades 3–8 Interpretive Guide (iGUIDE) Spring 2022. The *LEAP 2025 Grades 3–8 Interpretive Guide (iGUIDE) Spring 2022* was written to help school administrators, teachers and parents in the Louisiana school system, and the general public understand the LEAP Social Studies Grades 3–8 tests. The *LEAP 2025 Grades 3–8 Interpretive Guide (iGUIDE) Spring 2022* was developed collaboratively by the LDOE and DRC staff. LDOE staff had opportunities to review the guide, provide feedback, and give final approval. The elements of the table of contents are provided below:

- Introduction to the Interpretive Guide
    - Overview
        - Purpose of the Interpretive Guide
    - Test Design
    - Scoring
        - Item Types and Scoring
    - Interpreting Scores and Achievement Levels
        - Scale Score
        - Achievement Level Definitions
        - Student Rating by Reporting Category and Subcategory
- Student-Level Reports
    - Sample Student Report: Explanation of Results and Terms
    - Sample Student Report A
    - Sample Student Report B
    - Sample Student Report C
    - Sample Student Report D
- School Roster Report
    - Sample School Roster Report: Explanation of Results and Terms
    - Sample Social Studies School Roster Report

Achievement Level Policy Definitions

Achievement level policy definitions for the LEAP 2025 Social Studies tests are shown in Table 8.2. The titles and descriptions of the achievement levels were defined to be part of a cohesive assessment system, and the achievement levels indicate a student's ability to demonstrate proficiency on the LSSS defined for a specific course. The standard-setting section of the LEAP 2025 Social Studies 2016-2017 technical report contains comprehensive information.

Table 8.2
Achievement Level Policy Definitions for LEAP 2025

Achievement Level	Achievement Level Policy Definition
Advanced	Students performing at this level have **exceeded** college and career readiness expectations and are well prepared for the next level of studies in this content area.
Mastery	Students performing at this level have **met** college and career readiness expectations and are prepared for the next level of studies in this content area.
Basic	Students performing at this level have **nearly met** college and career readiness expectations and may need additional support to be fully prepared for the next level of studies in this content area.
Approaching Basic	Students performing at this level have **partially met** college and career readiness expectations and will need much support to be prepared for the next level of studies in this content area.
Unsatisfactory	Students performing at this level have **not yet met** the college and career readiness expectations and will need extensive support to be prepared for the next level of studies in this content area.

It should be noted that the overall purpose of reporting test results is to communicate information on student performance to stakeholders. These results are presented in the context of score reports that aid the user in understanding the meaning of the test scores. The reports and ancillary information address multiple best practices of the testing industry.

9. Reliability

Internal Consistency Reliability Estimation

Internal consistency methods use data from a single administration to estimate test score reliability. For state assessments where student testing time is at a premium, internal consistency procedures have a practical advantage over reliability estimation procedures that require multiple test administrations. One of the most frequently used internal consistency reliability estimates is coefficient alpha (Cronbach, 1951). Coefficient alpha is based on the assumption that inter-item covariances constitute true-score variance and the fact that the average true-score variance of items is greater than or equal to the average inter-item covariance. The formula for coefficient alpha is

$$\alpha = \left(\frac{N}{N-1}\right)\left(1 - \frac{\sum_{i=1}^{N} s_{Y_i}^2}{s_X^2}\right),$$

where N is the number of items on the test, $s_{Y_i}^2$ is the sample variance of the *ith* item or component, and s_X^2 is the observed score variance for the test. Coefficient alpha is appropriate for use when the items on the test are reasonably homogeneous. The homogeneity of LEAP 2025 Social Studies tests is evidenced through a dimensionality analysis. Dimensionality analysis results are discussed in Chapter 7, Data Analysis. The reliability and classification accuracy reports in Appendix F: Reliability and Classification Accuracy provide coefficient alpha and IRT model-based or "marginal reliability" (Thissen, Chen, & Bock, 2003) for the total test.

While coefficient alpha values were between 0.86 and 0.92, the marginal alpha values were between 0.86 and 0.93 for the Social Studies tests. Marginal reliability is described as "an average reliability over levels of θ or theta" (Thissen, 1990). Marginal reliability may be reproduced by squaring and subtracting from 1 each of the 31 "posterior standard

deviations" (SEMs) in the IRTPRO output file. Since the variance of the population is 1, each of these values represents the reliability at each of the 31 θs. Marginal reliability is the average of these computations weighted by the normal probabilities for each of the 31 quadrature intervals. The formula for marginal reliability is

$$\overline{\rho} = \frac{s_\theta^2 - E(SEM_\theta^2)}{s_\theta^2},$$

where s_θ^2 is the variance of a given θ (is 1 for standardized θ) and $E(SEM_\theta^2)$ is the average error variance or the mean of the squared posterior standard deviations by weighting population density. Marginal reliability can be interpreted in the same way as traditional internal consistency reliability estimates such as coefficient alpha.

Additional reliabilities were calculated for various demographics using the population of students. (Please refer to Table F 1.1) Included with coefficient alpha in the tables are the number of students responding to the test, the mean score obtained by this group of students, and the standard deviation of the scores obtained for this group.

Coefficient alpha estimates are computed for the entire test and each subscale by reporting category. Subscore reliability will generally be lower than total score reliability because reliability is influenced by the number of items as well as their covariation. In some cases, the number of items associated with a subscore is small (10 or fewer). Subscore results must be interpreted carefully when these measures reflect the limited number of items associated with the score.

Classical Standard Error of Measurement

The classical standard error of measurement (SEM) represents the amount of variance in a score that results from random factors other than what the assessment is intended to measure. Because underlying traits such as academic achievement cannot be measured with perfect precision, the SEM is used to quantify the margin of uncertainty in test scores. For example, factors such as chance error and differential testing conditions can cause a student's observed score (the score achieved on a test) to fluctuate above or

below his or her true score (the student's expected score). The SEM is calculated using both the standard deviation and the reliability of test scores, as follows:

$$SEM = \sigma_x\sqrt{(1 - P'_{xx})},$$

where P'_{xx} is the reliability estimate and σ_x is the standard deviation of raw scores on the test. A standard error provides some sense of the uncertainty or error in the estimate of the true score using the observed score. For example, suppose a student achieves a raw score of 50 on a test with an SEM of 3. Placing a one-SEM band around this student's score would result in a raw score range of 47 to 53. If the student took the test 100 times and 100 similar raw score ranges were computed, about 68 of those score ranges would include the student's true score.

It is important to note that the SEM provides an estimate of the average test score error for all students regardless of their individual proficiency levels. It is generally accepted that the SEM varies across the range of student proficiencies (Peterson, Kolen, & Hoover, 1989). For this reason, it is useful to report test-level SEM, and SEMs for 2022 Social Studies range between 3.02 and 3.56, as seen in Table B.4. In addition, SEMs by student group can be found in Appendix F.

Conditional Standard Error of Measurement and Cut Scores

It is important to note that the SEM index provides only an estimate of the average test score error for all students regardless of their individual levels of proficiency. By comparison, conditional standard error of measurement (CSEM) provides a reliability estimate at each score point on a test. Like the SEM, the CSEM reflects the amount of variance in a score resulting from random factors other than what the assessment is designed to measure, but it provides an estimate conditional on proficiency. The CSEM is usually smallest, and thus scores are most reliable, near the middle of the score distribution. Typically, achievement tests included relatively large numbers of moderately difficult items. Because these items are usually well matched to a majority of students' ability, they provide the most reliable estimates of ability. It is desirable, for an achievement test where students are classified into pass/fail categories, that the CSEM be lowest at the cut score for passing. The CSEMs at the four cut scores of each grade that define the performance levels are presented in Table 9.1. The standard-setting section of the LEAP 2025 Social Studies 2016–2017 technical report contains comprehensive information.

Table 9.1

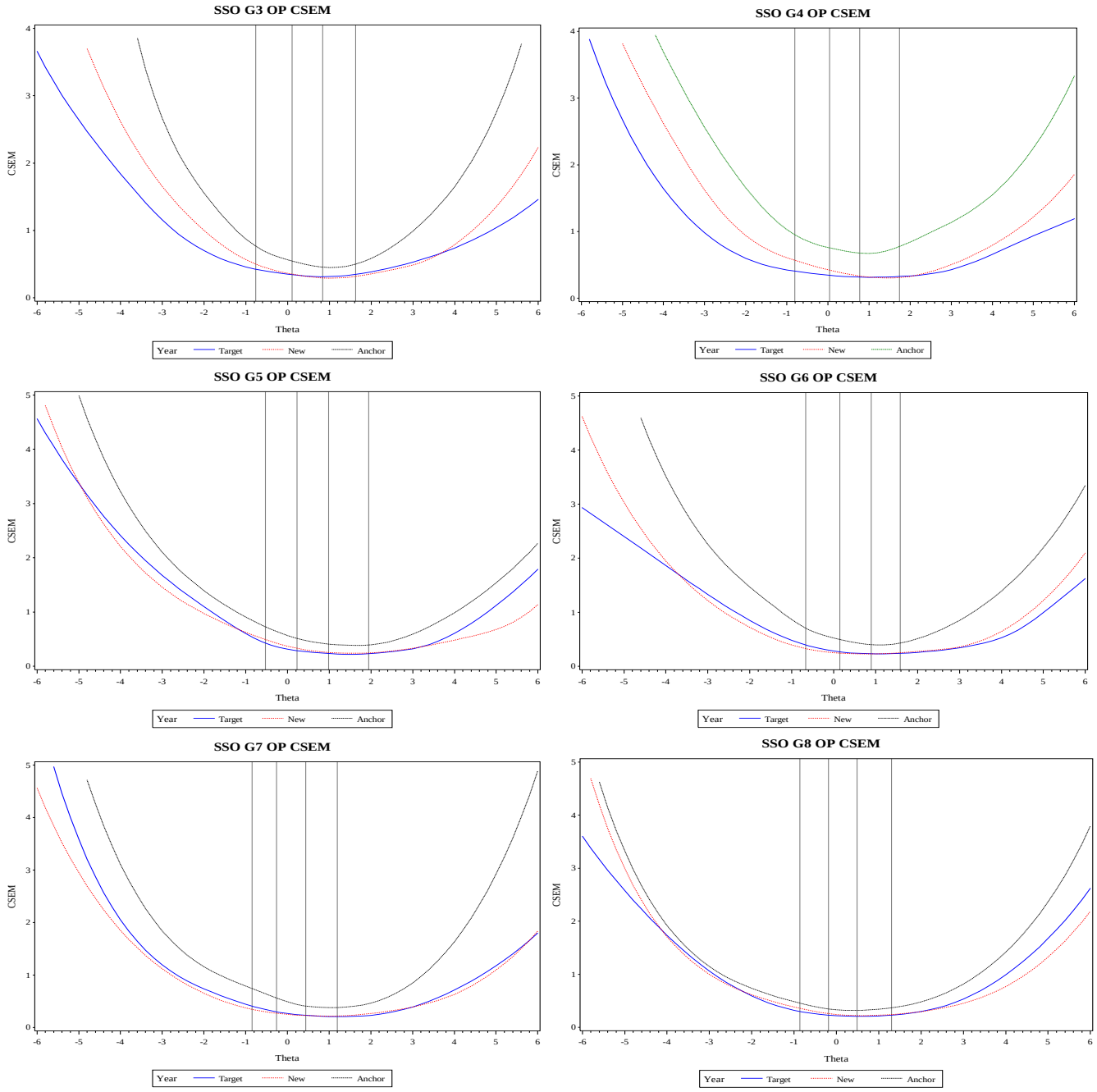
Conditional Standard Errors of Measurement at the Approaching Basic, Basic, Mastery, and Advanced Cut Scores: Operational 2022 LEAP Social Studies

Grade	*Approaching Basic*		*Basic*		*Mastery*		*Advanced*	
	Cut Score	CSEM	Cut Score	CSEM	Cut Score	CSEM	Cut Score	CSEM
3	695	16	725	12	750	10	777	11
4	696	19	725	14	750	11	783	11
5	700	15	725	11	750	8	782	8
6	698	11	725	8	750	8	773	8
7	704	12	725	10	750	8	777	8
8	700	13	725	9	750	8	780	9

IRT methods are used for estimating CSEM and are presented in the following graph. With fixed-form assessments, the estimates of measurement error tend to be higher at the low and high ends of the scale-score range, where few items measure the ability levels. Generally, there are few students with extreme scores, and these score levels cannot be estimated as accurately as levels toward the middle of the ability range. The middle of the ability range, where cut scores are located, shows lower measurement error than the low and high ends of the ability ranges. Plot 9.1 demonstrates that the tests are designed so that measurement error is minimized in the middle of the scale range, where most students are located.

Plot 9.1

Social Studies CSEM Curves: G3–8



Note: In the plot, although the CSEM values are placed on the theta scale, they are converted into the values shown in Table 9.1 when these are transformed using the scaling constants. For instance, for grade 3, the CSEM value on the theta scale translates into 16, 12, 10, and 11.; Target = 2017 test; New = 2022 OP form; Anchor = anchor items.

Student Classification Accuracy and Consistency

Students are classified into one of five performance levels based on their scale scores. It is important to know the reliability of student scores in any examination; assessing the reliability of the classification decisions based on these scores is of even greater importance. Classification decision reliability is estimated by the probabilities of correct and consistent classification of students. Procedures from Livingston and Lewis (1995) and Lee, Hanson, and Brennan (2000) were used to derive accuracy and consistency classification measures.

Accuracy of Classification. According to Livingston and Lewis (1995, p. 180), the classification accuracy is "the extent to which the actual classifications of the test takers . . . agree with those that would be made on the basis of their true scores, if their true scores could somehow be known." Accuracy estimates are calculated from cross-tabulations between "classifications based on an observable variable (scores on a test) and classifications based on an unobservable variable (the test takers' true scores)" (Livingston and & Lewis, 1995, p. 189). A true score is also referred to as a hypothetical mean of scores from all possible forms of the test if they could be somehow obtained (Young & Yoon, 1998).

Consistency of Classification. Classification consistency is "the agreement between classifications based on two non-overlapping, equally difficult forms of the test" (Livingston & Lewis, 1995, p. 180). Consistency is estimated using actual response data from a test and the test's reliability to statistically model two parallel forms of the test and compare the classifications on those alternate forms.

Accuracy and Consistency Indices. Three types of accuracy and consistency indices were generated: *overall*, *conditional-on-level*, and *cut point*, provided in Appendix F. The *overall accuracy* of performance-level classifications is computed as a sum of the proportions on the diagonal of the joint distribution of true score and observed score levels. It is a proportion (or percentage) of correct classification across all the levels. While the overall accuracy indices were between 0.608 and 0.719, the overall consistency indices were 0.508 and 0.633 for the LEAP 2025 Social Studies tests.

Another way to express overall consistency is to use Cohen's Kappa (κ) coefficient (Cohen, 1960). The overall coefficient Kappa when applying all cutoff scores together is

$$\kappa = \frac{P - P_c}{1 - P_c},$$

where P is the probability of consistent classification, and P_c is the probability of consistent classification by chance (Lee, Hanson, & Brennan, 2000). P is the sum of the diagonal elements, and P_c is the sum of the squared row totals. The PChance indices were between 0.218 and 0.256 for the 2022 Social Studies tests.

Kappa is a measure of "how much agreement exists beyond chance alone" (Fleiss, 1973), which means that it provides the proportion of consistent classifications between two forms after removing the proportion of consistent classifications expected by chance alone. The Kappa indices were between 0.365 and 0.509 for the 2022 Social Studies tests. *Consistency conditional-on-level* is computed as the ratio between the proportion of correct classifications at the selected level (diagonal entry) and the proportion of all the students classified into that level (marginal entry).

Accuracy conditional-on-level is analogously computed. The only difference is that in the consistency table, both row and column marginal sums are the same, whereas in the accuracy table, the sum that is based on true status is used as a total for computing accuracy conditional on the level.

Perhaps the most important indices for accountability systems are those for the accuracy and consistency of classification decisions made at specific cut points. To evaluate decisions at specific cut points, the joint distribution of all the performance levels is collapsed into a dichotomized distribution around that specific cut point.

10. Validity

“Validity is defined as … the degree to which evidence and theory support the interpretations of test scores entailed by proposed users of tests" (AERA/APA/NCME, 2014). The purpose of test score validation is not to validate the test itself but to validate interpretations of the test scores for particular purposes or uses. Test score validation is not a quantifiable property but an ongoing process, beginning at initial conceptualization and continuing throughout the entire assessment process.

The 2021–2022 LEAP 2025 Social Studies tests were designed and developed to provide fair and accurate scores that support appropriate, meaningful information for educational decisions. The knowledge, expertise, and professional judgment offered by Louisiana educators ultimately ensure that the content of the LEAP 2025 Social Studies tests is an adequate and representative sample of appropriate content, and that the content is a legitimate basis upon which to derive valid conclusions about student achievement.

Chapters 2, 3, and 4 provide a general discussion of test book creation and the editing process, describing the selection of operational test items, the content distribution of embedded field test items, and the process to obtain approvals from the LDOE. The test design process and participation by Louisiana educators throughout the process—from item development, content review, and bias review to test selection—reinforce confidence in the content and design of LEAP 2025 to derive valid inferences about Louisiana student performance. The data review process and results are also discussed. Chapter 5 of the technical report describes the process, procedures, and policies that guide the administration of the LEAP 2025 assessments, including accommodations, test security, and detailed written procedures provided to test administrators and school personnel. Chapter 6 describes scoring processes and activities for the LEAP 2025 Social Studies tests.

Chapter 7 describes classical data analysis and item response theoretic calibration, scaling, and equating methods, as well as processes and procedures to clean data to

ensure replicable, iterative calibrations and scaling of the 2022 Social Studies tests to derive scale scores from students' raw scores. Some references to introductory and advanced discussions of IRT are provided. Chapter 7 also describes an analysis of DIF. Complete tables of gender and ethnicity DIF results for all 2022 Social Studies operational items are presented in Appendix C. Chapter 8 of the technical report summarizes the test results, score distributions, score reports, and achievement level information. Chapter 9 addresses Cronbach's alpha and marginal alpha as measures of internal consistency and describes analysis procedures for classification consistency and classification accuracy. In addition, test validity is addressed in this chapter.

Evidence for Construct-Related Validity

Evidence for construct-related validity—the meaning of test scores and the inferences they support—is the central concept underlying the LEAP 2025 validation process. Validity evidence, from the design of the test to item development and scoring, is created throughout the entire assessment process. Therefore, evidence of validity is described throughout the LEAP 2025 technical report.

Internal Structure of Reporting Categories

The 2022 Social Studies tests contain three reporting categories: *Investigate, Evaluate, and Reason Scientifically.* Table D.1 shows that moderate correlations were observed among the reporting categories; since we used distinct items for each reporting category, a moderate correlation was anticipated.

Content-Related Evidence

Content validity is frequently defined in terms of the sampling adequacy of test items. That is, content validity is the extent to which the items in a test adequately represent the domain of items or the construct of interest (Suen, 1990). Consequently, content validity provides judgmental evidence in support of the domain relevance and representativeness of the content in the test (Messick, 1989). It should be noted that the 2022 Social Studies operational test forms were built exclusively using an ABBI bank program which

contained both content and statistical information about both operational and field test items.

Dimensionality and Principal Component Analysis

Appendix D: Dimensionality provides information about the principal component analysis of the Social Studies tests. Measurement implies order and magnitude along a single dimension (Andrich, 2004). Consequently, in the case of scholastic achievement, a one-dimensional scale is required to reflect this idea of measurement (Andrich, 1988, 1989). However, unidimensionality cannot be strictly met in a real testing situation because students' cognitive, personality, and test-taking factors usually have a unique influence on their test performance to some level (Andrich, 2004; Hambleton, Swaminathan, & Rogers, 1991).
Consequently, what is required for unidimensionality to be met is an investigation of the presence of a dominant factor that influences test performance. This dominant factor is considered as the ability measured by the test (Andrich, 1988; Hambleton et al., 1991; Ryan, 1983).

To check the unidimensionality of the Spring 2022 assessment, the relative sizes of the eigenvalues associated with a principal component analysis of the item set were examined using the Statistical Analysis System (SAS) program. The first and second principal component eigenvalues were compared *without rotation*. Table D.2 and Plot D.1 summarize the results of the first and second principal component eigenvalues of the assessments. A general rule of thumb in exploratory factor analysis suggests that a set of items may represent as many factors as there are eigenvalues greater than 1 because there is one unit of information per item and the eigenvalues sum to the total number of items. However, a set of items may have multiple eigenvalues greater than 1 and still be sufficiently unidimensional for analysis with IRT (Loehlin, 1987; Orlando, 2004). As seen from the tables and figures, the first component is substantially larger than the second eigenvalue for the 2022 Social Studies tests. Because the spring test was administered during the COVID-19 pandemic, great caution should be applied when any statistical inference is drawn.

Evidence Based on Relations to Other Variables

Evidence based on *relations to other variables* is a typical utility of criterion-related validity evidence to measure concurrent or predictive validity, as well as more comprehensive investigations of the relationships among test scores and other variables such as multitrait-multimethod studies (Campbell & Fiske, 1959). Thus, external variables can be used to evaluate hypothesized relationships between test scores and other measures of student achievement (e.g., test scores on other tests) to evaluate the degree to which different tests actually measure different skills and the utility of test scores for predicting specific criteria (e.g., college grades).

A significant number of students who took the Social Studies test also took the LEAP Science test. For the total student group, in general, moderate correlation was observed between the Social Studies and Science exams. In general, however, the English Learner group reported a slightly lower correlation coefficient than the other groups. A separate report, *External Validity Study: SPR 2022*, that was submitted to the LDOE has more specific information.

Item Development and Field Test Analysis

Test development for LEAP Social Studies tests is ongoing and continuous. Content specialists, teachers from across Louisiana, WestEd/Pearson, and the LDOE were greatly involved in developing and reviewing test items. Committees such as content review and bias review reviewed all of the items, which were finally stored in the item bank. Specifically, an internal review by the LDOE and WestEd/Pearson staff for alignment and quality required a great deal of time and energy. More specific information on item (test) development and review can be obtained in Chapter 3, Overview of the Test Development Process.

Various field test forms were used to administer the test items. Once these items were scored, the LDOE and WestEd/Pearson conducted additional item analysis and content review. Any field test items that exhibited statistical results that suggested potential problems were carefully reviewed by both the LDOE and WestEd/Pearson content specialists. A determination was then made as to whether an item should be accepted, rejected, or revised/re-field-tested. Information on statistical analyses for field test items can be obtained in Chapter 7, Data Analysis.

In summary, additional, corroborating evidence consistent with the validity, reliability, and consistency of the LEAP 2025 Social Studies assessment has been documented in the LEAP Grades 3–8 Social Studies assessment framework, test development plans, and the 2017 Social Studies standard-setting technical report. Table 10.1 summarizes the sources of validity evidence and indicates where the evidence can be found in the technical report.

Mode Effect Study

It is important to evaluate fairness in test administration in addition to evaluating fairness by examining performance among subgroups. Since two modes (i.e., paper-based tests and computer-based tests) were administered for grade 3, the following techniques (i.e., mode effect analysis and equating) were applied to operational test data to investigate the item mode effect. The mode effect analysis has been conducted, and the results indicate no items exhibiting C category DIF, suggesting no mode effect between online and paper tests; *all items* exhibited *A category DIF*.

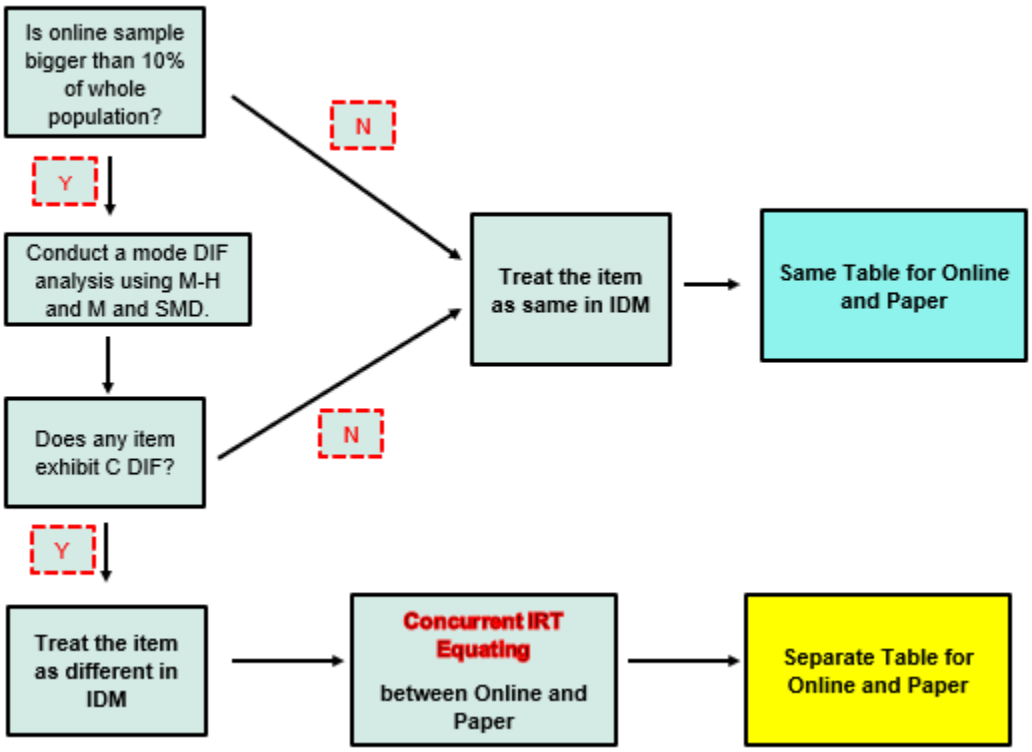


Figure 10.1 General overview of equating, including a mode effect analysis

Table 10.1

Evidence of Validity and the Corresponding Technical Report Chapter

Source of Validity	Related Information	Related Chapter/Source
Evidence Based on Test Content	Item Development Process	Chapter 3 LEAP 2025 Grades 3–8 Social Studies Assessment Frameworks
	Test Blueprint and Item Alignment to Curriculum and Standards	Chapters 2 & 3 Appendix A LEAP 2025 Grades 3–8 Social Studies Assessment Frameworks
	Item Bias, Sensitivity, and Content Appropriateness	Chapter 3
	Accommodations	Chapter 4
Evidence Based on Response Processes	Field Test Analysis Data Review	Chapters 3, 7, & 9 LEAP 2025 Grades 3–8 Social Studies Assessment Frameworks
	Classical Item Analysis IRT Analysis	Chapter 7
Evidence Based on Internal Structure	Differential Item Functioning	Chapter 7
	Reliability and Standard Errors of Measurement	Chapter 9
	Correlation among Reporting Categories	Chapter 10
	Dimensionality Analysis	Chapter 10
Evidence Based on Relations to Other Variables	Correlation Analyses between LEAP Social Studies and LEAP Science	Chapter 10
Evidence Based on the Consequences of Testing	Scale Score and Performance Level Information	Chapter 8
	Test Interpretive Guide	Chapter 8

References

AERA/APA/NCME. (2014). *Standards for educational and psychological testing.* Washington, DC: Author.

Andrich, A. (1988). *Rasch models for measurement*. Newbury Park, CA: Sage Publications.

Andrich, A. (1989). Distinctions between assumptions and requirements in measurement in the social sciences. In J. A. Keats, R. Taft, R. A. Heath, & H. H. Lovibond (Eds.), *Mathematical and theoretical systems*. North-Holland: Elsevier Science Publisher B.V.

Andrich, A. (2004). *Modern measurement and analysis in social science*. Murdoch University, Perth, Western Australia.

Angoff, W. (1993). Perspectives on differential item functioning methodology. In P. W. Holland & H. Warner (Eds.), *Differential item functioning* (pp. 3–24). Hillsdale, NJ: Lawrence Erlbaum Associates.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Newbury Park, CA: Sage Publications.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin,* 56, 81–105.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17, 31–44.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–47.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.

Dorans, N. J., & Schmitt, A. P. (1991). *Constructed response and differential item functioning: A pragmatic approach* (Research Report No. RR-91-47). Princeton, NJ: Educational Testing Service.

Fleiss, J. L. (1973). *Statistical methods for rates and proportions*. New York: Wiley.

Green, D. R. (1975, December). Procedures for assessing bias in achievement tests. Presented at the National Institute of Education Conference on Test Bias, Annapolis, MD.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item* response theory. Newbury Park, CA: Sage Publications.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel–Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lee, W., Hanson, B. A., & Brennan, R. L. (2000, October). *Procedures for computing classification consistency and accuracy indices with multiple categories* (ACT Research Report Series 2000–10). Iowa City: ACT, Inc.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32(2), 179–197.

Loehlin, J. C. (1987). *Latent variable models*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel–Haenszel procedure. *Journal of the American Statistical Association,* 58, 690–700.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719–748.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18, 5–11.

Mogilner, A., & Mogilner, T. (2006). *Children's Writer's Word Book*, Cincinnati, OH: Writer's Digest Books

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.

Orlando, M. (2004, June). Critical issues to address when applying item response theory (IRT) models. Paper presented at the Drug Information Association, Bethesda, MD.

Ryan, J. P. (1983). Introduction to latent trait analysis and item response theory. In W. E. Hathaway (Ed.), *Testing in the schools: New directions for testing and measurement* (p. 19). San Francisco: Jossey-Bass.

Taylor, Stanford (1989). EDL Core Vocabularies in Reading, Mathematics, Science, and Social Studies, Orlando, FL: Steck-Vaughn Company.

Suen, H. K. (1990). *Principles of test theories*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Thissen, D. (1990). Reliability and measurement precision. In H. Wainer (Ed.), *Computerized adaptive testing:* A *primer* (pp. 161–186). Hillsdale, NJ: Lawrence Erlbaum.

Thissen, D., Chen, W.-H., & Bock, R. D. (2003). MULTILOG (version 7) [Computer software]. In Mathilda du Toit (Ed.), *IRT from SSI: BILOG-MG MULTILOG PARSCALE TESTFACT*. Chicago: Scientific Software International.

Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement, 5(2),* 245-262.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8(2),* 125–145.

Young, M. J., & Yoon, B. (1998, April). Estimating the consistency and accuracy of classifications in a standards-referenced assessment (CSE Technical Report 475). Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing. Los Angeles: University of California, Los Angeles.

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337–348). Hillsdale, NJ: Lawrence Erlbaum Associates.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 26, 44–66.

Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). Descriptive and inferential procedures for assessing differential item functioning in polytomous items. *Applied Measurement inEducation*, 10(4), 321–344.

Appendix A: Training Agendas

LEAP 2025 Social Studies Grades 3–8 Source Search Training Agenda

I. Introductions

II. Source Set Overviews
   a. Item Set and Task Topics
      i. Themes of the item set or task that will need to be developed and supported by sources and items
      ii. Reporting Categories
      iii. Potential Assessable GLEs
         1. Sources should support these GLEs
      iv. Potential Types of Sources
         1. The overview contains recommended sources that will support the item set or task
         2. Searchers can propose other sources that will support the item set or task
      v. Source Internet Source Links
         1. The overview contains specific websites that can be used to find sources or specific sources
   b. Bias and Sensitivity
      i. Bias: Avoid stimuli that cannot support items aligned to GLEs and standards. The focus on content aligned to the GLEs reduces the potential for bias that can occur by including content that is not aligned to instruction. This could give an advantage to one student group over other student groups.
      ii. Sensitivity: Avoid topics in sources that may upset or offend students in items (e.g., references to graphic violence, nudity, alcohol, drugs, recent natural disasters, caricature representation of ethnic groups)
      iii. Universal design and visual impairments

III. Receiving Source Search Assignments

IV. Submitting Sources for Assignments
    a. Text-Based Sources
        i. Readability Measurements
            1. Lexile
                a. Lexile bands
            2. ATOS
        ii. Originals and marked-up copies of texts
        iii. Text Complexity
        iv. Range of Textual Evidence
        v. Levels of Inference
    b. Graphic-Based Sources
        i. PDFs with source of graphic and location
        ii. Word document with caption
        iii. Gifs and JPEGs

V. Completing Webforms

VI. Using Box

VII. Additional Resources

LEAP 2025 Grades 3–8 Item Writer and Editor Training Agenda

I. Louisiana Student Standards and GLEs
   a. Grades 3–8
      i. Reporting Categories (History, Geography, Civics, Economics)
      ii. Grade-Level Expectations (GLEs)

II. Item Types and Set Overviews
   a. Selected-Response Items (Multiple Choice, Multiple Select)
      i. Rules for numbers of answer options and number correct
   b. Constructed-Response Items (Item Sets Only)
   c. Extended-Response Items (Tasks Only)
   d. Technology-Enhanced Items (Item Sets Only)
   e. Item Sets
      i. Sources (Each set will have multiple sources)
      ii. Item Set Overviews
         1. Item stems provided for each item
         2. Metadata associated with each item
         3. Answer options and the nature of distractors
   f. Task Sets
      i. Sources (Each set will have multiple sources)
      ii. Task Set Overviews
         1. Item stems provided for each item
         2. Metadata associated with each item
         3. Answer options and nature of distractors
   g. Standalone Items
      i. Purpose
      ii. Sources

III. Writing and Editing Rubrics and Scoring Guides
   a. Constructed-Response Item Scoring Rubrics
   b. Constructed-Response Item Scoring Information
   c. Extended-Response Scoring Rubrics
      i. Content
      ii. Claims
   d. Extended-Response Scoring Information

IV. Item Metadata
    a. Range of Textual Evidence
    b. Levels of Inference
    c. Depth of Knowledge: Items should be DOK 2 or DOK 3
V. Item Writing and Editing Reminders
    a. Grade Appropriate Language: Make sure the vocabulary of the items does not exceed the grade level of the students (Exception: Content-specific vocabulary that is part of the state standards).
    b. Plausible and Logical Distractors: Distractors should address misconceptions that the students may have about the topic.
    c. Cueing and Clanging of answer options:
        i. Items should avoid using key terms from the sources or in the stem that direct students to specific answer options.
        ii. Items in sets should avoid cueing each other, either in the stems or in the answer options.
    d. Outliers in answer options. Answer options should not stand out because they appear different from the other answer options.
        i. Capitalized words, use of numerals
        ii. Grammatical differences in answer options
    e. Bias and Sensitivity
        i. Bias: Avoid information in items that may give an advantage to one group over another group in answering the item (e.g., information that is not part of the curriculum, standards).
        ii. Sensitivity: Avoid topics that may upset or offend students in items (e.g., references to graphic violence, nudity, alcohol, drugs, recent natural disasters, group stereotypes, representation of ethnic groups).
VI. ABBI Item Development Platform
    a. Functionality of the ABBI platform
    b. Writing Items in ABBI
    c. Editing Items in ABBI
    d. Attaching Scoring Information in ABBI
    e. Checking Scoring of Technology-Enhanced Items
VII. Receiving Item Assignments via Smartsheet

VIII. Graphic Art Requests (Editing only)
   a. Using the Smartsheet Form
   b. Attaching Marked-Up Graphics in ABBI
   c. Confirming graphic edits have been made

IX. Alerting the coordinator that you have completed the item-writing or item-editing assignment and are ready for another assignment

X. Constructed-Response Item Sample Prompt, Rubric, and Scoring Notes:

Scoring for SOXXXXXXXXXXXXXX

Stem: Based on the sources and your knowledge of social studies, describe two different ways that World War II affected Louisiana.

Scoring Information	
Score Points	Description
2	Student's response correctly describes two different ways that World War II affected Louisiana.
1	Student's response correctly describes one way that World War II affected Louisiana.
0	Student's response does not correctly describe one way that World War II affected Louisiana.

Scoring Notes:

- People in Louisiana migrated from rural to urban areas because many jobs in war industries were in the cities.
- The number of employees increased in Louisiana businesses that produced goods for the war.
- Louisiana helped train and mobilize U.S. forces.
- Individuals from Louisiana served in the war.

Accept other reasonable answers.

XI. Selected-Response (multiple-choice, multiple-select Items)

a. Reference sources in stems where appropriate. Use the language Sources 1 and 2 rather than Source 1 and Source 2. When referring to all of the sources, say “all of the sources.” Refer to the source in the stem, where it is most appropriate.
b. Make sure MS items are in the correct format:
   Which natural resources inspired Americans to migrate westward? Select the two correct answers.
c. Make sure the item scores correctly.

XII. Editorial Process

a. Move the items to Content Editor 2 or to Proofing 1, depending on the editorial status of the item or the direction of the coordinator.

LEAP 2025 U.S. History and Grades 3–8 Data Review Training Agenda

I. What is a Data Review?

a. Statistical Definition: Classical Test Theory
   1. P-value
   2. Point-Biserial
   3. Option/Distribution Analysis
   4. Differential Item Function (DIF)
   5. Flagging Value

Statistics	**Flagging Value**
P-value	≤ 0.25 or > 0.9
Omit Percentage	> 4%
Point-biserial Correlation	< 0.20
Distractor Percentage (MC only)	> 40%
Distractor Point-biserial Correlation (MC only)	> 0.00
DIF	B, C

b. Statistical Definition: Item Response Theory (IRT)
   1. IRT Discrimination (a-parameter)
   2. IRT Difficulty (b-parameter)
   3. IRT Guessing (c-parameter)
   4. Q1 (Zq1)
   5. Item Fit Plot
   6. Flagging Value

Flagging Value for IRT Item Parameters		
a (Discrimination)	b (Difficulty)	c (Guessing)
< 0.34	Lower than -3.0 or Higher than 3.0	> 0.35

II. Judgment Task in ABBI

a. Accept
b. Accept with Edits
c. Reject

Appendix B: Test Summary

Social Studies G3–8

Contents
Table B.1 Percentage of Points by Reporting Category (includes Task Items): Spring 2022 Operational SS G3–8
Table B.2 Standard Coverage: Spring 2022 Operational SS G3–8
Table B.3 Item Type Summary: Spring 2022 Operational SS G3–8
Table B.4 Raw Score Summary: Spring 2022 Operational SS G3–8
Table B.5 Raw Score Summary by Reporting Category: Spring 2022 Operational SS G3–8
Table B.6 Scale Score and Raw Score Summary: Spring 2022 Operational SS G3–8

- Because the spring 2022 test was administered during the 2022 COVID-19 pandemic, great caution should be applied when any statistical inference is drawn.

Table B.1

Percentage of Points by Reporting Category (includes Task Items): Spring 2022 Operational SS G3–8

Reporting Category	G3	G4	G5	G6	G7	G8
1 History	24.4%	24.4%	43.9%	47.0%	56.1%	51.5%
2 Geography	26.7%	26.7%	12.3%	18.2%	12.1%	12.1%
3 Civics	24.4%	24.4%	29.8%	22.7%	21.2%	24.2%
4 Economics	24.4%	24.4%	14.0%	12.1%	10.6%	12.1%

Table B.2

Standard Coverage: Spring 2022 Operational SS G3–8

Grade 3

Reporting Categories		No. of Items			% of Test
		MS	MC	CR	
		N	N	N	
1 History	3.1.2		4		9.30
	3.1.6		1		2.33
	3.1.7		1		2.33
	3.2.2		2		4.65
	3.2.3		1		2.33
	3.2.4	1	1		4.65
	Sub-Total	**1**	**10**		**25.58**
2 Geography	3.3.1		1		2.33
	3.3.2		1		2.33
	3.3.4		1		2.33
	3.4.1		2		4.65
	3.4.3		1		2.33
	3.4.4		3	1	9.30
	3.4.6		1		2.33
	Sub-Total		**10**	**1**	**25.58**
3 Civics	3.5.1	1	1		4.65
	3.5.2		1		2.33
	3.5.3		1		2.33
	3.5.4		1		2.33
	3.5.5		1		2.33
	3.6.1		1	1	4.65
	3.6.2		1		2.33
	3.6.3		1		2.33
	Sub-Total	**1**	**8**	**1**	**23.26**
4 Economics	3.7.1		1		2.33
	3.7.3		1		2.33
	3.8.1		1		2.33
	3.8.2	1			2.33
	3.8.4		1		2.33
	3.9.1		1		2.33
	3.9.2		2		4.65
	3.10.1		3		6.98
	Sub-Total	**1**	**10**		**25.58**
Total		**3**	**38**	**2**	**100.00**

Grade 4

Reporting Categories		No. of Items			% of Test
		MS	MC	CR	
		N	N	N	
1 History	4.1.2		1		2.33
	4.1.6		1		2.33
	4.2.1		2		4.65
	4.2.3	1	1		4.65
	4.2.4		2		4.65
	4.2.5		1		2.33
	4.3.1		2		4.65
	Sub-Total	**1**	**10**		**25.58**
2 Geography	4.4.1		2		4.65
	4.4.2		1		2.33
	4.4.4		1		2.33
	4.4.5		1		2.33
	4.4.6		1		2.33
	4.5.1			1	2.33
	4.5.3		1		2.33
	4.6.1		1		2.33
	4.6.2	1	1		4.65
	Sub-Total	**1**	**9**	**1**	**25.58**
3 Civics	4.7.1		1		2.33
	4.7.3		2		4.65
	4.7.4		2		4.65
	4.8.1		1		2.33
	4.8.2		1		2.33
	4.8.3		1	1	4.65
	4.8.4	1			2.33
	Sub-Total	**1**	**8**	**1**	**23.26**
4 Economics	4.9.1		1		2.33
	4.9.2		1		2.33
	4.9.4		3		6.98
	4.9.5		1		2.33
	4.9.6		2		4.65
	4.9.7		1		2.33
	4.9.8		1		2.33
	4.9.10		1		2.33
	Sub-Total		**11**		**25.58**
Total		**3**	**38**	**2**	**100.00**

Grade 5

Reporting Categories		No. of Items					% of Test
		TEI	MS	MC	ER	CR	
		N	N	N	N	N	
1 History	5.1.3			2			4.35
	5.2.1	1	1	2			8.70
	5.2.2			3			6.52
	5.2.3			1		1	4.35
	5.2.4			3			6.52
	5.3.2			2			4.35
	5.3.5		1	5			13.04
	5.3.7			1			2.17
	Sub-Total	**1**	**2**	**19**		**1**	**50.00**
2 Geography	5.4.3			3			6.52
	5.5.1		1				2.17
	5.5.2			3			6.52
	Sub-Total		**1**	**6**			**15.22**
3 Civics	5.6.1			6	1		15.22
	5.6.2			1			2.17
	5.7.1			2			4.35
	Sub-Total			**9**	**1**		**21.74**
4 Economics	5.8.1			1			2.17
	5.9.1	1				1	4.35
	5.9.2			1			2.17
	5.10.1	1		1			4.35
	Sub-Total	**2**		**3**		**1**	**13.04**
Total		**3**	**3**	**37**	**1**	**2**	**100.00**

Grade 6

Reporting Categories		No. of Items					% of Test
		TEI	MS	MC	ER	CR	
		N	N	N	N	N	
1 History	6.2.2			2			3.70
	6.2.3		1	5		1	12.96
	6.2.4		1	1			3.70
	6.2.5		1	2			5.56
	6.2.6	1		6			12.96
	6.2.7					1	1.85
	6.2.8			1			1.85
	6.2.9			4			7.41
	6.2.10			1			1.85
	Sub-Total	**1**	**3**	**22**		**2**	**51.85**
2 Geography	6.3.1			2			3.70
	6.3.3			1			1.85
	6.3.4		1				1.85
	6.4.2			2			3.70
	6.4.3		1	5			11.11
	Sub-Total		**2**	**10**			**22.22**
3 Civics	6.5.1	1		4			9.26
	6.5.2			1	1		3.70
	Sub-Total	**1**		**5**	**1**		**12.96**
4 Economics	6.6.1	1		1			3.70
	6.6.2			2			3.70
	6.6.4		1	2			5.56
	Sub-Total	**1**	**1**	**5**			**12.96**
Total		**3**	**6**	**42**	**1**	**2**	**100.00**

Grade 7

Reporting Categories		No. of Items					% of Test
		TEI	MS	MC	ER	CR	
		N	N	N	N	N	
1 History	7.1.3			4			7.55
	7.2.1		1	2		1	7.55
	7.2.3	1	1	4			11.32
	7.2.4			2			3.77
	7.3.1			1		1	3.77
	7.3.2			3			5.66
	7.3.3			1			1.89
	7.4.1			1			1.89
	7.4.2			2	1		5.66
	7.4.3			1			1.89
	Sub-Total	**1**	**2**	**21**	**1**	**2**	**50.94**
2 Geography	7.5.3		1	2			5.66
	7.6.1			1			1.89
	7.6.2		1				1.89
	7.6.3		1				1.89
	7.6.4			1			1.89
	7.7.1			1			1.89
	Sub-Total		**3**	**5**			**15.09**
3 Civics	7.8.2			1			1.89
	7.8.4			1			1.89
	7.8.5	1		2			5.66
	7.8.6			1			1.89
	7.8.8	1		3			7.55
	7.9.2			1			1.89
	7.10.1			1			1.89
	Sub-Total	**2**		**10**			**22.64**
4 Economics	7.11.1	1		1			3.77
	7.11.2			2			3.77
	7.11.3			2			3.77
	Sub-Total	**1**		**5**			**11.32**
Total		**4**	**5**	**41**	**1**	**2**	**100.00**

Grade 8

Reporting Categories		No. of Items					% of Test
		TEI	MS	MC	ER	CR	
		N	N	N	N	N	
1 History	8.2.1	1		3			7.41
	8.2.3			1			1.85
	8.2.4			1			1.85
	8.2.5			4			7.41
	8.2.6	1	2	1		1	9.26
	8.2.7	1	1	5		1	14.81
	8.2.8			3			5.56
	8.2.9			3			5.56
	Sub-Total	**3**	**3**	**21**		**2**	**53.70**
2 Geography	8.3.1			1			1.85
	8.3.2			1			1.85
	8.4.1			2			3.70
	8.4.2			2			3.70
	8.4.3		1				1.85
	8.5.1		1				1.85
	Sub-Total		**2**	**6**			**14.81**
3 Civics	8.6.1			1			1.85
	8.7.1		1	4			9.26
	8.7.2			1			1.85
	8.8.1			1	1		3.70
	Sub-Total		**1**	**7**	**1**		**16.67**
4 Economics	8.9.1			1			1.85
	8.9.3			3			5.56
	8.10.1			1			1.85
	8.10.2			1			1.85
	8.10.3		1				1.85
	8.10.4		1				1.85
	Sub-Total		**2**	**6**			**14.81**
Total		**3**	**8**	**40**	**1**	**2**	**100.00**

Table B.3

Item Type Summary: Spring 2022 Operational SS G3–8

Grade	MC	MS	TEI	CR	ER*
3	38	3	-	2	-
4	38	3	-	2	-
5	37	3	3	2	1
6	42	6	3	2	1
7	41	5	4	2	1
8	40	8	3	2	1

***** Classical analyses are calculated and estimated separately for each dimension of the ER item, and the result summarizes both dimensions.

Table B.4

Raw Score Summary: Spring 2022 Operational SS G3–8

Grade	*N*	Mean	SD	Min	Max	Mean_Pval	Mean_Pbis	Reliability*	SEM
3	≥49,310	19.22	8.15	0	44	0.44	0.38	0.86	3.05
4	≥48,880	19.67	7.82	0	44	0.45	0.37	0.85	3.03
5	≥48,880	19.89	9.55	1	56	0.38	0.40	0.89	3.17
6	≥49,270	27.11	12.19	3	65	0.45	0.43	0.92	3.45
7	≥50,930	26.78	12.53	0	65	0.43	0.44	0.92	3.54
8	≥50,660	29.61	13.25	0	66	0.47	0.45	0.93	3.51

* Reliability is Cronbach's alpha.

Table B.5

Raw Score Summary by Reporting Category: Spring 2022 Operational SS G3–8

Admin	Reporting Category	Mean	SD	Min	Max	Mean_Pval	Mean_Pbis	Reliability	SEM
3	History	4.54	2.31	0	11	0.41	0.36	0.57	1.51
	Geography	4.77	2.48	0	12	0.41	0.36	0.58	1.61
	Civics	5.10	2.61	0	11	0.49	0.45	0.71	1.41
	Economics	4.81	2.33	0	11	0.44	0.36	0.59	1.49
4	History	4.86	2.46	0	11	0.44	0.39	0.63	1.50
	Geography	4.87	2.45	0	12	0.43	0.36	0.60	1.55
	Civics	5.35	2.35	0	11	0.50	0.38	0.60	1.49
	Economics	4.59	2.26	0	11	0.42	0.34	0.54	1.53
5	History	9.58	4.56	0	25	0.39	0.37	0.76	2.23
	Geography	3.50	1.81	0	7	0.50	0.40	0.57	1.19
	Civics	4.18	2.87	0	17	0.32	0.42	0.71	1.55
	Economics	2.64	1.82	0	8	0.35	0.45	0.59	1.17
6	History	14.45	6.50	0	31	0.49	0.44	0.86	2.43
	Geography	5.25	2.54	0	12	0.44	0.36	0.61	1.59
	Civics	4.31	2.67	0	15	0.37	0.48	0.71	1.44
	Economics	3.10	1.91	0	8	0.40	0.42	0.58	1.24
7	History	13.6	7.01	0	37	0.41	0.44	0.87	2.53
	Geography	3.40	1.92	0	8	0.43	0.40	0.58	1.24
	Civics	6.43	3.13	0	14	0.47	0.44	0.73	1.63
	Economics	3.33	1.84	0	7	0.46	0.46	0.59	1.18
8	History	15.67	7.05	0	34	0.46	0.44	0.86	2.64
	Geography	4.31	2.29	0	8	0.54	0.50	0.72	1.21
	Civics	5.86	3.26	0	16	0.43	0.46	0.73	1.69
	Economics	3.77	2.01	0	8	0.47	0.42	0.63	1.22

Table B.6.1

Scale Score and Raw Score Summary: Spring 2022 Operational SS G3

Subgroup	*N*	Percent	Scale Score Mean	Scale Score SD	Raw Score Mean	Raw Score SD	Effect Size
Total	≥49,310	100.00	717.26	39.51	19.22	8.15	-
Female	≥24,090	48.85	717.49	38.61	19.22	7.96	**0.03**
Male	≥25,220	51.15	717.03	40.34	19.22	8.33	-
African American	≥20,380	41.34	703.82	35.95	16.41	6.93	**0.85**
American Indian or Alaska Native	≥270	0.55	721.92	35.84	20.04	7.65	**0.28**
Asian	≥830	1.69	741.25	39.97	24.47	8.81	**-0.22**
Hispanic/Latino	≥4,970	10.09	709.17	38.33	17.56	7.63	**0.60**
Multi-Racial	≥1,850	3.76	722.25	37.93	20.18	7.96	**0.24**
Native Hawaiian or Other PacificIslander	≥30	0.08	721.34	37.81	19.92	8.39	**0.31**
White	≥20,930	42.45	730.82	38.16	22.05	8.28	-
Economically Disadvantaged: No	≥13,960	28.31	738.16	38.37	23.72	8.47	**-0.83**
Economically Disadvantaged: Yes	≥35,190	71.37	709.06	36.79	17.46	7.29	-
EL: No	≥46,460	94.23	718.63	39.50	19.50	8.19	**-0.63**
EL: Yes	≥2,840	5.77	694.85	32.20	14.62	5.73	-
Regular Education	≥43,000	87.21	719.41	39.44	19.66	8.20	**-0.48**
Special Education	≥6,300	12.79	702.56	36.70	16.22	7.11	-
Section 504: No	≥45,770	92.82	717.94	39.69	19.37	8.21	**-0.26**
Section 504: Yes	≥3,540	7.18	708.41	35.95	17.28	7.09	-
Migrant: No	≥49,220	99.81	717.29	39.50	19.23	8.15	**-0.42**
Migrant: Yes	≥90	0.19	697.91	39.96	15.53	7.49	-
Homeless: No	≥47,890	97.13	717.69	39.51	19.31	8.16	**-0.39**
Homeless: Yes	≥1,410	2.87	702.67	36.35	16.19	7.02	-
Military Affiliation: No	≥48,360	98.07	716.92	39.44	19.15	8.13	**0.51**
Military Affiliation: Yes	≥950	1.93	734.21	39.10	22.86	8.50	-
Foster Care: No	≥49,150	99.68	717.29	39.51	19.23	8.15	**-0.21**
Foster Care: Yes	≥150	0.32	707.70	36.03	17.08	7.18	-

Table B.6.2

Scale Score and Raw Score Summary: Spring 2022 Operational SS G4

Subgroup	N	Percent	Scale Score Mean	Scale Score SD	Raw Score Mean	Raw Score SD	Effect Size
Total	≥48,880	100.00	722.49	41.97	19.67	7.82	-
Female	≥23,890	48.87	721.77	40.43	19.46	7.50	**0.11**
Male	≥24,990	51.13	723.18	43.37	19.88	8.10	-
African American	≥20,480	41.89	707.58	38.42	16.84	6.60	**0.90**
American Indian or Alaska Native	≥260	0.54	727.95	40.80	20.74	7.64	**0.23**
Asian	≥800	1.64	750.82	42.78	25.39	8.57	**-0.35**
Hispanic/Latino	≥5,080	10.40	716.17	41.42	18.49	7.52	**0.57**
Multi-Racial	≥1,680	3.44	729.60	39.34	20.92	7.52	**0.24**
Native Hawaiian or Other Pacific Islander	≥30	0.08	727.41	32.56	20.05	6.59	**0.34**
White	≥20,510	41.96	737.19	39.89	22.45	7.89	-
Economically Disadvantaged: No	≥13,980	28.60	744.80	38.91	23.99	7.91	**-0.83**
Economically Disadvantaged: Yes	≥34,610	70.80	713.67	39.74	17.96	7.07	-
EL: No	≥46,420	94.96	723.87	41.85	19.93	7.84	**-0.66**
EL: Yes	≥2,460	5.04	696.44	35.15	14.90	5.62	-
Regular Education	≥42,910	87.79	725.63	41.30	20.23	7.79	**-0.60**
Special Education	≥5,960	12.21	699.92	39.80	15.66	6.78	-
Section 504: No	≥44,740	91.52	723.39	42.07	19.85	7.86	**-0.24**
Section 504: Yes	≥4,140	8.48	712.71	39.60	17.78	7.04	-
Migrant: No	≥48,820	99.87	722.50	41.97	19.67	7.82	**-0.21**
Migrant: Yes	≥60	0.13	711.10	41.46	17.63	7.35	-
Homeless: No	≥47,570	97.32	722.92	41.95	19.75	7.83	**-0.38**
Homeless: Yes	≥1,300	2.68	706.71	39.55	16.77	6.84	-
Military Affiliation: No	≥47,930	98.05	722.06	41.92	19.59	7.79	**0.57**
Military Affiliation: Yes	≥950	1.95	744.14	38.59	23.81	7.90	-
Foster Care: No	≥48,750	99.73	722.51	41.98	19.68	7.82	**-0.22**
Foster Care: Yes	≥130	0.27	712.62	37.28	17.62	6.63	-

Table B.6.3

Scale Score and Raw Score Summary: Spring 2022 Operational SS G5

Subgroup	N	Percent	Scale Score Mean	Scale Score SD	Raw Score Mean	Raw Score SD	Effect Size
Total	≥48,880	100.00	720.28	38.57	19.89	9.55	-
Female	≥23,810	48.72	720.43	37.44	19.80	9.26	**-0.02**
Male	≥25,060	51.28	720.14	39.62	19.98	9.81	-
African American	≥20,650	42.26	706.37	35.41	16.43	7.73	**0.87**
American Indian or Alaska Native	≥250	0.52	719.96	35.19	19.46	8.42	**0.38**
Asian	≥740	1.53	749.54	39.43	28.12	11.27	**-0.41**
Hispanic/Latino	≥4,790	9.81	715.73	38.59	18.79	9.20	**0.55**
Multi-Racial	≥1,640	3.36	726.75	36.65	21.33	9.44	**0.21**
Native Hawaiian or Other Pacific Islander	≥40	0.09	717.75	42.13	19.61	10.37	**0.39**
White	≥20,730	42.41	733.65	36.36	23.20	9.86	-
Economically Disadvantaged: No	≥14,580	29.83	740.13	36.32	25.06	10.18	**-0.82**
Economically Disadvantaged: Yes	≥33,990	69.54	711.99	36.34	17.73	8.35	-
EL: No	≥46,800	95.75	721.52	38.41	20.18	9.58	**-0.77**
EL: Yes	≥2,070	4.25	692.56	31.15	13.45	5.74	-
Regular Education	≥42,880	87.72	723.52	37.87	20.62	9.57	**-0.75**
Special Education	≥6,000	12.28	697.19	35.57	14.69	7.57	-
Section 504: No	≥44,140	90.31	721.37	38.64	20.17	9.63	**-0.31**
Section 504: Yes	≥4,730	9.69	710.14	36.38	17.35	8.34	-
Migrant: No	≥48,820	99.88	720.30	38.58	19.90	9.55	**-0.45**
Migrant: Yes	≥50	0.12	707.32	34.04	16.29	7.05	-
Homeless: No	≥47,660	97.51	720.65	38.57	19.98	9.57	**-0.38**
Homeless: Yes	≥1,210	2.49	705.95	35.92	16.40	7.85	-
Military Affiliation: No	≥47,990	98.17	719.91	38.52	19.80	9.51	**0.55**
Military Affiliation: Yes	≥890	1.83	740.28	35.86	25.09	10.08	-
Foster Care: No	≥48,770	99.77	720.32	38.57	19.90	9.55	**-0.42**
Foster Care: Yes	≥110	0.23	703.40	34.54	15.81	7.23	-

Table B.6.4

Scale Score and Raw Score Summary: Spring 2022 Operational SS G6

Subgroup	N	Percent	Scale Score Mean	Scale Score SD	Raw Score Mean	Raw Score SD	Effect Size
Total	≥49,270	100.00	712.43	37.51	27.11	12.19	-
Female	≥23,930	48.58	713.72	36.28	27.42	11.86	**0.01**
Male	≥25,340	51.42	711.22	38.59	26.82	12.48	-
African American	≥20,680	41.98	699.59	33.79	22.86	10.33	**0.81**
American Indian or Alaska Native	≥280	0.57	716.03	34.47	28.01	11.39	**0.30**
Asian	≥790	1.61	742.60	39.77	37.46	13.58	**-0.41**
Hispanic/Latino	≥5,060	10.28	706.11	37.32	25.14	11.77	**0.56**
Multi-Racial	≥1,620	3.29	718.31	36.36	28.96	12.14	**0.25**
Native Hawaiian or Other Pacific Islander	≥20	0.06	729.00	34.03	32.00	12.15	**0.09**
White	≥20,770	42.16	725.11	36.12	31.28	12.29	-
Economically Disadvantaged: No	≥14,460	29.35	732.72	35.01	33.88	12.20	**-0.83**
Economically Disadvantaged: Yes	≥34,530	70.08	704.12	35.21	24.33	11.03	-
EL: No	≥47,210	95.80	713.79	37.26	27.53	12.18	**-0.81**
EL: Yes	≥2,060	4.20	681.43	28.67	17.58	7.78	-
Regular Education	≥43,710	88.71	715.87	36.73	28.16	12.11	**-0.72**
Special Education	≥5,560	11.29	685.43	32.23	18.86	9.30	-
Section 504: No	≥44,070	89.44	713.97	37.56	27.61	12.27	**-0.35**
Section 504: Yes	≥5,200	10.56	699.42	34.35	22.85	10.55	-
Migrant: No	≥49,210	99.87	712.44	37.50	27.11	12.19	**-0.01**
Migrant: Yes	≥60	0.13	711.79	39.10	27.06	12.58	-
Homeless: No	≥48,020	97.46	712.77	37.54	27.22	12.21	**-0.34**
Homeless: Yes	≥1,250	2.54	699.52	33.67	22.80	10.40	-
Military Affiliation: No	≥48,430	98.28	712.13	37.46	27.01	12.16	**0.50**
Military Affiliation: Yes	≥840	1.72	730.06	35.72	32.97	12.37	-
Foster Care: No	≥49,150	99.74	712.47	37.51	27.12	12.19	**-0.32**
Foster Care: Yes	≥120	0.26	698.00	32.95	22.27	10.16	-

Table B.6.5

Scale Score and Raw Score Summary: Spring 2022 Operational SS G7

Subgroup	N	Percent	Scale Score Mean	Scale Score SD	Raw Score Mean	Raw Score SD	Effect Size
Total	≥50,930	100.00	727.01	40.54	26.78	12.53	-
Female	≥25,070	49.23	727.95	39.10	26.91	12.17	**-0.04**
Male	≥25,850	50.77	726.10	41.86	26.65	12.86	-
African American	≥21,850	42.91	713.30	36.71	22.47	10.57	**0.81**
American Indian or Alaska Native	≥290	0.58	731.20	35.87	27.65	11.45	**0.28**
Asian	≥740	1.47	762.21	41.91	38.42	13.76	**-0.50**
Hispanic/Latino	≥4,830	9.50	720.89	40.58	24.99	12.21	**0.54**
Multi-Racial	≥1,690	3.32	733.66	39.29	28.77	12.50	**0.23**
Native Hawaiian or Other Pacific Islander	≥40	0.08	743.10	37.95	31.79	12.29	**0.05**
White	≥21,440	42.10	740.54	39.02	30.99	12.71	-
Economically Disadvantaged: No	≥15,130	29.72	747.62	38.69	33.35	12.87	**-0.77**
Economically Disadvantaged: Yes	≥35,500	69.71	718.41	38.08	24.03	11.29	-
EL: No	≥49,100	96.41	728.33	40.31	27.16	12.54	**-0.89**
EL: Yes	≥1,820	3.59	691.73	28.91	16.58	6.65	-
Regular Education	≥45,380	89.10	730.66	39.81	27.83	12.51	**-0.82**
Special Education	≥5,550	10.90	697.22	33.50	18.15	8.74	-
Section 504: No	≥45,550	89.44	728.80	40.52	27.32	12.61	**-0.37**
Section 504: Yes	≥5,370	10.56	711.87	37.40	22.13	10.77	-
Migrant: No	≥50,870	99.88	727.03	40.54	26.78	12.53	**-0.35**
Migrant: Yes	≥60	0.12	711.80	37.87	22.16	10.70	-
Homeless: No	≥49,770	97.72	727.34	40.54	26.88	12.55	**-0.32**
Homeless: Yes	≥1,160	2.28	712.75	37.87	22.43	10.91	-
Military Affiliation: No	≥50,070	98.30	726.66	40.48	26.66	12.49	**0.58**
Military Affiliation: Yes	≥860	1.70	747.57	38.38	33.32	12.76	-
Foster Care: No	≥50,810	99.76	727.05	40.54	26.79	12.53	**-0.38**
Foster Care: Yes	≥120	0.24	711.96	34.86	21.86	9.77	-

Table B.6.6

Scale Score and Raw Score Summary: Spring 2022 Operational SS G8

Subgroup	N	Percent	Scale Score Mean	Scale Score SD	Raw Score Mean	Raw Score SD	Effect Size
Total	≥50,660	100.00	726.97	42.44	29.61	13.25	-
Female	≥24,990	49.34	729.47	40.19	30.22	12.76	**0.00**
Male	≥25,660	50.66	724.53	44.38	29.01	13.69	-
African American	≥21,520	42.49	712.72	39.30	25.05	11.67	**0.96**
American Indian or Alaska Native	≥280	0.55	735.43	40.01	32.17	12.82	**0.27**
Asian	≥810	1.60	761.77	44.30	40.99	14.05	**-0.42**
Hispanic/Latino	≥4,820	9.52	720.55	43.84	27.80	13.38	**0.62**
Multi-Racial	≥1,570	3.11	733.28	40.77	31.50	13.02	**0.27**
Native Hawaiian or Other Pacific Islander	≥40	0.08	746.12	43.46	35.77	14.10	**0.02**
White	≥21,600	42.63	740.68	39.83	33.94	12.98	-
Economically Disadvantaged: No	≥16,040	31.67	747.93	38.53	36.31	12.79	**-0.83**
Economically Disadvantaged: Yes	≥34,310	67.72	717.37	40.61	26.54	12.29	-
EL: No	≥48,850	96.42	728.46	42.01	30.04	13.20	**-0.95**
EL: Yes	≥1,810	3.58	686.74	33.18	17.85	8.56	-
Regular Education	≥45,460	89.73	730.88	41.25	30.76	13.08	**-0.84**
Special Education	≥5,200	10.27	692.74	36.91	19.54	10.08	-
Section 504: No	≥45,500	89.82	728.73	42.37	30.17	13.29	**-0.38**
Section 504: Yes	≥5,150	10.18	711.38	39.78	24.67	11.85	-
Migrant: No	≥50,600	99.88	726.98	42.44	29.61	13.25	**-0.18**
Migrant: Yes	≥50	0.12	719.68	41.02	27.34	12.63	-
Homeless: No	≥49,550	97.80	727.27	42.41	29.70	13.26	**-0.32**
Homeless: Yes	≥1,110	2.20	713.47	41.67	25.41	12.48	-
Military Affiliation: No	≥49,750	98.20	726.53	42.38	29.47	13.22	**0.59**
Military Affiliation: Yes	≥910	1.80	750.78	38.57	37.23	12.83	-
Foster Care: No	≥50,530	99.74	727.03	42.42	29.63	13.25	**-0.48**
Foster Care: Yes	≥130	0.26	702.54	40.50	22.21	11.70	-

Appendix C: Item Analysis Summary Report

Social Studies G3–8

Contents
Table C.1 P-Value Summary by Grade: Spring 2022 Operational SS G3–8 Plot C.1 P-Value Summary by Grade: Spring 2022 Operational SS G3–8
Table C.2 Item-Total Correlation Summary by Grade: Spring 2022 Operational SS G3–8 Plot C.2 Item-Total Correlation Summary by Grade: Spring 2022 Operational SS G3–8
Table C.3 Corrected Point-Biserial Correlation Summary by Grade: Spring 2022 Operational SS G3–8 Plot C.3 Corrected Point-Biserial Correlation Summary by Grade: Spring 2022 Operational SS G3–8
Table C.4 Item-Total Correlation Summary by Reporting Category and Grade: Spring 2022 Operational SS G3–8
Table C.5.1 IRT-A Parameter Summary by Reporting Category: SS G3–8 Table C.5.2 IRT-B Parameter Summary by Reporting Category: SS G3–8 Table C.5.3 IRT Parameter Summary: Spring 2022 Operational SS G3–8 Plot C.5.1 IRT Parameter Summary: Spring 2022 Operational SS G3–8: A-Parameter Plot C.5.2 IRT Parameter Summary: Spring 2022 Operational SS G3–8: B-Parameter Plot C.5.3 IRT Parameter Summary: Spring 2022 Operational SS G3–8: C-Parameter
Table C.6 Statistically Flagged Items by Item Type: Spring 2022 Operational SS G3–8

- Because the spring 2022 test was administered during the 2022 COVID-19 pandemic, great caution should be applied when any statistical inference is drawn.

Table C.1.1

P-Value Summary by Grade: Spring 2022 Operational SS G3–8

Grade	No. of Items	0 ≤ p < 0.2	0.2 ≤ p < 0.4	0.4 ≤ p < 0.6	0.6 ≤ p < 0.8	0.8 ≤ p ≤ 1.0
3	43	0	17	20	6	0
4	43	1	13	25	4	0
5	47	4	22	20	1	0
6	55	3	17	27	8	0
7	54	2	17	31	4	0
8	55	2	12	34	7	0

Plot C.1.1

P-Value Summary by Grade: Spring 2022 Operational SS G3–8

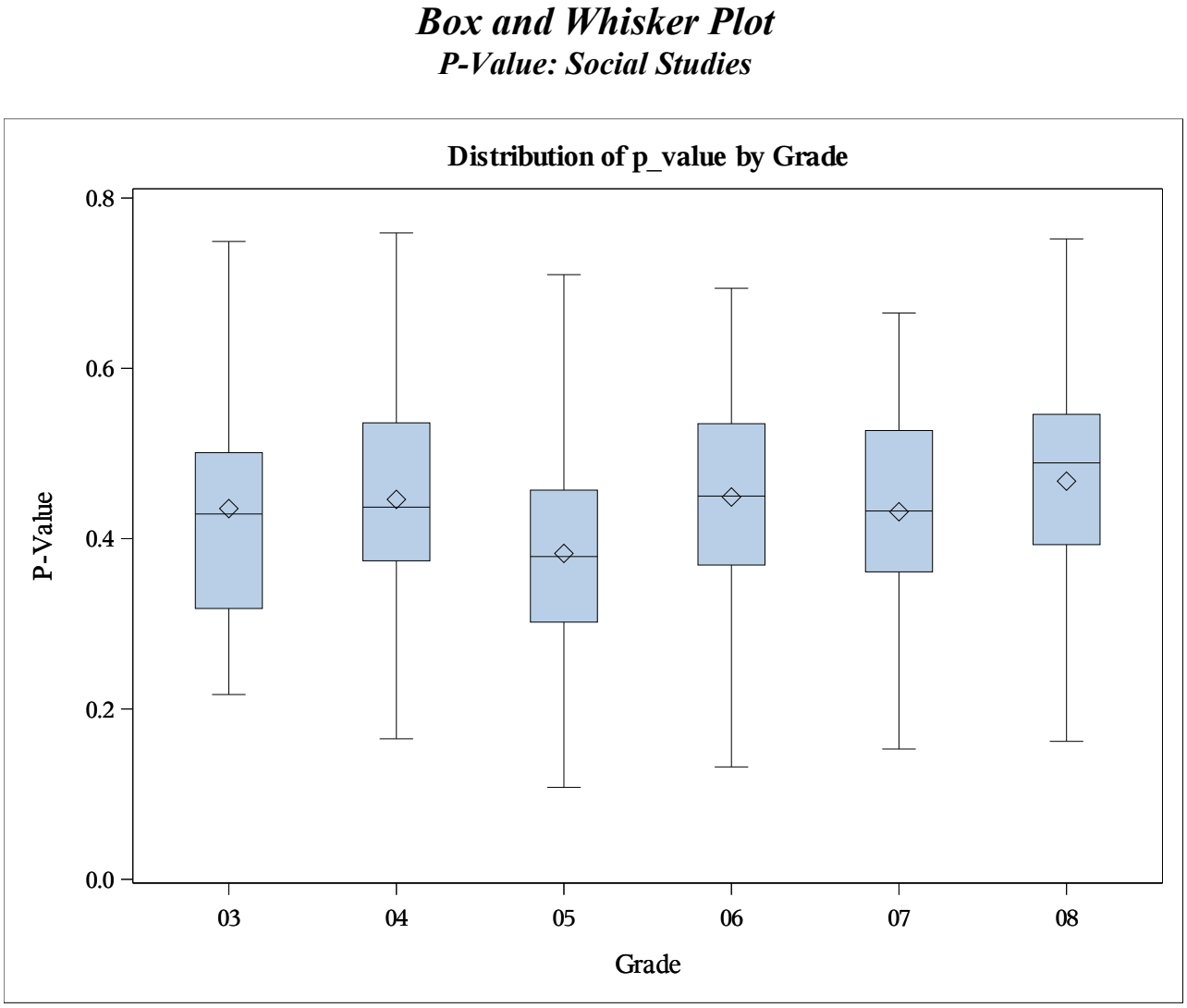


Table C.1.2

P-Value Summary by Item Type: Spring 2022 Operational SS G3–8

Grade 3

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.217	0.217	0.255	0.292	0.292
MC	38	0.247	0.371	0.446	0.515	0.749
MS	3	0.282	0.282	0.302	0.396	0.396

Grade 4

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.165	0.165	0.246	0.326	0.326
MC	38	0.240	0.395	0.456	0.536	0.634
MS	3	0.248	0.248	0.403	0.759	0.759

Grade 5

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.187	0.187	0.234	0.281	0.281
ER	2	0.108	0.108	0.110	0.112	0.112
MC	37	0.240	0.344	0.402	0.457	0.710
MS	3	0.137	0.137	0.271	0.561	0.561
TEI	3	0.264	0.264	0.269	0.507	0.507

Grade 6

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.164	0.164	0.231	0.298	0.298
ER	2	0.132	0.132	0.138	0.143	0.143
MC	42	0.248	0.400	0.490	0.561	0.694
MS	6	0.260	0.329	0.355	0.435	0.493
TEI	3	0.290	0.290	0.350	0.495	0.495

Grade 7

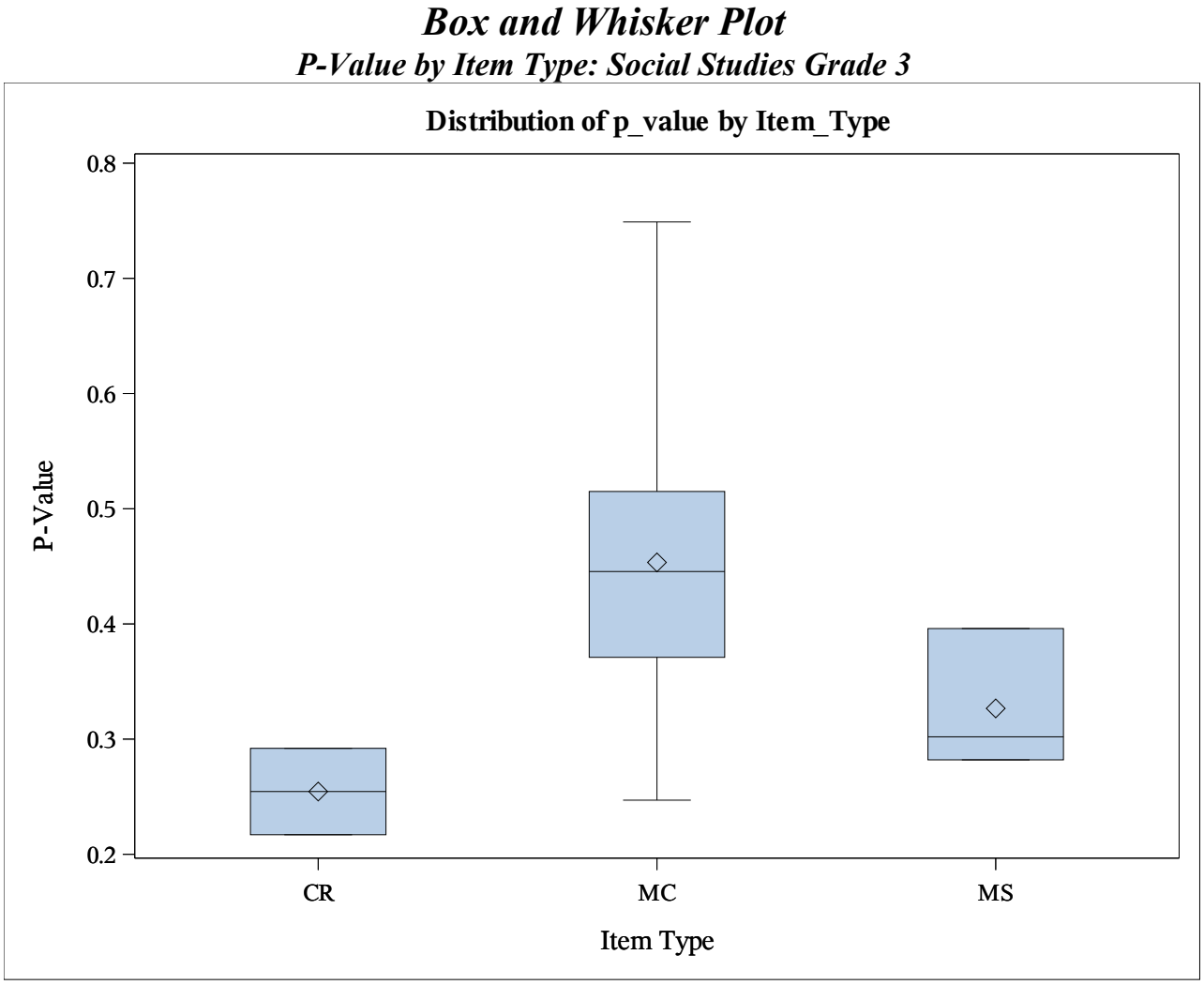
Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.303	0.303	0.344	0.384	0.384
ER	2	0.153	0.153	0.160	0.166	0.166
MC	41	0.249	0.395	0.450	0.535	0.665
MS	5	0.216	0.278	0.300	0.436	0.440
TEI	4	0.339	0.373	0.462	0.540	0.564

Grade 8

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.359	0.359	0.423	0.487	0.487
ER	2	0.257	0.257	0.259	0.260	0.260
MC	40	0.240	0.461	0.514	0.576	0.656
MS	8	0.162	0.187	0.377	0.445	0.519
TEI	3	0.305	0.305	0.441	0.752	0.752

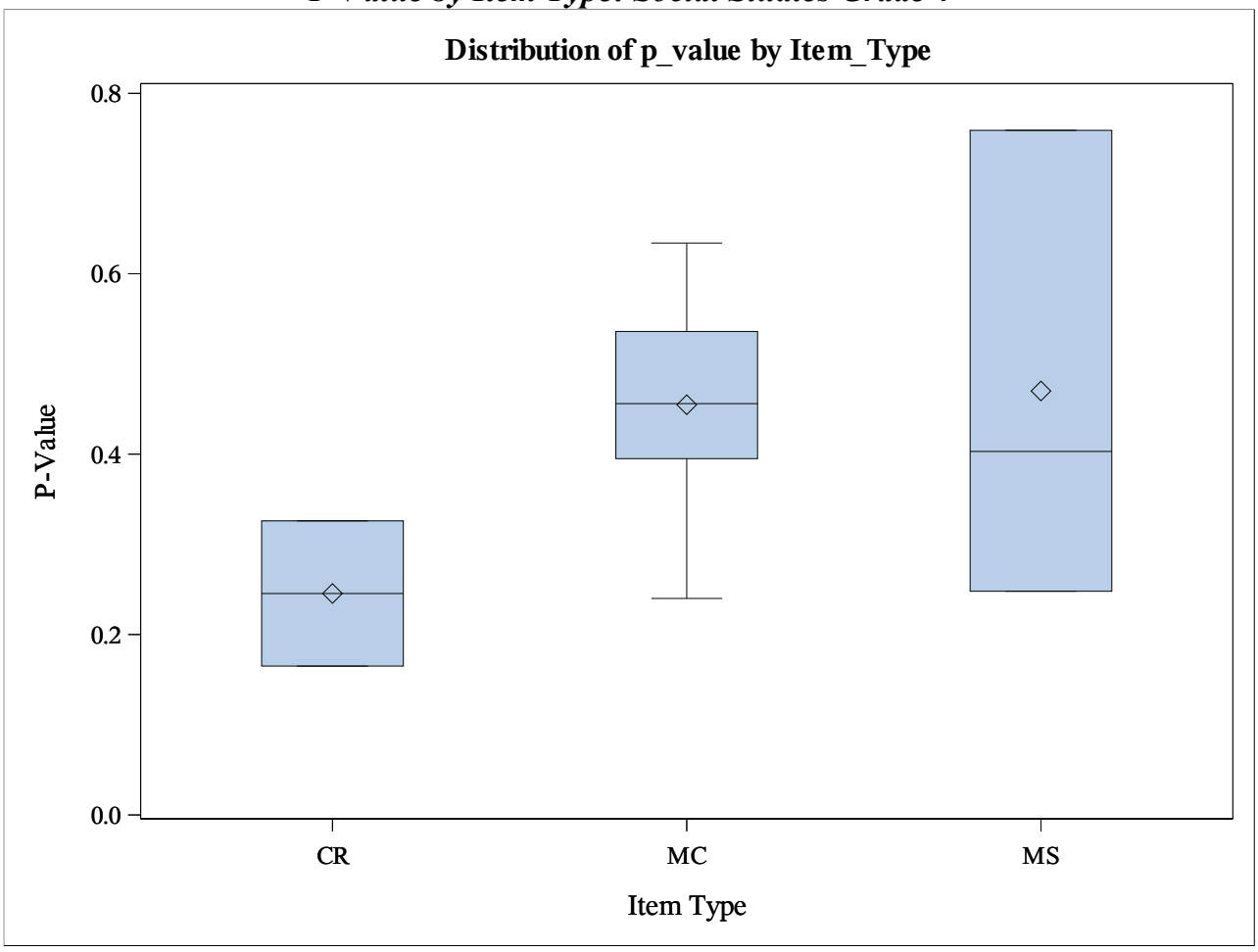
Plot C.1.2

P-Value Summary by Item Type: Spring 2022 Operational SS G3–8



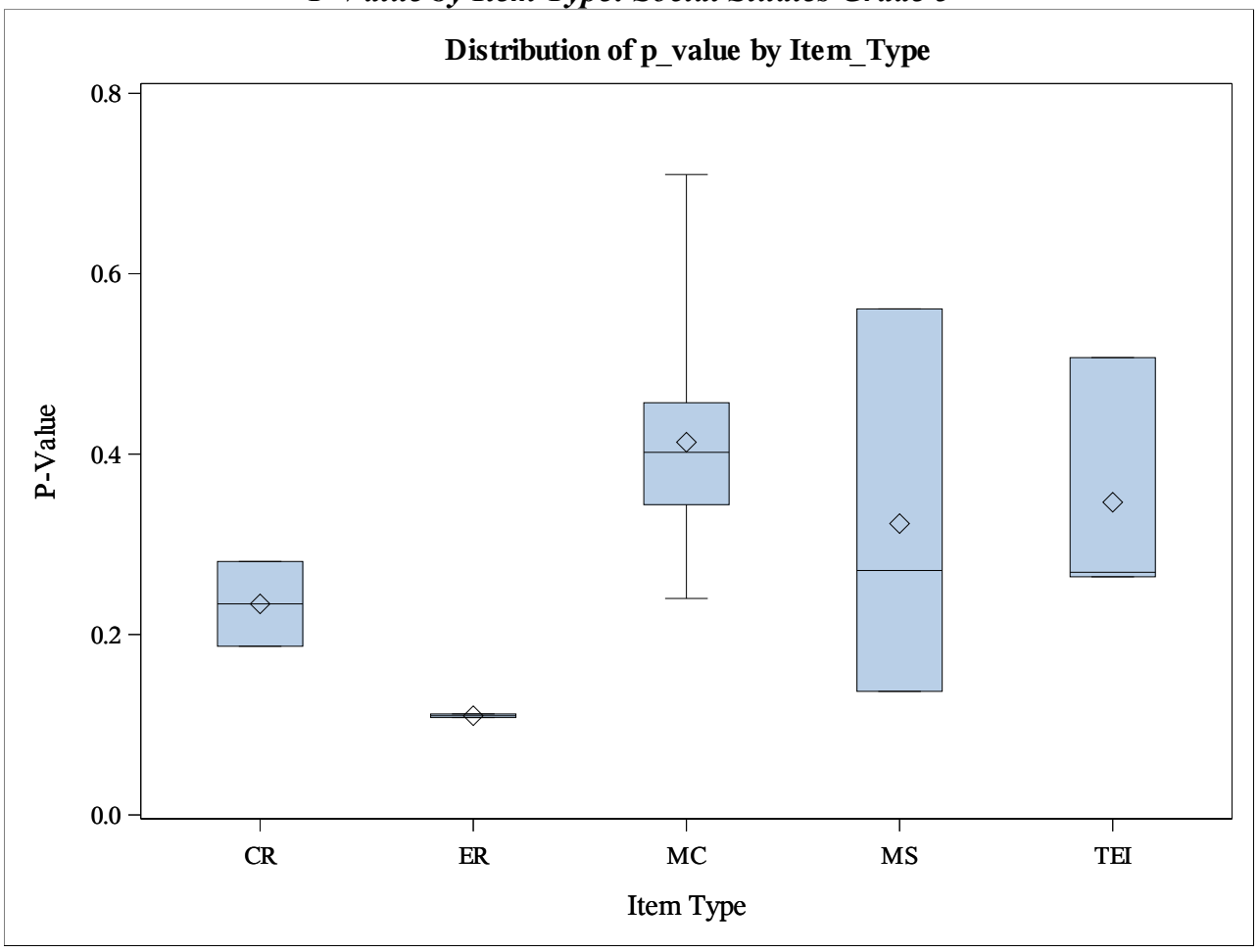
Box and Whisker Plot

P-Value by Item Type: Social Studies Grade 4



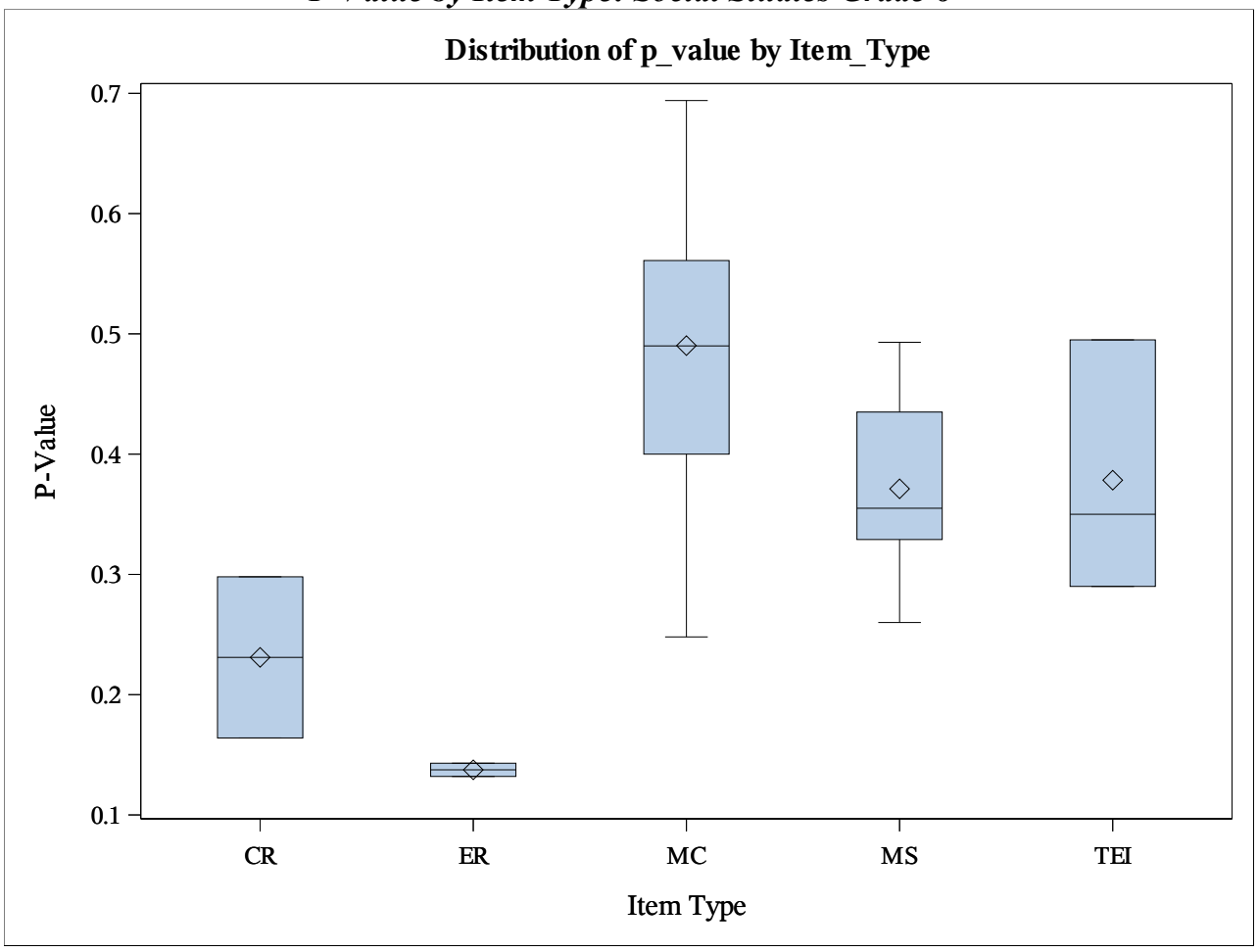
Box and Whisker Plot

P-Value by Item Type: Social Studies Grade 5



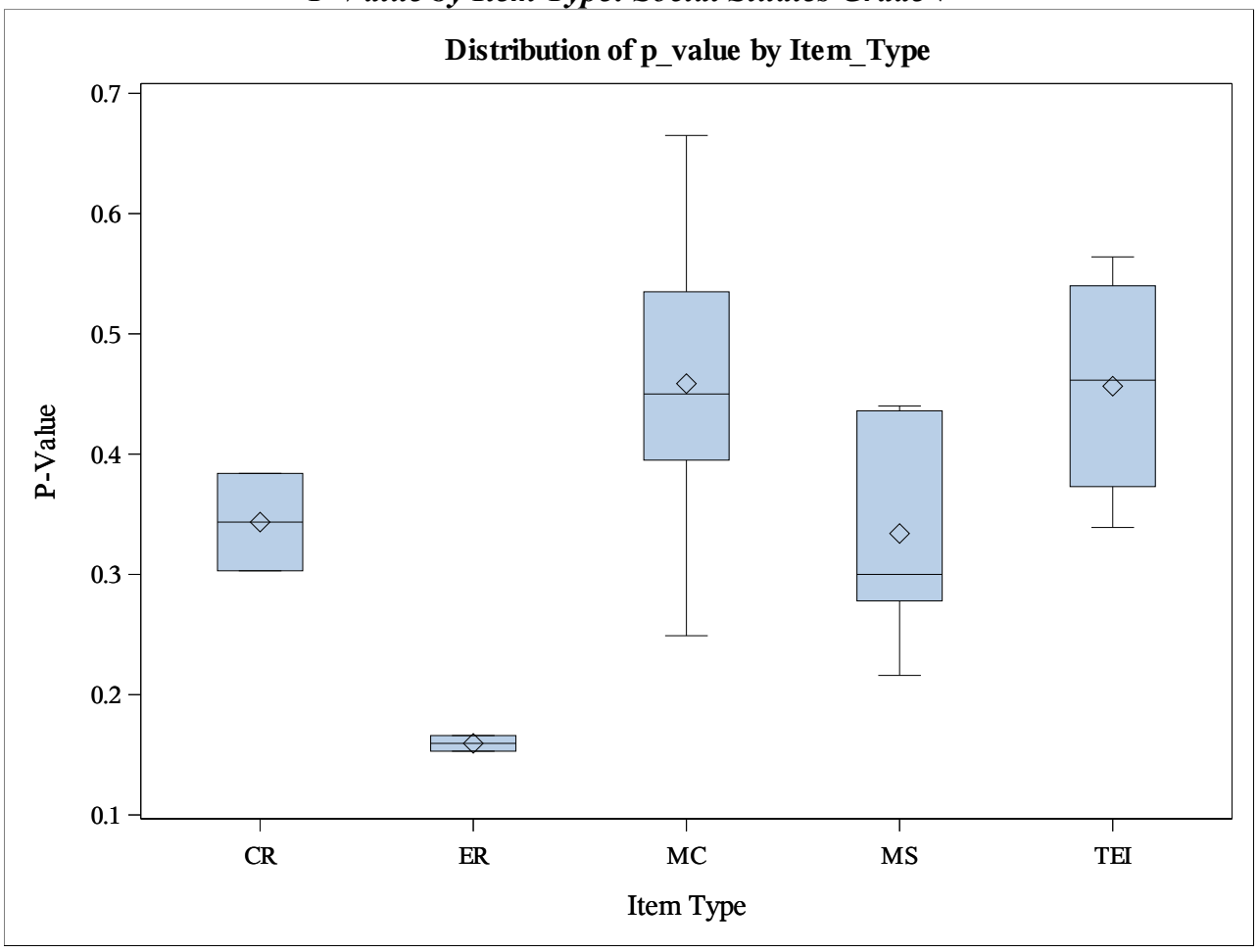
Box and Whisker Plot

P-Value by Item Type: Social Studies Grade 6



Box and Whisker Plot

P-Value by Item Type: Social Studies Grade 7



Box and Whisker Plot

P-Value by Item Type: Social Studies Grade 8

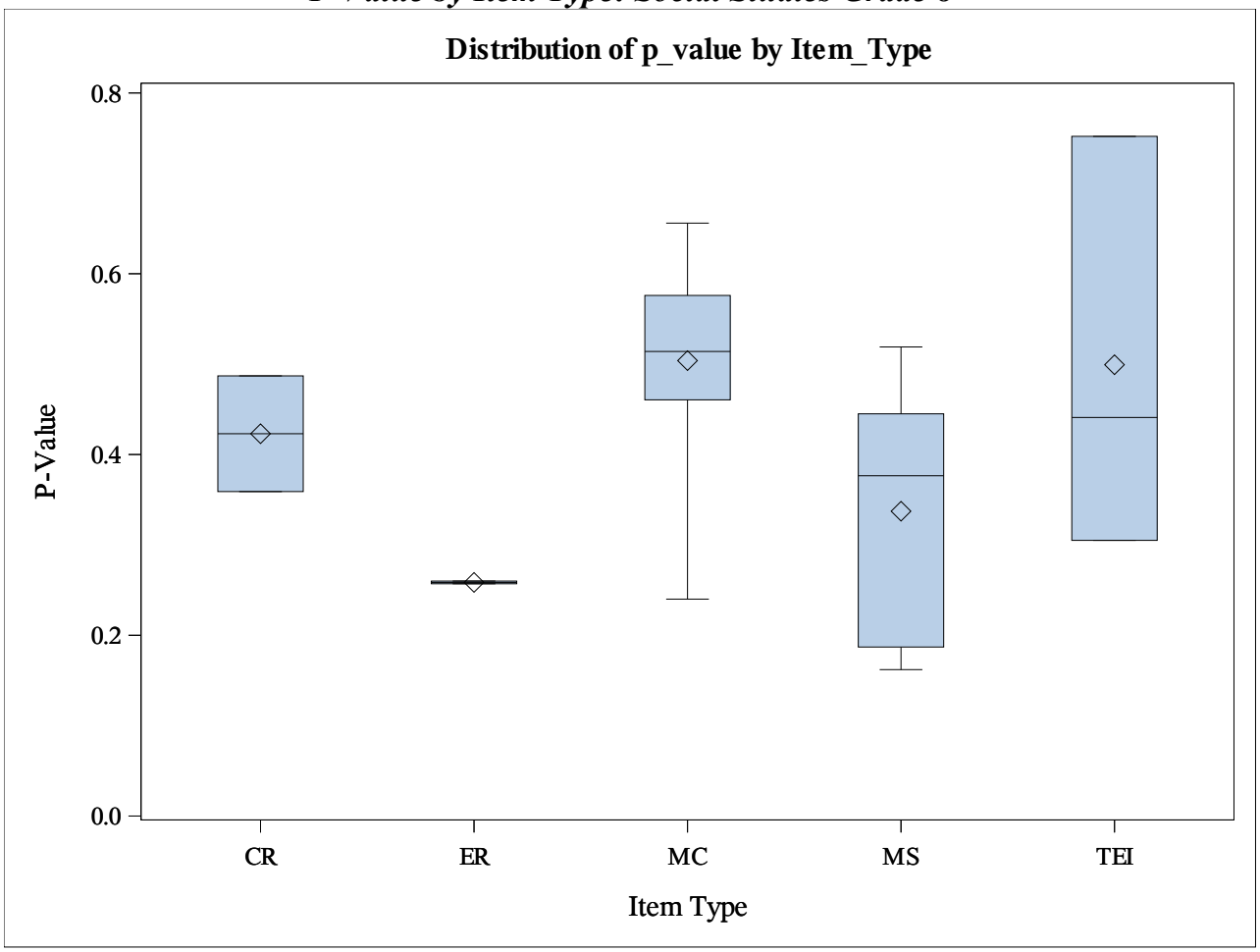


Table C.2.1

Item-Total Correlation by Grade: Spring 2022 Operational SS G3–8

Grade	No. of Items	r < 0	0.0 ≤ r < 0.2	0.2 ≤ r < 0.3	0.3 ≤ r < 0.4	0.4 ≤ r < 0.5	r ≥ 0.5
3	43	0	3	5	14	16	5
4	43	0	2	8	18	12	3
5	47	0	0	10	14	16	7
6	55	0	0	3	21	17	14
7	54	0	1	7	14	18	14
8	55	0	0	3	17	17	18

Plot C.2.1

Item-Total Correlation by Grade: Spring 2022 Operational SS G3–8

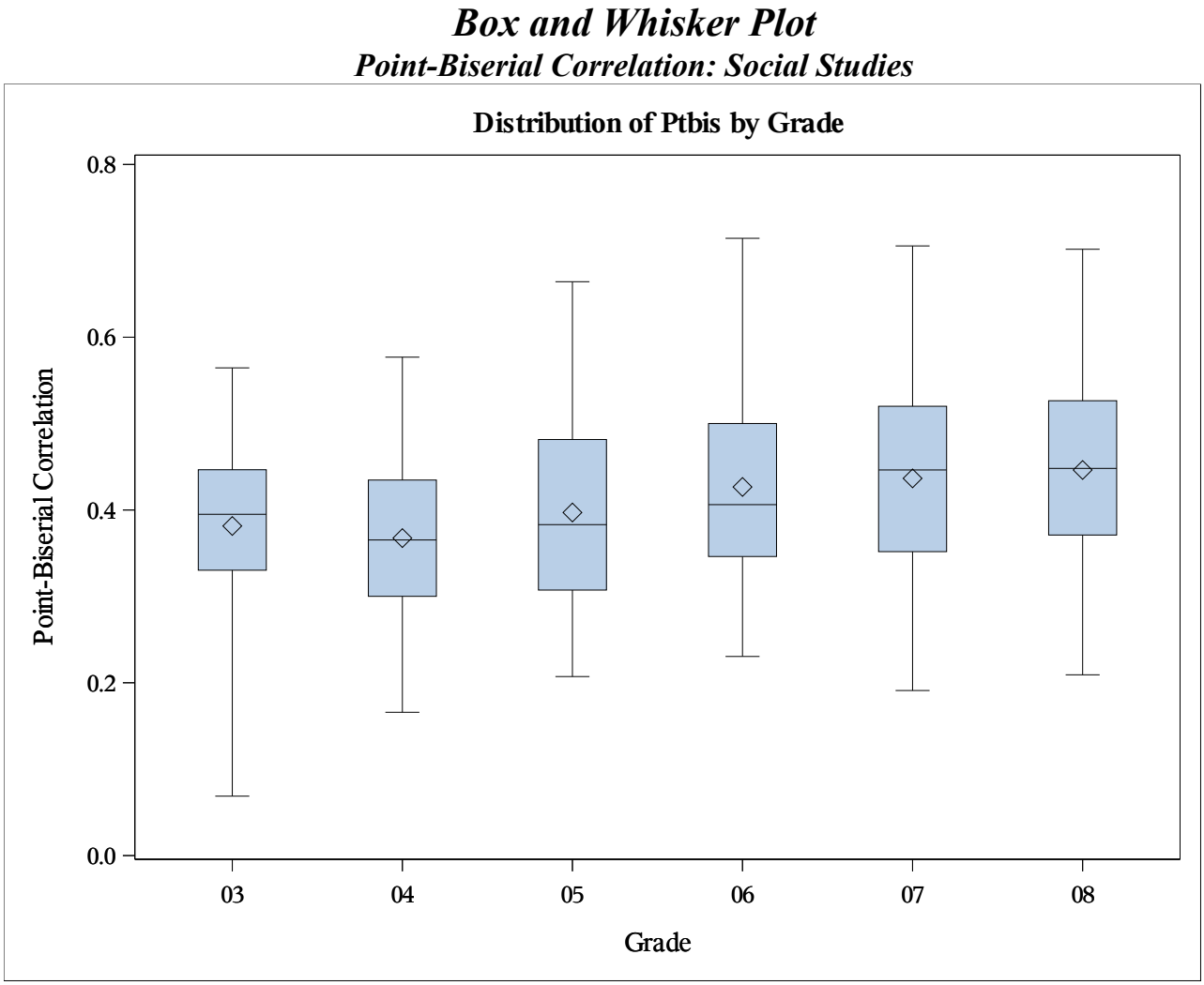


Table C.2.2

Item-Total Correlation Summary by Item Type: Spring 2022 Operational SS G3–8

Grade 3

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.490	0.490	0.527	0.565	0.565
MC	38	0.069	0.327	0.385	0.432	0.536
MS	3	0.403	0.403	0.412	0.453	0.453

Grade 4

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.490	0.490	0.526	0.562	0.562
MC	38	0.166	0.276	0.352	0.413	0.502
MS	3	0.461	0.461	0.493	0.577	0.577

Grade 5

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.564	0.564	0.569	0.573	0.573
ER	2	0.660	0.660	0.662	0.664	0.664
MC	37	0.207	0.296	0.365	0.442	0.502
MS	3	0.326	0.326	0.466	0.554	0.554
TEI	3	0.316	0.316	0.487	0.553	0.553

Grade 6

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.527	0.527	0.568	0.609	0.609
ER	2	0.682	0.682	0.698	0.714	0.714
MC	42	0.230	0.336	0.391	0.438	0.584
MS	6	0.312	0.368	0.404	0.500	0.548
TEI	3	0.478	0.478	0.499	0.619	0.619

Grade 7

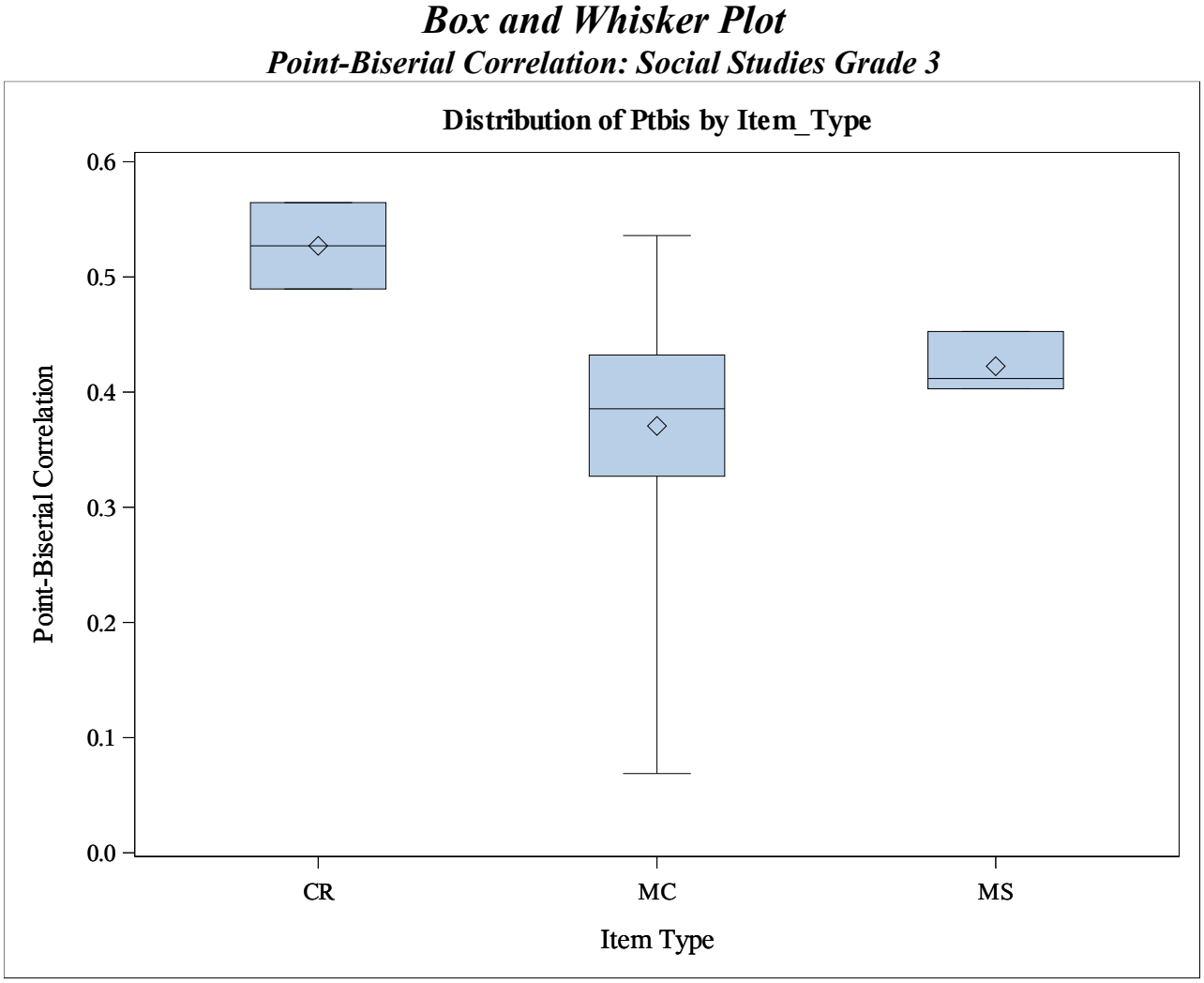
Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.635	0.635	0.643	0.652	0.652
ER	2	0.705	0.705	0.705	0.706	0.706
MC	41	0.191	0.330	0.422	0.462	0.556
MS	5	0.289	0.396	0.463	0.599	0.615
TEI	4	0.446	0.466	0.503	0.569	0.618

Grade 8

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.594	0.594	0.604	0.615	0.615
ER	2	0.696	0.696	0.699	0.702	0.702
MC	40	0.227	0.366	0.419	0.498	0.581
MS	8	0.209	0.352	0.412	0.452	0.532
TEI	3	0.547	0.547	0.556	0.588	0.588

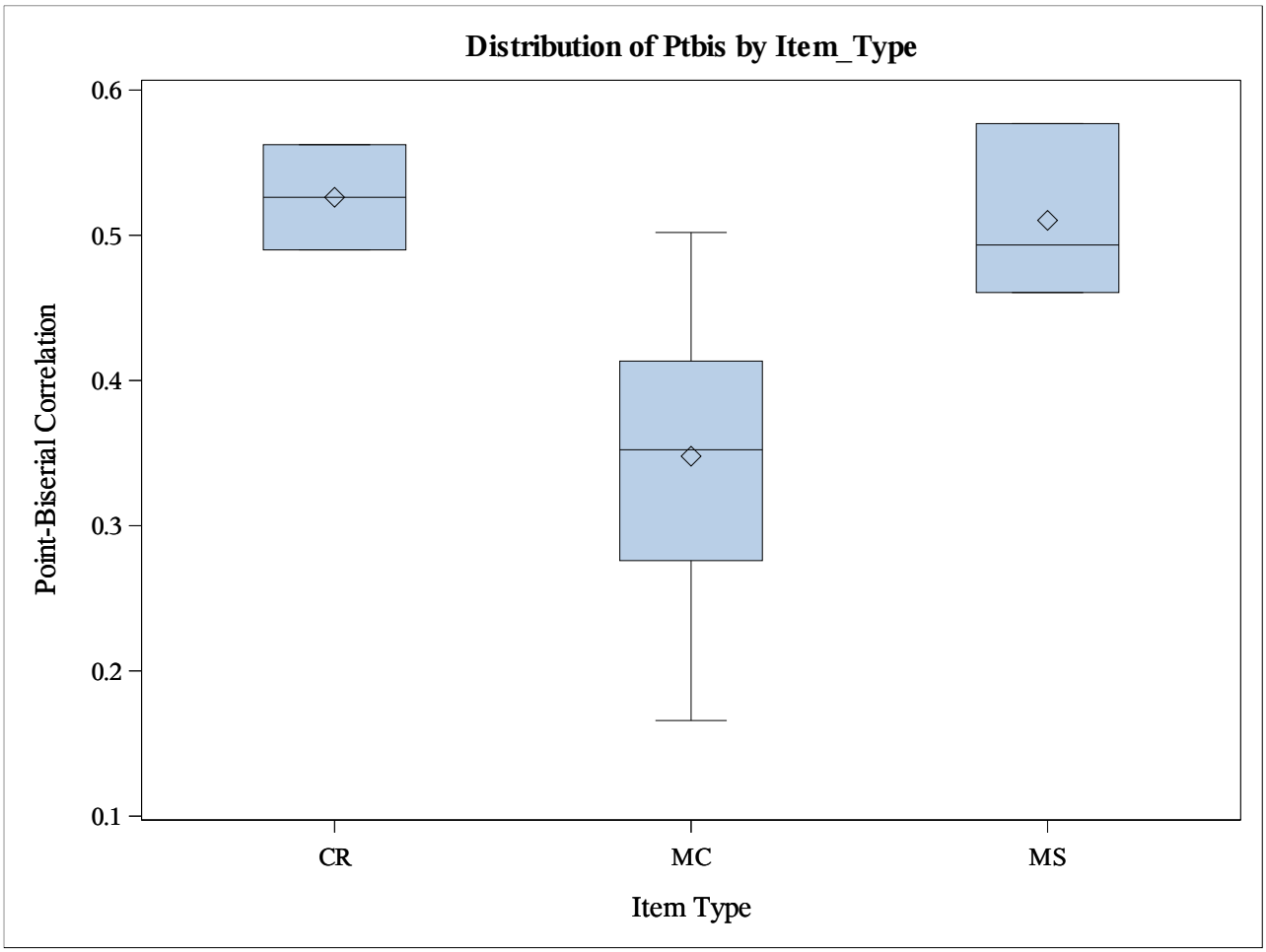
Plot C.2.2

Item-Total Correlation Summary by Item Type: Spring 2022 Operational SS G3–8



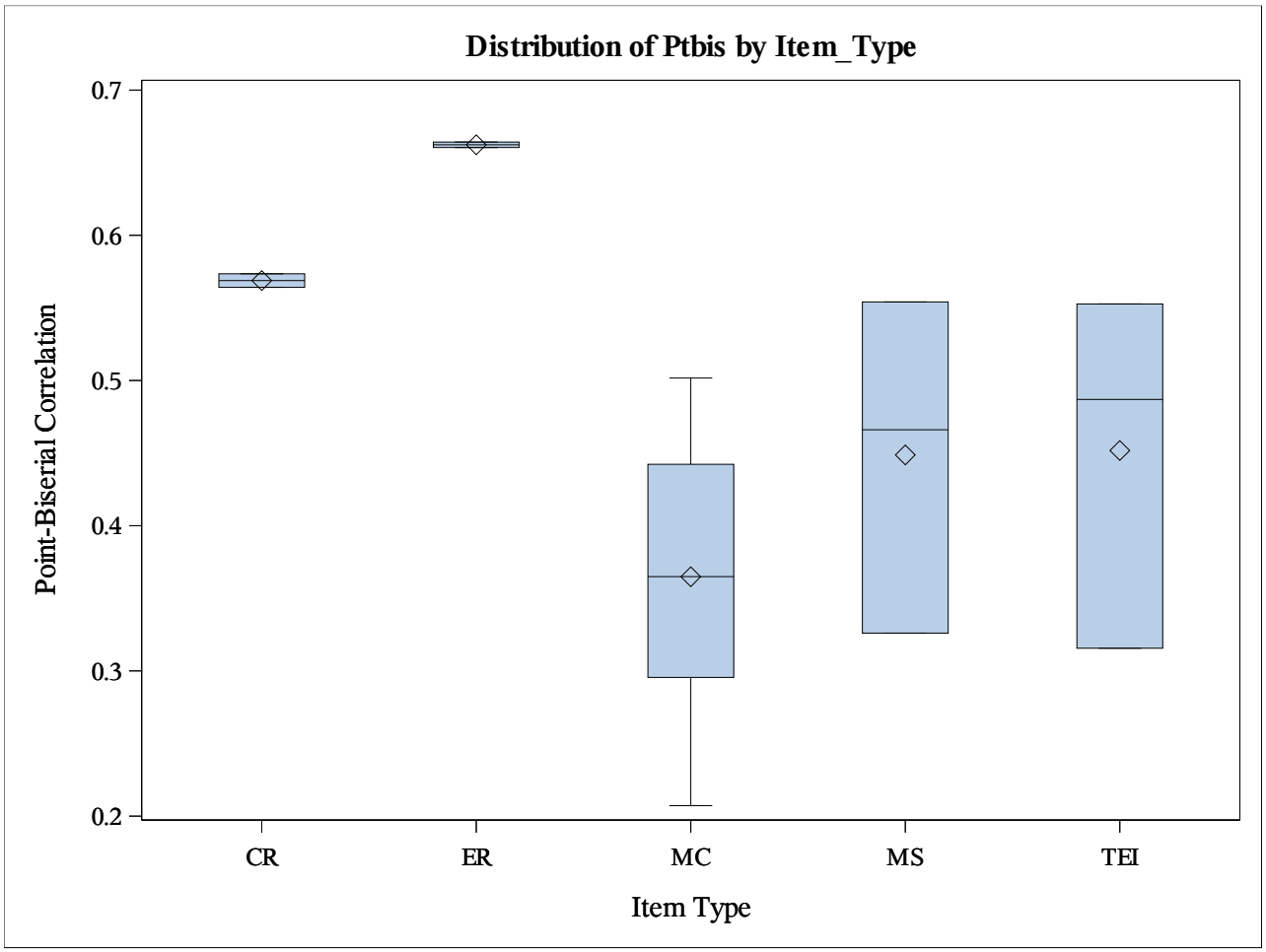
Box and Whisker Plot

Point-Biserial Correlation: Social Studies Grade 4



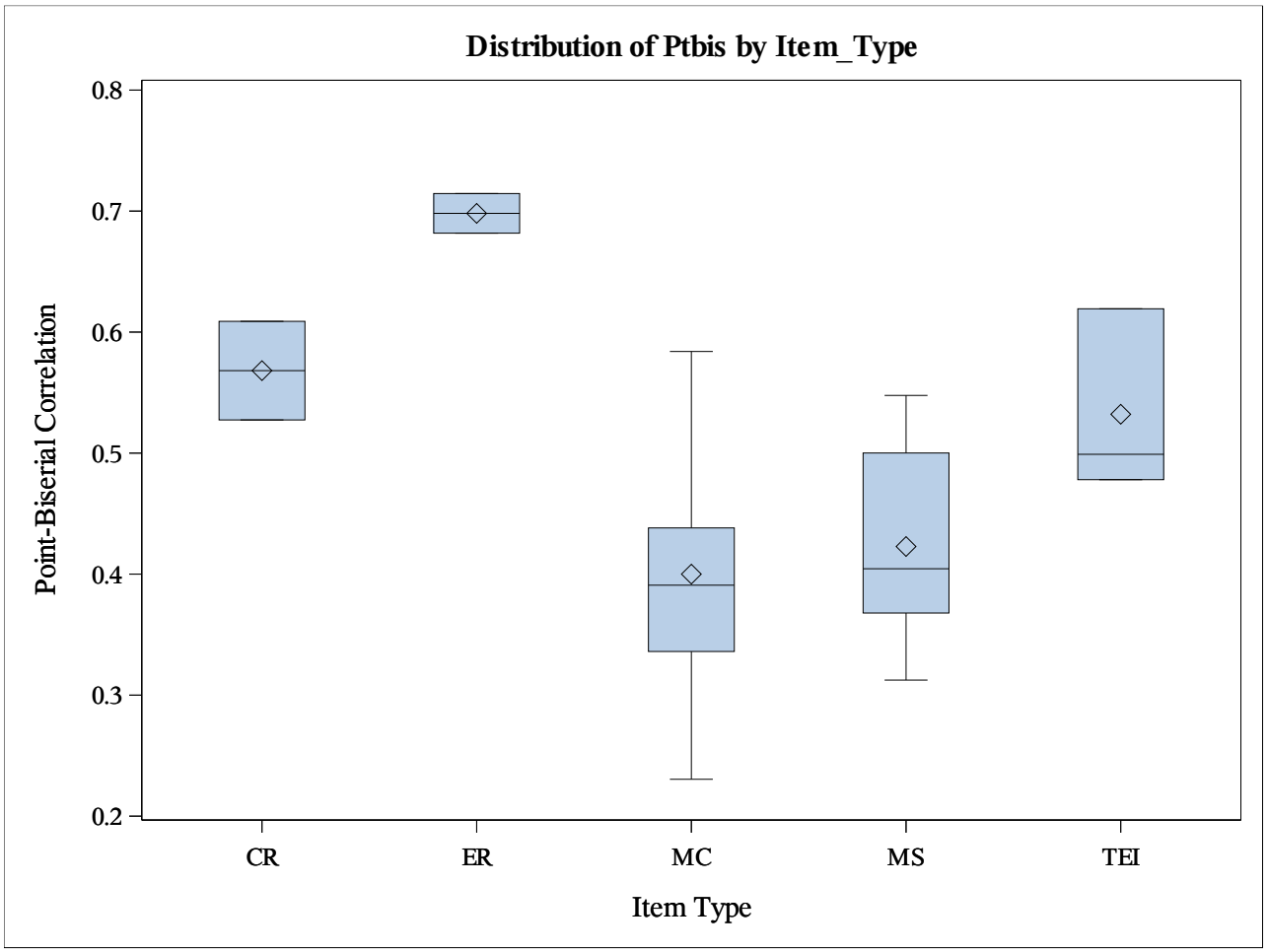
Box and Whisker Plot

Point-Biserial Correlation: Social Studies Grade 5



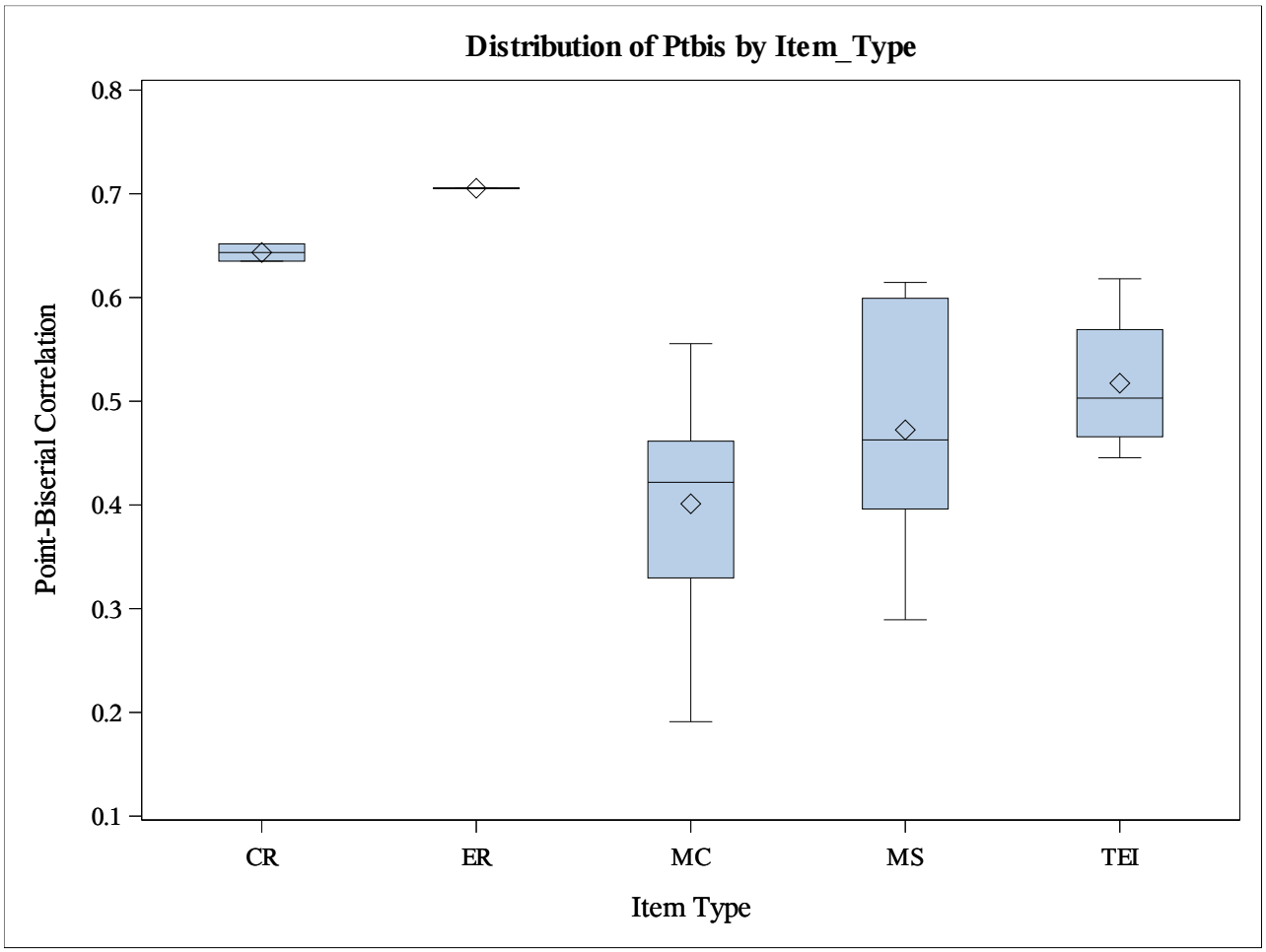
Box and Whisker Plot

Point-Biserial Correlation: Social Studies Grade 6



Box and Whisker Plot

Point-Biserial Correlation: Social Studies Grade 7



Box and Whisker Plot

Point-Biserial Correlation: Social Studies Grade 8

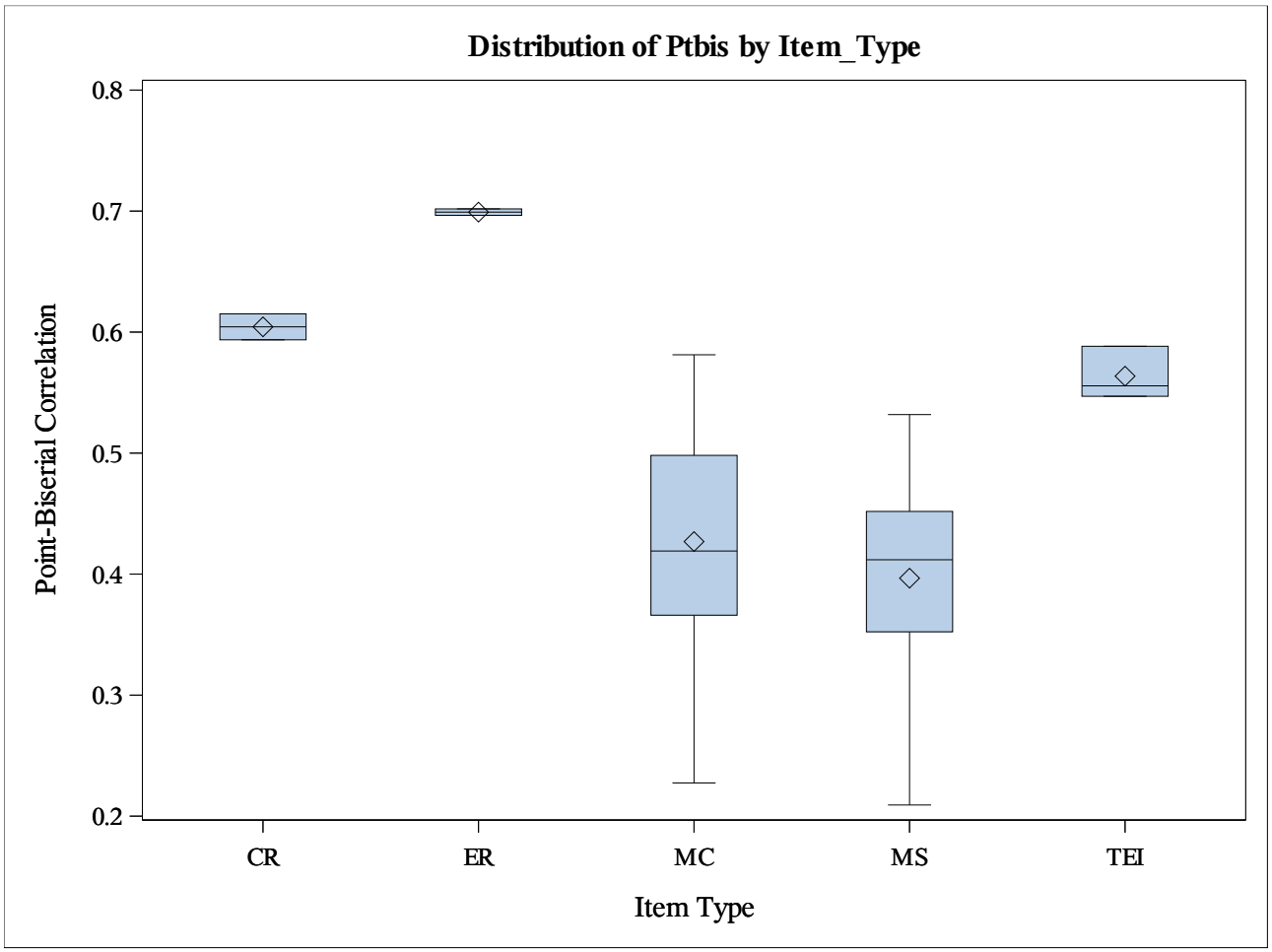


Table C.3.1

Corrected Point-Biserial Correlation Summary by Grade: Spring 2022 Operational SS G3–8*

Grade	No. of Items	r < 0	0.0 ≤ r < 0.2	0.2 ≤ r < 0.3	0.3 ≤ r < 0.4	0.4 ≤ r < 0.5	r ≥ 0.5
3	43	0	4	9	20	10	0
4	43	0	8	12	15	6	2
5	47	0	6	11	12	13	5
6	55	0	1	10	22	13	9
7	54	0	2	10	13	17	12
8	55	0	2	8	15	17	13

* Corrected point-biserial correlation, which is slightly more robust than point-biserial correlation, calculates the relationship between the item score and the total test score after removing the item score from the total test score.

Plot C.3.1

Corrected Point-Biserial Correlation Summary by Grade: Spring 2022 Operational SS G3–8*

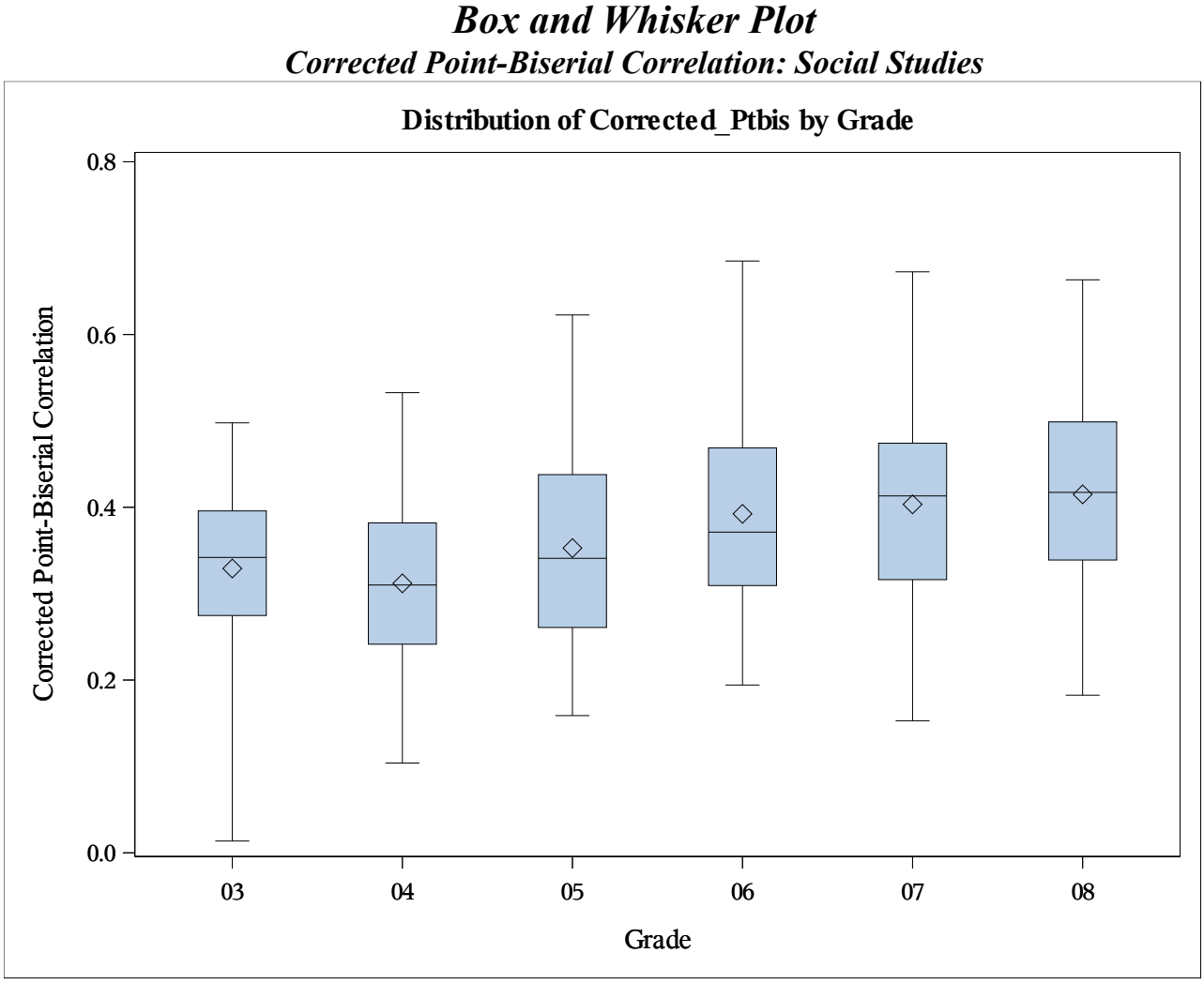


Table C.3.2

Corrected Point-Biserial Correlation Summary by Item Type: Spring 2022 Operational SS G3–8*

Grade 3

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.426	0.426	0.462	0.498	0.498
MC	38	0.014	0.272	0.333	0.381	0.492
MS	3	0.351	0.351	0.363	0.407	0.407

Grade 4

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.432	0.432	0.466	0.500	0.500
MC	38	0.104	0.217	0.295	0.362	0.454
MS	3	0.415	0.415	0.450	0.533	0.533

Grade 5

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.514	0.514	0.522	0.530	0.530
ER	2	0.618	0.618	0.621	0.623	0.623
MC	37	0.159	0.248	0.320	0.402	0.461
MS	3	0.293	0.293	0.424	0.521	0.521
TEI	3	0.273	0.273	0.438	0.492	0.492

Grade 6

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.493	0.493	0.532	0.571	0.571
ER	2	0.649	0.649	0.667	0.685	0.685
MC	42	0.194	0.300	0.356	0.406	0.556
MS	6	0.277	0.336	0.370	0.469	0.519
TEI	3	0.436	0.436	0.457	0.580	0.580

Grade 7

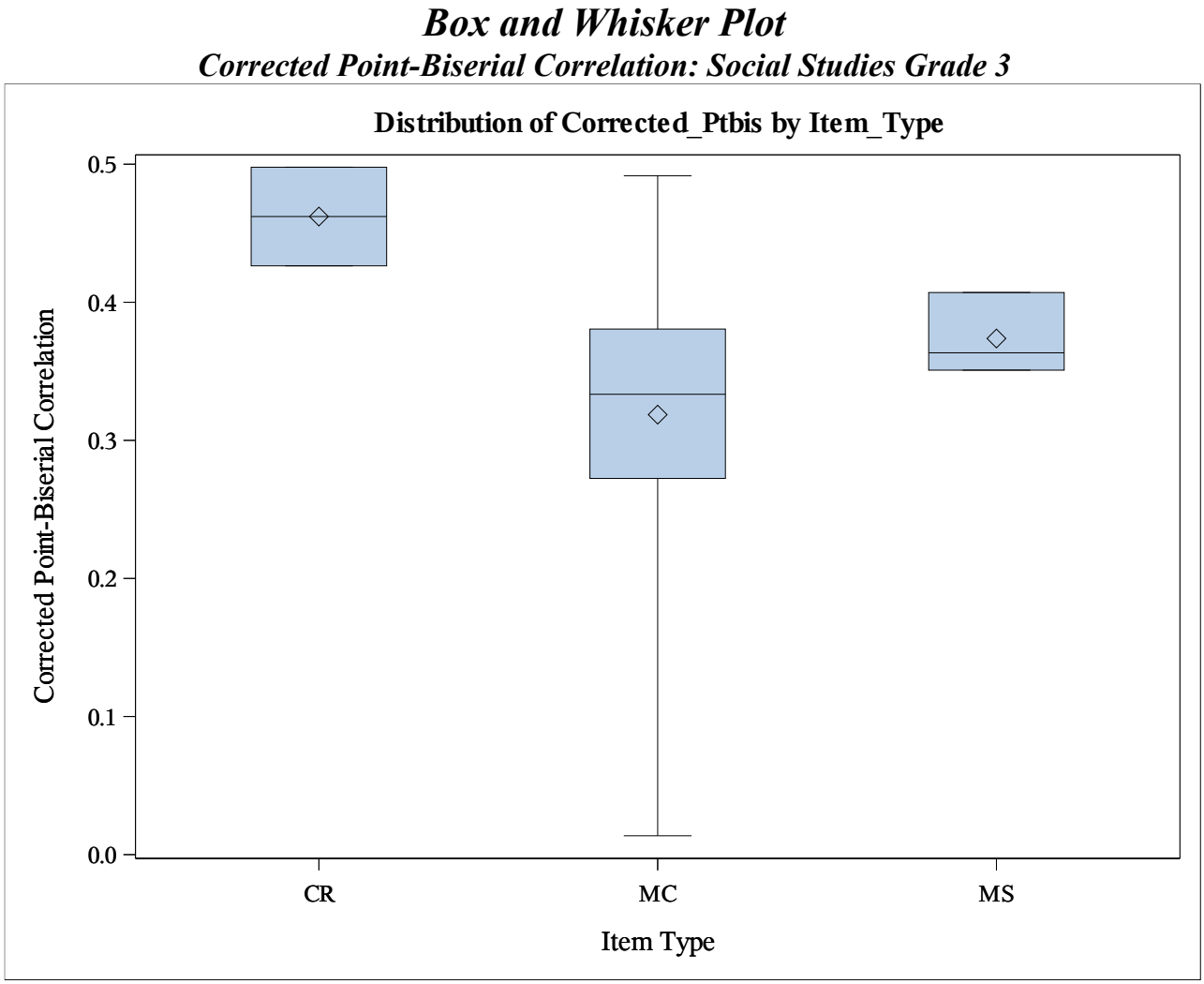
Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.599	0.599	0.609	0.618	0.618
ER	2	0.672	0.672	0.672	0.673	0.673
MC	41	0.153	0.295	0.389	0.430	0.528
MS	5	0.259	0.365	0.434	0.573	0.589
TEI	4	0.412	0.428	0.459	0.527	0.580

Grade 8

Type	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2	0.558	0.558	0.567	0.576	0.576
ER	2	0.657	0.657	0.660	0.663	0.663
MC	40	0.192	0.333	0.387	0.469	0.556
MS	8	0.182	0.323	0.383	0.422	0.504
TEI	3	0.505	0.505	0.521	0.553	0.553

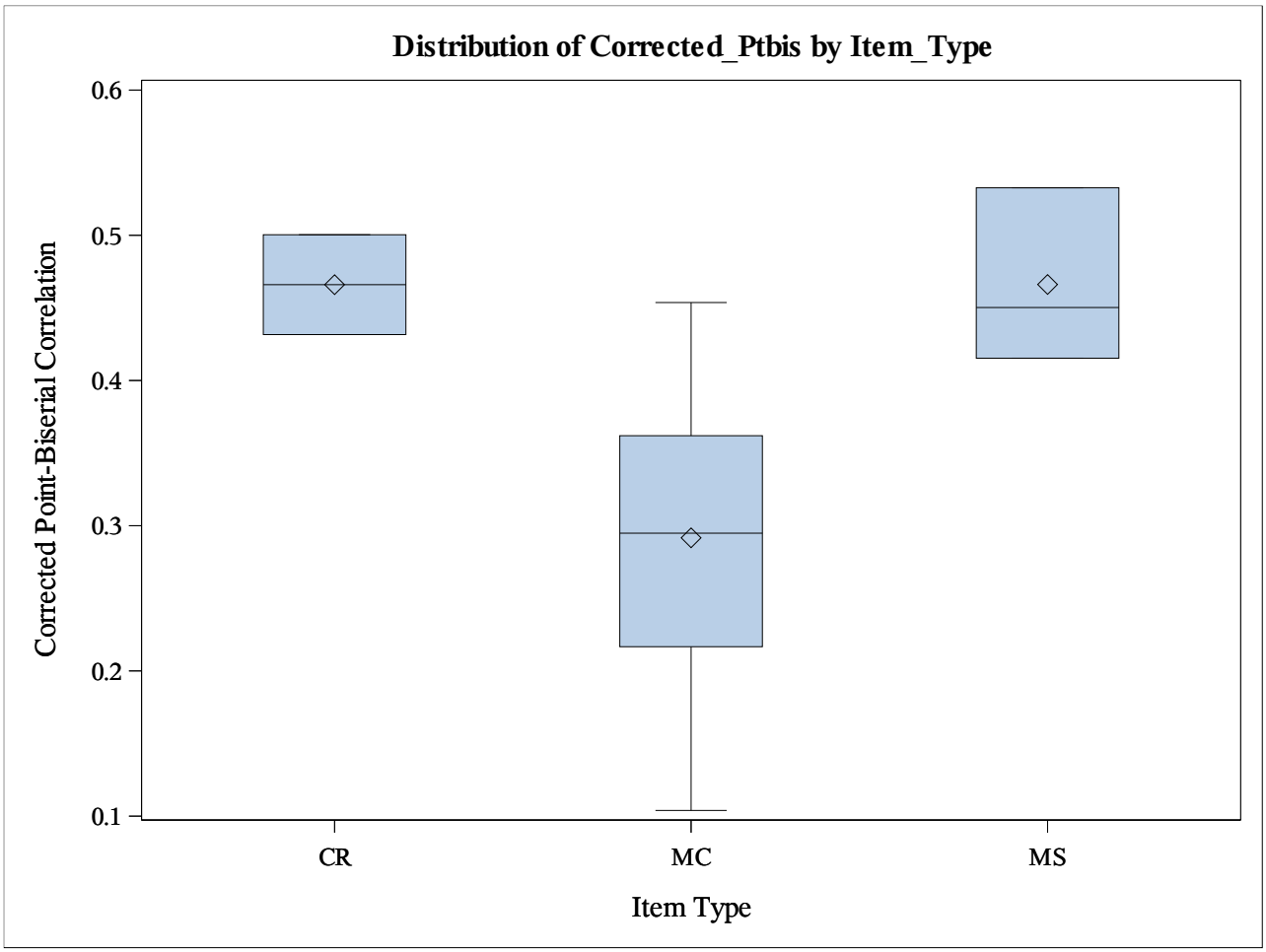
Plot C.3.2

Corrected Point-Biserial Correlation Summary by Item Type: Spring 2022 Operational SS G3–8



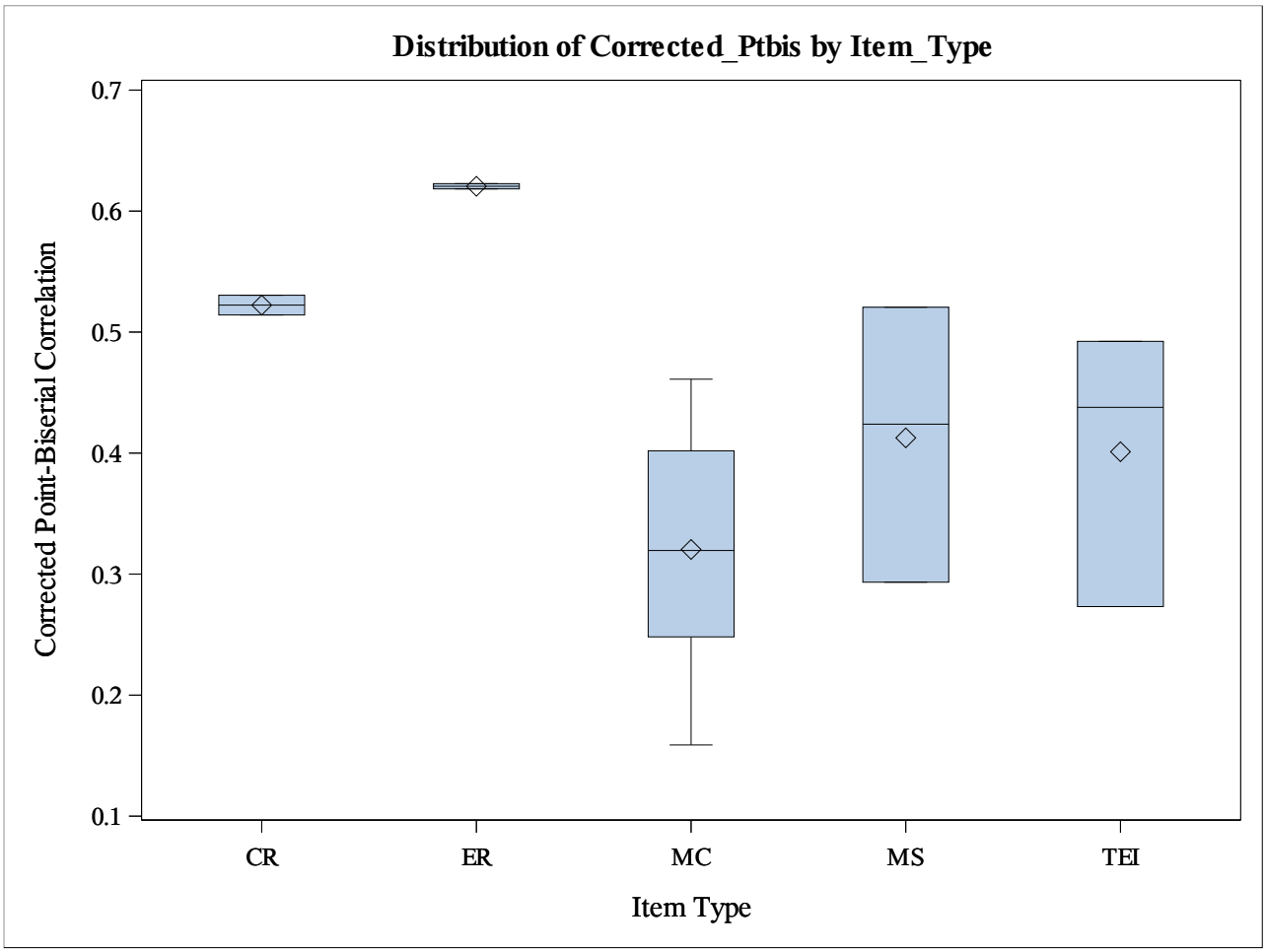
Box and Whisker Plot

Corrected Point-Biserial Correlation: Social Studies Grade 4



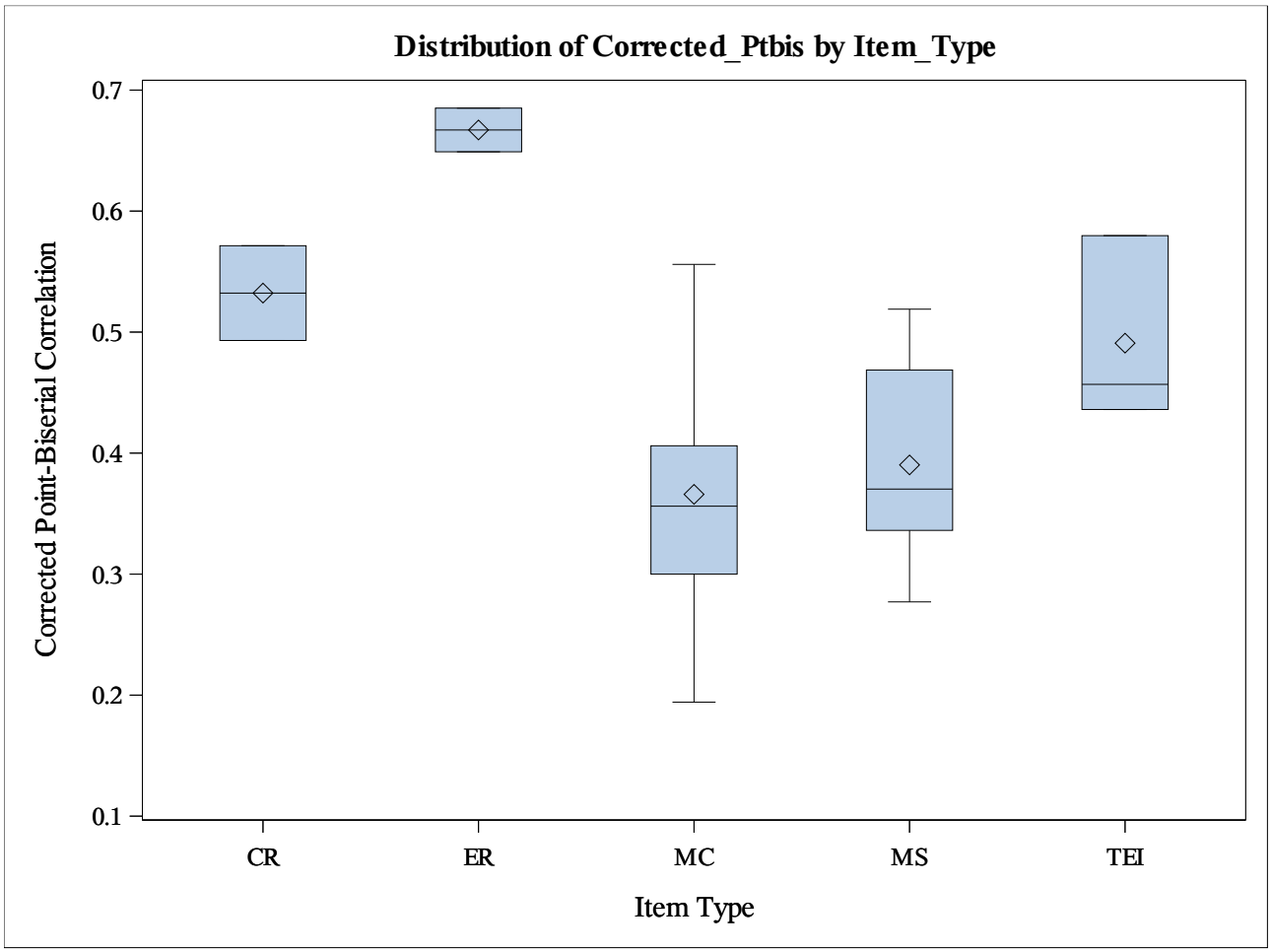
Box and Whisker Plot

Corrected Point-Biserial Correlation: Social Studies Grade 5



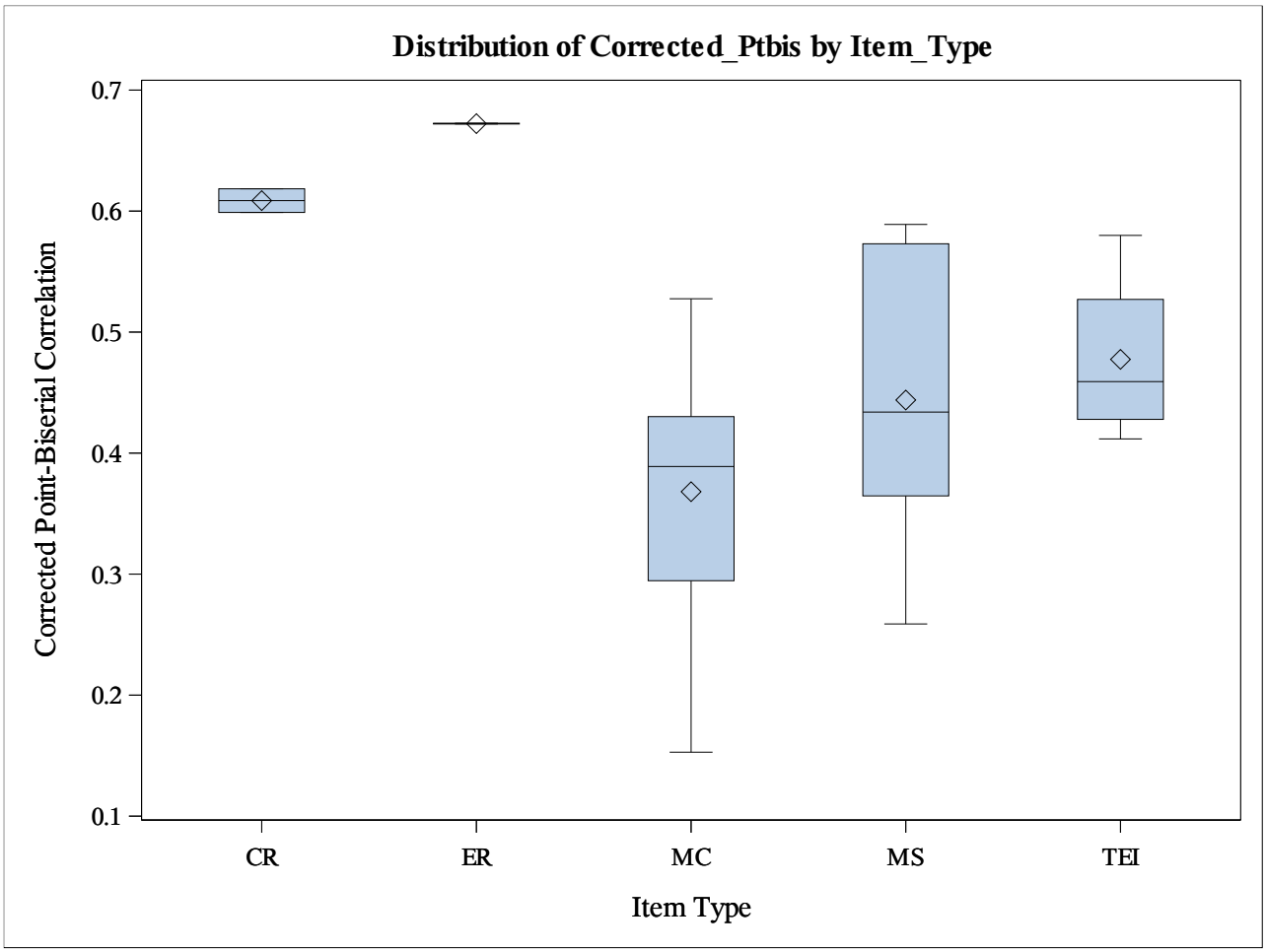
Box and Whisker Plot

Corrected Point-Biserial Correlation: Social Studies Grade 6



Box and Whisker Plot

Corrected Point-Biserial Correlation: Social Studies Grade 7



Box and Whisker Plot

Corrected Point-Biserial Correlation: Social Studies Grade 8

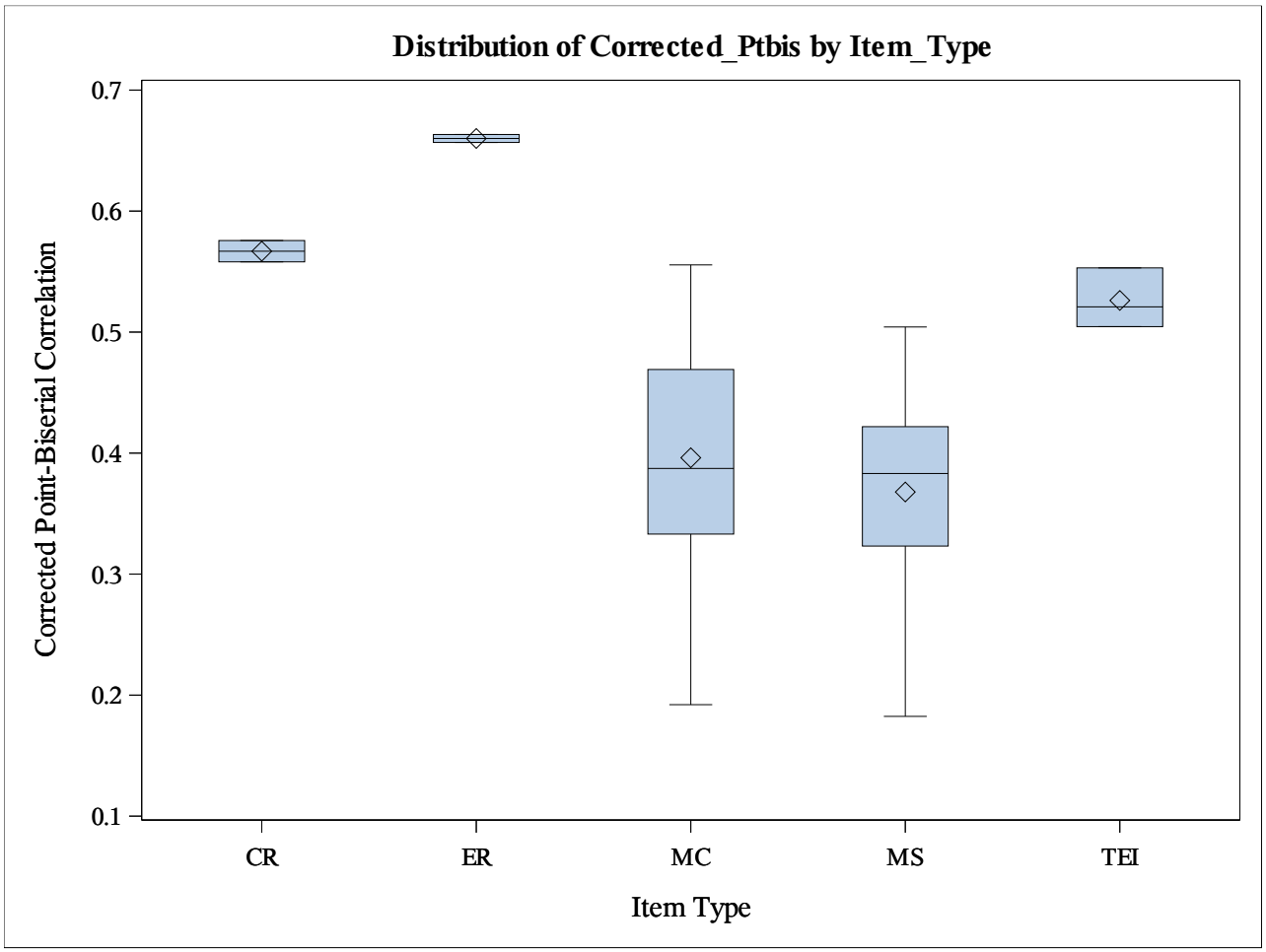


Table C.4.1

Item-Total Correlation Summary by Reporting Category: Spring 2022 Operational SS G3–8

Grade	Reporting Category	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
3	1 History	11	0.138	0.327	0.387	0.416	0.456
	2 Geography	11	0.174	0.286	0.363	0.419	0.565
	3 Civics	10	0.351	0.419	0.459	0.501	0.536
	4 Economics	11	0.069	0.314	0.393	0.432	0.513
4	1 History	11	0.300	0.346	0.365	0.457	0.502
	2 Geography	11	0.166	0.258	0.389	0.413	0.577
	3 Civics	10	0.214	0.241	0.408	0.470	0.562
	4 Economics	11	0.227	0.260	0.349	0.392	0.449
5	1 History	23	0.237	0.280	0.349	0.447	0.573
	2 Geography	7	0.332	0.352	0.372	0.466	0.482
	3 Civics	11	0.207	0.296	0.421	0.500	0.664
	4 Economics	6	0.316	0.383	0.481	0.502	0.564
6	1 History	28	0.315	0.372	0.413	0.522	0.609
	2 Geography	12	0.255	0.330	0.348	0.398	0.494
	3 Civics	8	0.230	0.314	0.481	0.651	0.714
	4 Economics	7	0.265	0.375	0.417	0.499	0.548
7	1 History	28	0.191	0.363	0.447	0.489	0.706
	2 Geography	8	0.289	0.302	0.389	0.454	0.615
	3 Civics	12	0.281	0.349	0.464	0.527	0.537
	4 Economics	6	0.296	0.390	0.456	0.534	0.618
8	1 History	29	0.285	0.371	0.411	0.522	0.615
	2 Geography	8	0.397	0.455	0.511	0.537	0.581
	3 Civics	10	0.227	0.353	0.445	0.522	0.702
	4 Economics	8	0.209	0.354	0.434	0.502	0.563

Table C.4.2

Item-Total Correlation Summary by Reporting Category and Item Type: Spring 2022 Operational SS G3–8

Grade 3

Type	Reporting Category	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2 Geography	1	0.565	0.565	0.565	0.565	0.565
	3 Civics	1	0.490	0.490	0.490	0.490	0.490
MC	1 History	10	0.138	0.327	0.385	0.416	0.456
	2 Geography	10	0.174	0.286	0.347	0.407	0.446
	3 Civics	8	0.351	0.394	0.456	0.502	0.536
	4 Economics	10	0.069	0.314	0.385	0.432	0.513
MS	1 History	1	0.403	0.403	0.403	0.403	0.403
	3 Civics	1	0.453	0.453	0.453	0.453	0.453
	4 Economics	1	0.412	0.412	0.412	0.412	0.412

Grade 4

Type	Reporting Category	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	2 Geography	1	0.490	0.490	0.490	0.490	0.490
	3 Civics	1	0.562	0.562	0.562	0.562	0.562
MC	1 History	10	0.300	0.346	0.360	0.435	0.502
	2 Geography	9	0.166	0.258	0.358	0.394	0.413
	3 Civics	8	0.214	0.232	0.355	0.441	0.470
	4 Economics	11	0.227	0.260	0.349	0.392	0.449
MS	1 History	1	0.461	0.461	0.461	0.461	0.461
	2 Geography	1	0.577	0.577	0.577	0.577	0.577
	3 Civics	1	0.493	0.493	0.493	0.493	0.493

Grade 5

Type	Reporting Category	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	1 History	1	0.573	0.573	0.573	0.573	0.573
	4 Economics	1	0.564	0.564	0.564	0.564	0.564
ER	3 Civics	2	0.660	0.660	0.662	0.664	0.664
MC	1 History	19	0.237	0.252	0.331	0.421	0.498
	2 Geography	6	0.332	0.352	0.367	0.453	0.482
	3 Civics	9	0.207	0.296	0.407	0.437	0.500
	4 Economics	3	0.383	0.383	0.476	0.502	0.502
MS	1 History	2	0.326	0.326	0.440	0.554	0.554
	2 Geography	1	0.466	0.466	0.466	0.466	0.466
TEI	1 History	1	0.553	0.553	0.553	0.553	0.553
	4 Economics	2	0.316	0.316	0.401	0.487	0.487

Grade 6

Type	Reporting Category	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	1 History	2	0.527	0.527	0.568	0.609	0.609
ER	3 Civics	2	0.682	0.682	0.698	0.714	0.714
MC	1 History	22	0.315	0.361	0.408	0.516	0.584
	2 Geography	10	0.255	0.332	0.348	0.405	0.494
	3 Civics	5	0.230	0.311	0.318	0.436	0.526
	4 Economics	5	0.265	0.375	0.413	0.417	0.437
MS	1 History	3	0.403	0.403	0.406	0.500	0.500
	2 Geography	2	0.312	0.312	0.340	0.368	0.368
	4 Economics	1	0.548	0.548	0.548	0.548	0.548
TEI	1 History	1	0.478	0.478	0.478	0.478	0.478
	3 Civics	1	0.619	0.619	0.619	0.619	0.619
	4 Economics	1	0.499	0.499	0.499	0.499	0.499

Grade 7

Type	Reporting Category	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	1 History	2	0.635	0.635	0.643	0.652	0.652
ER	1 History	2	0.705	0.705	0.705	0.706	0.706
MC	1 History	21	0.191	0.327	0.381	0.456	0.556
	2 Geography	5	0.297	0.307	0.382	0.450	0.458
	3 Civics	10	0.281	0.347	0.454	0.530	0.537
	4 Economics	5	0.296	0.390	0.443	0.469	0.534
MS	1 History	2	0.463	0.463	0.531	0.599	0.599
	2 Geography	3	0.289	0.289	0.396	0.615	0.615
TEI	1 History	1	0.486	0.486	0.486	0.486	0.486
	3 Civics	2	0.446	0.446	0.483	0.520	0.520
	4 Economics	1	0.618	0.618	0.618	0.618	0.618

Grade 8

Type	Reporting Category	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	1 History	2	0.594	0.594	0.604	0.615	0.615
ER	3 Civics	2	0.696	0.696	0.699	0.702	0.702
MC	1 History	21	0.285	0.322	0.390	0.464	0.531
	2 Geography	6	0.462	0.503	0.523	0.548	0.581
	3 Civics	7	0.227	0.353	0.425	0.477	0.522
	4 Economics	6	0.322	0.385	0.440	0.538	0.563
MS	1 History	3	0.385	0.385	0.426	0.532	0.532
	2 Geography	2	0.397	0.397	0.423	0.448	0.448
	3 Civics	1	0.319	0.319	0.319	0.319	0.319
	4 Economics	2	0.209	0.209	0.332	0.455	0.455
TEI	1 History	3	0.547	0.547	0.556	0.588	0.588

Table C.5.1.1

IRT-A Parameter Summary by Reporting Category: Grade 3

IRT-a Range	1 History	2 Geography	3 Civics	4 Economics	Number of Items
a < 0.0	0	0	0	0	0
0.0 <= a < 0.2	0	0	0	0	0
0.2 <= a < 0.4	0	1	0	1	2
0.4 <= a < 0.6	4	2	3	1	10
0.6 <= a < 0.8	1	5	2	2	10
0.8 <= a < 1.0	5	1	2	4	12
1.0 <= a < 1.2	0	2	3	1	6
1.2 <= a < 1.4	1	0	0	2	3
1.4 <= a < 1.6	0	0	0	0	0
1.6 <= a < 1.8	0	0	0	0	0
1.8 <= a < 2.0	0	0	0	0	0
2.0 <= a	0	0	0	0	0
Minimum	0.49	0.30	0.44	0.34	0.30
Maximum	1.22	1.07	1.15	1.30	1.30
Mean	0.80	0.72	0.83	0.87	0.80
SD	0.25	0.24	0.27	0.29	0.26
Number of Items	11	11	10	11	43

Table C.5.2.1

IRT-B Parameter Summary by Reporting Category: Grade 3

IRT-b Range	1 History	2 Geography	3 Civics	4 Economics	Number of Items
b < -3.5	0	0	0	0	0
-3.5 <= b < -3.0	0	0	0	0	0
-3.0 <= b < -2.5	0	0	0	0	0
-2.5 <= b < -2.0	0	0	0	0	0
-2.0 <= b < -1.5	0	0	0	0	0
-1.5 <= b < -1.0	0	0	1	0	1
-1.0 <= b < -0.5	1	0	0	2	3
-0.5 <= b < 0.0	0	0	3	1	4
0.0 <= b < 0.5	0	1	2	1	4
0.5 <= b < 1.0	5	6	1	2	14
1.0 <= b < 1.5	3	2	2	3	10
1.5 <= b < 2.0	0	1	1	1	3
2.0 <= b < 2.5	1	0	0	0	1
2.5 <= b < 3.0	1	1	0	1	3
3.0 <= b < 3.5	0	0	0	0	0
3.5 <= b	0	0	0	0	0
Minimum	-0.66	0.06	-1.32	-0.77	-1.32
Maximum	2.71	2.82	1.61	2.78	2.82
Mean	1.05	1.13	0.32	0.75	0.82
SD	0.90	0.72	0.89	1.06	0.92
Number of Items	11	11	10	11	43

Table C.5.1.2

IRT-A Parameter Summary by Reporting Category: Grade 4

IRT-a Range	1 History	2 Geography	3 Civics	4 Economics	Number of Items
a < 0.0	0	0	0	0	0
0.0 <= a < 0.2	0	0	0	0	0
0.2 <= a < 0.4	0	1	0	1	2
0.4 <= a < 0.6	4	2	3	1	10
0.6 <= a < 0.8	1	5	2	2	10
0.8 <= a < 1.0	5	1	2	4	12
1.0 <= a < 1.2	0	2	3	1	6
1.2 <= a < 1.4	1	0	0	2	3
1.4 <= a < 1.6	0	0	0	0	0
1.6 <= a < 1.8	0	0	0	0	0
1.8 <= a < 2.0	0	0	0	0	0
2.0 <= a	0	0	0	0	0
Minimum	0.49	0.30	0.44	0.34	0.30
Maximum	1.22	1.07	1.15	1.30	1.30
Mean	0.80	0.72	0.83	0.87	0.80
SD	0.25	0.24	0.27	0.29	0.26
Number of Items	11	11	10	11	43

Table C.5.2.2

IRT-B Parameter Summary by Reporting Category: Grade 4

IRT-b Range	1 History	2 Geography	3 Civics	4 Economics	Number of Items
b < -3.5	0	0	0	0	0
-3.5 <= b < -3.0	0	0	0	0	0
-3.0 <= b < -2.5	0	0	0	0	0
-2.5 <= b < -2.0	0	0	0	0	0
-2.0 <= b < -1.5	0	0	0	0	0
-1.5 <= b < -1.0	0	0	1	0	1
-1.0 <= b < -0.5	1	0	1	0	2
-0.5 <= b < 0.0	1	0	1	1	3
0.0 <= b < 0.5	2	3	3	0	8
0.5 <= b < 1.0	0	1	1	4	6
1.0 <= b < 1.5	6	3	1	3	13
1.5 <= b < 2.0	1	1	0	2	4
2.0 <= b < 2.5	0	2	1	0	3
2.5 <= b < 3.0	0	1	1	1	3
3.0 <= b < 3.5	0	0	0	0	0
3.5 <= b	0	0	0	0	0
Minimum	-0.70	0.23	-1.12	-0.02	-1.12
Maximum	1.54	2.70	2.67	2.58	2.70
Mean	0.75	1.24	0.56	1.19	0.94
SD	0.73	0.80	1.20	0.70	0.89
Number of Items	11	11	10	11	43

Table C.5.1.3

IRT-A Parameter Summary by Reporting Category: Grade 5

IRT-a Range	1 History	2 Geography	3 Civics	4 Economics	Number of Items
a < 0.0	0	0	0	0	0
0.0 <= a < 0.2	0	0	0	0	0
0.2 <= a < 0.4	3	0	1	0	4
0.4 <= a < 0.6	3	3	1	1	8
0.6 <= a < 0.8	5	1	0	2	8
0.8 <= a < 1.0	4	0	2	2	8
1.0 <= a < 1.2	6	2	5	1	14
1.2 <= a < 1.4	2	1	1	0	4
1.4 <= a < 1.6	0	0	1	0	1
1.6 <= a < 1.8	0	0	0	0	0
1.8 <= a < 2.0	0	0	0	0	0
2.0 <= a	0	0	0	0	0
Minimum	0.28	0.40	0.37	0.50	0.28
Maximum	1.31	1.22	1.55	1.03	1.55
Mean	0.80	0.75	1.01	0.78	0.84
SD	0.31	0.32	0.32	0.19	0.31
Number of Items	23	7	11	6	47

Table C.5.2.3

IRT-B Parameter Summary by Reporting Category: Grade 5

IRT-b Range	1 History	2 Geography	3 Civics	4 Economics	Number of Items
b < -3.5	0	0	0	0	0
-3.5 <= b < -3.0	0	0	0	0	0
-3.0 <= b < -2.5	0	0	0	0	0
-2.5 <= b < -2.0	0	0	0	0	0
-2.0 <= b < -1.5	0	0	0	0	0
-1.5 <= b < -1.0	0	0	0	0	0
-1.0 <= b < -0.5	0	1	0	0	1
-0.5 <= b < 0.0	2	2	0	0	4
0.0 <= b < 0.5	1	3	1	1	6
0.5 <= b < 1.0	2	0	0	1	3
1.0 <= b < 1.5	7	0	4	1	12
1.5 <= b < 2.0	7	1	3	3	14
2.0 <= b < 2.5	4	0	2	0	6
2.5 <= b < 3.0	0	0	1	0	1
3.0 <= b < 3.5	0	0	0	0	0
3.5 <= b	0	0	0	0	0
Minimum	-0.28	-0.63	0.35	0.47	-0.63
Maximum	2.33	1.78	2.71	1.97	2.71
Mean	1.32	0.26	1.62	1.32	1.23
SD	0.70	0.76	0.71	0.62	0.81
Number of Items	23	7	11	6	47

Table C.5.1.4

IRT-A Parameter Summary by Reporting Category: Grade 6

IRT-a Range	1 History	2 Geography	3 Civics	4 Economics	Number of Items
a < 0.0	0	0	0	0	0
0.0 <= a < 0.2	0	0	0	0	0
0.2 <= a < 0.4	0	2	0	0	2
0.4 <= a < 0.6	6	2	0	1	9
0.6 <= a < 0.8	6	4	2	3	15
0.8 <= a < 1.0	5	2	0	3	10
1.0 <= a < 1.2	5	2	3	0	10
1.2 <= a < 1.4	4	0	3	0	7
1.4 <= a < 1.6	2	0	0	0	2
1.6 <= a < 1.8	0	0	0	0	0
1.8 <= a < 2.0	0	0	0	0	0
2.0 <= a	0	0	0	0	0
Minimum	0.40	0.27	0.61	0.48	0.27
Maximum	1.48	1.05	1.35	0.88	1.48
Mean	0.90	0.68	1.06	0.72	0.85
SD	0.33	0.24	0.27	0.15	0.31
Number of Items	28	12	8	7	55

Table C.5.2.4

IRT-B Parameter Summary by Reporting Category: Grade 6

IRT-b Range	1 History	2 Geography	3 Civics	4 Economics	Number of Items
b < -3.5	0	0	0	0	0
-3.5 <= b < -3.0	0	0	0	0	0
-3.0 <= b < -2.5	0	0	0	0	0
-2.5 <= b < -2.0	0	0	0	0	0
-2.0 <= b < -1.5	0	0	0	0	0
-1.5 <= b < -1.0	1	0	0	0	1
-1.0 <= b < -0.5	3	0	0	0	3
-0.5 <= b < 0.0	5	2	2	0	9
0.0 <= b < 0.5	8	2	0	4	14
0.5 <= b < 1.0	7	3	1	1	12
1.0 <= b < 1.5	3	5	2	1	11
1.5 <= b < 2.0	1	0	3	1	5
2.0 <= b < 2.5	0	0	0	0	0
2.5 <= b < 3.0	0	0	0	0	0
3.0 <= b < 3.5	0	0	0	0	0
3.5 <= b	0	0	0	0	0
Minimum	-1.34	-0.39	-0.49	0.19	-1.34
Maximum	1.63	1.40	2.00	1.80	2.00
Mean	0.29	0.72	1.01	0.71	0.54
SD	0.72	0.53	0.93	0.60	0.73
Number of Items	28	12	8	7	55

Table C.5.1.5

IRT-A Parameter Summary by Reporting Category: Grade 7

IRT-a Range	1 History	2 Geography	3 Civics	4 Economics	Number of Items
a < 0.0	0	0	0	0	0
0.0 <= a < 0.2	0	0	0	0	0
0.2 <= a < 0.4	0	0	0	0	0
0.4 <= a < 0.6	2	2	5	0	9
0.6 <= a < 0.8	8	3	1	2	14
0.8 <= a < 1.0	4	1	3	2	10
1.0 <= a < 1.2	9	1	2	1	13
1.2 <= a < 1.4	2	1	0	1	4
1.4 <= a < 1.6	2	0	0	0	2
1.6 <= a < 1.8	0	0	1	0	1
1.8 <= a < 2.0	0	0	0	0	0
2.0 <= a	1	0	0	0	1
Minimum	0.45	0.49	0.45	0.69	0.45
Maximum	2.21	1.26	1.77	1.39	2.21
Mean	1.00	0.80	0.84	0.92	0.93
SD	0.37	0.28	0.39	0.26	0.35
Number of Items	28	8	12	6	54

Table C.5.2.5

IRT-B Parameter Summary by Reporting Category: Grade 7

IRT-b Range	1 History	2 Geography	3 Civics	4 Economics	Number of Items
b < -3.5	0	0	0	0	0
-3.5 <= b < -3.0	0	0	0	0	0
-3.0 <= b < -2.5	0	0	0	0	0
-2.5 <= b < -2.0	0	0	0	0	0
-2.0 <= b < -1.5	0	0	0	0	0
-1.5 <= b < -1.0	0	0	0	0	0
-1.0 <= b < -0.5	1	0	1	0	2
-0.5 <= b < 0.0	5	2	2	1	10
0.0 <= b < 0.5	3	2	4	2	11
0.5 <= b < 1.0	8	1	2	2	13
1.0 <= b < 1.5	6	1	2	0	9
1.5 <= b < 2.0	5	1	1	1	8
2.0 <= b < 2.5	0	1	0	0	1
2.5 <= b < 3.0	0	0	0	0	0
3.0 <= b < 3.5	0	0	0	0	0
3.5 <= b	0	0	0	0	0
Minimum	-0.67	-0.15	-0.56	-0.49	-0.67
Maximum	1.78	2.06	1.54	1.74	2.06
Mean	0.74	0.78	0.49	0.52	0.67
SD	0.71	0.82	0.68	0.77	0.72
Number of Items	28	8	12	6	54

Table C.5.1.6

IRT-A Parameter Summary by Reporting Category: Grade 8

IRT-a Range	1 History	2 Geography	3 Civics	4 Economics	Number of Items
a < 0.0	0	0	0	0	0
0.0 <= a < 0.2	0	0	0	0	0
0.2 <= a < 0.4	1	0	0	1	2
0.4 <= a < 0.6	5	2	1	1	9
0.6 <= a < 0.8	9	0	3	4	16
0.8 <= a < 1.0	4	1	3	0	8
1.0 <= a < 1.2	4	4	2	2	12
1.2 <= a < 1.4	4	0	1	0	5
1.4 <= a < 1.6	2	1	0	0	3
1.6 <= a < 1.8	0	0	0	0	0
1.8 <= a < 2.0	0	0	0	0	0
2.0 <= a	0	0	0	0	0
Minimum	0.35	0.55	0.57	0.30	0.30
Maximum	1.44	1.48	1.31	1.14	1.48
Mean	0.87	1.02	0.90	0.74	0.88
SD	0.31	0.33	0.23	0.26	0.29
Number of Items	29	8	10	8	55

Table C.5.2.6

IRT-B Parameter Summary by Reporting Category: Grade 8

IRT-b Range	1 History	2 Geography	3 Civics	4 Economics	Number of Items
b < -3.5	0	0	0	0	0
-3.5 <= b < -3.0	0	0	0	0	0
-3.0 <= b < -2.5	0	0	0	0	0
-2.5 <= b < -2.0	0	0	0	0	0
-2.0 <= b < -1.5	0	0	0	0	0
-1.5 <= b < -1.0	1	0	0	0	1
-1.0 <= b < -0.5	0	0	0	0	0
-0.5 <= b < 0.0	3	4	0	2	9
0.0 <= b < 0.5	11	2	5	3	21
0.5 <= b < 1.0	5	2	1	2	10
1.0 <= b < 1.5	8	0	2	0	10
1.5 <= b < 2.0	1	0	2	0	3
2.0 <= b < 2.5	0	0	0	0	0
2.5 <= b < 3.0	0	0	0	0	0
3.0 <= b < 3.5	0	0	0	1	1
3.5 <= b	0	0	0	0	0
Minimum	-1.35	-0.24	0.18	-0.25	-1.35
Maximum	1.70	0.76	1.85	3.42	3.42
Mean	0.56	0.13	0.82	0.65	0.56
SD	0.61	0.38	0.63	1.19	0.71
Number of Items	29	8	10	8	55

Table C.5.3

IRT Parameter Summary: Spring 2022 Operational SS G3–8

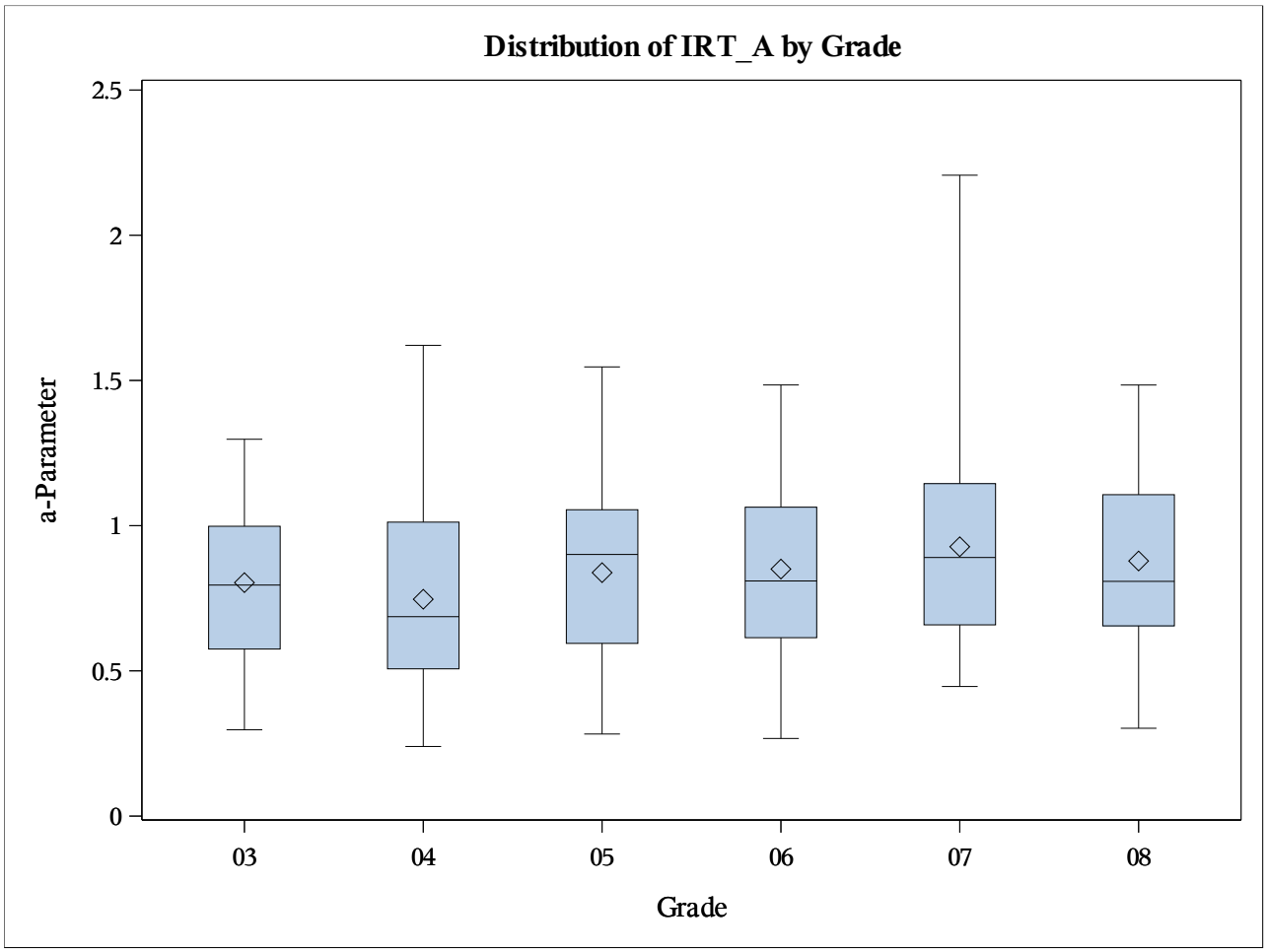
Grade	Parameter	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
3	a	43	0.297	0.575	0.795	0.998	1.298
	b	43	-1.316	0.418	0.818	1.255	2.816
	c	41	0.001	0.125	0.196	0.234	0.381
4	a	43	0.24	0.507	0.687	1.012	1.621
	b	43	-1.116	0.405	1.028	1.413	2.699
	c	41	0.013	0.146	0.199	0.252	0.356
5	a	47	0.283	0.595	0.901	1.055	1.547
	b	47	-0.634	0.701	1.322	1.775	2.715
	c	41	0.000*	0.084	0.159	0.251	0.313
6	a	55	0.267	0.614	0.81	1.064	1.484
	b	55	-1.337	0.03	0.571	1.037	1.997
	c	48	0.000**	0.149	0.191	0.241	0.374
7	a	54	0.446	0.658	0.89	1.145	2.207
	b	54	-0.667	0.128	0.675	1.254	2.055
	c	46	0.001	0.121	0.188	0.251	0.356
8	a	55	0.302	0.655	0.808	1.107	1.484
	b	55	-1.35	0.198	0.395	1.06	3.419
	c	48	0.001	0.136	0.198	0.277	0.345

* Actual value is 0.00009.

** Actual value is 0.00001.

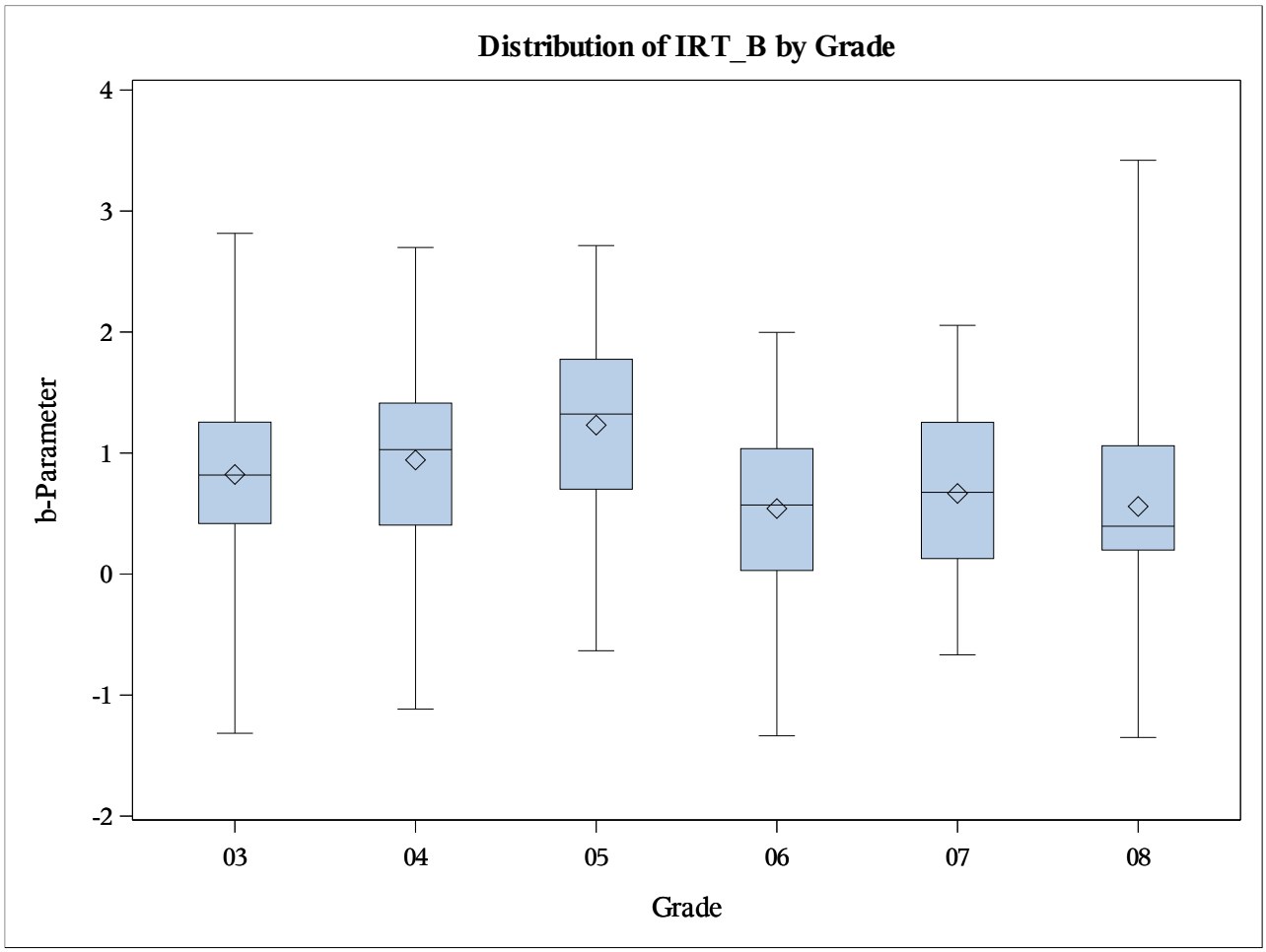
Plot C.5.1

IRT Item Parameter Summary for Spring 2022 Operational SS G3–8: A-Parameter



Plot C.5.2

IRT Item Parameter Summary for Spring 2022 Operational SS G3–8: B-Parameter



Plot C.5.3

IRT Item Parameter Summary for Spring 2022 Operational SS G3–8: C-Parameter

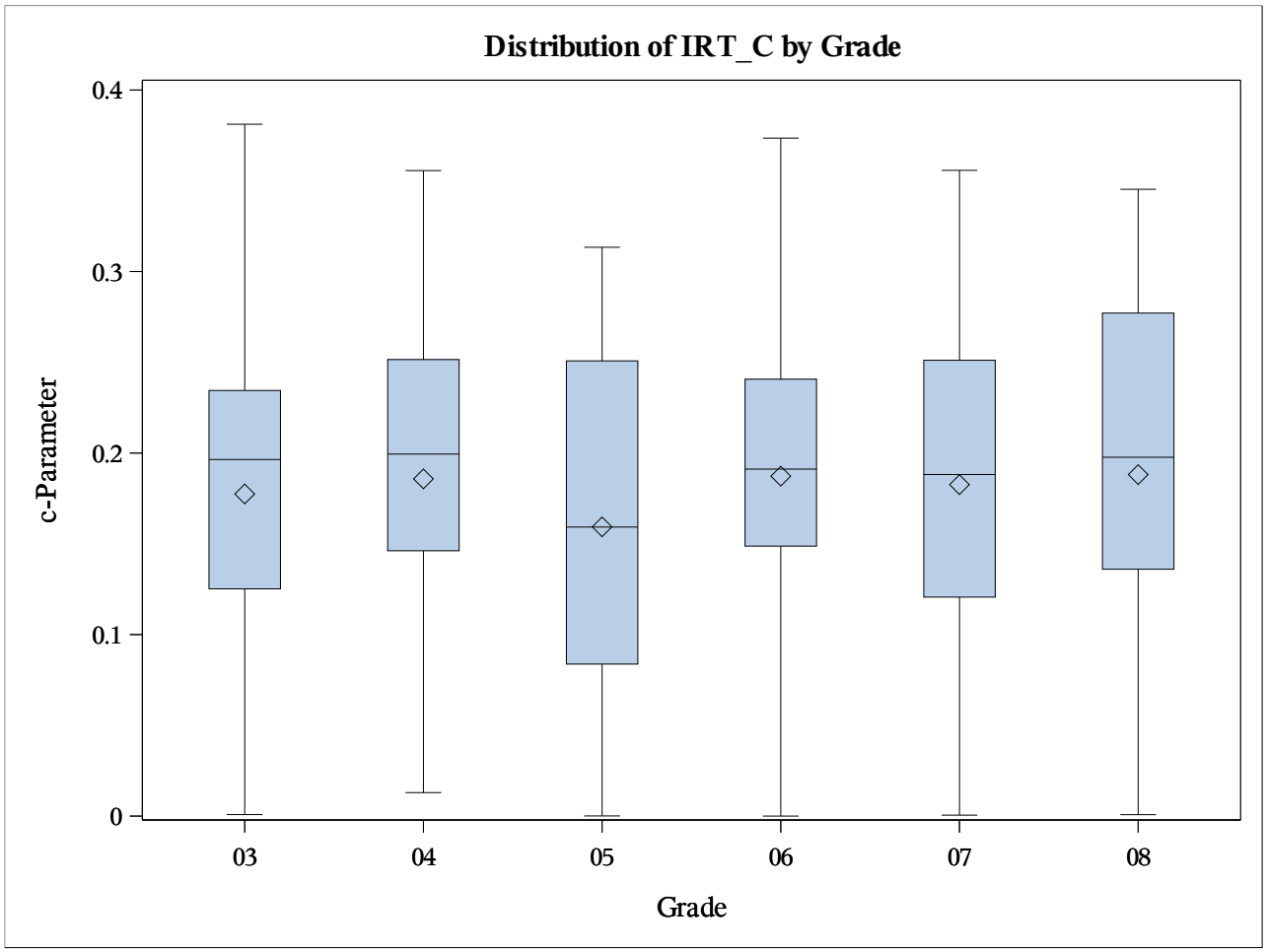


Table C.5.4

IRT Parameter Summary by Item Type: Spring 2022 Operational SS G3–8

Grade 3

Type	Parameter	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	a	2	0.485	0.485	0.506	0.528	0.528
	b	2	0.979	0.979	1.295	1.610	1.61
MC	a	38	0.297	0.650	0.879	1.000	1.298
	b	38	-1.316	0.210	0.816	1.255	2.816
	c	38	0.001	0.134	0.202	0.249	0.381
MS	a	3	0.507	0.507	0.516	0.965	0.965
	b	3	0.421	0.421	0.59	1.159	1.159
	c	3	0.001	0.001	0.025	0.094	0.094

Grade 4

Type	Parameter	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	a	2	0.543	0.543	0.574	0.604	0.604
	b	2	1.028	1.028	1.529	2.03	2.03
MC	a	38	0.24	0.502	0.685	0.938	1.621
	b	38	-0.705	0.405	1.03	1.413	2.699
	c	38	0.013	0.154	0.208	0.253	0.356
MS	a	3	0.996	0.996	1.012	1.135	1.135
	b	3	-1.116	-1.116	0.551	1.353	1.353
	c	3	0.018	0.018	0.075	0.078	0.078

Grade 5

Type	Parameter	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	a	2	0.637	0.637	0.702	0.766	0.766
	b	2	1.322	1.322	1.539	1.755	1.755
ER	a	2	1.041	1.041	1.05	1.059	1.059
	b	2	2.479	2.479	2.597	2.715	2.715
MC	a	37	0.283	0.595	0.906	1.069	1.547
	b	37	-0.634	0.701	1.287	1.607	2.375
	c	37	0.000*	0.116	0.166	0.268	0.313
MS	a	3	0.679	0.679	1.009	1.027	1.027
	b	3	-0.11	-0.11	1.134	2.22	2.22
	c	3	0.015	0.015	0.025	0.062	0.062
TEI	a	3	0.488	0.488	0.501	0.766	0.766
	b	3	0.097	0.097	1.88	1.969	1.969
	c	1	0.142	0.142	0.142	0.142	0.142

* Actual value is 0.00009.

Grade 6

Type	Parameter	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	a	2	0.686	0.686	0.699	0.712	0.712
	b	2	0.793	0.793	1.21	1.627	1.627
ER	a	2	1.039	1.039	1.124	1.209	1.209
	b	2	1.784	1.784	1.826	1.869	1.869
MC	a	42	0.267	0.62	0.877	1.127	1.484
	b	42	-1.337	-0.008	0.411	0.981	1.997
	c	42	0.003	0.159	0.199	0.267	0.374
MS	a	6	0.379	0.557	0.662	0.681	0.832
	b	6	-0.489	0.112	0.662	1.192	1.396
	c	6	0.000*	0.022	0.059	0.087	0.099
TEI	a	3	0.447	0.447	0.477	0.701	0.701
	b	3	-0.19	-0.19	0.921	1.155	1.155

* Actual value is 0.00001.

Grade 7

Type	Parameter	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	a	2	0.775	0.775	0.811	0.848	0.848
	b	2	0.379	0.379	0.56	0.742	0.742
ER	a	2	1.003	1.003	1.022	1.041	1.041
	b	2	1.69	1.69	1.734	1.778	1.778
MC	a	41	0.504	0.703	0.923	1.178	2.207
	b	41	-0.667	0.128	0.624	1.21	1.768
	c	41	0.027	0.144	0.195	0.254	0.356
MS	a	5	0.491	0.52	0.661	1.024	1.261
	b	5	-0.437	0.208	0.94	0.999	2.055
	c	5	0.001	0.005	0.007	0.059	0.08
TEI	a	4	0.446	0.447	0.502	0.633	0.709
	b	4	-0.494	-0.395	0.041	0.959	1.54

Grade 8

Type	Parameter	No. of Items	Minimum	25th Percentile	Median	75th Percentile	Maximum
CR	a	2	0.613	0.613	0.634	0.655	0.655
	b	2	-0.049	-0.049	0.297	0.643	0.643
ER	a	2	0.756	0.756	0.782	0.808	0.808
	b	2	1.175	1.175	1.233	1.291	1.291
MC	a	40	0.346	0.735	0.998	1.177	1.484
	b	40	-0.25	0.188	0.382	0.903	1.705
	c	40	0.028	0.166	0.222	0.29	0.345
MS	a	8	0.302	0.524	0.57	0.76	0.916
	b	8	-0.198	0.282	0.494	1.633	3.419
	c	8	0.001	0.008	0.019	0.072	0.08
TEI	a	3	0.484	0.484	0.595	0.744	0.744
	b	3	-1.35	-1.35	0.264	0.903	0.903

Table C.6

Statistically Flagged Operational Items: Spring 2022 Operational SS G3–8

Grade 3

Item Type	*N* of OP Items	*N* of Items Flagged for *P*-Value	*N* of Items Flagged for Point-Biserial Correlation	*N* of Items Flagged for DIF*	*N* of Items Flagged for Omitting
CR	2	1	0	0	0
MC	38	1	3	0	0
MS	3	0	0	0	0

* The number of flagged DIF items includes both B and C DIF items.

Grade 4

Item Type	*N* of OP Items	*N* of Items Flagged for *P*-Value	*N* of Items Flagged for Point-Biserial Correlation	*N* of Items Flagged for DIF*	*N* of Items Flagged for Omitting
CR	2	1	0	0	0
MC	38	1	2	0	0
MS	3	1	0	0	0

* The number of flagged DIF items includes both B and C DIF items.

Grade 5

Item Type	*N* of OP Items	*N* of Items Flagged for *P*-Value	*N* of Items Flagged for Point-Biserial Correlation	*N* of Items Flagged for DIF*	*N* of Items Flagged for Omitting
CR	2	1	0	0	0
ER**	1	1	0	0	0
MC	37	1	0	0	0
MS	3	1	0	0	0
TEI	3	0	0	0	0

* The number of flagged DIF items includes both B and C DIF items.

** Classical and IRT analyses are calculated and estimated separately for each dimension of the ER item,and the result summarizes both dimensions.

Grade 6

Item Type	*N* of OP Items	*N* of Items Flagged for *P*-Value	*N* of Items Flagged for Point-Biserial Correlation	*N* of Items Flagged for DIF*	*N* of Items Flagged for Omitting
CR	2	1	0	0	0
ER**	1	1	0	0	0
MC	42	1	0	0	0
MS	6	0	0	0	0
TEI	3	0	0	0	0

* The number of flagged DIF items includes both B and C DIF items.

** Classical and IRT analyses are calculated and estimated separately for each dimension of the ER item, and the result summarizes both dimensions.

Grade 7

Item Type	*N* of OP Items	*N* of Items Flagged for *P*-Value	*N* of Items Flagged for Point-Biserial Correlation	*N* of Items Flagged for DIF*	*N* of Items Flagged for Omitting
CR	2	0	0	1	0
ER**	1	1	0	0	0
MC	41	1	1	1	0
MS	5	1	0	0	0
TEI	4	0	0	0	0

* The number of flagged DIF items includes both B and C DIF items.

** Classical analyses are calculated and estimated separately for each dimension of the ER item, and the result summarizes both dimensions.

Grade 8

Item Type	*N* of OP Items	*N* of Items Flagged for *P*-Value	*N* of Items Flagged for Point-Biserial Correlation	*N* of Items Flagged for DIF*	*N* of Items Flagged for Omitting
CR	2	0	0	0	0
ER**	1	0	0	1	0
MC	40	1	0	2	0
MS	8	3	0	0	0
TEI	3	0	0	0	0

* The number of flagged DIF items includes both B and C DIF items.

** Classical analyses are calculated and estimated separately for each dimension of the ER item, and the result summarizes both dimensions.

Appendix D: Dimensionality

Social Studies G3–8

Contents
Table D.1 Zq1 Statistics and Summary Data: Spring 2022 Operational SS G3–8
Table D.2 Q3 Statistics and Summary Data: Spring 2022 Operational SS G3–8
Table D.3 Reporting Category Intercorrelation Coefficients: Spring 2022 Operational SS G3–8
Table D.4 First and Second Eigenvalues: Spring 2022 Operational SS G3–8 Plot D.1 Principal Component Analysis: Spring 2022 Operational SS G3–8

- Because the spring 2022 test was administered during the 2022 COVID-19 pandemic, great caution should be applied when any statistical inference is drawn.

Table D.1

Zq1 Statistics and Summary Data: Spring 2022 Operational SS G3–8

Grade 3

Item Type	Minimum	25th Percentile	Median	75th Percentile	Maximum	Num. of Items with Poor Fit
CR	71.05	71.05	78.09	85.12	85.12	0
MC	8.60	18.87	25.73	72.60	437.43	5
MS	22.81	22.81	60.16	726.63	726.63	1

Grade 4

Item Type	Minimum	25th Percentile	Median	75th Percentile	Maximum	Num. of Items with Poor Fit
CR	114.98	114.98	152.92	190.86	190.86	1
MC	4.59	18.01	29.31	52.49	342.49	2
MS	40.58	40.58	70.89	453.43	453.43	1

Grade 5

Item Type	Minimum	25th Percentile	Median	75th Percentile	Maximum	Num. of Items with Poor Fit
CR	75.02	75.02	92.00	108.97	108.97	0
ER	99.57	99.57	102.93	106.30	106.30	0
MC	3.06	15.38	21.45	39.76	710.16	3
MS	17.71	17.71	28.85	226.08	226.08	1
TEI	5.98	5.98	128.44	257.09	257.09	1

Grade 6

Item Type	Minimum	25th Percentile	Median	75th Percentile	Maximum	Num. of Items with Poor Fit
CR	30.61	30.61	78.61	126.62	126.62	0
ER	41.48	41.48	47.98	54.49	54.49	0
MC	5.78	9.39	13.98	22.56	554.61	1
MS	12.10	13.20	28.60	30.64	1176.87	1
TEI	132.11	132.11	134.16	338.94	338.94	3

Grade 7

Item Type	Minimum	25th Percentile	Median	75th Percentile	Maximum	Num. of Items with Poor Fit
CR	140.80	140.80	157.95	175.09	175.09	2
ER	79.92	79.92	85.01	90.11	90.11	0
MC	4.58	10.09	14.72	30.85	105.57	0
MS	6.25	14.97	35.71	45.36	1620.27	1
TEI	61.34	68.32	112.45	372.41	595.22	2

Grade 8

Item Type	Minimum	25th Percentile	Median	75th Percentile	Maximum	Num. of Items with Poor Fit
CR	52.16	52.16	96.06	139.96	139.96	1
ER	161.69	161.69	177.46	193.24	193.24	2
MC	3.50	8.65	12.75	16.81	47.72	0
MS	6.82	15.32	22.45	45.67	358.02	1
TEI	16.31	16.31	100.98	107.91	107.91	0

Table D.2

Q3 Statistics and Summary Data: Spring 2022 Operational SS G3–8

Grade	Average Zero-Order Correlation	Minimum	5th Percentile	Median	95th Percentile	Maximum
3	0.125	-0.072	-0.051	-0.026	0.015	0.117
4	0.115	-0.066	-0.046	-0.025	0.006	0.089
5	0.140	-0.072	-0.043	-0.021	0.008	0.892
6	0.167	-0.068	-0.039	-0.018	0.010	0.785
7	0.176	-0.071	-0.042	-0.017	0.011	0.850
8	0.185	-0.095	-0.042	-0.017	0.014	0.933

Table D.3

Reporting Category Intercorrelation Coefficients: Spring 2022 Operational SS G3–8

Grade	Reporting Category	1	2	3	4
3	1 History	1.00			
	2 Geography	0.60	1.00		
	3 Civics	0.61	0.61	1.00	
	4 Economics	0.57	0.57	0.64	1.00
4	1 History	1.00			
	2 Geography	0.58	1.00		
	3 Civics	0.61	0.55	1.00	
	4 Economics	0.58	0.53	0.54	1.00
5	1 History	1.00			
	2 Geography	0.62	1.00		
	3 Civics	0.69	0.56	1.00	
	4 Economics	0.68	0.55	0.61	1.00
6	1 History	1.00			
	2 Geography	0.72	1.00		
	3 Civics	0.78	0.64	1.00	
	4 Economics	0.71	0.60	0.62	1.00
7	1 History	1.00			
	2 Geography	0.70	1.00		
	3 Civics	0.77	0.65	1.00	
	4 Economics	0.72	0.61	0.67	1.00
8	1 History	1.00			
	2 Geography	0.78	1.00		
	3 Civics	0.78	0.69	1.00	
	4 Economics	0.72	0.66	0.63	1.00

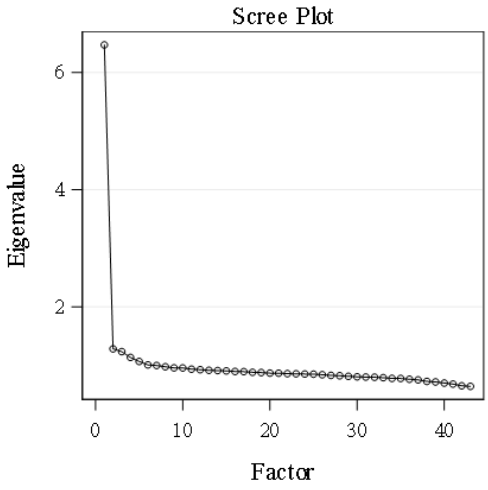
Table D.4

First and Second Eigenvalue: Spring 2022 Operational SS G3–8

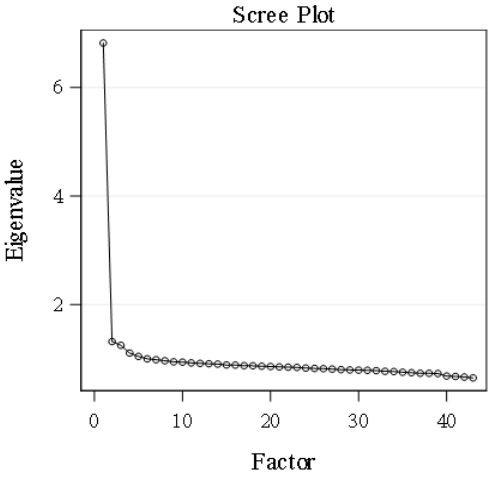
Grade	Mode	First Eigenvalue	Second Eigenvalue	Ratio
3	Online	6.473	1.281	5.054
	Paper	6.821	1.319	5.173
4	Online	6.277	1.373	4.572
5	Online	8.070	1.335	6.046
6	Online	10.658	1.342	7.943
7	Online	11.090	1.446	7.669
8	Online	11.611	1.437	8.077

Plot D.1

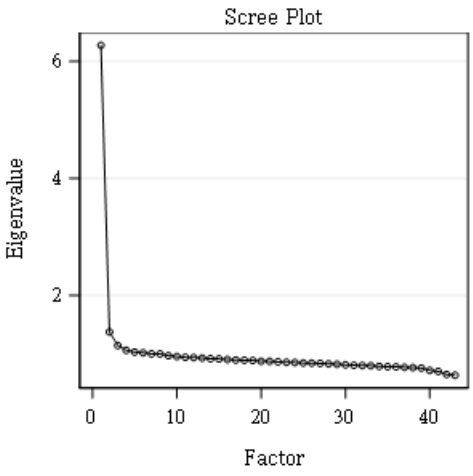
Principal Component Analysis Plot: Spring 2022 Operational SS G3-8



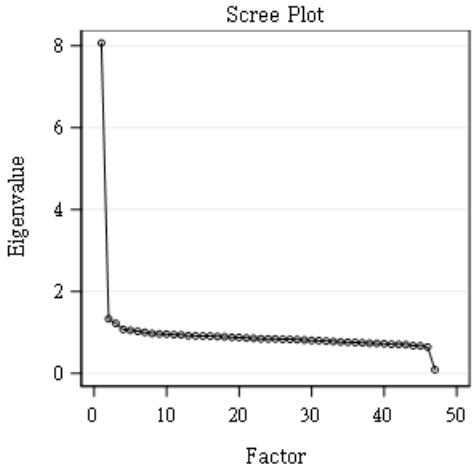
Grade 3: Online



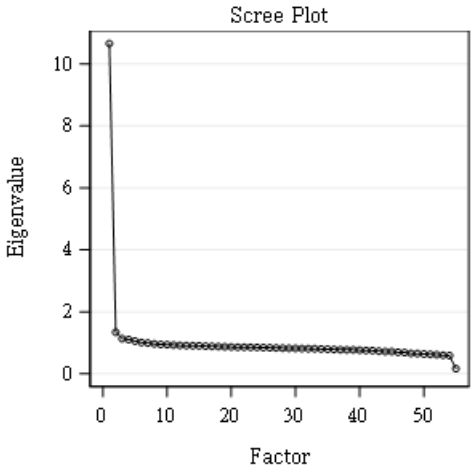
Grade 3: Paper



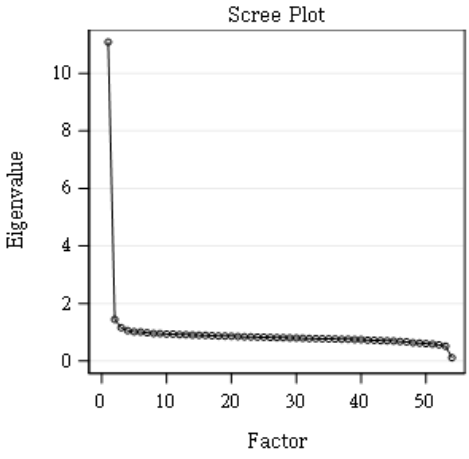
Grade 4



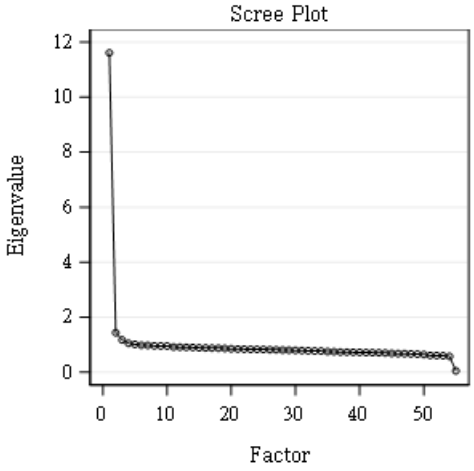
Grade 5



Grade 6



Grade 7



Grade 8

Appendix E: Scale Distribution and Statistical Report

Social Studies G3–8

Contents
Table E.1.1 Scale Score Descriptive Statistics and Plots: Spring 2022 Social Studies Grade 3 Table E.1.2 Frequency Distribution of Scale Scores: Spring 2022 Social Studies Grade 3
Table E.2.1 Scale Score Descriptive Statistics and Plots: Spring 2022 Social Studies Grade 4 Table E.2.2 Frequency Distribution of Scale Scores: Spring 2022 Social Studies Grade 4
Table E.3.1 Scale Score Descriptive Statistics and Plots: Spring 2022 Social Studies Grade 5 Table E.3.2 Frequency Distribution of Scale Scores: Spring 2022 Social Studies Grade 5
Table E.4.1 Scale Score Descriptive Statistics and Plots: Spring 2022 Social Studies Grade 6 Table E.4.2 Frequency Distribution of Scale Scores: Spring 2022 Social Studies Grade 6
Table E.5.1 Scale Score Descriptive Statistics and Plots: Spring 2022 Social Studies Grade 7 Table E.5.2 Frequency Distribution of Scale Scores: Spring 2022 Social Studies Grade 7
Table E.6.1 Scale Score Descriptive Statistics and Plots: Spring 2022 Social Studies Grade 8 Table E.6.2 Frequency Distribution of Scale Scores: Spring 2022 Social Studies Grade 8

- Because the spring 2022 test was administered during the 2022 COVID-19 pandemic, great caution should be applied when any statistical inference is drawn.

Table E.1.1

Scale Score Descriptive Statistics and Plots: Spring 2022 Social Studies Grade 3

DESCRIPTIVE STATISTICS - SCALE SCORES
Social Studies
ALL STUDENTS
GRADE 03

N	≥49310		
Mean	717.26	Median	719.00
Std deviation	39.51	Variance	1560.68
Skewness	0.0347	Kurtosis	-0.5784
Mode	650.00	Std Error Mean	0.1779
Range	200.00	Interquartile Range	55.00

Quantile	Estimate
100% Max	850
99%	807
95%	779
90%	767
75% Q3	746
50% Median	719
25% Q1	691
10%	660
5%	650
1%	650
0% Min	650

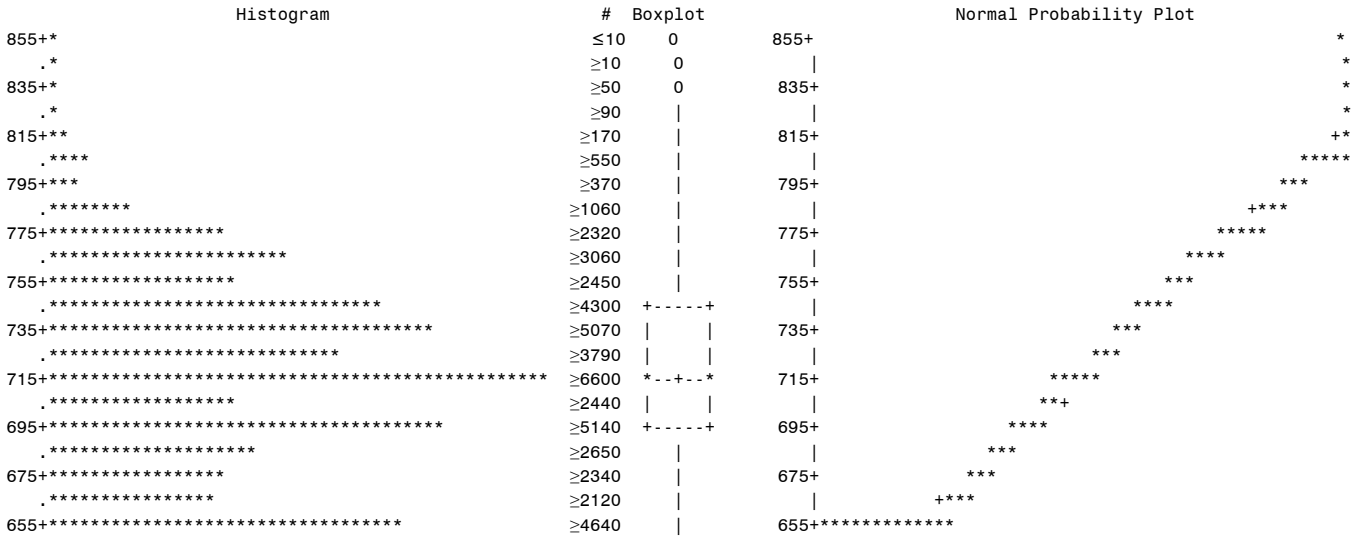


Table E.1.2

Frequency Distribution of Scale Scores: Spring 2022 Social Studies Grade 3

FREQUENCY DISTRIBUTION - SCALE SCORES
Social Studies
ALL STUDENTS
GRADE 03

Scale_Score		Freq	Cum. Freq	Percent	Cum. Percent
650	**********	≥4640	≥4640	9.42	9.42
660	**********	≥2120	≥6760	4.31	13.73
674	**********	≥2340	≥9110	4.75	18.48
683	**********	≥2650	≥11760	5.38	23.86
691	**********	≥2600	≥14370	5.29	29.14
698	**********	≥2540	≥16910	5.15	34.29
704	**********	≥2440	≥19350	4.95	39.25
710	**********	≥2360	≥21710	4.80	44.04
715	**********	≥2150	≥23870	4.36	48.41
719	**********	≥2080	≥25950	4.23	52.63
723	**********	≥1910	≥27860	3.88	56.51
728	**********	≥1880	≥29750	3.82	60.33
731	**********	≥1770	≥31520	3.60	63.93
735	**********	≥1670	≥33190	3.39	67.32
739	**********	≥1620	≥34820	3.30	70.62
742	**********	≥1530	≥36360	3.12	73.74
746	**********	≥1390	≥37750	2.83	76.57
749	**********	≥1370	≥39130	2.78	79.35
753	**********	≥1250	≥40390	2.55	81.91
756	**********	≥1190	≥41580	2.42	84.32
760	**********	≥1120	≥42700	2.27	86.60
764	**********	≥1040	≥43740	2.11	88.71
767	**********	≥900	≥44640	1.84	90.54
771	**********	≥870	≥45510	1.76	92.31
775	**********	≥780	≥46300	1.59	93.90
779	**********	≥660	≥46970	1.35	95.25
784	**********	≥590	≥47560	1.20	96.46
789	**********	≥470	≥48040	0.96	97.42
794	**********	≥370	≥48410	0.76	98.19
800	**********	≥330	≥48750	0.67	98.86
807	**********	≥210	≥48960	0.44	99.30
814	**********	≥170	≥49140	0.35	99.66
823	**********	≥90	≥49230	0.19	99.85
834	**********	≥50	≥49280	0.10	99.95
849	**********	≥10	≥49300	0.04	99.99
850	**********	≤10	≥49310	0.01	100.00

600 1200 1800 2400 3000 3600 4200

Table E.2.1

Scale Score Descriptive Statistics and Plots: Spring 2022 Social Studies Grade 4

DESCRIPTIVE STATISTICS - SCALE SCORES
Social Studies
ALL STUDENTS
GRADE 04

N	≥48880		
Mean	722.49	Median	727.00
Std deviation	41.97	Variance	1761.37
Skewness	-0.0788	Kurtosis	-0.6463
Mode	650.00	Std Error Mean	0.1898
Range	200.00	Interquartile Range	56.00

Quantile	Estimate
100% Max	850
99%	808
95%	788
90%	775
75% Q3	753
50% Median	727
25% Q1	697
10%	650
5%	650
1%	650
0% Min	650

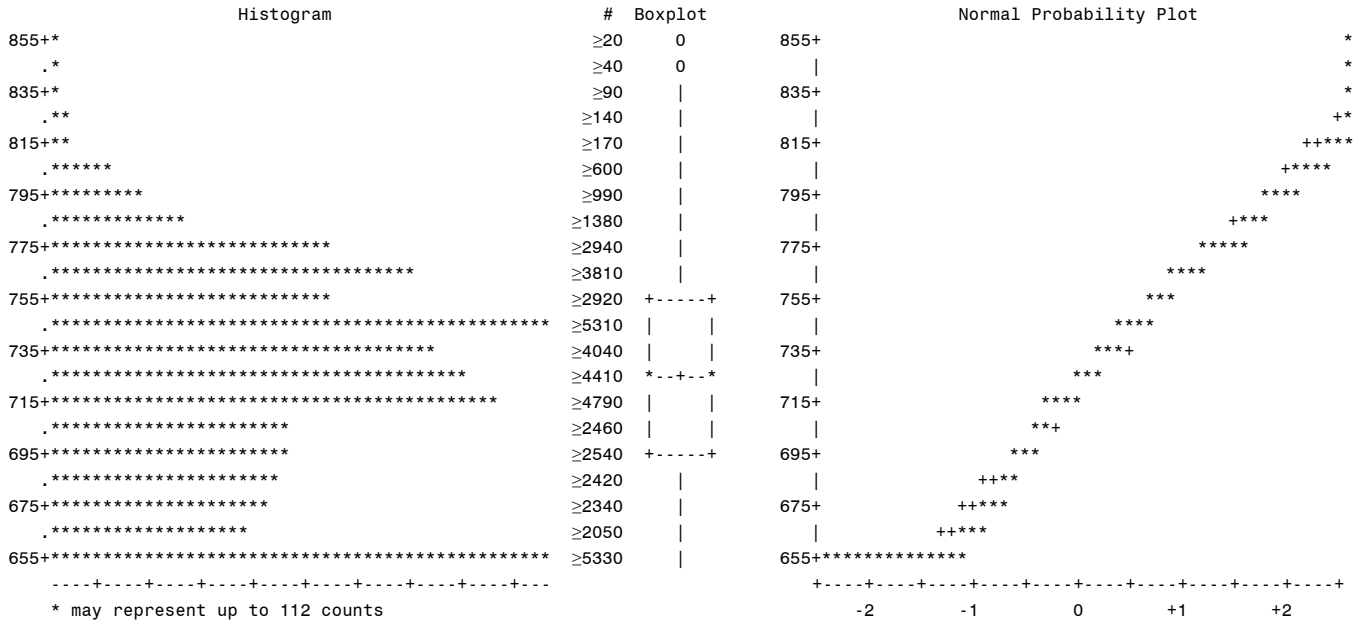


Table E.2.2

Frequency Distribution of Scale Scores: Spring 2022 Social Studies Grade 4

FREQUENCY DISTRIBUTION ╤ SCALE SCORES
Social Studies
ALL STUDENTS
GRADE 04

Scale_Score		Freq	Cum. Freq	Percent	Cum. Percent
650	**********	≥5330	≥5330	10.91	10.91
666	**********	≥2050	≥7380	4.19	15.11
678	**********	≥2340	≥9720	4.79	19.90
688	**********	≥2420	≥12140	4.95	24.85
697	**********	≥2540	≥14690	5.20	30.05
704	**********	≥2460	≥17150	5.04	35.10
711	**********	≥2450	≥19600	5.02	40.11
717	**********	≥2340	≥21950	4.80	44.91
722	**********	≥2230	≥24190	4.58	49.48
727	**********	≥2180	≥26370	4.46	53.95
732	**********	≥2110	≥28480	4.32	58.26
737	**********	≥1930	≥30420	3.97	62.23
741	**********	≥1890	≥32310	3.88	66.11
745	**********	≥1710	≥34030	3.51	69.61
749	**********	≥1700	≥35730	3.48	73.09
753	**********	≥1490	≥37220	3.05	76.14
757	**********	≥1430	≥38650	2.93	79.07
760	**********	≥1370	≥40020	2.80	81.87
764	**********	≥1250	≥41270	2.57	84.44
768	**********	≥1190	≥42470	2.44	86.87
772	**********	≥1070	≥43540	2.21	89.08
775	**********	≥960	≥44510	1.98	91.06
779	**********	≥890	≥45410	1.83	92.90
783	**********	≥710	≥46130	1.47	94.37
788	**********	≥660	≥46800	1.37	95.74
792	**********	≥520	≥47320	1.07	96.81
797	**********	≥470	≥47790	0.96	97.77
802	**********	≥340	≥48140	0.71	98.48
808	**********	≥260	≥48400	0.54	99.02
815	**********	≥170	≥48580	0.36	99.38
823	**********	≥140	≥48720	0.29	99.67
834	**********	≥90	≥48820	0.20	99.86
847	**********	≥40	≥48860	0.09	99.95
850	**********	≥20	≥48880	0.05	100.00

600 1200 1800 2400 3000 3600 4200 4800

Table E.3.1

Scale Score Descriptive Statistics and Plots: Spring 2022 Social Studies Grade 5

DESCRIPTIVE STATISTICS - SCALE SCORES
Social Studies
ALL STUDENTS
GRADE 05

N	≥48880		
Mean	720.28	Median	723.00
Std deviation	38.57	Variance	1487.85
Skewness	-0.0778	Kurtosis	-0.6058
Mode	650.00	Std Error Mean	0.1745
Range	200.00	Interquartile Range	57.00

Quantile	Estimate
100% Max	850
99%	802
95%	780
90%	769
75% Q3	748
50% Median	723
25% Q1	691
10%	656
5%	650
1%	650
0% Min	650

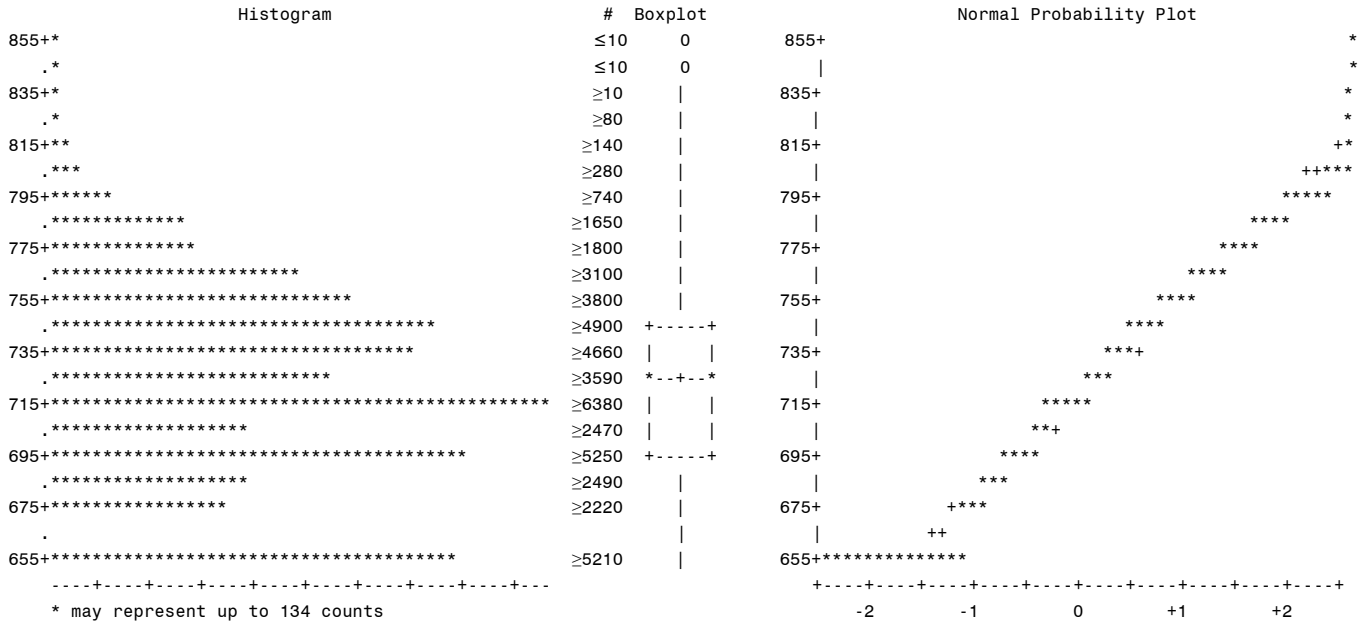


Table E.3.2

Frequency Distribution of Scale Scores: Spring 2022 Social Studies Grade 5

FREQUENCY DISTRIBUTION - SCALE SCORES
Social Studies
ALL STUDENTS
GRADE 05

Scale_Score		Freq	Cum. Freq	Percent	Cum. Percent
650	**********	≥3280	≥3280	6.72	6.72
656	*******	≥1930	≥5210	3.95	10.67
671	*******	≥2220	≥7440	4.55	15.22
682	*******	≥2490	≥9930	5.11	20.33
691	*******	≥2660	≥12590	5.44	25.77
698	*******	≥2590	≥15190	5.30	31.07
705	*******	≥2470	≥17660	5.07	36.14
710	*******	≥2280	≥19950	4.67	40.82
715	*******	≥2120	≥22070	4.34	45.16
719	*******	≥1980	≥24050	4.05	49.21
723	*******	≥1820	≥25880	3.74	52.95
727	*******	≥1770	≥27650	3.62	56.57
730	*******	≥1650	≥29310	3.39	59.96
734	*******	≥1560	≥30870	3.19	63.15
737	*******	≥1450	≥32320	2.97	66.12
740	*******	≥1300	≥33620	2.67	68.79
743	*******	≥1250	≥34870	2.56	71.35
746	*******	≥1180	≥36050	2.41	73.76
748	*******	≥1170	≥37230	2.40	76.16
751	*******	≥1080	≥38310	2.22	78.38
754	******	≥960	≥39270	1.97	80.35
756	******	≥910	≥40190	1.88	82.23
759	******	≥840	≥41030	1.72	83.95
761	******	≥850	≥41890	1.75	85.70
764	******	≥770	≥42660	1.58	87.28
766	******	≥770	≥43430	1.58	88.86
769	******	≥700	≥44140	1.43	90.29
772	******	≥660	≥44800	1.37	91.66
774	******	≥590	≥45390	1.21	92.87
777	******	≥540	≥45940	1.11	93.98
780	******	≥520	≥46460	1.07	95.04
783	******	≥430	≥46890	0.89	95.93
785	******	≥350	≥47250	0.73	96.66
788	******	≥340	≥47590	0.70	97.37
792	******	≥300	≥47890	0.62	97.98
795	******	≥240	≥48140	0.51	98.49
798	******	≥190	≥48340	0.41	98.90
802	******	≥160	≥48500	0.33	99.23
806	******	≥120	≥48620	0.25	99.47
811	******	≥90	≥48720	0.20	99.67
815	******	≥50	≥48770	0.11	99.77
821	******	≥50	≥48830	0.12	99.89
827	******	≥20	≥48850	0.05	99.94
833	******	≥10	≥48860	0.03	99.97
842	******	≤10	≥48870	0.02	99.98
850	******	≤10	≥48880	0.02	100.00

400 800 1200 1600 2000 2400 2800 3200

Frequency

Table E.4.1

Scale Score Descriptive Statistics and Plots: Spring 2022 Social Studies Grade 6

DESCRIPTIVE STATISTICS - SCALE SCORES
Social Studies
ALL STUDENTS
GRADE 06

N	≥49270		
Mean	712.43	Median	713.00
Std deviation	37.51	Variance	1406.65
Skewness	0.0430	Kurtosis	-0.6929
Mode	650.00	Std Error Mean	0.1690
Range	200.00	Interquartile Range	55.00

Quantile	Estimate
100% Max	850
99%	792
95%	772
90%	761
75% Q3	740
50% Median	713
25% Q1	685
10%	659
5%	650
1%	650
0% Min	650

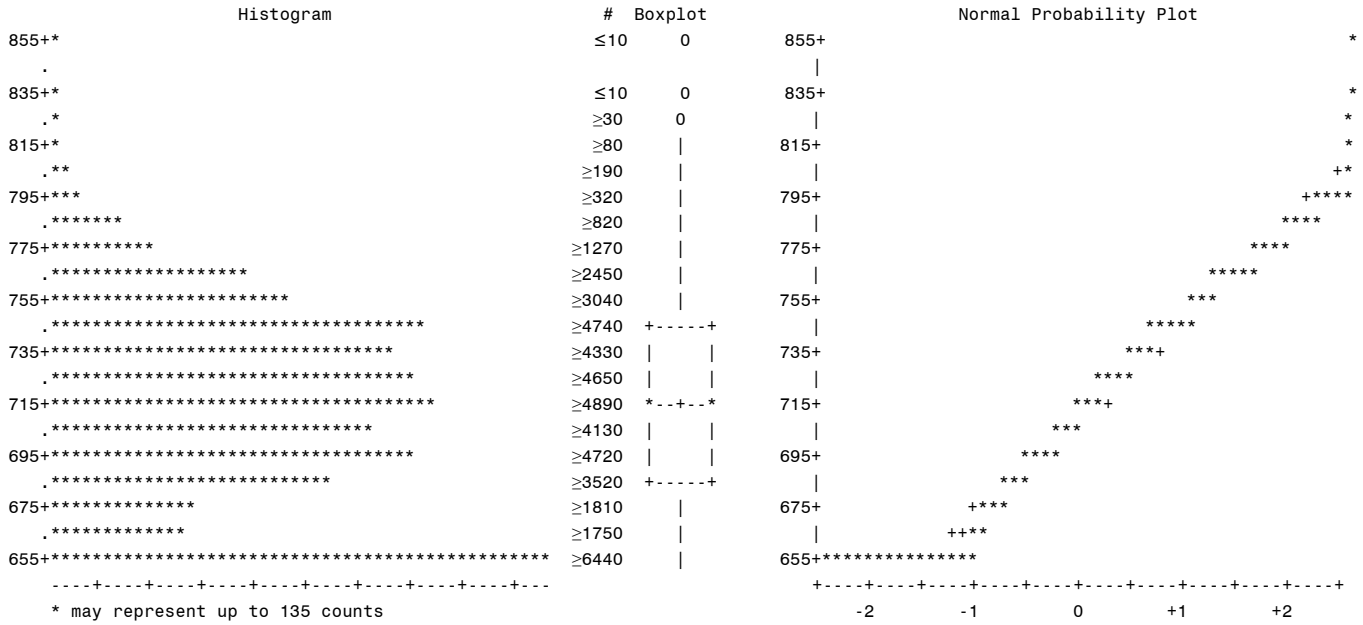


Table E.4.2

Frequency Distribution of Scale Scores: Spring 2022 Social Studies Grade 6

FREQUENCY DISTRIBUTION - SCALE SCORES
Social Studies
ALL STUDENTS
GRADE 06

Scale_Score		Freq	Cum. Freq	Percent	Cum. Percent
650	**********	≥4810	≥4810	9.77	9.77
659	**********	≥1630	≥6440	3.31	13.08
668	**********	≥1750	≥8200	3.56	16.64
674	**********	≥1810	≥10010	3.68	20.32
680	**********	≥1810	≥11820	3.68	23.99
685	**********	≥1700	≥13530	3.47	27.46
690	**********	≥1630	≥15170	3.32	30.78
694	**********	≥1610	≥16780	3.27	34.06
698	**********	≥1470	≥18250	2.99	37.05
701	**********	≥1440	≥19700	2.93	39.98
704	**********	≥1340	≥21050	2.74	42.72
707	**********	≥1340	≥22390	2.73	45.45
710	**********	≥1290	≥23690	2.63	48.08
713	**********	≥1220	≥24920	2.49	50.57
716	**********	≥1240	≥26160	2.52	53.09
718	**********	≥1130	≥27290	2.30	55.39
721	**********	≥1230	≥28530	2.51	57.90
723	**********	≥1170	≥29710	2.39	60.29
726	**********	≥1130	≥30850	2.31	62.61
728	**********	≥1090	≥31940	2.22	64.83
731	**********	≥1080	≥33020	2.19	67.02
733	**********	≥1130	≥34160	2.31	69.33
735	**********	≥1070	≥35240	2.18	71.51
737	**********	≥1030	≥36270	2.10	73.62
740	**********	≥990	≥37270	2.02	75.64
742	**********	≥960	≥38240	1.96	77.60
744	**********	≥960	≥39210	1.97	79.57
747	**********	≥910	≥40120	1.86	81.43
749	**********	≥890	≥41020	1.82	83.25
751	**********	≥850	≥41880	1.74	84.99
754	**********	≥720	≥42600	1.46	86.45
756	**********	≥760	≥43370	1.56	88.01
759	**********	≥690	≥44060	1.40	89.42
761	**********	≥710	≥44770	1.44	90.86
764	**********	≥630	≥45400	1.28	92.15
767	**********	≥610	≥46020	1.25	93.39
769	**********	≥490	≥46510	0.99	94.39
772	**********	≥470	≥46980	0.96	95.35
775	**********	≥420	≥47410	0.87	96.22
778	**********	≥370	≥47780	0.76	96.98
782	**********	≥300	≥48090	0.62	97.60
785	**********	≥280	≥48380	0.59	98.18
789	**********	≥230	≥48610	0.48	98.66
792	**********	≥180	≥48800	0.38	99.04
796	**********	≥130	≥48940	0.27	99.31
801	**********	≥120	≥49060	0.24	99.56
805	**********	≥70	≥49130	0.16	99.72
810	**********	≥50	≥49190	0.12	99.83
815	**********	≥30	≥49220	0.06	99.89
821	**********	≥20	≥49250	0.05	99.94
829	**********	≥10	≥49260	0.03	99.97
838	**********	≤10	≥49270	0.01	99.98
850	**********	≤10	≥49270	0.02	100.00

600 1200 1800 2400 3000 3600 4200 4800

Frequency

Table E.5.1

Scale Score Descriptive Statistics and Plots: Spring 2022 Social Studies Grade 7

DESCRIPTIVE STATISTICS - SCALE SCORES
Social Studies
ALL STUDENTS
GRADE 07

N	≥50930		
Mean	727.01	Median	726.00
Std deviation	40.54	Variance	1643.09
Skewness	0.0469	Kurtosis	-0.5622
Mode	685.00	Std Error Mean	0.1796
Range	200.00	Interquartile Range	61.00

Quantile	Estimate
100% Max	850
99%	814
95%	792
90%	779
75% Q3	757
50% Median	726
25% Q1	696
10%	677
5%	655
1%	650
0% Min	650

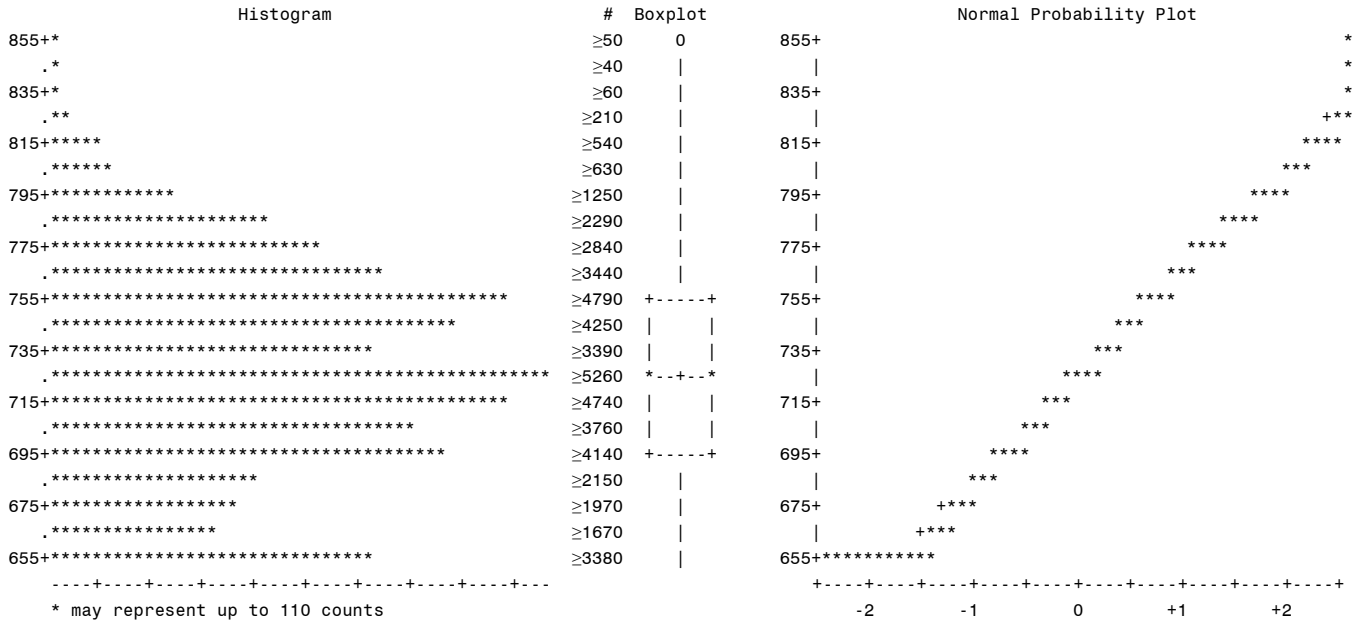


Table E.5.2

Frequency Distribution of Scale Scores: Spring 2022 Social Studies Grade 7

FREQUENCY DISTRIBUTION - SCALE SCORES
Social Studies
ALL STUDENTS
GRADE 07

Scale_Score		Freq	Cum. Freq	Percent	Cum. Percent
650	**********	≥2070	≥2070	4.07	4.07
655	**********	≥1310	≥3380	2.58	6.64
668	**********	≥1670	≥5060	3.29	9.94
677	**********	≥1970	≥7030	3.87	13.81
685	**********	≥2150	≥9190	4.24	18.05
691	**********	≥2090	≥11290	4.12	22.17
696	**********	≥2040	≥13330	4.01	26.18
701	**********	≥1960	≥15290	3.85	30.03
706	**********	≥1800	≥17100	3.55	33.58
710	**********	≥1770	≥18870	3.48	37.06
714	**********	≥1570	≥20450	3.10	40.16
717	**********	≥1390	≥21840	2.73	42.89
720	**********	≥1340	≥23180	2.63	45.52
723	**********	≥1380	≥24570	2.72	48.24
726	**********	≥1320	≥25890	2.61	50.85
729	**********	≥1210	≥27110	2.38	53.23
732	**********	≥1150	≥28260	2.28	55.50
735	**********	≥1170	≥29440	2.30	57.80
737	**********	≥1060	≥30500	2.09	59.90
740	**********	≥1110	≥31610	2.18	62.08
743	**********	≥1050	≥32670	2.07	64.15
745	**********	≥1060	≥33730	2.09	66.23
747	**********	≥1020	≥34750	2.01	68.24
750	**********	≥1010	≥35770	1.99	70.24
752	**********	≥1010	≥36790	2.00	72.23
755	**********	≥910	≥37700	1.79	74.02
757	**********	≥930	≥38630	1.84	75.86
759	**********	≥910	≥39550	1.80	77.66
762	**********	≥880	≥40430	1.74	79.39
764	**********	≥890	≥41330	1.77	81.16
766	**********	≥840	≥42180	1.66	82.82
769	**********	≥810	≥43000	1.60	84.42
771	**********	≥790	≥43790	1.56	85.98
774	**********	≥690	≥44480	1.35	87.34
776	**********	≥650	≥45140	1.29	88.63
779	**********	≥700	≥45840	1.37	90.00
781	**********	≥640	≥46480	1.27	91.27
784	**********	≥580	≥47070	1.14	92.41
787	**********	≥550	≥47620	1.09	93.51
789	**********	≥500	≥48130	0.99	94.50
792	**********	≥470	≥48610	0.94	95.44
796	**********	≥420	≥49030	0.83	96.27
799	**********	≥350	≥49380	0.70	96.97
802	**********	≥350	≥49740	0.69	97.66
806	**********	≥270	≥50020	0.55	98.21
810	**********	≥210	≥50230	0.41	98.62
814	**********	≥190	≥50420	0.38	99.00
818	**********	≥130	≥50560	0.26	99.27
823	**********	≥130	≥50690	0.26	99.52
829	**********	≥80	≥50770	0.17	99.70
835	**********	≥60	≥50840	0.12	99.82
843	**********	≥40	≥50880	0.08	99.90
850	**********	≥50	≥50930	0.10	100.00

200 400 600 800 1000 1200 1400 1600 1800 2000

Frequency

Table E.6.1

Scale Score Descriptive Statistics and Plots: Spring 2022 Social Studies Grade 8

DESCRIPTIVE STATISTICS - SCALE SCORES
Social Studies
ALL STUDENTS
GRADE 08

N	≥50660		
Mean	726.97	Median	729.00
Std deviation	42.44	Variance	1800.99
Skewness	-0.0978	Kurtosis	-0.6493
Mode	650.00	Std Error Mean	0.1885
Range	200.00	Interquartile Range	62.00

Quantile	Estimate
100% Max	850
99%	816
95%	791
90%	779
75% Q3	758
50% Median	729
25% Q1	696
10%	663
5%	650
1%	650
0% Min	650

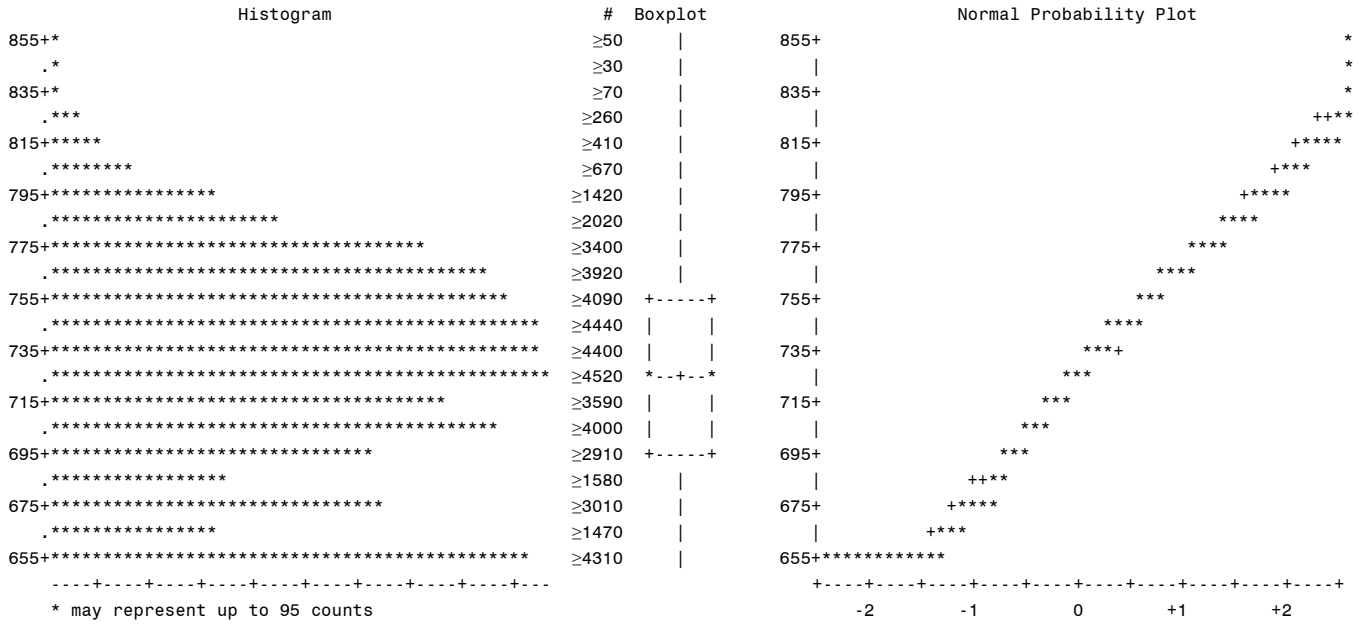


Table E.6.2

Frequency Distribution of Scale Scores: Spring 2022 Social Studies Grade 8

FREQUENCY DISTRIBUTION - SCALE SCORES
Social Studies
ALL STUDENTS
GRADE 08

Scale_Score		Freq	Cum. Freq	Percent	Cum. Percent
650	**********	≥3040	≥3040	6.02	6.02
652	**********	≥1260	≥4310	2.49	8.51
663	**********	≥1470	≥5780	2.90	11.41
672	**********	≥1470	≥7250	2.91	14.32
679	**********	≥1530	≥8790	3.03	17.35
686	**********	≥1580	≥10370	3.12	20.48
691	**********	≥1480	≥11850	2.92	23.40
696	**********	≥1430	≥13290	2.83	26.23
700	**********	≥1330	≥14620	2.64	28.87
704	**********	≥1370	≥15990	2.70	31.57
708	**********	≥1300	≥17290	2.57	34.14
712	**********	≥1230	≥18530	2.44	36.58
715	**********	≥1200	≥19740	2.38	38.97
718	**********	≥1150	≥20890	2.27	41.24
721	**********	≥1160	≥22060	2.30	43.54
724	**********	≥1130	≥23190	2.23	45.77
726	**********	≥1090	≥24280	2.16	47.94
729	**********	≥1120	≥25410	2.22	50.16
732	**********	≥1050	≥26470	2.08	52.24
734	**********	≥1130	≥27600	2.24	54.49
737	**********	≥1080	≥28690	2.15	56.63
739	**********	≥1120	≥29820	2.23	58.86
741	**********	≥1030	≥30860	2.05	60.91
744	**********	≥1140	≥32000	2.26	63.17
746	**********	≥1110	≥33120	2.21	65.38
749	**********	≥1140	≥34260	2.25	67.63
751	**********	≥1070	≥35340	2.13	69.76
753	**********	≥1030	≥36370	2.03	71.79
756	**********	≥990	≥37370	1.97	73.76
758	**********	≥980	≥38350	1.94	75.70
760	**********	≥1020	≥39370	2.02	77.72
763	**********	≥1010	≥40390	2.01	79.73
765	**********	≥960	≥41360	1.90	81.63
768	**********	≥920	≥42280	1.82	83.46
771	**********	≥890	≥43180	1.77	85.23
773	**********	≥830	≥44010	1.64	86.87
776	**********	≥830	≥44840	1.65	88.51
779	**********	≥840	≥45680	1.66	90.17
782	**********	≥790	≥46480	1.57	91.74
785	**********	≥650	≥47140	1.30	93.04
788	**********	≥570	≥47710	1.13	94.17
791	**********	≥520	≥48240	1.04	95.22
795	**********	≥490	≥48730	0.97	96.18
799	**********	≥410	≥49140	0.81	96.99
803	**********	≥360	≥49500	0.71	97.70
807	**********	≥310	≥49820	0.63	98.33
811	**********	≥230	≥50050	0.47	98.80
816	**********	≥170	≥50230	0.35	99.15
822	**********	≥150	≥50390	0.31	99.46
828	**********	≥100	≥50500	0.21	99.67
836	**********	≥70	≥50570	0.15	99.82
844	**********	≥30	≥50610	0.07	99.90
850	**********	≥50	≥50660	0.10	100.00

400 800 1200 1600 2000 2400 2800

Frequency

Appendix F: Reliability and Classification Accuracy

Social Studies G3–8

Contents
Tables F.1.1-F.1.2 Reliability and SEM for Overall and Subgroups: Spring 2022 Operational SS G3-8
Table F.2 Cronbach's Alpha and Marginal Reliability: Spring 2022 Operational SS G3-8
Table F.3 Classification Accuracy and Decision Consistency: Spring 2022 Operational SS G3-8

- Because the spring 2022 test was administered during the 2022 COVID-19 pandemic, great caution should be applied when any statistical inference is drawn.

Table F.1.1

Reliability for Overall and Subgroups: Spring 2022 Operational Social Studies

Grade	3	4	5	6	7	8
All Students	**0.863**	**0.849**	**0.888**	**0.919**	**0.923**	**0.928**
Female	0.855	0.835	0.879	0.913	0.917	0.921
Male	0.870	0.860	0.895	0.924	0.928	0.934
African American	0.814	0.790	0.838	0.889	0.894	0.907
American Indian or Alaska Native	0.842	0.841	0.852	0.906	0.905	0.919
Asian	0.886	0.879	0.916	0.936	0.938	0.940
Hispanic/Latino	0.845	0.837	0.881	0.914	0.919	0.929
Multi-Racial	0.854	0.835	0.882	0.919	0.921	0.924
Native Hawaiian or Other Pacific Islander	0.871	0.777	0.905	0.913	0.916	0.938
White	0.866	0.851	0.890	0.920	0.924	0.925
Economically Disadvantaged: No	0.873	0.853	0.896	0.919	0.926	0.924
Economically Disadvantaged: Yes	0.829	0.815	0.857	0.902	0.906	0.916
English Learner: No	0.864	0.849	0.888	0.919	0.923	0.927
English Learner: Yes	0.733	0.715	0.730	0.821	0.759	0.844
Regular Education	0.864	0.847	0.887	0.917	0.922	0.926
Special Education	0.823	0.806	0.841	0.873	0.860	0.887
Section 504: No	0.865	0.850	0.889	0.920	0.923	0.928
Section 504: Yes	0.819	0.814	0.860	0.895	0.899	0.912
Migrant: No	0.863	0.849	0.888	0.919	0.923	0.928
Migrant: Yes	0.847	0.828	0.807	0.922	0.899	0.921
Homeless: No	0.863	0.849	0.888	0.919	0.923	0.928
Homeless: Yes	0.821	0.806	0.843	0.891	0.902	0.920
Military Affiliation: No	0.862	0.848	0.887	0.919	0.922	0.927
Military Affiliation: Yes	0.874	0.853	0.895	0.921	0.924	0.925
Foster Care: No	0.863	0.849	0.888	0.919	0.923	0.928
Foster Care: Yes	0.825	0.788	0.815	0.885	0.877	0.913

Table F.1.2

SEM for Overall and Subgroups: Spring 2022 Operational Social Studies

Grade	3	4	5	6	7	8
All Students	**3.02**	**3.04**	**3.20**	**3.47**	**3.48**	**3.56**
Female	3.03	3.05	3.22	3.50	3.51	3.59
Male	3.00	3.03	3.18	3.44	3.45	3.52
African American	2.99	3.02	3.11	3.44	3.44	3.56
American Indian or Alaska Native	3.04	3.05	3.24	3.49	3.53	3.65
Asian	2.97	2.98	3.27	3.44	3.43	3.44
Hispanic/Latino	3.00	3.04	3.17	3.45	3.48	3.57
Multi-Racial	3.04	3.05	3.24	3.46	3.51	3.59
Native Hawaiian or Other Pacific Islander	3.01	3.11	3.20	3.58	3.56	3.51
White	3.03	3.05	3.27	3.48	3.50	3.55
Economically Disadvantaged: No	3.02	3.03	3.28	3.47	3.50	3.53
Economically Disadvantaged: Yes	3.01	3.04	3.16	3.45	3.46	3.56
English Learner: No	3.02	3.05	3.21	3.47	3.48	3.57
English Learner: Yes	2.96	3.00	2.98	3.29	3.26	3.38
Regular Education	3.02	3.05	3.22	3.49	3.49	3.56
Special Education	2.99	2.99	3.02	3.31	3.27	3.39
Section 504: No	3.02	3.04	3.21	3.47	3.50	3.57
Section 504: Yes	3.02	3.04	3.12	3.42	3.42	3.52
Migrant: No	3.02	3.04	3.20	3.47	3.48	3.56
Migrant: Yes	2.93	3.05	3.10	3.51	3.40	3.55
Homeless: No	3.02	3.04	3.20	3.48	3.48	3.56
Homeless: Yes	2.97	3.01	3.11	3.43	3.42	3.53
Military Affiliation: No	3.02	3.04	3.20	3.46	3.49	3.57
Military Affiliation: Yes	3.02	3.03	3.27	3.48	3.52	3.51
Foster Care: No	3.02	3.04	3.20	3.47	3.48	3.56
Foster Care: Yes	3.00	3.05	3.11	3.45	3.43	3.45

Table F.2

Cronbach's Alpha and Marginal Reliability: Spring 2022 Operational SS G3–8

Grade	Cronbach's Alpha	Marginal Reliability
3	0.863	0.860
4	0.849	0.850
5	0.888	0.880
6	0.919	0.920
7	0.923	0.930
8	0.928	0.920

Table F.3

Classification Accuracy and Decision Consistency: Spring 2022 Operational SS G3–8

Accuracy Matrix: Grade 3

Grade	Level	Unsatisfactory (1)	Approaching Basic (2)	Basic (3)	Mastery (4)	Advanced (5)	Total
3	1	0.25	0.04	0.00	0.00	0.00	0.29
	2	0.04	0.16	0.05	0.00	0.00	0.26
	3	0.00	0.06	0.12	0.04	0.00	0.23
	4	0.00	0.01	0.06	0.10	0.06	0.22
	5	0.00	0.00	0.00	0.00	0.00	0.00
	Total	0.29	0.27	0.23	0.15	0.06	1.00

Accuracy Matrix: Grade 4

Grade	Level	Unsatisfactory (1)	Approaching Basic (2)	Basic (3)	Mastery (4)	Advanced (5)	Total
4	1	0.21	0.04	0.00	0.00	0.00	0.25
	2	0.04	0.14	0.05	0.00	0.00	0.22
	3	0.00	0.06	0.11	0.05	0.00	0.22
	4	0.00	0.01	0.08	0.15	0.07	0.30
	5	0.00	0.00	0.00	0.00	0.00	0.00
	Total	0.25	0.25	0.24	0.20	0.07	1.00

Accuracy Matrix: Grade 5

Grade	Level	Unsatisfactory (1)	Approaching Basic (2)	Basic (3)	Mastery (4)	Advanced (5)	Total
5	1	0.27	0.04	0.00	0.00	0.00	0.32
	2	0.03	0.12	0.04	0.00	0.00	0.20
	3	0.00	0.06	0.13	0.04	0.00	0.23
	4	0.00	0.00	0.06	0.15	0.05	0.26
	5	0.00	0.00	0.00	0.00	0.00	0.00
	Total	0.31	0.22	0.23	0.19	0.05	1.00

Accuracy Matrix: Grade 6

Grade	Level	Unsatisfactory (1)	Approaching Basic (2)	Basic (3)	Mastery (4)	Advanced (5)	Total
6	1	0.31	0.04	0.00	0.00	0.00	0.34
	2	0.03	0.18	0.04	0.00	0.00	0.25
	3	0.00	0.05	0.14	0.03	0.00	0.22
	4	0.00	0.00	0.05	0.09	0.04	0.18
	5	0.00	0.00	0.00	0.00	0.00	0.00
	Total	0.34	0.26	0.23	0.12	0.05	1.00

Accuracy Matrix: Grade 7

Grade	Level	Unsatisfactory (1)	Approaching Basic (2)	Basic (3)	Mastery (4)	Advanced (5)	Total
7	1	0.27	0.04	0.00	0.00	0.00	0.30
	2	0.03	0.10	0.04	0.00	0.00	0.17
	3	0.00	0.04	0.12	0.04	0.00	0.21
	4	0.00	0.00	0.04	0.14	0.03	0.21
	5	0.00	0.00	0.00	0.02	0.08	0.11
	Total	0.30	0.18	0.20	0.20	0.11	1.00

Accuracy Matrix: Grade 8

Grade	Level	Unsatisfactory (1)	Approaching Basic (2)	Basic (3)	Mastery (4)	Advanced (5)	Total
8	1	0.23	0.03	0.00	0.00	0.00	0.26
	2	0.03	0.13	0.04	0.00	0.00	0.19
	3	0.00	0.04	0.13	0.03	0.00	0.21
	4	0.00	0.00	0.05	0.16	0.04	0.25
	5	0.00	0.00	0.00	0.03	0.06	0.09
	Total	0.26	0.20	0.22	0.23	0.10	1.00

Consistency Matrix: Grade 3

Grade	Level	Unsatisfactory (1)	Approaching Basic (2)	Basic (3)	Mastery (4)	Advanced (5)	Total
3	1	0.24	0.06	0.01	0.00	0.00	0.31
	2	0.04	0.13	0.06	0.01	0.00	0.25
	3	0.01	0.06	0.09	0.05	0.01	0.22
	4	0.00	0.02	0.06	0.07	0.03	0.17
	5	0.00	0.00	0.01	0.02	0.02	0.05
	Total	0.29	0.27	0.23	0.15	0.06	1.00

Consistency Matrix: Grade 4

Grade	Level	Unsatisfactory (1)	Approaching Basic (2)	Basic (3)	Mastery (4)	Advanced (5)	Total
4	1	0.20	0.06	0.01	0.00	0.00	0.27
	2	0.04	0.11	0.06	0.01	0.00	0.22
	3	0.01	0.06	0.09	0.06	0.01	0.22
	4	0.00	0.02	0.07	0.10	0.04	0.23
	5	0.00	0.00	0.01	0.03	0.02	0.06
	Total	0.25	0.25	0.24	0.20	0.07	1.00

Consistency Matrix: Grade 5

Grade	Level	Unsatisfactory (1)	Approaching Basic (2)	Basic (3)	Mastery (4)	Advanced (5)	Total
5	1	0.26	0.06	0.01	0.00	0.00	0.33
	2	0.04	0.09	0.05	0.01	0.00	0.19
	3	0.01	0.06	0.10	0.05	0.01	0.22
	4	0.00	0.01	0.07	0.11	0.03	0.23
	5	0.00	0.00	0.00	0.02	0.01	0.03
	Total	0.31	0.22	0.23	0.19	0.05	1.00

Consistency Matrix: Grade 6

Grade	Level	Unsatisfactory (1)	Approaching Basic (2)	Basic (3)	Mastery (4)	Advanced (5)	Total
6	1	0.30	0.05	0.00	0.00	0.00	0.36
	2	0.04	0.15	0.05	0.00	0.00	0.24
	3	0.00	0.06	0.11	0.04	0.01	0.22
	4	0.00	0.01	0.06	0.06	0.03	0.15
	5	0.00	0.00	0.01	0.02	0.01	0.03
	Total	0.34	0.26	0.23	0.12	0.05	1.00

Consistency Matrix: Grade 7

Grade	Level	Unsatisfactory (1)	Approaching Basic (2)	Basic (3)	Mastery (4)	Advanced (5)	Total
7	1	0.26	0.05	0.01	0.00	0.00	0.31
	2	0.04	0.08	0.04	0.01	0.00	0.17
	3	0.01	0.05	0.10	0.05	0.00	0.20
	4	0.00	0.01	0.05	0.11	0.04	0.20
	5	0.00	0.00	0.00	0.04	0.07	0.12
	Total	0.30	0.18	0.20	0.20	0.11	1.00

Consistency Matrix: Grade 8

Grade	Level	Unsatisfactory (1)	Approaching Basic (2)	Basic (3)	Mastery (4)	Advanced (5)	Total
8	1	0.23	0.04	0.00	0.00	0.00	0.27
	2	0.03	0.10	0.05	0.00	0.00	0.19
	3	0.00	0.05	0.10	0.04	0.00	0.20
	4	0.00	0.01	0.06	0.13	0.05	0.24
	5	0.00	0.00	0.00	0.05	0.05	0.10
	Total	0.26	0.20	0.22	0.23	0.10	1.00

Table F.3.1

Estimates of Accuracy and Consistency of Achievement Level Classification

Grade	Accuracy	Consistency	PChance	Kappa
3	0.634	0.540	0.238	0.397
4	0.608	0.508	0.225	0.365
5	0.665	0.572	0.247	0.431
6	0.719	0.633	0.256	0.507
7	0.711	0.617	0.221	0.509
8	0.708	0.612	0.218	0.504

Table F.3.2

Accuracy of Classification at Each Achievement Level

Grade	Unsatisfactory (1)	Approaching Basic (2)	Basic (3)	Mastery (4)	Advanced (5)
3	0.850	0.641	0.523	0.452	0.000
4	0.836	0.616	0.498	0.493	0.000
5	0.862	0.595	0.560	0.568	0.000
6	0.896	0.711	0.641	0.492	0.000
7	0.879	0.590	0.590	0.660	0.767
8	0.889	0.654	0.638	0.638	0.648

Table F.3.3

Accuracy of Dichotomous Categorizations by Form (PAC Metric)

Grade	1 / 2+3+4+5	1+2 / 3+4+5	1+2+3 / 4+5	1+2+3+4 / 5
3	0.915	0.879	0.885	0.939
4	0.921	0.877	0.861	0.929
5	0.919	0.889	0.894	0.950
6	0.932	0.910	0.919	0.953
7	0.929	0.913	0.920	0.943
8	0.943	0.919	0.916	0.927

Table F.3.4

Consistency of Dichotomous Categorizations by Form (PAC Metric)

Grade	1 / 2+3+4+5	1+2 / 3+4+5	1+2+3 / 4+5	1+2+3+4 / 5
3	0.878	0.833	0.839	0.918
4	0.885	0.831	0.806	0.898
5	0.884	0.849	0.849	0.935
6	0.903	0.875	0.885	0.942
7	0.899	0.879	0.887	0.919
8	0.918	0.887	0.882	0.902

Table F.3.5

Kappa of Dichotomous Categorizations by Form (PAC Metric)

Grade	1 / 2+3+4+5	1+2 / 3+4+5	1+2+3 / 4+5	1+2+3+4 / 5
3	0.716	0.662	0.538	0.174
4	0.710	0.661	0.530	0.152
5	0.740	0.697	0.608	-0.009
6	0.788	0.740	0.620	0.083
7	0.766	0.758	0.740	0.608
8	0.794	0.773	0.737	0.444

Table F.3.6

Accuracy of Dichotomous Categorizations: False Positive Rates (PAC Metric)

Grade	1/ 2+3+4+5	1+2 / 3+4+5	1+2+3 / 4+5	1+2+3+4 / 5
3	0.044	0.054	0.050	0.061
4	0.042	0.052	0.054	0.071
5	0.044	0.048	0.042	0.050
6	0.036	0.041	0.032	0.047
7	0.037	0.040	0.040	0.032
8	0.029	0.040	0.035	0.043

Table F.3.7

Accuracy of Dichotomous Categorizations: False Negative Rates (PAC Metric)

Grade	1 / 2+3+4+5	1+2 / 3+4+5	1+2+3 / 4+5	1+2+3+4 / 5
3	0.041	0.067	0.065	0.000
4	0.037	0.072	0.085	0.000
5	0.037	0.063	0.064	0.000
6	0.032	0.049	0.048	0.000
7	0.034	0.047	0.040	0.025
8	0.028	0.042	0.049	0.030

Appendix G: Guidelines for Accommodated Print and Braille

Louisiana believes that all students requiring test accommodations should be presented with the same rigor as students taking tests without accommodations. To ensure this, Louisiana accommodates the operational test form for each test administration, allowing all students to take the same items regardless of the need for an accommodated presentation. Careful consideration is given to all items that are used for Louisiana assessments for their ability to be faithfully represented in accommodated print (AP) and/or braille formats. Fairness for all populations, item integrity, and student-item interaction for technology-enhanced (TE) items are all factors when selecting the items that will appear on a Louisiana form. TE items are modified so that students who interact with an item on an AP or braille form will have a similar and equivalent experience to students who interact with that same item in the online environment. This maintains both the rigor and the content being assessed. Some examples of the modification process are provided below.

- Drag-and-drop items in the online environment require a student to place the answer options in an interactive table. For the AP and braille forms, the student is presented with a table with the same information as the interactive table (column or row headers, any completed cells, and blank spaces) and the answer options are listed below the table (similar to the online form in which the options are listed either below or to the right of the table). The directions are modified to ask the student to write the correct answer in its corresponding box. Students are also able to circle the text and draw arrows to indicate where it should be placed or add labels to the answer choices and write only the label in the box, as long as the intended response is clear to the test administrator who will transcribe the answers into the online system.
- Matching items in the online environment require a student to select a checkbox in one or more columns for each of multiple rows. In the AP and braille forms, the student is provided with a table and asked to mark an X in the correct places.
- Highlight-text items or item parts in the online environment require a student to click on the selected text, which highlights the selected word, phrase, or sentence. In the AP and braille forms, the text is presented in the same format

and the student is asked to circle the answer. Where only certain words or phrases are selectable in the online system, those options are underlined in the AP and braille forms to indicate which words and/or phrases the student should select from.

- Drop-down menu items in the online environment have answer options in a drop-down menu format, oftentimes as part of a complete sentence. The AP and braille forms display the item with a blank line in place of the drop-down menu in the sentence, with all the answer options for the drop-down menu presented vertically below the sentence. The directions are then modified to ask the student to circle the word/phrase that belongs in the blank.
- Short-answer items in the online environment require a student to type the answer in a box. In the AP and braille forms, a box is provided for the student to write the response.
- Keypad input items in the online environment require a student to enter a numeric response including all rational and irrational numbers as well as expressions and equations. In the AP and braille forms, a box is provided for the student to write the response.
- Graphing items, including coordinate planes, number lines, line plots, and bar graphs, in the online environment require a student to complete a graph by plotting points, adding Xs to create a line plot, or raising/lowering bars to create a bar graph or histogram. In the AP and braille forms, the student is provided with the same coordinate plane, number line, line plot, or bar graph as in the online item, including titles, axis labels, and keys, and is asked to complete the graph.

Displaying items similarly in accommodated print and braille forms and in the online environment (and allowing students to interact with the items in a similar manner) maintains item integrity by assessing a similar construct in a similar manner regardless of where a student encounters an item. This provides students who are unable to access the assessment online with an assessment at the same level of rigor as the online test.

AP forms are thoroughly reviewed by LDOE and DRC content experts, and braille forms are reviewed by an outside third-party braille expert. Students respond to their accommodated print and braille test using the same online test as used by the general population, either through use of a scribe or by themselves if able. This ensures a valid and reliable assessment for students who are unable to participate in the online assessment.

Appendix H: Ongoing Quality Control

A system for monitoring, maintaining, and increasing the quality of its assessment system, including precise and technically sound criteria for the analyses of all of the assessments in its assessment system, is crucial and critical for keeping a high quality of assessments.

The places where information about monitoring, maintaining, and improving quality is incorporated are included in the following table.

Related Information		**Related Chapter/Source**
Test Materials		
Item development quality procedures	Content alignment Cognitive complexity Bias, fairness, and sensitivity Technical design	Chapter 3
Form development quality procedures	Test specifications Review of statistical quality of items	Chapter 4
Test Administration		
Test administration training and procedures	Training and monitoring of test administrators Security Checklists Test Security Measurements	Chapter 5
Monitoring test administrations	LDOE site audits Data Forensics Analysis Response-Change Analysis Web Monitoring Plagiarism Detection	Chapter 5
Scoring		
Scorer recruitment, training and security procedures	Recruitment and interview process Security Training process, including material development and qualifying procedures	Chapter 6
Monitoring scoring quality	Inter-rater reliability studies Validity Reader monitoring	Chapter 6
Psychometric Processes		
Psychometric quality procedures	Specifications document for operational analysis	Internal document between Pearson and the LDOE
Monitoring psychometric quality	Key verification Calibration Scoring table generation Psychometric quality checks on the data	Chapter 7
Cuts based on Performance-Level Setting	Quality-controlled procedures for performance-level setting Derivation of the cut scores	Chapter 8