

Appendices for
CMAS Technical Report
Spring 2014

List of Appendices

APPENDIX A: STANDARD SETTING REPORT3

APPENDIX B: SAMPLE CMAS SCORE REPORTS61

APPENDIX C: IRT CURVES.....68

APPENDIX D: COGNITIVE LAB REPORT81

APPENDIX A: STANDARD SETTING REPORT

Colorado Measures of Academic Success (CMAS) Spring 2014 Standard Setting Report



COLORADO DEPARTMENT *of* EDUCATION

&

PEARSON

October 18, 2014

Table of Contents

OVERVIEW.....	4
PREPARATION FOR STANDARD SETTING.....	4
PANELIST SELECTION AND COMPOSITION.....	4
DEVELOPMENT OF PLDs.....	5
CREATION OF MATERIALS	5
<i>Slides and Script</i>	<i>5</i>
<i>OIB.....</i>	<i>5</i>
<i>Rubrics and Sample Responses.....</i>	<i>5</i>
<i>P-value Reports.....</i>	<i>6</i>
<i>External Data.....</i>	<i>6</i>
<i>Forms</i>	<i>6</i>
TRAINING OF FACILITATORS AND DATA ANALYSTS	6
STANDARD SETTING MEETING ACTIVITIES.....	7
GENERAL SESSION	7
THE STANDARD SETTING PROCESS	7
<i>Review and Discuss Performance Level Descriptors</i>	<i>7</i>
<i>Development of Threshold Descriptors</i>	<i>7</i>
<i>Review Test Questions</i>	<i>8</i>
<i>Standard Setting Training and Practice Round.....</i>	<i>8</i>
<i>Readiness Survey</i>	<i>9</i>
<i>Round 1</i>	<i>9</i>
<i>Round 1 Feedback.....</i>	<i>9</i>
<i>Round 2.....</i>	<i>10</i>
<i>Round 2 Feedback.....</i>	<i>10</i>
<i>Round 3.....</i>	<i>10</i>
<i>Evaluation.....</i>	<i>10</i>
ROUND 3 RECOMMENDED CUT SCORES.....	11
VERTICAL ARTICULATION.....	12
PARTICIPANTS.....	13
VERTICAL ARTICULATION PROCESS.....	13
<i>Review of PLDs.....</i>	<i>13</i>
<i>Discuss Expectations</i>	<i>13</i>
<i>Review and Discuss Impact Data across Grades</i>	<i>13</i>
<i>Establish Shared Recommendation.....</i>	<i>13</i>
<i>Evaluation.....</i>	<i>14</i>

VERTICAL ARTICULATION RECOMMENDED CUT SCORES	14
REFERENCES.....	16
APPENDIX A: PANEL COMPOSITION.....	17
APPENDIX B: PERFORMANCE LEVEL DESCRIPTORS	19
APPENDIX C: OIB METADATA SPREADSHEETS	24
APPENDIX D: P-VALUE REPORTS	37
APPENDIX E: EXTERNAL DATA	42
APPENDIX F: READINESS SURVEY	45
APPENDIX G: BOOKMARK RECOMMENDATION FORMS.....	47
APPENDIX H: STANDARD SETTING EVALUATION	49
APPENDIX I: VERTICAL ARTICULATION PARTICIPANT COMPOSITION	53
APPENDIX J: VERTICAL ARTICULATION EVALUATION FORM.....	55

OVERVIEW

Colorado Measures of Academic Success (CMAS) is a newly developed standards-based assessment designed to measure what students should know and be able to demonstrate at each grade level. It was first administered in spring 2014 and standards were subsequently set in July in order to aid the interpretability of scores. The purpose of this document is to provide a detailed report of the standard setting process for the spring 2014 administration of grades 4, 5, 7, and 8.

CMAS is aligned with the Colorado Academic Standards (CAS) for Science and Social Studies (located at <http://www.cde.state.co.us/coscience/statestandards> and <http://www.cde.state.co.us/cosocialstudies/statestandards>, respectively). Each test contained selected-response items (SR), a variety of technology-enhanced items (TEI), and constructed-response items (CR). The subject and grade combinations for CMAS are shown in Table 1. The first operational administration for grades 4, 5, 7, and 8 was in April 2014 and for high school in November of 2014. The majority of students took the assessment online with the paper test serving as an accommodated form for a very small percentage of students.

Table 1: CMAS Subjects and Grades

	Grade				
	4	5	7	8	HS
Science		X		X	X
Social Studies	X		X		X

To support the interpretation of student results, student performance on the CMAS is described in terms of four performance levels: Limited Command, Moderate Command, Strong Command, and Distinguished Command. The standard setting meeting was held in order to obtain cut score recommendations to assist the state in delineating thresholds for each of these four levels.

The methodology implemented was the commonly-used Bookmark, or Item-Mapping, method (Lewis, Mitzel, Green & Patz, 1999). This method is an item response theory-based item mapping procedure and makes use of an Ordered Item Book (OIB)—a collection of items ordered by difficulty. Panelists use performance level descriptors (PLDs) to conceptualize “threshold” students (those students just barely in a particular performance level) in order to determine the appropriate location of each cut score.

PREPARATION FOR STANDARD SETTING

Preparation for the standard setting started months before the actual meeting. This section provides details about the selection of panelists, the development of performance level descriptors (PLDs), the various materials that were created for the meeting, and the training of those who facilitated the meeting and analyzed the data.

Panelist Selection and Composition

The standard setting meeting included between 11 and 13 panelists for each grade grouped in tables of three or four. Panelists were selected for participation by the Colorado Department of

Education (CDE) to represent the state in terms of gender and ethnicity as well as relevant demographic characteristics (e.g., school size, geographic location). The majority of panelists for a given grade were teachers, and while most were experts at that particular grade level, teachers a grade level below and a grade level above also participated. In addition, there was administrator, Special Education, and English Language Learners (ELL) representation at each grade level. Appendix A describes panel composition for each grade level.

Development of PLDs

PLDs are an important tool for the Bookmark method. Prior to the standard setting meeting, PLDs were developed by Pearson content experts and then reviewed and edited by a committee of Colorado educators. PLDs are provided in Appendix B.

Creation of Materials

A standard setting meeting requires a considerable amount of materials, some paper-based and some electronic. This section outlines the primary materials and points to where the documents are provided.

Slides and Script

There were two main components of the meeting: the general session and the breakout sessions. For the general session, a PowerPoint presentation was created to provide a general overview of the meeting for all panelists in a large-group setting.

For the breakout sessions (where each grade/subject is in a separate room), slides and accompanying detailed scripts were developed. Because it is important that the process be standardized for each grade/subject, slides and associated script allow for the breakout sessions to be run in parallel fashion.

OIB

Since CMAS is primarily an online assessment and contains item types that require an online format to fully experience them (i.e., technology-enhanced items and simulations), the OIB was presented to panelists online. All operational items that appeared on the spring 2014 assessment were included in the OIB along with a handful of field test items to fill any gaps. Each item was presented on a separate page in item difficulty order according to its scale location using a response probability (RP) of 0.67. In addition, a metadata spreadsheet was provided indicating each page number, item ID, item type, content alignment, key (for MC items), and maximum points. In addition, space was provided for panelists to record their “yes” or “no” for each round. The metadata spreadsheet for each grade can be found in Appendix C.

Rubrics and Sample Responses

A booklet of rubrics and sample responses was created for each grade. The booklet included the rubric for each constructed-response item along with a sample response of each score point.

P-value Reports

As part of the feedback provided to panelists after Round 1 recommendations, p-value reports were provided. For one-point items, the p-value provided indicates the percentage of students who got the item correct during the spring 2014 administration. For constructed-response items, the p-value indicates the percentage of students who earned at least a particular score point during the spring 2014 administration. P-value reports for each grade can be found in Appendix D.

External Data

As part of the feedback provided to panelists after Round 2 recommendations, some external data in the form of percentages by performance level were shared with panelists to provide a point of reference. Transitional Colorado Assessment Program (TCAP) data were provided for all grades. For science, science data were provided; for social studies, both reading and writing data were provided since there is no social studies component of TCAP. Science data from the Trends in International Math and Science Study (TIMSS) were provided for grades 5 and 8. For Grade 8, science data from the National Assessment of Educational Progress (NAEP) were also presented. Data can be seen in Appendix E.

Forms

Numerous forms were created for panelists to complete and include the following:

- **Panelist Information Sheet:** While some demographic information was already included in the database of Colorado educators, the panelist information sheet was used to collect some additional information.
- **Readiness Survey:** A brief questionnaire was provided to panelists before each round of the standard setting process, in which panelists are asked to verify that they understand the task at hand and are ready to move forward. The readiness survey is provided in Appendix F.
- **Bookmark Recommendation Form:** This form was used to collect a panelist's recommendations for each round. It is provided in Appendix G.
- **Standard Setting Evaluation:** An evaluation was administered after the standard setting had been completed to gather information on panelists' perceptions on the meeting. The evaluation and its results are provided in Appendix H.

Training of Facilitators and Data Analysts

Several meetings were held with the facilitators and data analysts to properly train and prepare them for the meeting. For the facilitator training, the breakout session slides and script were walked through in detail and discussed to ensure that all four facilitators were in sync in terms of how to lead the panelists through the standard setting process and the logistics of the meeting. For data analysts, it was important the spreadsheets be set up properly to ensure accurate and rapid analysis of panelists' recommendations. Although not specifically trained for the meeting,

it should be noted that content specialists attended the meeting and were available to answer any content-related questions.

STANDARD SETTING MEETING ACTIVITIES

The CMAS standard setting took place July 14–16. During the first two days, panelists were responsible for placing the bookmark in the Ordered Item Book (OIB) to establish proposed standards, reviewing feedback data, and making final cut-score recommendations. On the morning of the third day, the vertical articulation was held. The specific procedures involved in the implementation of the Bookmark method are described in the sections that follow.

General Session

The meeting began with a session in which all panelists from both subjects convened to listen to introductory comments and receive directions for the meeting. First, a representative from CDE provided the context for the meeting by presenting details on CMAS and describing the importance of standard setting in the assessment development process. Next, a member of Pearson Psychometric Services staff (Dr. Jennifer Beimers) provided a brief overview of the Bookmark standard setting process including the rationale behind the procedure and the types of decisions panelists will be asked to make. Once the general overview was completed, panelists were dismissed to their designated committee rooms.

The Standard Setting Process

The standard setting specific tasks took place over the course of two days as outlined in this section of the report. Each grade was facilitated independently but the same standardized process was used across all grades.

Review and Discuss Performance Level Descriptors

After introductions and general housekeeping tasks were completed, each panelist was provided with a document listing the Performance Level Descriptors (Appendix B). Panelists were asked to review the labels and specific performance level descriptors in light of the content frameworks.

Development of Threshold Descriptors

Panelists were reminded that the main purpose behind reviewing and discussing PLDs was to operationalize the performance levels to *support the standard setting task*. The focus was on the threshold student: those who “just barely” make it into a particular performance level. The goal was to gain a common understanding so that when panelists were asked to think about a threshold student, they were all in agreement regarding what such a student can/cannot do.

To develop the threshold descriptors, panelists were asked to identify concepts and skills in a given PLD that should describe the threshold student. Questions that helped guide the discussion included:

- Do any concepts and skills listed in the PLD do this outright?
- How could you modify or constrain the PLD to better reflect the limited capabilities of the “just-barely” student?
- What should the “threshold” student be able to do relative to these particular skills?

Each table worked together to create specific descriptions that separate students who are just barely in a particular performance level (threshold students) from students who are at the top of the previous performance level. Once drafted at the table level, the entire room shared and discussed their threshold descriptors and agreed on a final set of threshold descriptors for their specific grade. Once final, the threshold descriptors were printed for each panelist to use throughout the remainder of the standard setting activity.

Review Test Questions

Panelists were given time to review the OIB in order to familiarize themselves with the nature of the assessment. This provided an opportunity for panelists to gain an appreciation of the assessment experience, understand the manner in which the content standards are operationalized in test items, and get an overall feel for the difficulty of the test. Panelists were instructed to work on their own to review each of the items in the OIB keeping in mind the concepts and skills required to answer each item correctly. Upon completion, scoring keys for multiple-choice items were provided so that panelists could score their work.

Standard Setting Training and Practice Round

Panelists received detailed training on how to place a bookmark in the ordered item book in order to determine the transition from one performance level to the next. For each performance level, panelists were instructed to work through the OIB to determine the last “yes” page where all preceding items would define the concepts and skills that a *just barely Strong Command* student, for example, is expected to know. It is equivalent to the place in the OIB that accurately divides the items into those that all students at a given level SHOULD, with 2/3 chance or greater, answer correctly from those that they are not expected to answer correctly. The following outlines the specific steps that were to be followed for the “Moderate Command” cut.

1. Think about the skills that characterize a threshold “Moderate Command” student.
2. Start on page 1 of OIB and ask yourself, “SHOULD a threshold ‘Moderate Command’ student have at least a 2/3 chance of answering this item correctly?”
3. If yes, move on to the next item.
4. Do this until you get to your first “no.”
5. Continue on to a couple more items to make sure these are also “no.”
6. Record page associated with last “yes” on your recommendation form.

The same steps were repeated for all “Strong Command” and “Distinguished Command.” Panelists were reminded that since the content standards are new, they may not yet be fully implemented so it was important that panelists consider threshold students who have been instructed in the new standards.

Following the training session, panelists engaged in a practice round of standard setting using a small set of sample items. The purpose of this exercise was to have panelists get a chance to practice placing of their bookmarks and to make sure everyone is comfortable with the task. This practice and training session was followed by a brief group discussion where panelists discussed their ratings and the general process employed. Based on discussion, facilitators provided additional instruction/guidance as needed.

Readiness Survey

To evaluate whether the training activities successfully helped panelists understand the task, a readiness survey was completed by each panelist prior to each round of recommendations (Appendix F). The readiness survey asked panelists to report if they understood the task Pearson facilitators asked of them as well as any feedback data provided. Results of the readiness survey indicated that panelists unanimously understood their tasks for each round and the data presented.

Round 1

After completing the readiness survey, the panelists began Round 1 of the standard setting. Panelists worked independently to determine which items in the OIB separated the performance levels. In reviewing each item, panelists were reminded to ask themselves, “Given the skill required to answer this item correctly, SHOULD a threshold level student answer the item correctly two thirds of the time?” Panelists recorded the page of their recommendation for each level on their Bookmark Recommendation Form (Appendix G), submitted it to the facilitator, and were dismissed for the day.

Round 1 Feedback

To begin Day 2, panelists were provided with several pieces of feedback information. With each piece of data, the panelists were reminded that the data was intended to inform their decisions, but not to dictate them.

First, each table was provided with a summary of their table’s recommendations including the minimum, maximum, mean, standard deviation, and median. Panelists were instructed to consider how close their recommendation is to that of others in the group and discuss why they placed the bookmark where they did. Table-level discussions were had around this information and then the facilitators projected the same statistics at the room level. In addition, a bar chart reflecting the panelist agreement was displayed. During both table-level and room-level discussions, the group tried to determine the factors underlying the variability in recommendations by discussing the items associated with and around the recommended cuts. While panelists were encouraged to reassess their cut recommendations based on these discussions, the main purpose of this activity was to allow panelists to think through and discuss the recommendation process; it was not to arrive at a consensus.

The second report provided to the panelists before Round 2 was the item difficulty (p-values) report (Appendix D). For selected-response items, this report showed the percentage of spring 2014 examinees who answered each item correctly; for constructed-response items, it showed the percentage of spring 2014 examinees who earned at least a particular score point. This report

was intended to be used to validate panelists' perceptions of item difficulty. Panelists were cautioned not to modify their ratings based on the item difficulty data alone.

Round 2

After discussing Round 1 feedback and completing the readiness survey for Round 2, panelists worked independently to re-evaluate their recommendations and decide whether they wanted to revise them. Panelists then recorded their Round 2 recommendations on their Bookmark Recommendation Form and submitted them to the facilitator.

Round 2 Feedback

Three pieces of feedback data were provided based on Round 2 recommendations. As before, panelists were reminded that their recommendations should be grounded in content and what students should know and be able to do, not what they can do or are currently doing.

First, panelists received the same summary statistics as in Round 1, but this time they were based on the page recommendations from Round 2. Table-level and group-level discussions were again conducted around these data.

Second, impact data were provided. Based on Round 2 recommendations, graphs indicating the percentage of students who would score in each of the performance level was displayed. Overall spring 2014 test taker impact was provided but it was also disaggregated by ethnicity (African American, Hispanic, White, and other), gender, socio-economic status (SES), students in special education, and students who are English Language Learners (ELL). Panelists were asked to discuss whether the percentage of students falling in each performance level meets their expectations given what they know about the population of students tested and the test content. Impact data were intended to provide a reasonableness check but panelists were reminded that any modifications to cut score recommendations should be based in content and not driven by impact data.

Third, several sources of external benchmark data were provided where available. To serve as a point of reference, the distribution across performance levels for TCAP, NAEP, and TIMSS was provided (Appendix E). For social studies, TCAP writing and reading information was shared; for grade 5 science, TCAP science and TIMSS science were provided; and for grade 8 science, TCAP, TIMSS, and NAEP science data were displayed. These data were discussed at the room level.

Round 3

After discussing Round 2 feedback and completing the readiness survey for Round 3, panelists worked independently to again re-evaluate their recommendations and decide whether they wanted to revise them. Panelists then recorded their Round 3 recommendations on their Bookmark Recommendation Form and submitted them to the facilitator.

Evaluation

After all panelists were finished and final results were determined, panelists were asked to complete a short evaluation. The evaluation asked about panelists' level of comfort with the

standard setting procedure, their understanding of the performance levels, and their satisfaction with final cut scores. The evaluation and results can be found in Appendix H. Upon completing the evaluations, panelists were thanked for their time and participation and dismissed.

Round 3 Recommended Cut Scores

This section provides results from the standard setting portion of the meeting. Table 2 shows the median of panelists' recommendations by round. There was relatively little fluctuation across rounds.

Table 2. Panelist Recommendations by Round

		Moderate Command	Strong Command	Distinguished Command
Grade 4	Round 1	10	55	71
	Round 2	10	55	69
	Round 3	10	43	62
Grade 7	Round 1	21	50	68
	Round 2	22	48	64
	Round 3	18	47	63
Grade 5	Round 1	28	56	84
	Round 2	27	53	77
	Round 3	23	52	77
Grade 8	Round 1	15	51	77
	Round 2	15	51	74
	Round 3	15	48	73

Based on Round 3 recommendations, Tables 3 and 4 show the percentages of students who would fall into each performance level based on the spring 2014 administration.

Table 3. Round 3 Impact for Social Studies

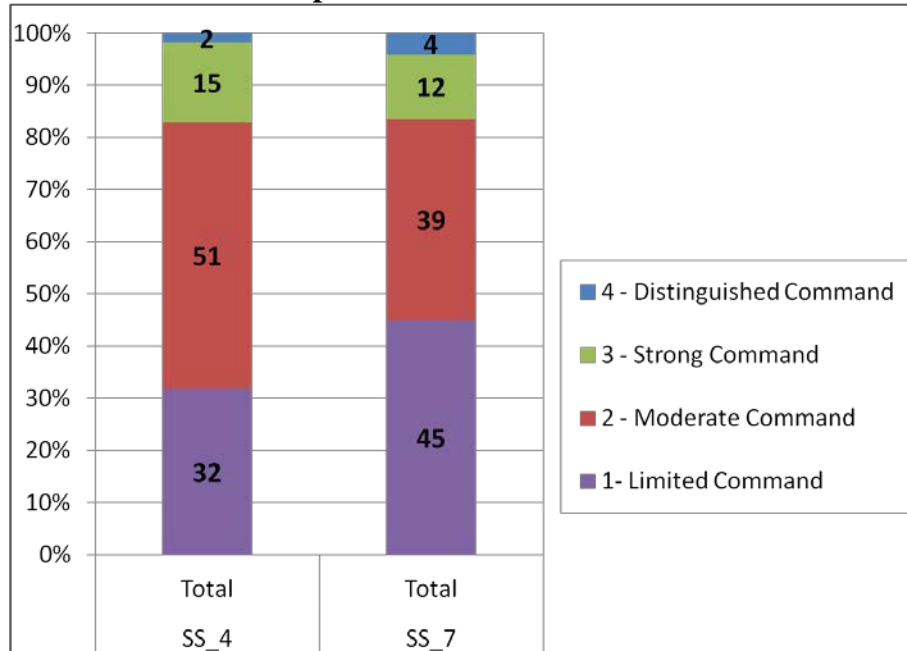
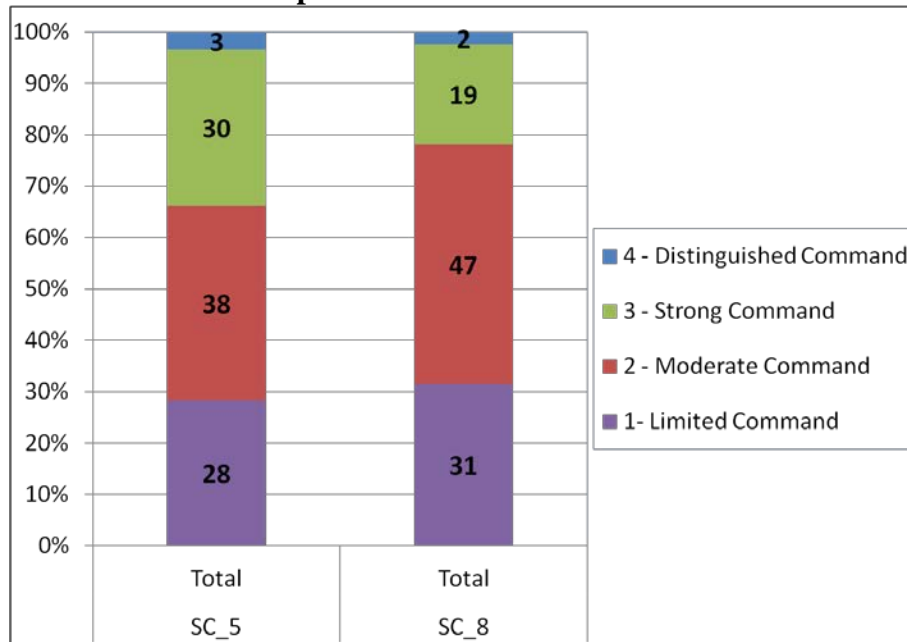


Table 4. Round 3 Impact for Science



VERTICAL ARTICULATION

A subset of standard setting panelists returned the morning of Day 3 to participate in vertical articulation. The purpose of the vertical articulation was to review the impact data associated with the recommended cut scores across both grades within a subject to see if the trend of the

impact data is reasonable given the performance level descriptors, the test-taking population, and the skills/tasks presented on the various assessments.

Participants

One vertical articulation committee was established for each subject by selecting four to five standard setting panelists from each grade level. From each grade, three content experts were selected with the remaining panelists being either administrators or special education representatives. Demographic breakdowns of the committees can be found in Appendix I.

Vertical Articulation Process

The social studies and science committees convened in separate rooms but were facilitated with a standardized process. Parallel slides and scripts were prepared ahead of time to ensure that the same process was used across subjects. The following section outlines the steps of the process.

Review of PLDs

After a brief introduction to the vertical articulation process, participants spent some time reviewing PLDs for both grades within the content area, focusing especially on the grade in which they did not participate in the standard setting. The review of both PLDs helped provide a complete picture of the developmental continuum for the content area.

Discuss Expectations

After reviewing the PLDs, the expectations for impact across the grade levels were discussed as a group. The following questions were posed to the group:

- What are your expectations of the student performance data progression across the grades?
 - Do you expect similar percentages of students in performance levels across grades? Why or why not?
 - Is there a progression of skills in PLDs that suggest differential impact from elementary to middle school?
 - Do populations differ significantly as you move from grade to grade?
 - What other trends might you expect to see and why?

Review and Discuss Impact Data across Grades

After discussing expectations, the impact data associated with the Round 3 recommended cuts from standard setting for each grades were provided in a side-by-side chart. Panelists were then encouraged to discuss how/if cut scores should change to be consistent with impact expectations.

Establish Shared Recommendation

After the discussion, the facilitator discussed the vertical articulation impact recommendation task. Throughout this discussion, it was stressed to panelists that the intent is not to undo all that was done in the standard setting workshops. Rather, the goal was to provide reasonable cut-score

recommendations to policy makers that consider both the content-based recommendations and the expectations about how students should perform across performance levels. However, any desire to change the cuts needed to be justified based on the PLDs and the items in the OIB. Once the group reached a shared recommendation, results were displayed.

Evaluation

To end the meeting, participants completed a brief evaluation. This evaluation asked about participants' level of comfort with the vertical articulation procedure and their satisfaction with final cut score recommendations. The evaluation and results can be seen in Appendix J. Upon completing the evaluations, panelists were thanked for their time and participation and dismissed.

Vertical Articulation Recommended Cut Scores

Few changes were made between Round 3 and Vertical Articulation recommendations. For social studies, no changes were made, as reflected in Table 5. For science, no change was made to grade 5 but the “strong command” cut was adjusted for grade 8, as is reflected in Table 6.

Table 5. Post-Vertical Articulation Impact for Social Studies

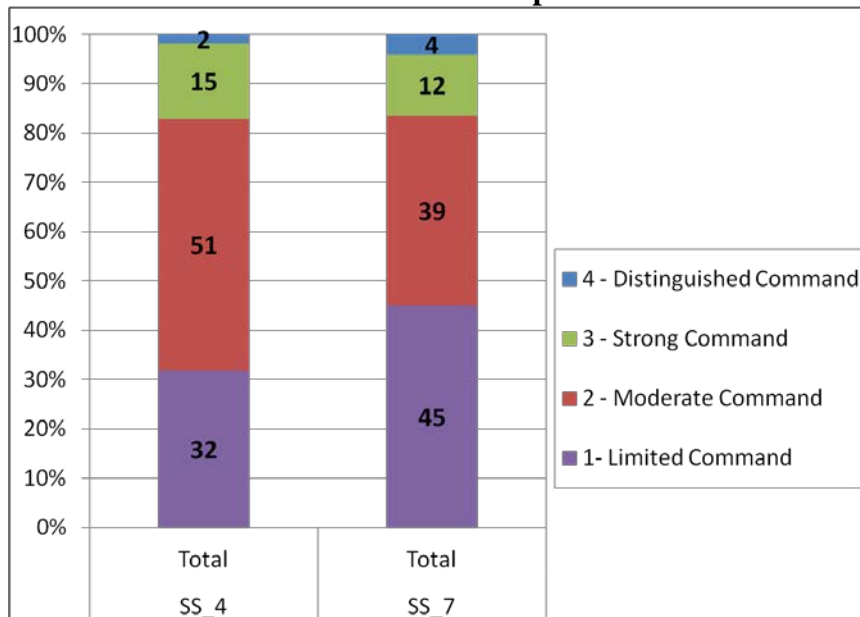
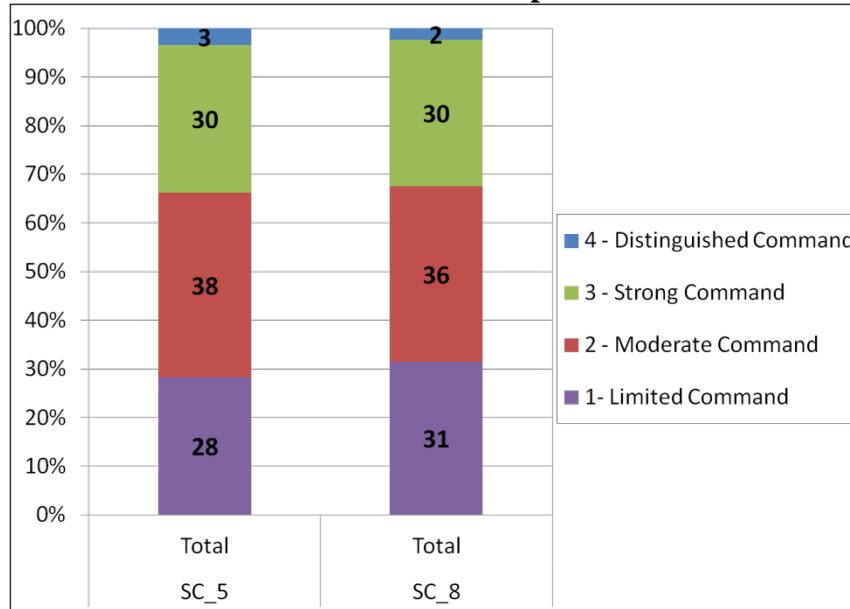


Table 6. Post-Vertical Articulation Impact for Science



Based on the vertical articulation outcomes, the resulting scale score ranges for each performance level can be seen in Table 7.

Table 7. Scale Score Ranges

	Limited Command	Moderate Command	Strong Command	Distinguished Command
Grade 4 Social Studies	300–556	557–698	699–792	793–900
Grade 7 Social Studies	300–591	592–700	701–769	770–900
Grade 5 Science	300–545	546–649	650–770	771–900
Grade 8 Science	300–555	556–651	652–784	785–900

REFERENCES

- Cizek, G.J., & Agger, C.A. (2012). Vertically moderated standard setting. In G.J. Cizek (2nd Ed.) *Setting performance standards: Foundations, methods, and innovations* (pp. 467–484), New York, NY: Routledge.
- Lewis, D. M., Mitzel, H. C., Green, D. R., & Patz, R. J. (1999). *The bookmark standard setting procedure*. Monterey, CA: McGraw-Hill.
- National Academies of Sciences. (2005). *Measuring literacy: Performance levels for adults*, interim report. Washington DC: Author.
- U.S. Department of Education. (2007, December). *Standards and assessments peer review guidance: Information and examples for meeting the requirements of the No Child Left Behind Act of 2001*. Washington DC: Author.
- Williams, N. J., & Schulz, E.M. (2005). *An investigation of response probability (RP) values used in standard setting*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Montreal, Canada.

APPENDIX A: PANEL COMPOSITION

Table 8. Panelist Breakdown by Expertise

	Content Expert	Administrator	Special Ed/ELL	Total
Grade 4 Social Studies	8	1	2	11
Grade 5 Science	7	1	3	11
Grade 7 Social Studies	9	1	2	12
Grade 8 Science	9	2	2	13
Total	33	5	9	47

Table 9. Panelists Breakdown by School Setting

	Rural	Suburban	Urban	Total
Grade 4 Social Studies	4	5	2	11
Grade 5 Science	4	4	3	11
Grade 7 Social Studies	3	4	5	12
Grade 8 Science	3	6	4	13
Total	14	19	14	47

Table 10. Panelists Breakdown by School Type

	Charter School	Innovation School	Neither Charter nor Innovation	District Level	Total
Grade 4 Social Studies	0	1	4	6	11
Grade 5 Science	1	2	6	2	11
Grade 7 Social Studies	2	1	5	4	12
Grade 8 Science	1	1	7	4	13
Total	4	5	22	16	47

Table 11. Panelists Breakdown by Region

	Grade 4 Social Studies	Grade 5 Science	Grade 7 Social Studies	Grade 8 Science	Total
Denver Metro	2	3	5	6	16
North Central	1	0	2	1	4
Northeast	1	1	1	2	5
Northwest	2	1	0	1	4
Pikes Peak	4	3	2	2	11
Southeast	0	1	1	1	3
Southwest	0	2	1	0	3
West Central	1	0	0	0	1
Total	11	11	12	13	47

APPENDIX B: PERFORMANCE LEVEL DESCRIPTORS

Colorado Measures of Academic Success: Grade 4 Social Studies Performance Level Descriptors (PLDs)

Students demonstrate mastery of social studies concepts and 21st century skills aligned to the Colorado Academic Standards at various performance levels. The performance level descriptors are organized in a manner that assumes students demonstrating higher levels of command have mastered the concepts and skills within the lower levels. For example, a student at moderate command also masters the concepts and skills of limited command.

At Distinguished Command, a student typically can

- analyze primary source documents and connect the various eras and events in Colorado history to events in U.S. and World History;
- use geographic tools to investigate and analyze settlement patterns, how people adapt to and modify the physical environment, and how places in Colorado have changed over time;
- analyze opportunity costs and ways to reduce financial risk to make financial decisions; and
- analyze multiple perspectives on an issue and provide solutions.

At Strong Command, a student typically can

- explain cause-and-effect relationships present in Colorado history using historical tools such as organizing and sequencing events and reading primary sources;
- create and investigate questions about Colorado in relation to other places and examine the connections between the physical environment and human activities such as migration;
- explain how the natural, human, and capital resources of Colorado have influenced the types of goods and services provided;
- analyze opportunity costs and risk to make financial decisions;
- compare arguments for both sides of a public policy debate; and
- explain the origins, structure, and functions of the Colorado government and its relationship with local and federal governments.

At Moderate Command, a student typically can

- describe how the people and cultures who have lived in Colorado have interacted with each other and have affected the development of Colorado;
- describe how Colorado's political structure developed, including the Colorado Constitution and the relationship between state and national government;
- compare the physical geography of Colorado with that of neighboring states and describe how places in Colorado are connected by technology and the movement of goods and services;
- identify and define types of economic incentives, choices, opportunity costs, and risks that individuals face;
- connect goods and services produced throughout Colorado's history to economic incentives; and
- provide examples of civic and political issues faced by the state.

At Limited Command, a student typically can

- recognize that major political and cultural groups have affected the development of Colorado;
- use maps, grids, and other geographic tools to answer questions about Colorado;
- describe various technological developments, including those that affect Colorado industries;
- identify goods and services produced in Colorado; and
- identify the structure and functions of the Colorado government and the services it provides.

Colorado Measures of Academic Success: Grade 5 Science Performance Level Descriptors (PLDs)

Students demonstrate mastery of science concepts and 21st century skills aligned to the Colorado Academic Standards at various performance levels. The performance level descriptors are organized in a manner that assumes students demonstrating higher levels of command have mastered the concepts and skills within the lower levels. For example, a student at moderate command also masters the concepts and skills of limited command.

At Distinguished Command, a student typically can

- evaluate and provide feedback on scientific evidence and reasoning about the separation of mixtures and how separation affects the total weight/mass;
- develop hypotheses about why similarities and differences exist between the body systems and parts of humans, plants, and animals;
- evaluate scientific claims about natural resources, in terms of reasonability and validity; and
- assess and provide feedback, through reasoning based on evidence, on scientific explanations about weather and factors that change Earth's surface.

At Strong Command, a student typically can

- explain why certain procedures that are used to separate simple mixtures work and discuss any unexpected results;
- evaluate evidence and models of the structure and functions of human, plant, and animal organs and organ systems;
- investigate and generate evidence that human systems are interdependent;
- analyze and interpret data to explore concerns associated with natural resources; and
- formulate testable questions and scientific explanations around weather and factors that change Earth's surface.

At Moderate Command, a student typically can

- discuss how the mass/weight of a mixture is a sum of its parts and design a procedure to separate simple mixtures based on physical properties;
- create models of human, plant, and animal organ systems, and compare and contrast similarities and differences between the organisms;
- explore and describe the origins and usage of natural resources in Colorado; and
- interpret data about Earth, including weather and changes to Earth's surface.

At Limited Command, a student typically can

- select appropriate tools and follow procedures to separate simple mixtures;
- identify how humans, plants, and animals address basic survival needs;
- identify the functions of human body systems;
- distinguish between renewable and nonrenewable resources; and
- use appropriate tools and resources to gather data regarding weather conditions and Earth processes.

Colorado Measures of Academic Success: Grade 7 Social Studies Performance Level Descriptors (PLDs)

Students demonstrate mastery of social studies concepts and 21st century skills aligned to the Colorado Academic Standards at various performance levels. The performance level descriptors are organized in a manner that assumes students demonstrating higher levels of command have mastered the concepts and skills within the lower levels. For example, a student at moderate command also masters the concepts and skills of limited command.

At Distinguished Command, a student typically can

- analyze historical sources while formulating historical questions and defending a thesis;
- use geographic tools to investigate and analyze data to make inferences and predictions regarding regional issues and perspectives in the Eastern Hemisphere;
- demonstrate how supply and demand influence changes in equilibrium price and quantity;
- evaluate how various governments interact and investigate examples of global collaboration; and
- apply various definitions of good government to evaluate the actions of different governments.

At Strong Command, a student typically can

- explain the historical time periods, individuals, groups, ideas, perspectives, themes, and how people are interconnected within regions of the Eastern Hemisphere;
- summarize the development of early civilizations, including Greece, Rome, China, Africa, and the medieval world;
- describe how the physical environment influences economy, culture, and trade patterns;
- explain how resources, production, choices, supply, demand, price, profit, and taxes are related;
- analyze how national and international government policies influence the global community; and
- compare the rights, roles, and responsibilities of citizens in various governments.

At Moderate Command, a student typically can

- describe the contributions of various peoples and cultures in the Eastern Hemisphere;
- compare different physical systems and cultural patterns to describe how different regions and places are interconnected;
- examine multiple points of view and issues in various regions in the Eastern Hemisphere;
- recognize how supply and demand influence price, profit, and production in a market economy;
- compare how taxes affect individual income and spending;
- compare different forms of government in the world and their sources of authority; and
- explain the rights and roles of citizens in various governments.

At Limited Command, a student typically can

- recognize the contributions of various peoples and cultures to the Eastern Hemisphere;
- use geographic tools to answer questions and identify patterns in the Eastern Hemisphere;
- identify factors that cause changes in supply, demand, and price;
- define resources and identify trade patterns based on the distribution of resources; and
- list the responsibilities and roles of citizens in various governments.

Colorado Measures of Academic Success: Grade 8 Science Performance Level Descriptors (PLDs)

Students demonstrate mastery of science concepts and 21st century skills aligned to the Colorado Academic Standards at various performance levels. The performance level descriptors are organized in a manner that assumes students demonstrating higher levels of command have mastered the concepts and skills within the lower levels. For example, a student at moderate command also masters the concepts and skills of limited command.

At Distinguished Command, a student typically can

- design an investigation to predict the movement of an object by examining the forces applied to it;
- use models to predict amounts of energy transferred;
- analyze data and models to support claims about genetic reproduction and traits of individuals;
- use observations and models to develop and communicate a weather prediction; and
- evaluate scientific theories and investigations that explain how the solar system was formed.

At Strong Command, a student typically can

- use mathematical expressions and appropriate information from sources to describe the movement of an object;
- analyze different forms of energy and energy transfer using tools;
- construct an experiment to show mass is conserved;
- investigate the characteristics and behaviors of waves using models, technology, and basic rules of waves;
- analyze human impact on local ecosystems;
- use mathematics to predict the physical traits and genetic makeup of offspring; and
- relate tides, eclipses, lunar phases, and seasons to the motion and positions of the Sun, Earth, and the Moon, using the basic rules of the solar system.

At Moderate Command, a student typically can

- analyze speed and acceleration of moving objects;
- describe different forms of energy and energy transfer;
- use a variety of sources, including popular media and peer-generated explanations, to investigate and describe an environmental issue;
- analyze data and historical research for various weather conditions and compare to historical data for that date and location; and
- investigate and ask testable questions about Earth's different climates using various techniques.

At Limited Command, a student typically can

- distinguish between physical and chemical changes;
- recognize the relationship between pitch and frequency in sound;
- identify human activities that alter the ecosystem;
- recognize that genetic information is passed from one generation to the next;
- compare basic and severe weather conditions and develop an action plan for safety; and
- use tools and simulations to explore the solar system.

APPENDIX C: OIB METADATA SPREADSHEETS

Grade 4 Social Studies OIB Spreadsheet

Page	Item Identifier	Item Type	Max Points	Key	Standard	GLE	Round 1	Round 2	Round 3
1	COSS120007	MC	1	A	Geography	1			
2	COSS120363_1	CR	3		Civics	1			
3	COSS120006	MC	1	B	Geography	1			
4	COSS130066_1	CR	3		History	2			
5	COSS120312	MC	1	A	Economics	1			
6	COSS120362_1	CR	3		Civics	1			
7	SS040015-SSS04002	MC	1	B	Geography	2			
8	COSS120363_2	CR	3		Civics	1			
9	COSS120005	MC	1	A	Geography	1			
10	SS040080	XI	1		Economics	1			
11	COSS120325_1	CR	3		Civics	2			
12	COSS130095_1	CR	3		Geography	2			
13	COSS130101_1	CR	3		History	1			
14	COSS130091	XI	1		History	1			
15	COSS120318	XI	1		History	2			
16	SS040079	XI	1		Economics	1			
17	COSS130013	XI	1		Economics	2			
18	SS040067	XI	1		History	2			
19	SS040013-SSS04002	MC	1	A	History	1			
20	COSS130073	MC	1	D	Geography	2			
21	COSS130004	MC	1	B	Economics	2			
22	SS040076	MC	1	B	Economics	2			
23	COSS120320_1	CR	3		Economics	1			
24	SS040016-SSS04002	MC	1	C	Economics	1			
25	COSS120181	MC	1	D	Economics	1			
26	COSS130041	MC	1	A	Geography	1			
27	COSS130105_1	CR	3		History	1			
28	COSS130300_1	CR	3		Geography	1			
29	COSS130095_2	CR	3		Geography	2			
30	COSS130097_1	CR	3		History	2			
31	COSS120309	MC	1	C	Geography	2			
32	COSS120008	MC	1	D	Economics	2			
33	COSS130099	MC	1	A	History	1			
34	COSS130055	XI	1		Geography	1			
35	COSS120069	MC	1	A	History	2			
36	COSS120304	XI	1		History	1			

Page	Item Identifier	Item Type	Max Points	Key	Standard	GLE	Round 1	Round 2	Round 3
37	COSS130066_2	CR	3		History	2			
38	SS040088	MC	1	C	History	2			
39	COSS120004	MC	1	C	History	2			
40	COSS130104_1	CR	3		Economics	2			
41	COSS120310	XI	1		Geography	2			
42	SS040068	XI	1		Civics	2			
43	COSS120184	XI	1		Civics	1			
44	COSS120010	MC	1	D	Civics	1			
45	COSS120325_2	CR	3		Civics	2			
46	COSS120314	XI	1		Economics	2			
47	COSS120324_1	CR	3		Civics	2			
48	COSS120180	MC	1	D	Economics	1			
49	COSS130043	MC	1	B	Economics	1			
50	COSS130053	XI	1		Economics	1			
51	COSS120315	MC	1	A	Economics	2			
52	COSS130044	MC	1	A	Geography	2			
53	COSS120057	XI	1		Geography	2			
54	COSS130048	MC	1	C	Civics	2			
55	COSS130097_2	CR	3		History	2			
56	COSS130300_2	CR	3		Geography	1			
57	COSS120362_2	CR	3		Civics	1			
58	COSS130101_2	CR	3		History	1			
59	COSS130100	XI	1		Geography	1			
60	COSS120320_2	CR	3		Economics	1			
61	COSS130008	MC	1	C	Civics	1			
62	COSS130049	MC	1	C	Civics	2			
63	COSS120362_3	CR	3		Civics	1			
64	COSS130095_3	CR	3		Geography	2			
65	COSS120324_2	CR	3		Civics	2			
66	COSS130104_2	CR	3		Economics	2			
67	COSS120179	MC	1	B	Geography	2			
68	COSS120325_3	CR	3		Civics	2			
69	COSS130105_2	CR	3		History	1			
70	COSS130066_3	CR	3		History	2			
71	COSS120363_3	CR	3		Civics	1			
72	COSS130300_3	CR	3		Geography	1			
73	COSS130101_3	CR	3		History	1			

Page	Item Identifier	Item Type	Max Points	Key	Standard	GLE	Round 1	Round 2	Round 3
74	COSS120320_3	CR	3		Economics	1			
75	COSS130097_3	CR	3		History	2			
76	COSS130104_3	CR	3		Economics	2			
77	COSS120324_3	CR	3		Civics	2			
78	COSS130047	MC	1	B	Civics	2			
79	COSS130096	MC	1	A	Geography	1			
80	COSS130105_3	CR	3		History	1			
81	COSS130102	XI	1		History	1			

Grade 5 Science OIB Spreadsheet

Page	Item Identifier	Item Type	Max Points	Key	Standard	GLE	Round 1	Round 2	Round 3
1	COSC130292	XI	1		Earth Systems	3			
2	COSC120146	XI	1		Life	1			
3	COSC130086_1	CR	2		Earth Systems	1			
4	COSC120264	MC	1	D	Physical	1			
5	COSC130048_1	CR	2		Life	1			
6	SC050144	MC	1	D	Earth Systems	1			
7	COSC130069	MC	1	D	Life	1			
8	COSC130154	MC	1	D	Life	2			
9	COSC130051	MC	1	B	Life	2			
10	COSC120027	MC	1	D	Earth Systems	1			
11	COSC130089	MC	1	D	Physical	1			
12	COSC130088_1	CR	2		Life	1			
13	COSC130294	XI	1		Life	2			
14	COSC120015	MC	1	C	Life	2			
15	COSC130046_1	CR	2		Life	1			
16	COSC130066_1	CR	2		Earth Systems	2			
17	COSC120050	MC	1	B	Earth Systems	3			
18	COSC120012	MC	1	D	Life	1			
19	COSC120024	MC	1	A	Life	1			
20	COSC120104	MC	1	B	Earth Systems	2			
21	COSC130163	MC	1	B	Earth Systems	3			
22	SC050115	MC	1	A	Life	2			
23	COSC130196_1	CR	2		Life	2			
24	COSC130291	XI	1		Earth Systems	3			
25	COSC130289	XI	1		Life	1			
26	COSC120122	MC	1	B	Physical	1			
27	SC050074	MC	1	C	Earth Systems	3			
28	COSC120170_1	CR	2		Physical	1			
29	COSC130157	MC	1	A	Earth Systems	1			
30	COSC130087_1	CR	2		Earth Systems	1			
31	SC050065	MC	1	B	Life	1			
32	SC050134	XI	1		Life	2			
33	COSC130290	XI	1		Physical	1			
34	COSC120160_1	CR	3		Earth Systems	3			
35	COSC120253	MC	1	B	Physical	1			
36	COSC120155_1	CR	3		Physical	1			

Page	Item Identifier	Item Type	Max Points	Key	Standard	GLE	Round 1	Round 2	Round 3
37	SC050143	MC	1	D	Earth Systems	2			
38	COSC130068_1	CR	2		Life	2			
39	COSC130111	MC	1	A	Earth Systems	1			
40	COSC120007	MC	1	C	Life	1			
41	COSC130072	MC	1	A	Earth Systems	3			
42	COSC130202_1	CR	3		Life	2			
43	SC050102	MC	1	B	Physical	1			
44	SC050132	MC	1	C	Earth Systems	1			
45	COSC130106_1	CR	2		Earth Systems	2			
46	SC050057	XI	1		Earth Systems	1			
47	SC050002	MC	1	D	Life	1			
48	COSC120145	XI	1		Earth Systems	2			
49	SC050101	MC	1	A	Earth Systems	2			
50	COSC120142	XI	1		Life	1			
51	COSC120134	MC	1	B	Life	2			
52	COSC130088_2	CR	2		Life	1			
53	COSC120108	MC	1	A	Physical	1			
54	COSC120257_1	CR	2		Physical	1			
55	COSC130108_1	CR	2		Earth Systems	2			
56	COSC120254	MC	1	D	Earth Systems	1			
57	SC050052	MC	1	B	Physical	1			
58	COSC130146	MC	1	C	Physical	1			
59	COSC120019	MC	1	D	Life	2			
60	COSC120115	MC	1	D	Earth Systems	3			
61	COSC120153	XI	1		Life	2			
62	SC050059	XI	1		Earth Systems	2			
63	COSC120160_2	CR	3		Earth Systems	3			
64	COSC120259_1	CR	2		Physical	1			
65	COSC120154	XI	1		Physical	1			
66	COSC130092	MC	1	B	Earth Systems	1			
67	COSC130196_2	CR	2		Life	2			
68	SC050108	XI	1		Life	1			
69	COSC130158	MC	1	B	Earth Systems	2			
70	COSC130086_2	CR	2		Earth Systems	1			
71	COSC130068_2	CR	2		Life	2			
72	COSC130202_2	CR	3		Life	2			
73	COSC120155_2	CR	3		Physical	1			

Page	Item Identifier	Item Type	Max Points	Key	Standard	GLE	Round 1	Round 2	Round 3
74	SC050146	MC	1	D	Life	2			
75	COSC130087_2	CR	2		Earth Systems	1			
76	COSC130046_2	CR	2		Life	1			
77	COSC120005	MC	1	B	Physical	1			
78	SC050145_1	CR	2		Physical	1			
79	COSC130287	XI	1		Life	2			
80	COSC130110	MC	1	B	Earth Systems	2			
81	COSC130048_2	CR	2		Life	1			
82	COSC130202_3	CR	3		Life	2			
83	COSC130066_2	CR	2		Earth Systems	2			
84	COSC120170_2	CR	2		Physical	1			
85	COSC130106_2	CR	2		Earth Systems	2			
86	COSC130108_2	CR	2		Earth Systems	2			
87	SC050145_2	CR	2		Physical	1			
88	COSC120257_2	CR	2		Physical	1			
89	COSC120155_3	CR	3		Physical	1			
90	COSC120160_3	CR	3		Earth Systems	3			
91	COSC130052	MC	1	C	Life	2			
92	COSC120259_2	CR	2		Physical	1			

Grade 7 Social Studies OIB Spreadsheet

Page	Item Identifier	Item Type	Max Points	Key	Standard	GLE	Round 1	Round 2	Round 3
1	COSS120090	XI	1		Geography	2			
2	COSS120192_1	CR	3		History	1			
3	SS070105	MC	1	B	Geography	2			
4	COSS120329_1	CR	3		Geography	2			
5	COSS120382_1	CR	3		Civics	2			
6	COSS120345	MC	1	C	Economics	2			
7	COSS120142	MC	1	C	Economics	1			
8	COSS130125_1	CR	3		History	2			
9	COSS120327_1	CR	3		History	1			
10	COSS120349_1	CR	3		Economics	1			
11	SS070073	MC	1	D	Geography	1			
12	COSS130130	MC	1	A	Geography	1			
13	COSS130111_1	CR	3		Civics	1			
14	COSS130016	MC	1	A	Geography	1			
15	COSS120204	MC	1	B	Geography	2			
16	COSS130256_1	CR	3		Geography	2			
17	COSS120192_2	CR	3		History	1			
18	SS070092	MC	1	C	Economics	1			
19	SS070112	MC	1	A	Civics	2			
20	COSS130022	MC	1	D	Civics	1			
21	SS070080	XI	1		Civics	1			
22	COSS120086	MC	1	C	Economics	1			
23	SS070081	MC	1	B	Economics	1			
24	COSS130035	MC	1	C	Civics	1			
25	COSS130110_1	CR	3		Geography	1			
26	COSS120333_1	CR	3		Civics	2			
27	COSS120154	MC	1	C	Economics	1			
28	COSS130112_1	CR	3		History	2			
29	COSS130021	MC	1	B	Economics	2			
30	COSS130114	MC	1	B	Geography	1			
31	COSS120024	XI	1		History	2			
32	COSS130129_1	CR	3		Economics	2			
33	SS070089	MC	1	A	Economics	1			
34	COSS130062	XI	1		Economics	2			
35	COSS120349_2	CR	3		Economics	1			
36	COSS130015	MC	1	C	History	2			

Page	Item Identifier	Item Type	Max Points	Key	Standard	GLE	Round 1	Round 2	Round 3
37	COSS120205	MC	1	A	Civics	2			
38	SS070050-SSS07007	MC	1	C	Geography	1			
39	COSS120025	XI	1		History	2			
40	SS070090	MC	1	B	Economics	1			
41	SS070095	MC	1	B	Civics	1			
42	COSS120329_2	CR	3		Geography	2			
43	COSS130018	MC	1	B	Geography	1			
44	COSS120033	MC	1	C	Economics	2			
45	COSS120343	MC	1	D	Economics	1			
46	COSS120366	XI	1		History	2			
47	COSS130036	MC	1	A	Civics	1			
48	COSS130115	MC	1	B	History	2			
49	COSS120076	XI	1		Economics	2			
50	COSS120084	MC	1	A	Civics	1			
51	COSS120138	MC	1	B	Geography	2			
52	COSS120382_2	CR	3		Civics	2			
53	COSS130056	XI	1		History	2			
54	COSS130131	MC	1	D	Economics	1			
55	COSS130038	MC	1	C	Civics	2			
56	COSS130125_2	CR	3		History	2			
57	COSS130111_2	CR	3		Civics	1			
58	COSS120192_3	CR	3		History	1			
59	COSS120333_2	CR	3		Civics	2			
60	COSS120327_2	CR	3		History	1			
61	COSS130037	MC	1	D	Civics	1			
62	COSS130078	XI	1		History	1			
63	COSS130126	XI	1		Geography	2			
64	COSS130077	MC	1	C	History	1			
65	COSS130129_2	CR	3		Economics	2			
66	COSS130256_2	CR	3		Geography	2			
67	COSS120382_3	CR	3		Civics	2			
68	COSS130058	XI	1		Civics	2			
69	COSS130112_2	CR	3		History	2			
70	COSS130110_2	CR	3		Geography	1			
71	COSS120333_3	CR	3		Civics	2			
72	COSS120349_3	CR	3		Economics	1			
73	COSS130111_3	CR	3		Civics	1			

Page	Item Identifier	Item Type	Max Points	Key	Standard	GLE	Round 1	Round 2	Round 3
74	COSS120329_3	CR	3		Geography	2			
75	COSS130125_3	CR	3		History	2			
76	COSS120342	MC	1	C	Geography	1			
77	COSS130256_3	CR	3		Geography	2			
78	COSS130129_3	CR	3		Economics	2			
79	COSS120327_3	CR	3		History	1			
80	COSS130112_3	CR	3		History	2			
81	COSS130110_3	CR	3		Geography	1			

Grade 8 Science OIB Spreadsheet

Page	Item Identifier	Item Type	Max Points	Key	Standard	GLE	Round 1	Round 2	Round 3
1	COSC130044	XI	1		Life	2			
2	COSC130267_1	CR	2		Life	1			
3	COSC130260	MC	1	B	Earth Systems	1			
4	COSC120251	XI	1		Physical	3			
5	COSC130273_1	CR	2		Life	1			
6	COSC130028	MC	1	C	Physical	2			
7	SC080074	XI	1		Life	2			
8	COSC130007	MC	1	D	Earth Systems	1			
9	COSC130038_1	CR	2		Physical	4			
10	COSC130261	MC	1	C	Earth Systems	1			
11	COSC130005	MC	1	A	Earth Systems	3			
12	COSC130271	MC	1	A	Life	1			
13	SC080141	MC	1	A	Earth Systems	2			
14	COSC120306_1	CR	3		Earth Systems	2			
15	COSC130247_1	CR	2		Physical	3			
16	COSC130026	MC	1	B	Physical	2			
17	SC080095	MC	1	C	Earth Systems	2			
18	COSC120061	MC	1	A	Physical	1			
19	COSC130186	XI	1		Physical	4			
20	SC080023-SCS08003	MC	1	D	Physical	4			
21	COSC120244	XI	1		Life	2			
22	COSC130012_1	CR	3		Physical	2			
23	SC080091	MC	1	B	Earth Systems	1			
24	SC080106	MC	1	B	Physical	1			
25	SC080028-SCS08004	MC	1	C	Physical	4			
26	COSC130037_1	CR	2		Physical	1			
27	COSC130263_1	CR	2		Earth Systems	1			
28	SC080124	MC	1	D	Physical	3			
29	COSC120098	MC	1	A	Earth Systems	3			
30	COSC130242_1	CR	2		Earth Systems	4			
31	COSC130241	MC	1	D	Earth Systems	4			
32	COSC130027	MC	1	D	Physical	2			
33	COSC120252	XI	1		Life	2			
34	COSC130011_1	CR	2		Earth Systems	3			
35	COSC120088	MC	1	C	Physical	2			
36	COSC130272	MC	1	C	Life	1			

Page	Item Identifier	Item Type	Max Points	Key	Standard	GLE	Round 1	Round 2	Round 3
37	SC080142	MC	1	B	Earth Systems	2			
38	COSC120306_2	CR	3		Earth Systems	2			
39	COSC130249_1	CR	2		Physical	3			
40	COSC130240	MC	1	A	Earth Systems	4			
41	COSC130040	XI	1		Earth Systems	2			
42	COSC120087	MC	1	D	Physical	1			
43	COSC130015	MC	1	A	Life	1			
44	COSC120295	MC	1	B	Life	1			
45	COSC120298	MC	1	B	Physical	1			
46	SC080132	XI	1		Physical	2			
47	COSC130001	MC	1	B	Earth Systems	3			
48	COSC120055	MC	1	C	Earth Systems	3			
49	COSC120272	MC	1	D	Life	2			
50	COSC130011_2	CR	2		Earth Systems	3			
51	COSC130244_1	CR	2		Earth Systems	4			
52	COSC130025	MC	1	B	Physical	1			
53	COSC130022_1	CR	3		Life	2			
54	COSC120275_1	CR	2		Life	2			
55	COSC120250	XI	1		Life	2			
56	COSC130035	MC	1	D	Physical	4			
57	COSC120274_1	CR	2		Life	2			
58	COSC130247_2	CR	2		Physical	3			
59	COSC120078_1	CR	2		Physical	4			
60	COSC130273_2	CR	2		Life	1			
61	COSC130022_2	CR	3		Life	2			
62	COSC130031	MC	1	B	Physical	3			
63	COSC120306_3	CR	3		Earth Systems	2			
64	COSC130012_2	CR	3		Physical	2			
65	COSC130254	MC	1	D	Physical	1			
66	COSC120078_2	CR	2		Physical	4			
67	COSC130038_2	CR	2		Physical	4			
68	COSC120273	MC	1	B	Life	2			
69	COSC130042	XI	1		Life	1			
70	COSC130041	XI	1		Earth Systems	3			
71	COSC130037_2	CR	2		Physical	1			
72	COSC130249_2	CR	2		Physical	3			
73	COSC130242_2	CR	2		Earth Systems	4			

Page	Item Identifier	Item Type	Max Points	Key	Standard	GLE	Round 1	Round 2	Round 3
74	COSC130267_2	CR	2		Life	1			
75	COSC120243	XI	1		Life	1			
76	COSC130262_1	CR	2		Earth Systems	1			
77	COSC130244_2	CR	2		Earth Systems	4			
78	COSC130252	MC	1	D	Physical	3			
79	COSC130022_3	CR	3		Life	2			
80	COSC120274_2	CR	2		Life	2			
81	COSC120246	XI	1		Earth Systems	4			
82	COSC130016	MC	1	D	Life	1			
83	COSC120090	MC	1	A	Physical	4			
84	COSC130263_2	CR	2		Earth Systems	1			
85	COSC130012_3	CR	3		Physical	2			
86	COSC130262_2	CR	2		Earth Systems	1			
87	COSC120275_2	CR	2		Life	2			

APPENDIX D: P-VALUE REPORTS

Grade 4 Social Studies

Page	Item Identifier	P-value
1	COSS120007	0.93
2	COSS120363_1	0.92
3	COSS120006	0.88
4	COSS130066_1	0.82
5	COSS120312	0.80
6	COSS120362_1	0.80
7	SS040015-SSS04002	0.84
8	COSS120363_2	0.77
9	COSS120005	0.71
10	SS040080	0.78
11	COSS120325_1	0.72
12	COSS130095_1	0.72
13	COSS130101_1	0.71
14	COSS130091	0.68
15	COSS120318	0.66
16	SS040079	0.69
17	COSS130013	0.66
18	SS040067	0.65
19	SS040013-SSS04002	0.64
20	COSS130073	0.58
21	COSS130004	0.62
22	SS040076	0.65
23	COSS120320_1	0.53
24	SS040016-SSS04002	0.59
25	COSS120181	0.58
26	COSS130041	0.65
27	COSS130105_1	0.54
28	COSS130300_1	0.47
29	COSS130095_2	0.43
30	COSS130097_1	0.50
31	COSS120309	0.48
32	COSS120008	0.54
33	COSS130099	0.56
34	COSS130055	0.47
35	COSS120069	0.54
36	COSS120304	0.50
37	COSS130066_2	0.44
38	SS040088	0.46
39	COSS120004	0.44
40	COSS130104_1	0.45
41	COSS120310	0.52
42	SS040068	0.57
43	COSS120184	0.37
44	COSS120010	0.47
45	COSS120325_2	0.34
46	COSS120314	0.44
47	COSS120324_1	0.36
48	COSS120180	0.36
49	COSS130043	0.36
50	COSS130053	0.30

Page	Item Identifier	P-value
51	COSS120315	0.37
52	COSS130044	0.40
53	COSS120057	0.42
54	COSS130048	0.57
55	COSS130097_2	0.28
56	COSS130300_2	0.21
57	COSS120362_2	0.19
58	COSS130101_2	0.21
59	COSS130100	0.30
60	COSS120320_2	0.18
61	COSS130008	0.31
62	COSS130049	0.25
63	COSS120362_3	0.09
64	COSS130095_3	0.10
65	COSS120324_2	0.10
66	COSS130104_2	0.12
67	COSS120179	0.26
68	COSS120325_3	0.11
69	COSS130105_2	0.17
70	COSS130066_3	0.08
71	COSS120363_3	0.17
72	COSS130300_3	0.04
73	COSS130101_3	0.03
74	COSS120320_3	0.03
75	COSS130097_3	0.06
76	COSS130104_3	0.02
77	COSS120324_3	0.01
78	COSS130047	0.29
79	COSS130096	0.33
80	COSS130105_3	0.02
81	COSS130102	0.11

Grade 5 Science

Page	Item Identifier	P-value
1	COSC130292	0.90
2	COSC120146	0.92
3	COSC130086_1	0.94
4	COSC120264	0.88
5	COSC130048_1	0.80
6	SC050144	0.83
7	COSC130069	0.92
8	COSC130154	0.88
9	COSC130051	0.84
10	COSC120027	0.87
11	COSC130089	0.85
12	COSC130088_1	0.83
13	COSC130294	0.77
14	COSC120015	0.80
15	COSC130046_1	0.79
16	COSC130066_1	0.77
17	COSC120050	0.74
18	COSC120012	0.79
19	COSC120024	0.79
20	COSC120104	0.77
21	COSC130163	0.79
22	SC050115	0.82
23	COSC130196_1	0.76
24	COSC130291	0.75
25	COSC130289	0.73
26	COSC120122	0.76
27	SC050074	0.75
28	COSC120170_1	0.73
29	COSC130157	0.74
30	COSC130087_1	0.72
31	SC050065	0.75
32	SC050134	0.73
33	COSC130290	0.69
34	COSC120160_1	0.69
35	COSC120253	0.68
36	COSC120155_1	0.66
37	SC050143	0.67
38	COSC130068_1	0.65
39	COSC130111	0.69
40	COSC120007	0.65
41	COSC130072	0.66
42	COSC130202_1	0.64
43	SC050102	0.67
44	SC050132	0.66
45	COSC130106_1	0.63
46	SC050057	0.65
47	SC050002	0.51
48	COSC120145	0.61
49	SC050101	0.62
50	COSC120142	0.54

Page	Item Identifier	P-value
51	COSC120134	0.59
52	COSC130088_2	0.50
53	COSC120108	0.56
54	COSC120257_1	0.51
55	COSC130108_1	0.49
56	COSC120254	0.52
57	SC050052	0.62
58	COSC130146	0.49
59	COSC120019	0.57
60	COSC120115	0.49
61	COSC120153	0.45
62	SC050059	0.46
63	COSC120160_2	0.40
64	COSC120259_1	0.42
65	COSC120154	0.44
66	COSC130092	0.50
67	COSC130196_2	0.37
68	SC050108	0.38
69	COSC130158	0.43
70	COSC130086_2	0.31
71	COSC130068_2	0.33
72	COSC130202_2	0.23
73	COSC120155_2	0.25
74	SC050146	0.35
75	COSC130087_2	0.27
76	COSC130046_2	0.29
77	COSC120005	0.34
78	SC050145_1	0.19
79	COSC130287	0.24
80	COSC130110	0.42
81	COSC130048_2	0.49
82	COSC130202_3	0.11
83	COSC130066_2	0.27
84	COSC120170_2	0.16
85	COSC130106_2	0.23
86	COSC130108_2	0.06
87	SC050145_2	0.03
88	COSC120257_2	0.04
89	COSC120155_3	0.03
90	COSC120160_3	0.09
91	COSC130052	0.40
92	COSC120259_2	0.03

Grade 7 Social Studies

Page	Item Identifier	P-value
1	COSS120090	0.96
2	COSS120192_1	0.85
3	SS070105	0.86
4	COSS120329_1	0.84
5	COSS120382_1	0.81
6	COSS120345	0.80
7	COSS120142	0.78
8	COSS130125_1	0.76
9	COSS120327_1	0.75
10	COSS120349_1	0.74
11	SS070073	0.74
12	COSS130130	0.70
13	COSS130111_1	0.70
14	COSS130016	0.69
15	COSS120204	0.69
16	COSS130256_1	0.67
17	COSS120192_2	0.67
18	SS070092	0.69
19	SS070112	0.64
20	COSS130022	0.64
21	SS070080	0.69
22	COSS120086	0.62
23	SS070081	0.66
24	COSS130035	0.64
25	COSS130110_1	0.61
26	COSS120333_1	0.59
27	COSS120154	0.61
28	COSS130112_1	0.57
29	COSS130021	0.64
30	COSS130114	0.62
31	COSS120024	0.60
32	COSS130129_1	0.53
33	SS070089	0.61
34	COSS130062	0.52
35	COSS120349_2	0.49
36	COSS130015	0.56
37	COSS120205	0.55
38	SS070050-SSS07007	0.57
39	COSS120025	0.54
40	SS070090	0.55
41	SS070095	0.54
42	COSS120329_2	0.41
43	COSS130018	0.53
44	COSS120033	0.53
45	COSS120343	0.47
46	COSS120366	0.48
47	COSS130036	0.56
48	COSS130115	0.45
49	COSS120076	0.45
50	COSS120084	0.44

Page	Item Identifier	P-value
51	COSS120138	0.44
52	COSS120382_2	0.26
53	COSS130056	0.48
54	COSS130131	0.39
55	COSS130038	0.39
56	COSS130125_2	0.24
57	COSS130111_2	0.20
58	COSS120192_3	0.31
59	COSS120333_2	0.18
60	COSS120327_2	0.29
61	COSS130037	0.35
62	COSS130078	0.22
63	COSS130126	0.37
64	COSS130077	0.27
65	COSS130129_2	0.13
66	COSS130256_2	0.15
67	COSS120382_3	0.12
68	COSS130058	0.31
69	COSS130112_2	0.14
70	COSS130110_2	0.13
71	COSS120333_3	0.06
72	COSS120349_3	0.13
73	COSS130111_3	0.04
74	COSS120329_3	0.06
75	COSS130125_3	0.05
76	COSS120342	0.24
77	COSS130256_3	0.02
78	COSS130129_3	0.02
79	COSS120327_3	0.07
80	COSS130112_3	0.01
81	COSS130110_3	0.01

Grade 8 Science

Page	Item Identifier	P-value
1	COSC130044	0.90
2	COSC130267_1	0.87
3	COSC130260	0.86
4	COSC120251	0.84
5	COSC130273_1	0.84
6	COSC130028	0.79
7	SC080074	0.82
8	COSC130007	0.80
9	COSC130038_1	0.81
10	COSC130261	0.77
11	COSC130005	0.78
12	COSC130271	0.75
13	SC080141	0.77
14	COSC120306_1	0.75
15	COSC130247_1	0.74
16	COSC130026	0.74
17	SC080095	0.73
18	COSC120061	0.71
19	COSC130186	0.69
20	SC080023-SCS08003	0.69
21	COSC120244	0.65
22	COSC130012_1	0.65
23	SC080091	0.67
24	SC080106	0.69
25	SC080028-SCS08004	0.66
26	COSC130037_1	0.61
27	COSC130263_1	0.59
28	SC080124	0.63
29	COSC120098	0.63
30	COSC130242_1	0.58
31	COSC130241	0.58
32	COSC130027	0.61
33	COSC120252	0.60
34	COSC130011_1	0.51
35	COSC120088	0.59
36	COSC130272	0.58
37	SC080142	0.62
38	COSC120306_2	0.50
39	COSC130249_1	0.51
40	COSC130240	0.57
41	COSC130040	0.52
42	COSC120087	0.53
43	COSC130015	0.52
44	COSC120295	0.58
45	COSC120298	0.59
46	SC080132	0.51
47	COSC130001	0.55
48	COSC120055	0.54
49	COSC120272	0.43
50	COSC130011_2	0.33

Page	Item Identifier	P-value
51	COSC130244_1	0.45
52	COSC130025	0.48
53	COSC130022_1	0.32
54	COSC120275_1	0.44
55	COSC120250	0.33
56	COSC130035	0.44
57	COSC120274_1	0.34
58	COSC130247_2	0.36
59	COSC120078_1	0.30
60	COSC130273_2	0.40
61	COSC130022_2	0.21
62	COSC130031	0.47
63	COSC120306_3	0.25
64	COSC130012_2	0.24
65	COSC130254	0.35
66	COSC120078_2	0.23
67	COSC130038_2	0.29
68	COSC120273	0.38
69	COSC130042	0.26
70	COSC130041	0.26
71	COSC130037_2	0.19
72	COSC130249_2	0.16
73	COSC130242_2	0.21
74	COSC130267_2	0.24
75	COSC120243	0.19
76	COSC130262_1	0.16
77	COSC130244_2	0.19
78	COSC130252	0.33
79	COSC130022_3	0.07
80	COSC120274_2	0.09
81	COSC120246	0.33
82	COSC130016	0.36
83	COSC120090	0.22
84	COSC130263_2	0.09
85	COSC130012_3	0.04
86	COSC130262_2	0.02
87	COSC120275_2	0.05

APPENDIX E: EXTERNAL DATA

Grade 4 External Data

© Pearson Education Ltd 2014.

PEARSON

TCAP

	Unsatisfactory	Partially Proficient	Proficient	Advanced
4th Grade Writing	7%	39%	45%	8%
4th Grade Reading	11%	21%	63%	5%

	Limited Command	Moderate Command	Strong Command	Distinguished Command
Round 2 CMAS Cuts	##	##	##	##

© Pearson Education Ltd 2014.

PEARSON

Grade 5 External Data

© Pearson Education Ltd 2014.

PEARSON

TCAP

	Unsatisfactory	Partially Proficient	Proficient	Advanced
5th Grade Science	14%	37%	35%	13%

	Limited Command	Moderate Command	Strong Command	Distinguished Command
Round 2 CMAS Cuts	##	##	##	##

© Pearson Education Ltd 2014.

PEARSON

TIMSS-Grade 4 Science

	Low	Intermediate	High	Advanced
US	15%	32%	34%	15%
Singapore	8%	21%	35%	33%
Finland	7%	27%	45%	20%
Hong Kong	14%	37%	36%	9%

© Pearson Education Ltd 2014.

PEARSON

Grade 7 External Data

© Pearson Education Ltd 2014.

PEARSON

TCAP

	Unsatisfactory	Partially Proficient	Proficient	Advanced
7th Grade Writing	3%	35%	47%	15%
7th Grade Reading	11%	21%	59%	9%

	Limited Command	Moderate Command	Strong Command	Distinguished Command
Round 2 CMAS Cuts	##	##	##	##

© Pearson Education Ltd 2014.

PEARSON

Grade 8
External Data

© Pearson Education Ltd 2014.

PEARSON

External Data-TCAP

	Unsatisfactory	Partially Proficient	Proficient	Advanced
8th Grade Science	21%	27%	43%	9%

	Limited Command	Moderate Command	Strong Command	Distinguished Command
Round 2 CMAS Cuts	##	##	##	##

© Pearson Education Ltd 2014.

PEARSON

NAEP-Grade 8 Science

	Percent Proficient
Colorado	42%
US	32%

© Pearson Education Ltd 2014.

PEARSON

TIMSS-Grade 8 Science

	Low	Intermediate	High	Advanced
Colorado	16%	32%	34%	14%
US	20%	33%	30%	10%
Singapore	9%	18%	29%	40%
Finland	11%	35%	40%	13%
Hong Kong	15%	33%	38%	9%

© Pearson Education Ltd 2014.

PEARSON

APPENDIX F: READINESS SURVEY

Colorado Measures of Academic Success (CMAS)

Standard-Setting Round Readiness Survey

Panelist ID: _____

Instructions: Please circle your response to the following questions.

Round 1		
I understand that my task for Round 1 is to use my content expertise, my experience with Colorado students, the threshold student descriptors, and the ordered item book to make cut score recommendations. To make my recommendation, I will indicate the last “yes” page on the recommendation sheet.	No	Yes
I am ready to begin Round 1.	No	Yes

Round 2		
I understand that my task for Round 2 is to use my content expertise, my experience with Colorado students, the threshold student descriptors, and the ordered item book to make cut score recommendations. To make my recommendation, I will indicate the last “yes” page on the recommendation sheet.	No	Yes
I understand the panelist feedback data that were presented from Round 1.	No	Yes
I understand the item difficulty data (i.e., p-values) that were provided.	No	Yes
I am ready to begin Round 2.	No	Yes

Round 3		
I understand that my task for Round 3 is to use my content expertise, my experience with Colorado students, the threshold student descriptors, and the ordered item book to make cut score recommendations. To make my recommendation, I will indicate the last “yes” page on the recommendation sheet.	No	Yes
I understand the impact data that were presented from Round 2.	No	Yes
I am ready to begin Round 3.	No	Yes

APPENDIX G: BOOKMARK RECOMMENDATION FORMS

Bookmark Recommendation Form

Directions: For each level, write down the **page number** corresponding to the **last YES** item. No cells should be left blank within a given round.

Panelist ID: _____

Table Number: _____

	Page Number of LAST YES Item		
	Moderate Command	Strong Command	Distinguished Command
Round 1			
Round 2			
Round 3			

APPENDIX H: STANDARD SETTING EVALUATION

Colorado Measures of Academic Success (CMAS)
Standard Setting Evaluation Form

The purpose of this evaluation form is to collect information about your experience in recommending performance cut scores for CMAS. Your opinions provide an important part of our evaluation of this meeting. Please do not write your name on this evaluation form as we want your comments to be anonymous. Thank you for your willingness to participate in this survey.

In which standard setting meeting did you participate?

- ☐ Grade 4 Social Studies
- ☐ Grade 5 Science
- ☐ Grade 7 Social Studies
- ☐ Grade 8 Science

Indicate your response by checking the appropriate box.

		Do not support	Support with some reservation	Moderately support	Strongly support	(Omit)
1. To what degree do you support the recommended cut score for "Moderate Command?"						
	Grade 4	0%	9%	18%	73%	
	Grade 5	9%	9%	36%	45%	
	Grade 7	0%	17%	17%	67%	
	Grade 8	0%	0%	8%	92%	
If you cannot support, please explain why not:						
2. To what degree do you support the recommended cut score for "Strong Command?"						
	Grade 4	0%	0%	64%	36%	
	Grade 5	9%	9%	18%	55%	9%
	Grade 7	0%	0%	25%	75%	
	Grade 8	23%	8%	15%	54%	
If you cannot support, please explain why not:						
3. To what degree do you support the recommended cut score for "Distinguished Command?"						
	Grade 4	0%	0%	9%	91%	
	Grade 5	9%	9%	27%	55%	
	Grade 7	0%	0%	25%	75%	
	Grade 8	0%	0%	31%	69%	
If you cannot support, please explain why not:						

		Way too low	A bit low	Appropriate	A bit high	Way too high	(Omit)
4. The recommended cut score for "Moderate Command" is:							
	Grade 4	0%	0%	82%	18%	0%	
	Grade 5	9%	18%	45%	18%	0%	9%
	Grade 7	0%	17%	67%	17%	0%	
	Grade 8	0%	0%	92%	8%	0%	
5. The recommended cut score for "Strong Command" is:							
	Grade 4	0%	0%	91%	0%	9%	
	Grade 5	0%	0%	55%	27%	9%	9%
	Grade 7	0%	25%	67%	8%	0%	
	Grade 8	0%	0%	46%	46%	8%	
6. The recommended cut score for "Distinguished Command" is:							
	Grade 4	0%	0%	91%	9%	0%	
	Grade 5	9%	9%	36%	36%	0%	9%
	Grade 7	0%	0%	75%	25%	0%	
	Grade 8	0%	0%	62%	38%	0%	

		Strongly Disagree	Disagree	Agree	Strongly Agree	(Omit)
7. The Bookmark Method was explained clearly by the group facilitator.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Grade 4	0%	0%	82%	18%	
	Grade 5	0%	0%	45%	55%	
	Grade 7	0%	0%	17%	83%	
	Grade 8	0%	8%	46%	46%	
8. I had a solid understanding of what the test was intended to measure.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Grade 4	0%	0%	36%	64%	
	Grade 5	0%	0%	27%	73%	
	Grade 7	0%	0%	25%	75%	
	Grade 8	0%	0%	38%	62%	
9. I could clearly distinguish between performance levels.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Grade 4	0%	0%	82%	18%	
	Grade 5	0%	0%	91%	9%	
	Grade 7	0%	0%	75%	25%	
	Grade 8	0%	0%	77%	15%	8%
10. After the first round of recommendations, I felt comfortable with the standard setting procedure.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Grade 4	0%	0%	55%	45%	
	Grade 5	0%	0%	55%	45%	
	Grade 7	0%	8%	67%	25%	
	Grade 8	0%	0%	54%	46%	
11. I found the feedback on the comparison of all panelists' recommendations to be useful in standard setting.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Grade 4	0%	0%	27%	73%	
	Grade 5	0%	0%	36%	64%	
	Grade 7	0%	0%	17%	83%	
	Grade 8	0%	0%	46%	46%	8%
12. I found the p-value information to be useful in standard setting.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Grade 4	0%	0%	9%	91%	
	Grade 5	0%	0%	36%	64%	
	Grade 7	0%	0%	42%	58%	
	Grade 8	0%	0%	54%	38%	8%
13. I found the feedback on the percentage of the students tested that would be classified at each performance level to be useful in standard setting.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Grade 4	0%	0%	36%	64%	
	Grade 5	0%	18%	27%	45%	9%
	Grade 7	0%	8%	25%	67%	
	Grade 8	0%	15%	31%	54%	
14. Table and group discussions were open and honest.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Grade 4	0%	0%	9%	91%	
	Grade 5	0%	0%	9%	91%	
	Grade 7	0%	0%	0%	100%	
	Grade 8	0%	0%	46%	54%	
15. I believe that my opinions were considered and valued by my group.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Grade 4	0%	0%	9%	91%	
	Grade 5	0%	0%	18%	82%	
	Grade 7	0%	0%	8%	92%	
	Grade 8	0%	0%	46%	54%	

		Strongly Disagree	Disagree	Agree	Strongly Agree	(Omit)
16. The facilitator led the group through the standard setting process without imposing ideas about where cut scores should be.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Grade 4	0%	0%	27%	73%	
	Grade 5	0%	0%	0%	100%	
	Grade 7	0%	0%	8%	92%	
	Grade 8	0%	0%	46%	54%	
17. I am confident that the final cut score recommendations reflect the performance level descriptors associated with CMAS.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Grade 4	0%	0%	64%	27%	9%
	Grade 5	0%	27%	45%	27%	
	Grade 7	0%	0%	58%	42%	
	Grade 8	0%	8%	46%	31%	15%
18. I am confident that the final cut score recommendations reflect high expectations consistent with the Colorado Academic Standards.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Grade 4	0%	0%	36%	55%	9%
	Grade 5	0%	9%	45%	36%	9%
	Grade 7	0%	0%	33%	67%	
	Grade 8	0%	0%	23%	54%	23%

Please use the back of this page to provide any additional comments.

APPENDIX I: VERTICAL ARTICULATION PARTICIPANT COMPOSITION

Table 15. Participants Breakdown by Expertise

	Content Expert	Administrator	Special Ed/ELL	Total
Social Studies	6	1	2	9
Science	6	2	2	10
Total	12	3	4	19

Table 16. Participants Breakdown by School Setting

	Rural	Suburban	Urban	Total
Social Studies	2	4	3	9
Science	4	4	2	10
Total	6	8	5	19

Table 17. Participants Breakdown by School Type

	Charter School	Innovation School	Neither Charter nor Innovation	District Level	Total
Social Studies	1	1	2	5	9
Science	0	1	6	3	10
Total	1	2	8	8	19

Table 18. Participants Breakdown by Region

	Social Studies	Science	Total
Denver Metro	4	1	5
North Central	2	1	3
Northeast	2	2	4
Northwest	0	1	1
Pikes Peak	1	3	4
Southeast	0	1	1
Southwest	0	1	1
West Central	0	0	0
Total	9	10	19

APPENDIX J: VERTICAL ARTICULATION EVALUATION FORM

Colorado Measures of Academic Success (CMAS)
Vertical Articulation Evaluation Form

The purpose of this evaluation form is to collect information about your experience in recommending performance cut scores for CMAS. Your opinions provide an important part of our evaluation of this meeting. Please do not write your name on this evaluation form as we want your comments to be anonymous. Thank you for your willingness to participate in this survey.

In which vertical articulation meeting did you participate?

☐ Social Studies

☐ Science

Indicate your response by checking the appropriate box.

		Do not support	Support with some reservation	Moderately support	Strongly support	
1. To what degree do you support the recommended cut score for Elementary School "Moderate Command?"		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Science	0%	10%	10%	80%	
	Social Studies	0%	0%	33%	67%	
If you cannot support, please explain why not:						
2. To what degree do you support the recommended cut score for Elementary School "Strong Command?"		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Science	0%	10%	10%	80%	
	Social Studies	0%	0%	33%	67%	
If you cannot support, please explain why not:						
3. To what degree do you support the recommended cut score for Elementary School "Distinguished Command?"		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Science	10%	20%	0%	70%	
	Social Studies	0%	0%	22%	78%	
If you cannot support, please explain why not:						
4. To what degree do you support the recommended cut score for Middle School "Moderate Command?"		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Science	0%	10%	10%	80%	
	Social Studies	0%	0%	67%	33%	
If you cannot support, please explain why not:						
5. To what degree do you support the recommended cut score for Middle School "Strong Command?"		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Science	0%	10%	30%	60%	
	Social Studies	0%	0%	56%	44%	
If you cannot support, please explain why not:						
6. To what degree do you support the recommended cut score for Middle School "Distinguished Command?"		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Science	0%	20%	0%	80%	
	Social Studies	0%	0%	22%	78%	
If you cannot support, please explain why not:						
		Way too low	A bit low	Appropriate	A bit high	Way too high
7. The recommended cut score for Elementary School "Moderate Command" is:		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Science	0%	0%	80%	20%	0%
	Social Studies	0%	0%	78%	22%	0%
8. The recommended cut score for Elementary School "Strong Command" is:		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Science	0%	0%	80%	20%	0%
	Social Studies	0%	0%	78%	22%	0%
9. The recommended cut score for Elementary School "Distinguished Command" is:		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Science	0%	0%	70%	30%	0%
	Social Studies	0%	0%	100%	0%	0%

	Way too low	A bit low	Appropriate	A bit high	Way too high	(Omit)
10. The recommended cut score for Middle School "Moderate Command" is:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Science	0%	0%	80%	20%	0%	
Social Studies	0%	0%	56%	44%	0%	
11. The recommended cut score for Middle School "Strong Command" is:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Science	0%	0%	80%	20%	0%	
Social Studies	0%	0%	67%	33%	0%	
12. The recommended cut score for Middle School "Distinguished Command" is:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Science	0%	0%	80%	10%	0%	10%
Social Studies	0%	0%	89%	11%	0%	

	Strongly Disagree	Disagree	Agree	Strongly Agree	(Omit)
13. Table and group discussions were open and honest.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Science	0%	0%	20%	80%	
Social Studies	0%	0%	22%	78%	
14. I believe that my opinions were considered and valued by my group.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Science	0%	0%	30%	70%	
Social Studies	0%	0%	11%	89%	
15. The facilitator led the group through the vertical articulation process without imposing ideas about where cut scores should be.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Science	0%	0%	20%	80%	
Social Studies	0%	0%	11%	89%	
16. I am confident that the final cut score recommendations reflect the performance level descriptors associated with CMAS.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Science	0%	0%	30%	70%	
Social Studies	0%	0%	33%	67%	
17. I am confident that the final cut score recommendations reflect high expectations consistent with the Colorado Academic Standards.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Science	0%	0%	30%	70%	
Social Studies	0%	0%	33%	67%	

Please use the back of this page to provide any additional comments.

APPENDIX B: SAMPLE CMAS SCORE REPORTS

Science Performance Level Descriptions

Students demonstrate mastery of science concepts and 21st century skills aligned to the Colorado Academic Standards at various performance levels. The performance level descriptors are organized in a manner that assumes students demonstrating higher levels of command have mastered the concepts and skills within the lower levels. For example, a student at moderate command also masters the concepts and skills of limited command.

At Distinguished Command, a student typically can

- evaluate and provide feedback on scientific evidence and reasoning about the separation of mixtures and how separation affects the total weight/mass;
- develop hypotheses about why similarities and differences exist between the body systems and parts of humans, plants, and animals;
- evaluate scientific claims about natural resources, in terms of reasonability and validity; and
- assess and provide feedback, through reasoning based on evidence, on scientific explanations about weather and factors that change Earth’s surface.

At Strong Command, a student typically can

- explain why certain procedures that are used to separate simple mixtures work and discuss any unexpected results;
- evaluate evidence and models of the structure and functions of human, plant, and animal organs and organ systems;
- investigate and generate evidence that human systems are interdependent;
- analyze and interpret data to explore concerns associated with natural resources; and
- formulate testable questions and scientific explanations around weather and factors that change Earth’s surface.

At Moderate Command, a student typically can

- discuss how the mass/weight of a mixture is a sum of its parts and design a procedure to separate simple mixtures based on physical properties;
- create models of human, plant, and animal organ systems, and compare and contrast similarities and differences between the organisms;
- explore and describe the origins and usage of natural resources in Colorado; and
- interpret data about Earth, including weather and changes to Earth’s surface.

At Limited Command, a student typically can

- select appropriate tools and follow procedures to separate simple mixtures;
- identify how humans, plants, and animals address basic survival needs;
- identify the functions of human body systems;
- distinguish between renewable and nonrenewable resources; and
- use appropriate tools and resources to gather data regarding weather conditions and Earth processes.



Student Performance Report

Colorado Measures of Academic Success

Student: STUDFIRSTNAME1 M. LASTNAME1

SASID: 9999999999 Birthdate: 08/14/2003
 School: SAMPLE SCHOOL
 District: SAMPLE DISTRICT

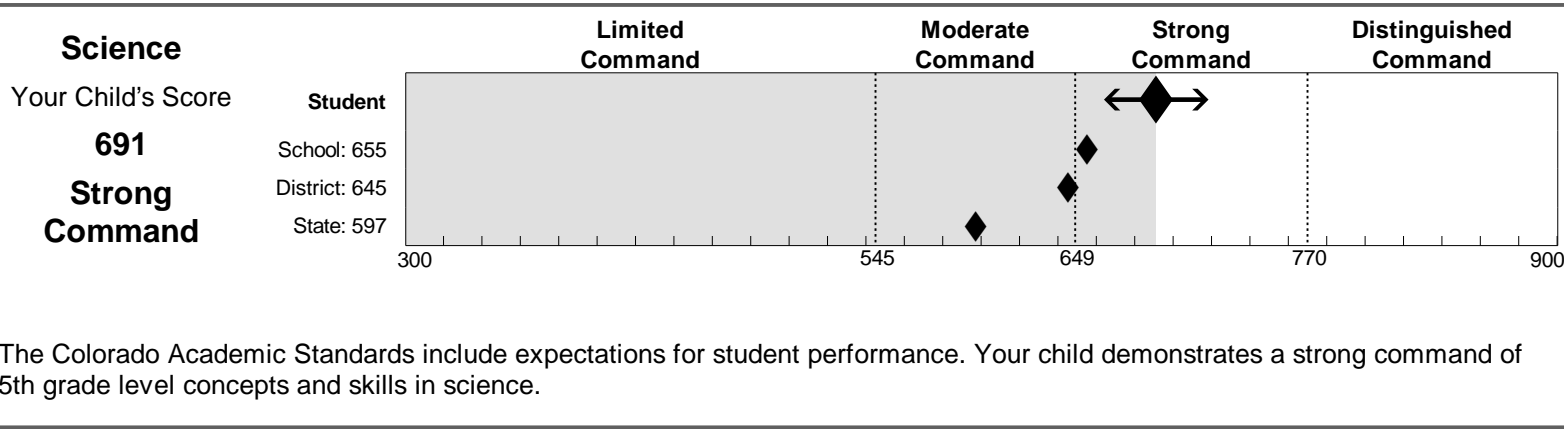
Spring 2014

Science

Grade 5

This score report provides information about your student’s performance on the Colorado Measures of Academic Success (CMAS) Science Assessment.

- Your student’s performance is represented by a scale score. Scores are placed on a scale so that student performance can be compared across years.
- School, district and state averages are provided so that you can compare your student’s performance to the performance of others.
- Scores are represented by diamonds. The arrows around the student’s diamond show the range of scores that your student would likely receive if the assessment was taken multiple times.
- Dotted lines show where the range of scores is divided into performance levels. Descriptions of the performance levels can be found at the end of this report.



Subscale Performance

- The shaded areas in the table below represent approximately 70% of student scores across the state.
- Scores outside of the shaded area indicate a weakness or a strength compared to the state.

Reporting Category Description	Subscale Score		Potential Relative Weakness	Typical	Potential Relative Strength
			300		900
Physical Science			478		718
Students know and understand common properties, forms, and changes in matter and energy.	672	Student			
	657	School			
	640	District			
Life Science			489		716
Students know and understand the characteristics and structure of living things, the processes of life, and how living things interact with each other and their environment.	723	Student			
	656	School			
	644	District			
Earth Systems Science			488		716
Students know and understand the processes and interactions of Earth's systems and the structure and dynamics of Earth and other objects in space.	676	Student			
	657	School			
	654	District			
Scientific Investigation and the Nature of Science			485		714
Students understand the processes of scientific investigation and design, conducting and evaluating, as well as communicating about, such investigations. Students understand that the nature of science involves a particular way of building knowledge and making meaning of the natural world.	674	Student			
	661	School			
	645	District			

Purpose

This report describes your child's mastery of the Colorado Academic Standards in science.

More information on the CMAS assessment program: www.cde.state.co.us/assessment



Content Standards Roster

Colorado Measures of Academic Success

Spring 2014

School: [REDACTED]
District: [REDACTED]

Science

CONFIDENTIAL - DO NOT DISTRIBUTE

Grade 5

Purpose: This report presents each student's performance on the overall test, content standards, prepared graduate competencies and grade level expectations for your school or district.

Performance Levels (PL)	Scale Score Ranges
Distinguished Command	771-900
Strong Command	650-770
Moderate Command	546-649
Limited Command	300-545

● = Potential Relative Strength (PRS)
 ○ = Typical
 ○ = Potential Relative Weakness (PRW)

Student	Overall Performance Level	Overall Scale Score	SEM Range	Content Standard Scale Score (SS) and Performance Indicator (PI)					
				Physical Science		Life Science		Earth Systems Science	
				SS	PI	SS	PI	SS	PI
1 AGGLNAAAAA, AGGAVAAAAA	Moderate Command	524	498-550	513	●	538	●	506	●
2 AGGLNAAAAAB, AGGAVAAAAAB P.	Moderate Command	521	495-547	440	●	542	●	534	●
3 AGGLNAAAAAC, AGGAVAAAAAC I.	Moderate Command	574	552-596	586	●	555	●	596	●
4 AGGLNAAAAAD, AGGAVAAAAAD	Moderate Command	518	491-545	436	●	518	●	566	●
5 AGGLNAAAAAE, AGGAVAAAAAE U.	Moderate Command	511	483-539	551	●	512	●	461	○
6 AGGLNAAAAAG, AGGAVAAAAAG B.	Moderate Command	542	518-566	535	●	549	●	534	●
7 AGGLNAAAAAH, AGGAVAAAAAH O.	Moderate Command	484	453-515	546	●	403	○	494	●
8 AGGLNAAAAAI, AGGAVAAAAAI I.	Moderate Command	533	508-558	587	●	542	●	427	○
9 AGGLNAAAAAJ, AGGAVAAAAAJ P.	Moderate Command	531	506-556	530	●	531	●	532	●
10 AGGLNAAAAAK, AGGAVAAAAAK K.	Moderate Command	514	487-541	537	●	525	●	461	○
11 AGGLNAAAAAL, AGGAVAAAAAL A.	Moderate Command	523	497-549	494	●	511	●	562	●
12 AGGLNAAAAAM, AGGAVAAAAAM E.	Moderate Command	498	469-527	523	●	527	●	380	○
13 AGGLNAAAAAN, AGGAVAAAAAN Q.	Moderate Command	540	516-564	628	●	490	●	522	●
14 AGGLNAAAAAO, AGGAVAAAAAO C.	Moderate Command	534	509-559	573	●	492	●	560	●
15 AGGLNAAAAAP, AGGAVAAAAAP B.	Moderate Command	514	487-541	347	○	555	●	496	●
16 AGGLNAAAAAQ, AGGAVAAAAAQ	Moderate Command	539	515-563	591	●	532	●	495	●
17 AGGLNAAAAAR, AGGAVAAAAAR G.	Moderate Command	505	476-534	566	●	447	○	513	●
18 AGGLNAAAAAS, AGGAVAAAAAS M.	Moderate Command	534	509-559	550	●	549	●	482	●

Note: Students with no scores are not included in summary calculations.



Content Standards Roster

Colorado Measures of Academic Success

Spring 2014

School: [REDACTED]
District: [REDACTED]

Science

CONFIDENTIAL - DO NOT DISTRIBUTE

Grade 5

Purpose: This report presents each student's performance on the prepared graduate competencies and grade level expectations for your school or district. Percent correct for each PGC is presented. If there is more than one GLE within a PGC then percent correct by GLE is also provided.

		Prepared Graduate Competencies (PGC) and Grade Level Expectations (GLE) Performance						
		Physical Science	Life Science		Earth Systems Science			
		Points Possible						
		20	13	17	10	11	11	9
		PGC1 GLE1	PGC1 GLE1	PGC2 GLE2	PGC1 GLE1	PGC2	GLE2	GLE3
		State Average	39%	35%	43%	42%	37%	37%
District Average	37%	37%	43%	46%	37%	38%	36%	
School Average	30%	30%	46%	51%	38%	39%	36%	
Student								
1	AGGLNAAAAA, AGGAVAAAAA	25%	62%	35%	50%	35%	36%	9%
2	AGGLNAAAAAB, AGGAVAAAAAB P.	15%	46%	47%	50%	40%	45%	9%
3	AGGLNAAAAAC, AGGAVAAAAAC I.	40%	54%	41%	70%	55%	55%	9%
4	AGGLNAAAAAD, AGGAVAAAAAD	15%	31%	47%	70%	45%	55%	9%
5	AGGLNAAAAAE, AGGAVAAAAAE U.	30%	0%	59%	40%	30%	27%	9%
6	AGGLNAAAAAG, AGGAVAAAAAG B.	25%	46%	53%	60%	40%	18%	9%
7	AGGLNAAAAAH, AGGAVAAAAAH O.	35%	31%	18%	40%	35%	45%	9%
8	AGGLNAAAAAI, AGGAVAAAAAI I.	45%	23%	59%	50%	25%	36%	9%
9	AGGLNAAAAAJ, AGGAVAAAAAJ P.	25%	46%	41%	60%	40%	27%	9%
10	AGGLNAAAAAK, AGGAVAAAAAK K.	30%	31%	35%	50%	30%	45%	9%
11	AGGLNAAAAAL, AGGAVAAAAAL A.	20%	23%	41%	50%	50%	64%	9%
12	AGGLNAAAAAM, AGGAVAAAAAM E.	30%	15%	53%	40%	20%	18%	9%
13	AGGLNAAAAAN, AGGAVAAAAAN Q.	55%	23%	35%	50%	40%	27%	9%
14	AGGLNAAAAAO, AGGAVAAAAAO C.	35%	23%	41%	60%	50%	55%	9%
15	AGGLNAAAAAP, AGGAVAAAAAP B.	10%	31%	65%	50%	35%	36%	9%
16	AGGLNAAAAAQ, AGGAVAAAAAQ	45%	8%	65%	50%	35%	45%	9%
17	AGGLNAAAAAR, AGGAVAAAAAR G.	35%	15%	35%	30%	45%	36%	9%
18	AGGLNAAAAAS, AGGAVAAAAAS M.	30%	31%	59%	50%	30%	27%	9%

Note: Students with no scores are not included in summary calculations.



District Performance Level Summary

Colorado Measures of Academic Success

Spring 2014

District: **EXAMPLE DISTRICT (1234)**

Science

CONFIDENTIAL - DO NOT DISTRIBUTE

Grade 8

Purpose: This report describes group achievement in terms of performance levels.

Purpose: This report describes group achievement in terms of performance levels.	Total Number Tested	Average Scale Score	Performance Levels								Strong and Distinguished		No Scores Reported
			Limited Command		Moderate Command		Strong Command		Distinguished Command				
			#	%	#	%	#	%	#	%	#	%	#
State	63,282	595	19,943	32%	22,149	36%	18,496	30%	1,471	2%	19,967	32%	1,223
District	205	456	9	82%	1	9%	1	9%	0	0%	1	9%	194
Gender													
Female	93	431	4	100%	0	0%	0	0%	0	0%	0	0%	89
Male	112	470	5	71%	1	14%	1	14%	0	0%	1	14%	105
Ethnicity/Race													
Hispanic or Latino	56	420	5	100%	0	0%	0	0%	0	0%	0	0%	51
American Indian or Alaska Native	4	0	0	0%	0	0%	0	0%	0	0%	0	0%	4
Asian	4	0	0	0%	0	0%	0	0%	0	0%	0	0%	4
Black or African-American	18	0	0	0%	0	0%	0	0%	0	0%	0	0%	18
White	85	585	1	50%	0	0%	1	50%	0	0%	1	50%	83
Native Hawaiian or Other Pacific Islander	3	0	0	0%	0	0%	0	0%	0	0%	0	0%	3
Two or more races	5	0	0	0%	0	0%	0	0%	0	0%	0	0%	5
Not Indicated	30	436	3	75%	1	25%	0	0%	0	0%	0	0%	26
Language Background													
English	137	566	3	75%	0	0%	1	25%	0	0%	1	25%	133
Spanish	0	0	0	0%	0	0%	0	0%	0	0%	0	0%	0
Other	0	0	0	0%	0	0%	0	0%	0	0%	0	0%	0
Not Indicated	42	393	6	86%	1	14%	0	0%	0	0%	0	0%	35
Language Proficiency													
Not Applicable	155	495	5	83%	0	0%	1	17%	0	0%	1	17%	149
NEP	2	0	0	0%	0	0%	0	0%	0	0%	0	0%	2
LEP	2	0	0	0%	0	0%	0	0%	0	0%	0	0%	2
FEP	8	0	0	0%	0	0%	0	0%	0	0%	0	0%	8
PHLOTE	1	0	0	0%	0	0%	0	0%	0	0%	0	0%	1
FELL	0	0	0	0%	0	0%	0	0%	0	0%	0	0%	0
Not Indicated	37	409	4	80%	1	20%	0	0%	0	0%	0	0%	32

This report is NOT for public review. Distribution within your school/district must be in accordance with state and federal privacy laws, and local school board policy.



School Item Analysis Report

Colorado Measures of Academic Success

Spring 2014

School: [REDACTED]
District: [REDACTED]

Social Studies

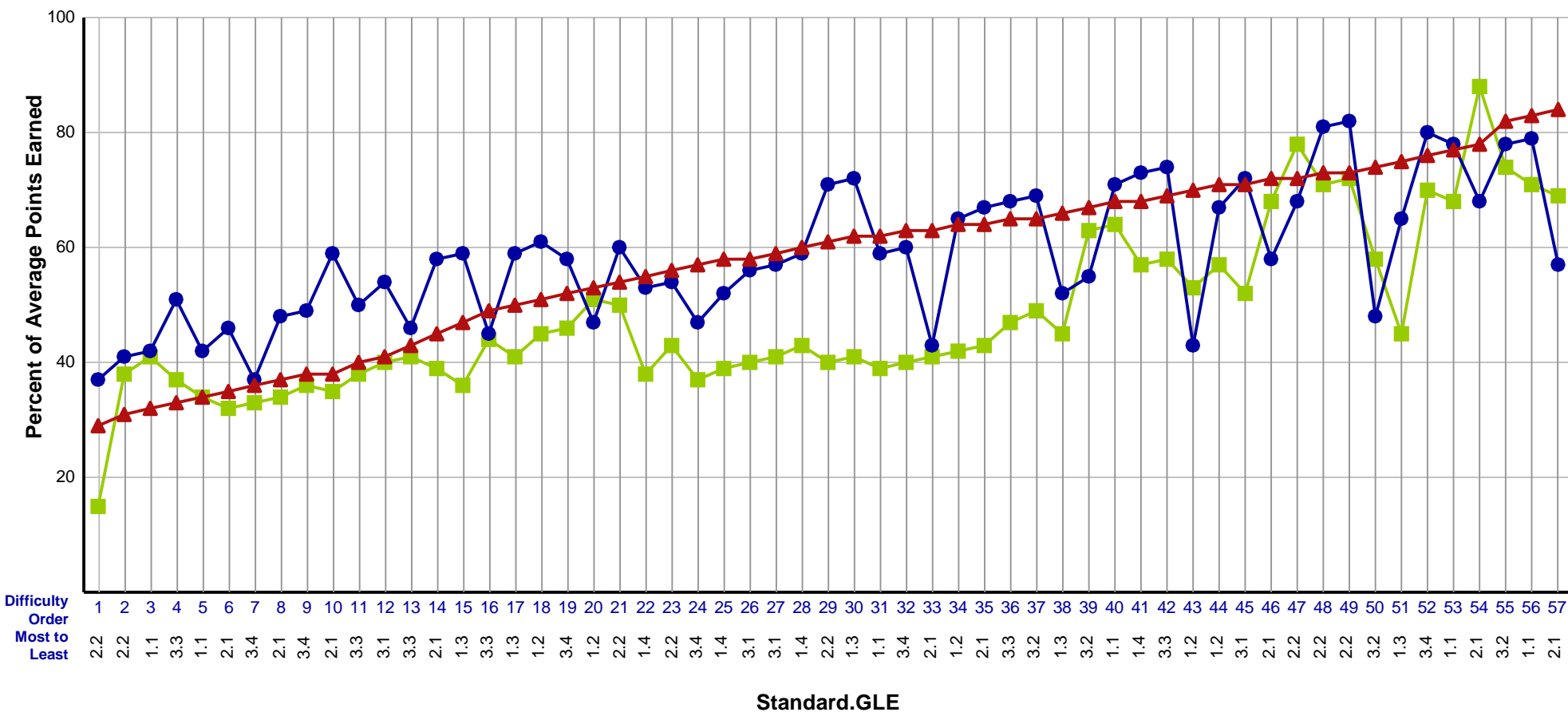
CONFIDENTIAL - DO NOT DISTRIBUTE

Grade 99

Students with Valid Scores (300)

Students testing on paper are not included.

- State
- District
- School



This report is NOT for public review. Distribution within your school/district must be in accordance with state and federal privacy laws, and local school board policy.



Item Analysis Report Detail

Colorado Measures of Academic Success

Spring 2014

This report shows the operational items for the given grade and subject sorted by difficulty.

Science

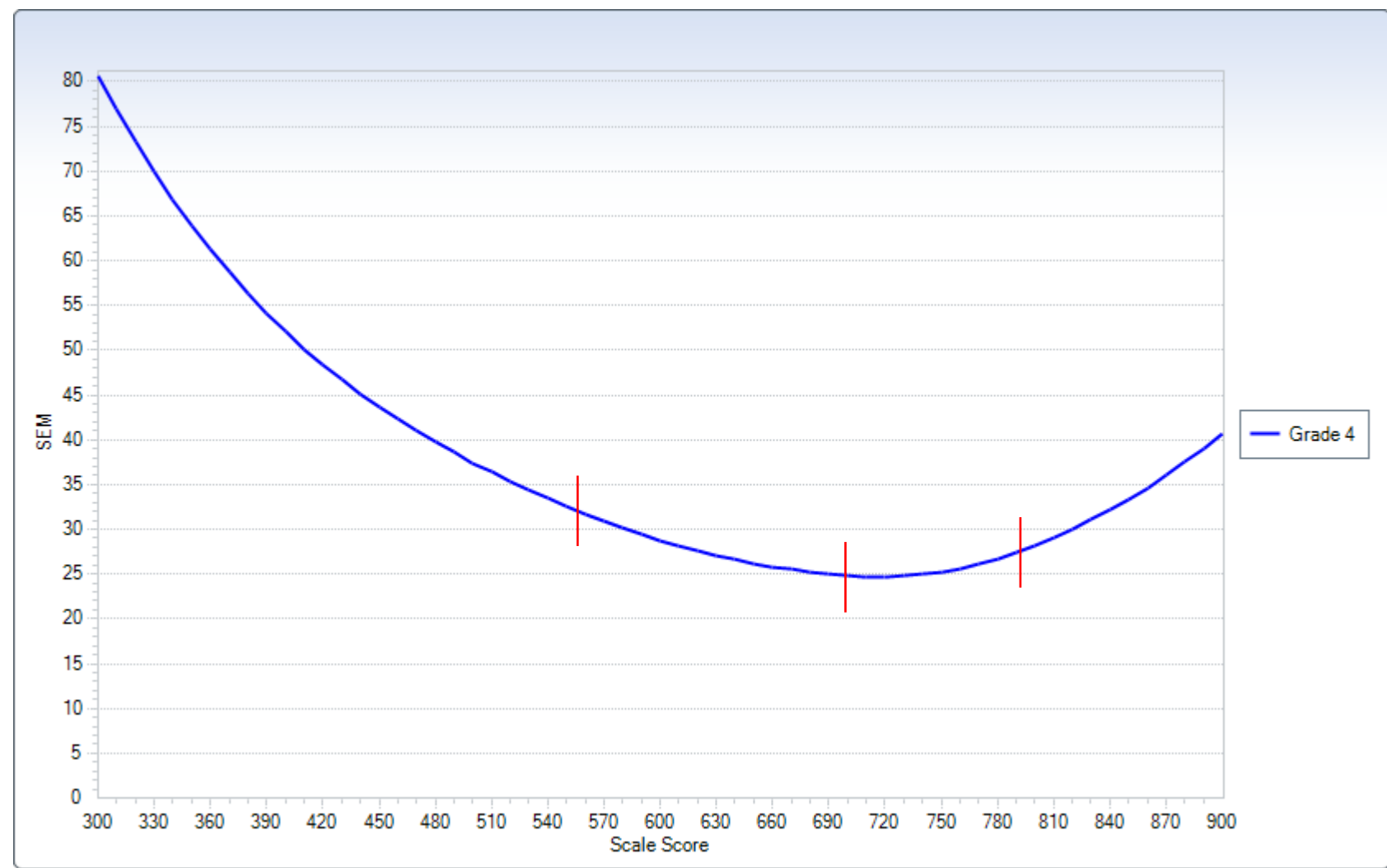
CONFIDENTIAL - DO NOT DISTRIBUTE

Grade 8

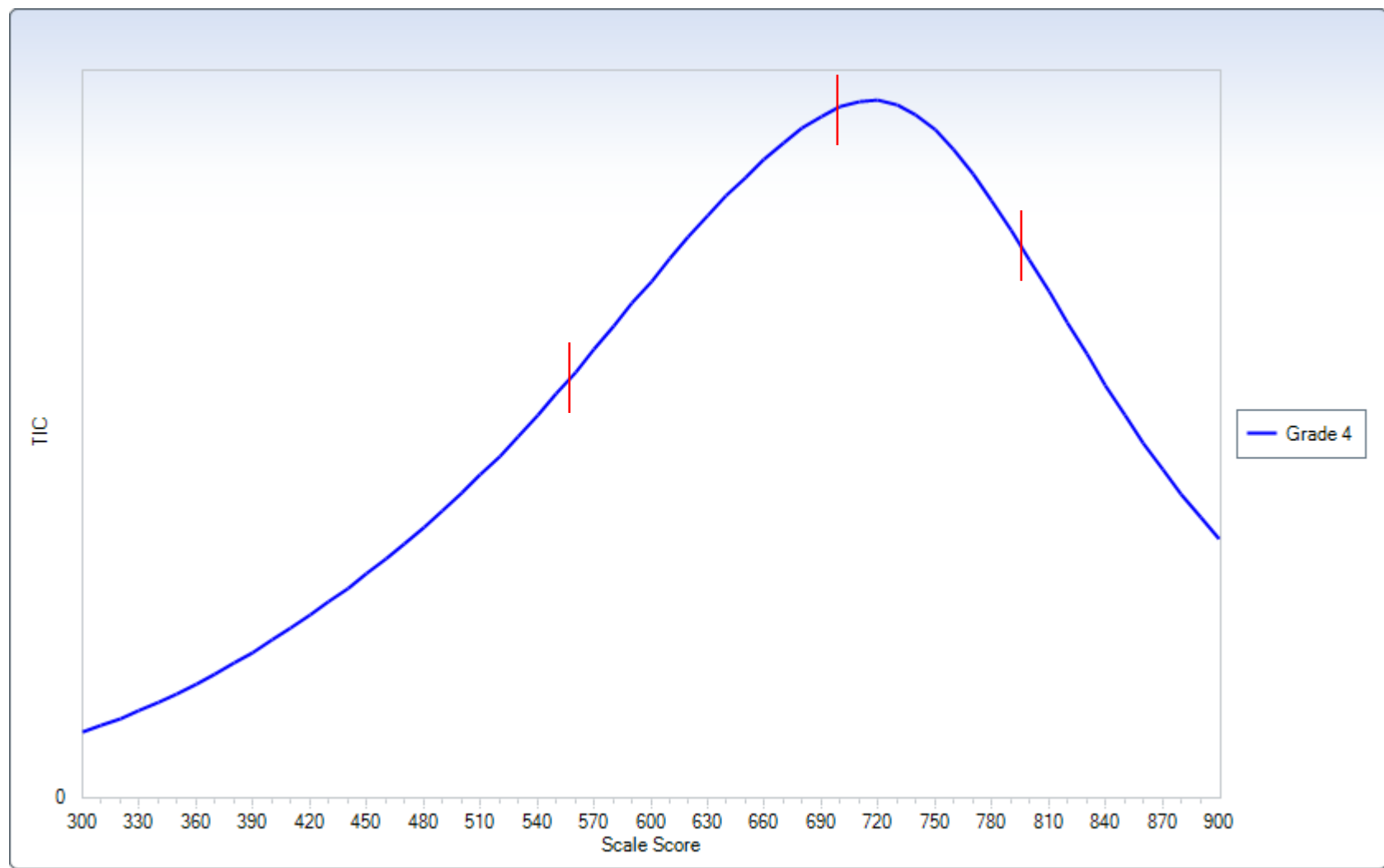
Difficulty Order Most to Least	Standard.GLE	Section-Item Number	Standard	Prepared Graduate Competencies (PGCs)	Grade Level Expectations (GLEs)	Item Type Selected Response (SR) Constructed Responses (CR)
1	2.2	1-5	Life Science	PGC 2	GLE 2	SR
2	2.2	1-13	Life Science	PGC 2	GLE 2	CR-3
3	1.1	3-10	Physical Science	PGC 1	GLE 1	SR
4	3.3	3-22	Earth Systems Science	PGC 2	GLE 3	CR-2
5	1.1	3-20	Physical Science	PGC 1	GLE 1	SR
6	2.1	3-14	Life Science	PGC 1	GLE 1	SR
7	3.4	3-6	Earth Systems Science	PGC 2	GLE 4	SR
8	2.1	3-4	Life Science	PGC 1	GLE 1	SR
9	3.3	1-21	Earth Systems Science	PGC 2	GLE 3	SR
10	3.1	2-18	Earth Systems Science	PGC 1	GLE 1	CR-2
11	3.3	1-14	Earth Systems Science	PGC 2	GLE 3	SR
12	2.1	1-6	Life Science	PGC 1	GLE 1	CR-2
13	1.3	1-16	Physical Science	PGC 3	GLE 3	SR
14	3.3	1-24	Earth Systems Science	PGC 2	GLE 3	SR
15	3.1	2-4	Earth Systems Science	PGC 1	GLE 1	SR
16	2.2	2-12	Life Science	PGC 2	GLE 2	SR
17	1.3	2-21	Physical Science	PGC 3	GLE 3	SR
18	1.2	2-24	Physical Science	PGC 2	GLE 2	SR
19	3.4	3-9	Earth Systems Science	PGC 2	GLE 4	CR-2
20	1.2	3-15	Physical Science	PGC 2	GLE 2	SR
21	2.2	3-17	Life Science	PGC 2	GLE 2	CR-2
22	1.4	3-21	Physical Science	PGC 2	GLE 4	SR
23	2.2	3-12	Life Science	PGC 2	GLE 2	SR
24	3.4	3-7	Earth Systems Science	PGC 2	GLE 4	SR
25	1.4	2-22	Physical Science	PGC 2	GLE 4	CR-2
26	3.1	2-19	Earth Systems Science	PGC 1	GLE 1	CR-2
27	3.1	2-16	Earth Systems Science	PGC 1	GLE 1	SR
28	1.4	2-10	Physical Science	PGC 2	GLE 4	SR
29	2.2	1-23	Life Science	PGC 2	GLE 2	SR
30	1.3	1-18	Physical Science	PGC 3	GLE 3	CR-2
31	1.1	1-15	Physical Science	PGC 1	GLE 1	SR
32	3.4	1-11	Earth Systems Science	PGC 2	GLE 4	SR
33	2.1	1-8	Life Science	PGC 1	GLE 1	SR
34	1.2	1-4	Physical Science	PGC 2	GLE 2	SR
35	2.1	1-7	Life Science	PGC 1	GLE 1	SR
36	3.3	1-10	Earth Systems Science	PGC 2	GLE 3	SR
37	3.2	1-12	Earth Systems Science	PGC 1	GLE 2	SR
38	1.3	1-17	Physical Science	PGC 3	GLE 3	CR-2
39	3.2	1-20	Earth Systems Science	PGC 1	GLE 2	SR
40	1.1	1-22	Physical Science	PGC 1	GLE 1	CR-2
41	1.4	2-3	Physical Science	PGC 2	GLE 4	CR-2
42	3.3	2-11	Earth Systems Science	PGC 2	GLE 3	SR
43	1.2	2-13	Physical Science	PGC 2	GLE 2	CR-3
44	1.2	2-15	Physical Science	PGC 2	GLE 2	SR
45	3.1	2-17	Earth Systems Science	PGC 1	GLE 1	SR
46	2.1	2-23	Life Science	PGC 1	GLE 1	SR
47	2.2	3-19	Life Science	PGC 2	GLE 2	CR-2
48	2.2	3-18	Life Science	PGC 2	GLE 2	SR
49	2.2	3-16	Life Science	PGC 2	GLE 2	SR
50	3.2	3-13	Earth Systems Science	PGC 1	GLE 2	CR-3
51	1.3	3-5	Physical Science	PGC 3	GLE 3	SR
52	3.4	3-8	Earth Systems Science	PGC 2	GLE 4	CR-2
53	1.1	2-20	Physical Science	PGC 1	GLE 1	SR
54	2.1	2-14	Life Science	PGC 1	GLE 1	SR
55	3.2	2-5	Earth Systems Science	PGC 1	GLE 2	SR
56	1.1	1-19	Physical Science	PGC 1	GLE 1	SR
57	2.1	1-9	Life Science	PGC 1	GLE 1	CR-2

APPENDIX C: IRT CURVES

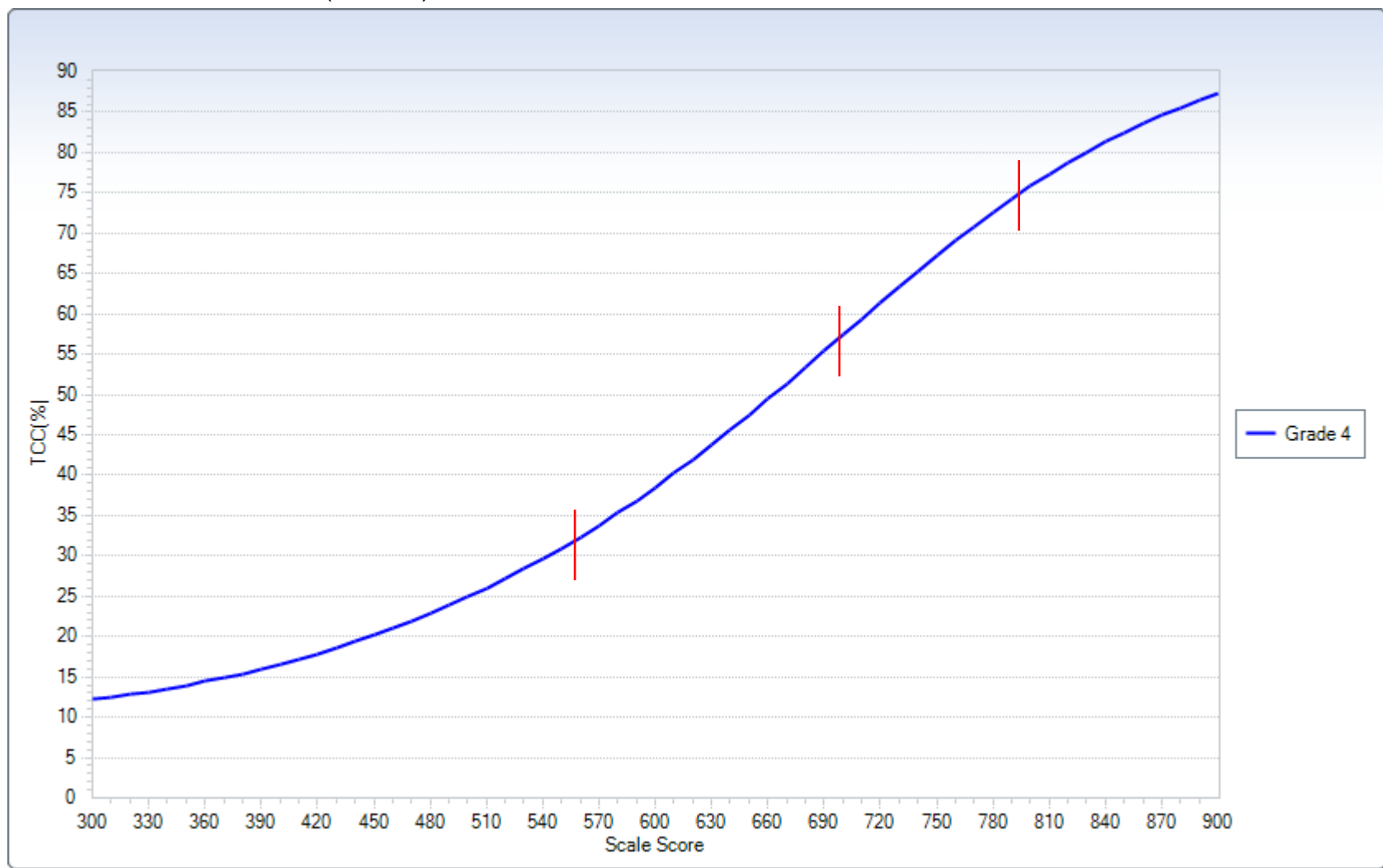
Test Summary Curves
Conditional Standard Error of Measurement Curves



Test Summary Curves
Test Information Curves

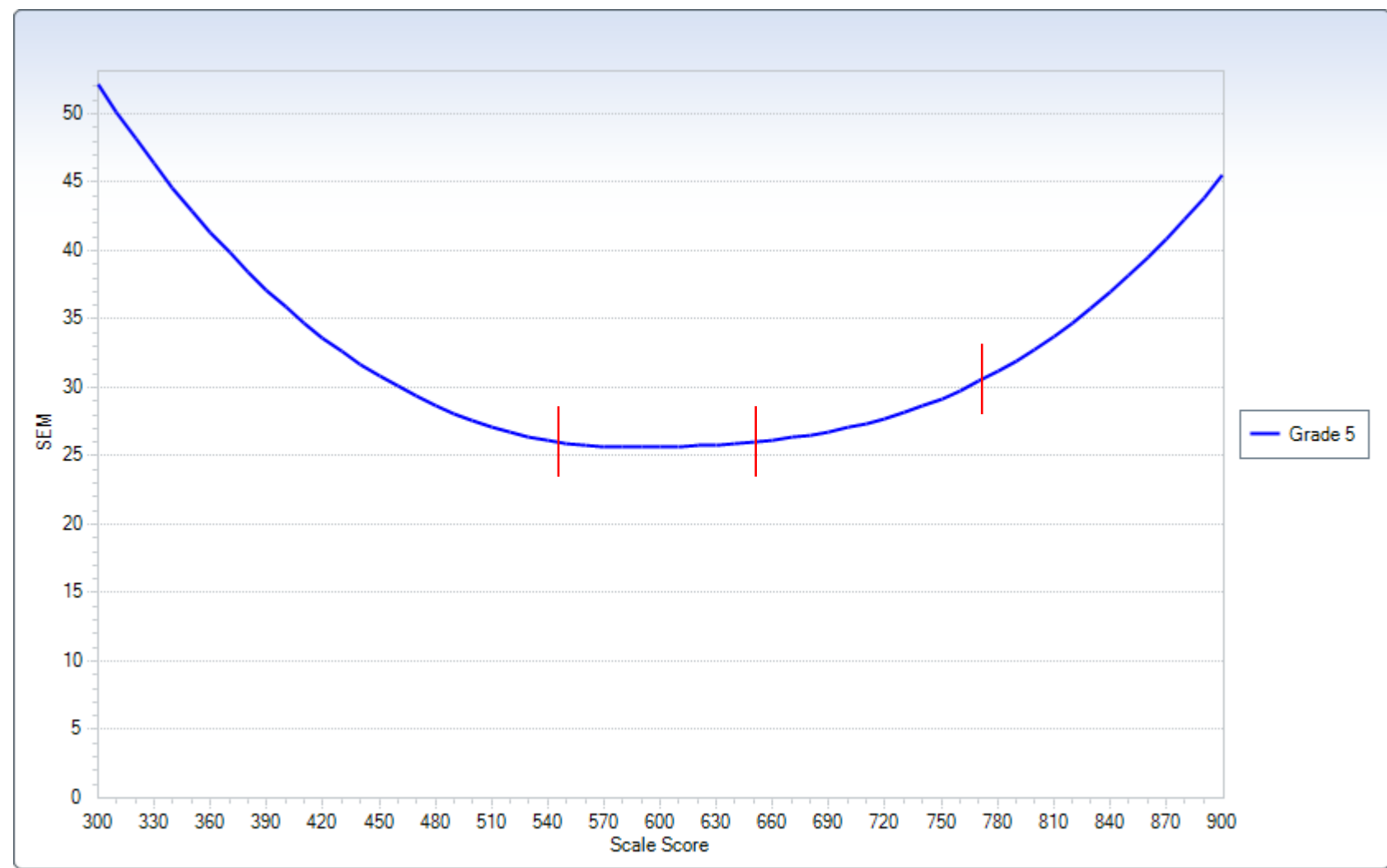


Test Summary Curves
Test Characteristic Curves (Percent)

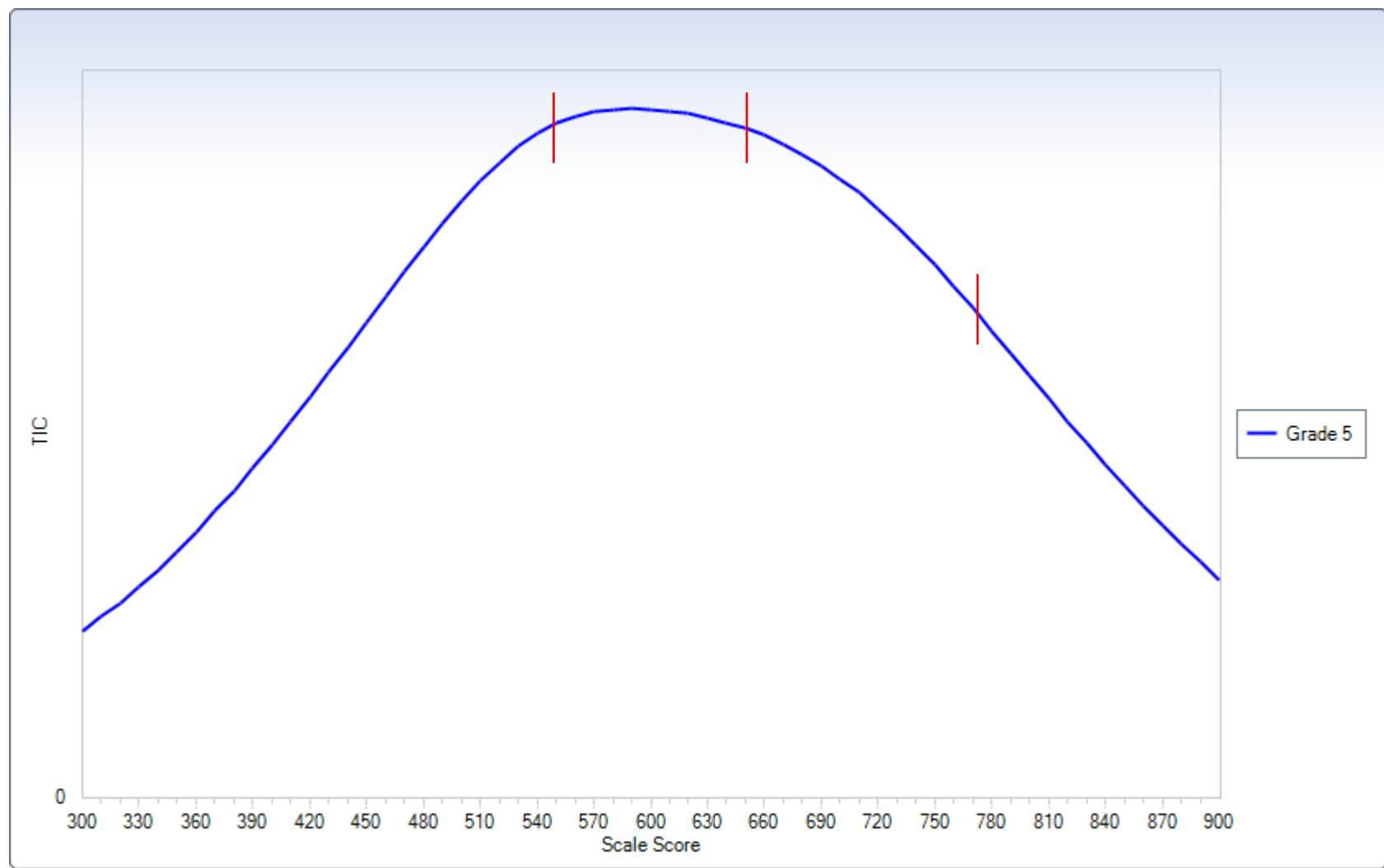


Test Summary Curves

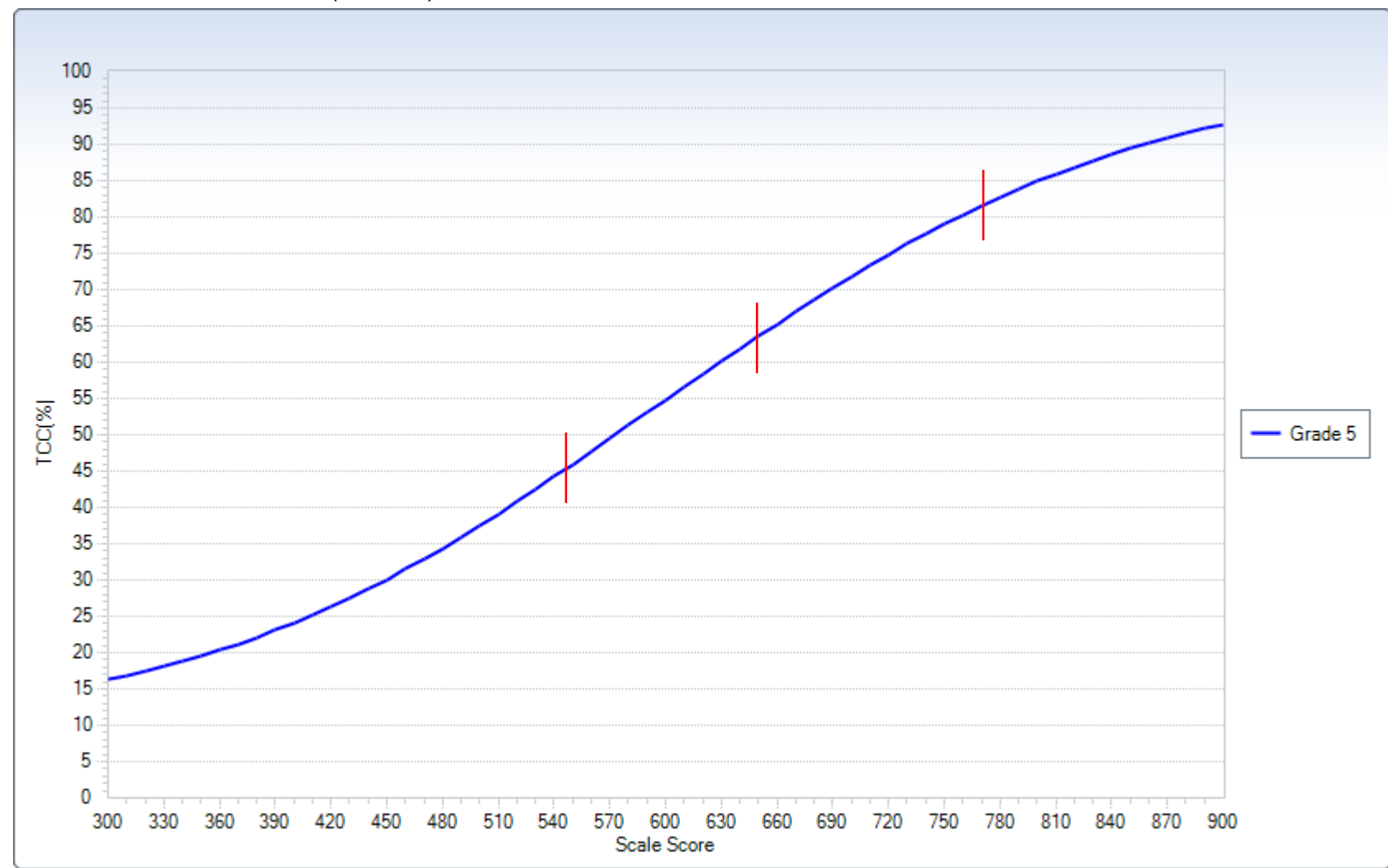
Conditional Standard Error of Measurement Curves



Test Summary Curves
Test Information Curves

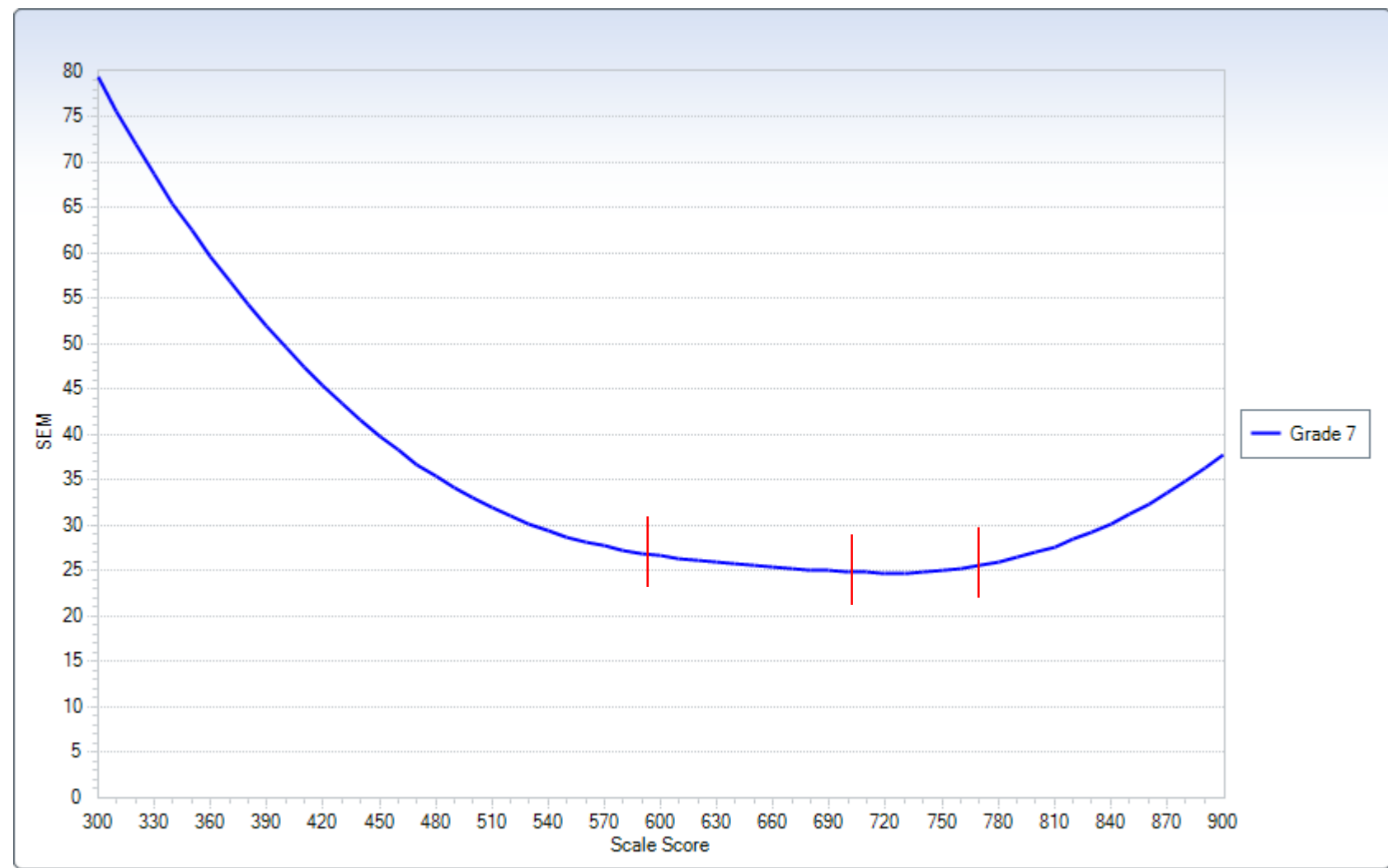


Test Summary Curves
Test Characteristic Curves (Percent)

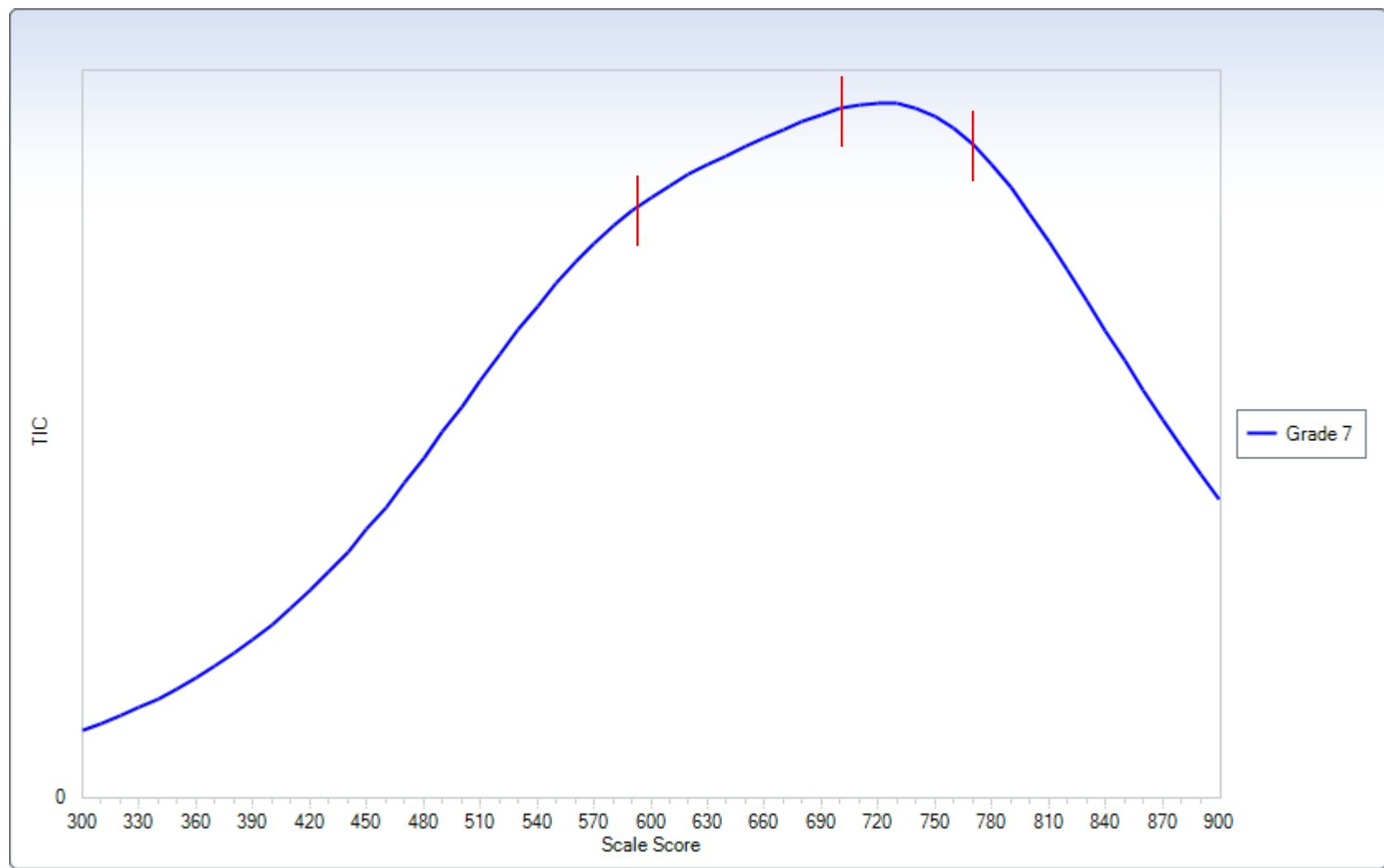


Test Summary Curves

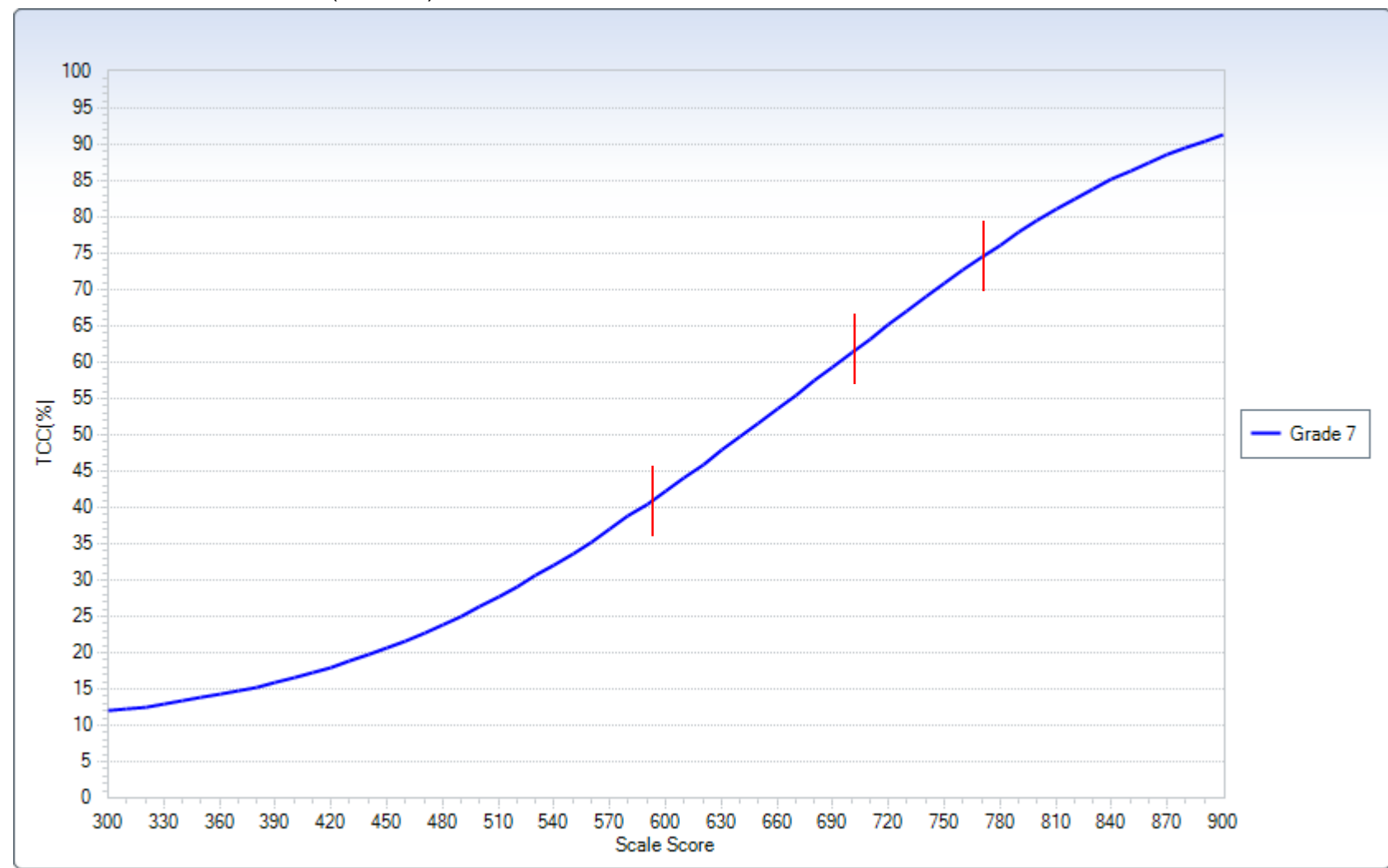
Conditional Standard Error of Measurement Curves



Test Summary Curves
Test Information Curves

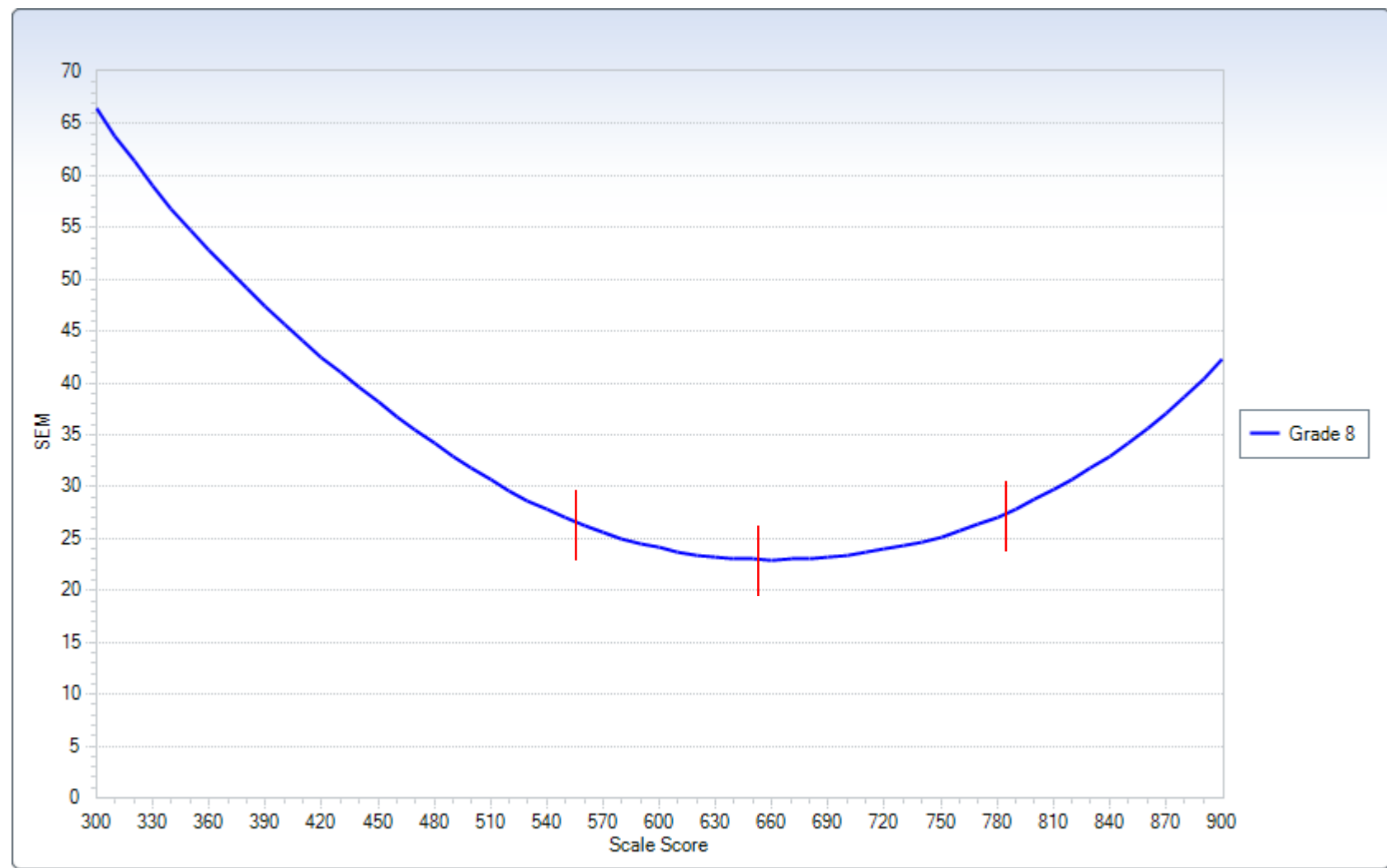


Test Summary Curves
Test Characteristic Curves (Percent)

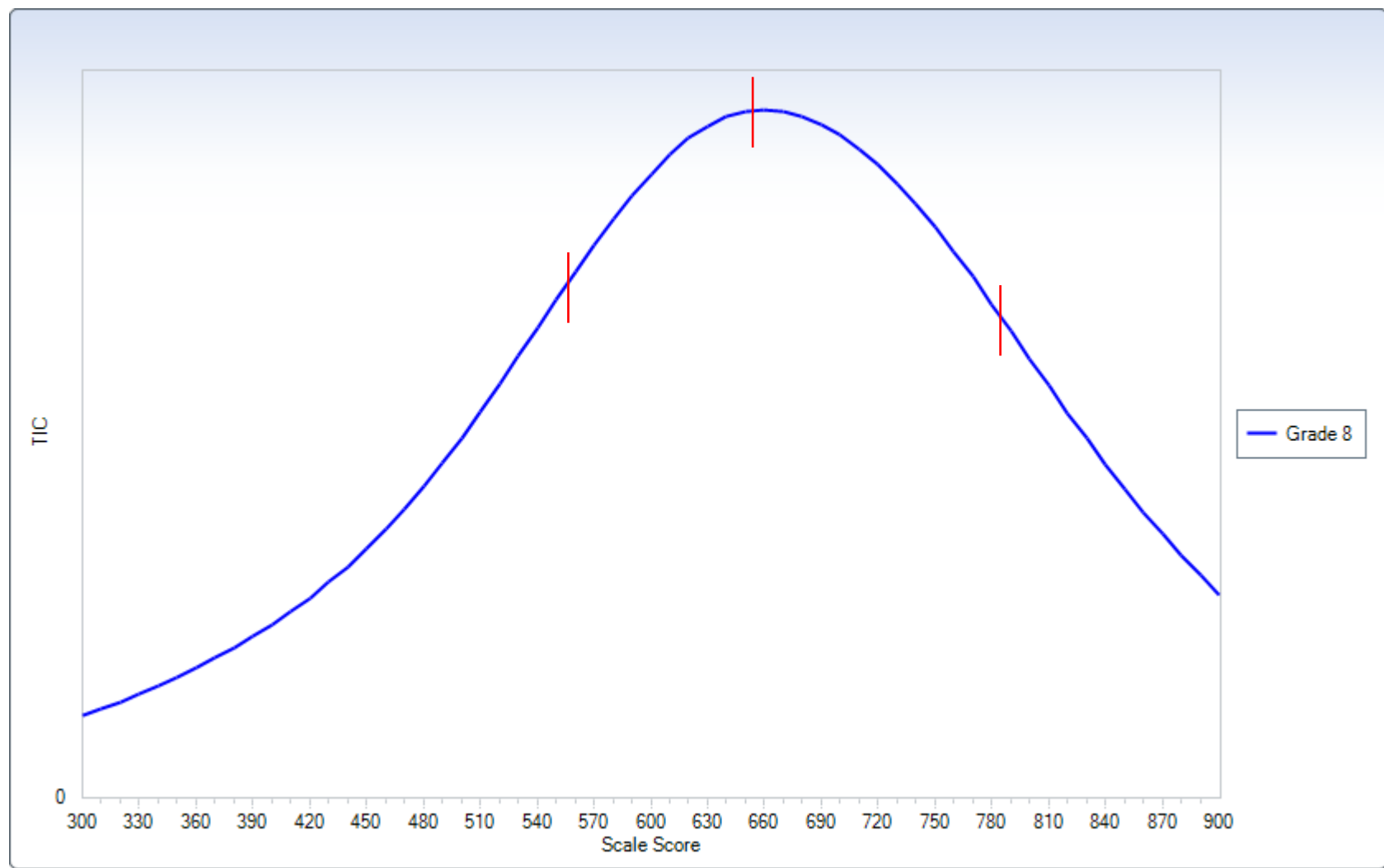


Test Summary Curves

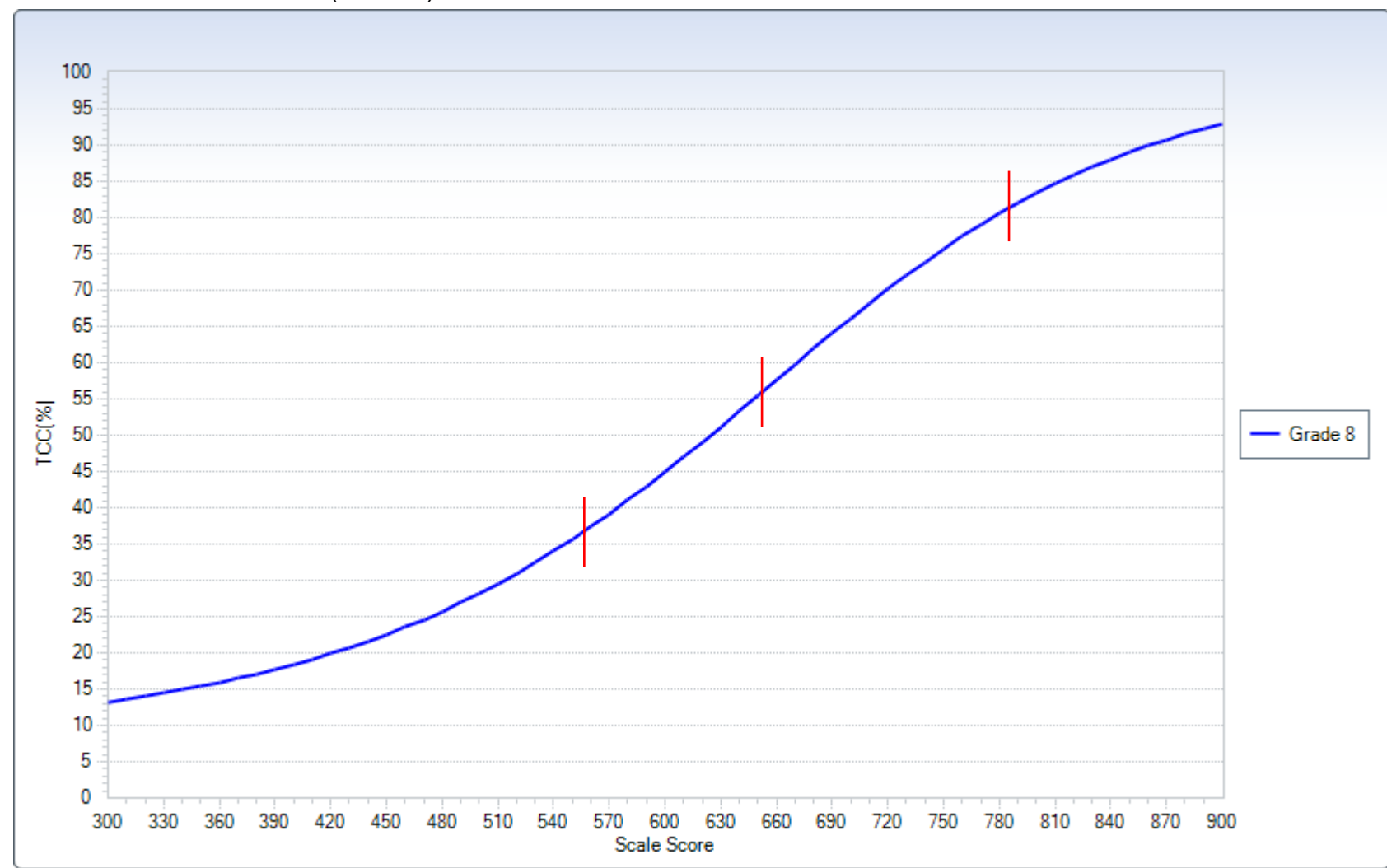
Conditional Standard Error of Measurement Curves



Test Summary Curves
Test Information Curves



Test Summary Curves
Test Characteristic Curves (Percent)



APPENDIX D: COGNITIVE LAB REPORT

Report on Cognitive Labs for CO Summative Assessment

Pearson

May 2013

Introduction

Pearson is currently developing summative statewide assessments under a contract with the Colorado Department of Education (CDE) for science and social studies. These assessments will become operational beginning with the spring 2014 administration. The social studies assessment will be administered to students in grades 4, 7, and 12; the science assessment will be given to students in grades 5, 8, and 12. The administration of both assessments at grade 12 will take place in the fall, while the lower grades will take the assessments in the spring.

These assessments will differ from the assessments currently in place in that they will be administered via computer (except for accommodated paper-based versions for students to whom the computer-based version is inaccessible), while all current statewide tests (TCAP) are administered in paper-and-pencil format. Because these computer-based tests are new to students, there is the concern that these new item types may contain unforeseen obstacles to students being able to fully demonstrate their science or social studies ability on these assessments. This concern is particularly acute with respect to limited English proficiency (LEP) and disabled students. In addition, when items are administered via computer, any sorts of manipulatives or tools (such as rulers or calculators) required to answer the item also must be implemented on the computer. While students are likely familiar with the use of some of these tools, they may not have much or any experience in using others in an online setting. Finally, the CDE is interested in confirming that students are able to successfully navigate through the test using Pearson's TestNav browser-based testing system. Cognitive labs were therefore planned and executed to allow the preceding areas to be researched. The CDE was also concerned about TestNav's help system (for the online tools), but there was little opportunity for students to interact with it during the cognitive labs, and consequently little data about its usability was generated.

The cognitive labs were conducted with students in the grades for which the new tests will be implemented (grades 4, 7, and high school [social studies], and grades 5, 8, and high school [science]). The cognitive labs with the high school students took place from April 30 through May 2, 2013, while those with the elementary and middle school students were conducted between May 15 and May 17, 2013. The high school cognitive labs used grade 7 social studies and grade 8 science items, while the labs for the lower grades all used items targeted to that specific grade. The use of off-grade items for the high school cognitive labs was approved by the Colorado Technical Advisory Committee (TAC), given that the focus of the cognitive labs was on the usability of the items, tools, and system rather than their grade-specific content. Students were administered between 7 (high school social studies) and 16 (grade 7 social studies) items, and in most cases completed them within 15 minutes to an hour. Students who were unable to complete all items within an hour were dismissed at the end of that time.

To capture as wide a sample of Colorado students as possible, students were sampled from one school in each of rural, urban, and suburban settings. Four students were sampled at each grade in the elementary and middle schools, and eight students were sampled in each of the high schools visited (four for social studies and four for science). This allowed four students to participate in each of the two subjects tested. There were only two exceptions. We were unable to secure any grade 7 students in one of the participating schools, and so the four participating grade 8 students at that school were split into two groups, and two students each were administered the grade 7 social studies and grade 8 science items. In addition, one student scheduled to participate in the high school social studies lab was absent the day of

the lab, leaving only three participants in the lab for that school. Participating schools are listed in Table 1.

Table 1. Participating schools and districts for the May 2013 cognitive labs.

Grade	Setting	District	School
4/5	Urban	Adams 14	Hanson Elementary School
7/8			Adams City Middle School
High School			Adams High School
4/5	Suburban	Cherry Creek	Arrowhead Elementary School
7/8		Sheridan 2	Sheridan Middle School
High School		Cherry Creek	Grandview High School
4/5	Rural	Weld Re 5J	Milliken Elementary School
7/8		Greeley	Franklin Middle School
High School			Jefferson High School

Schools were asked to select the students to participate. They were specifically asked to select LEP, learning disabled, and low achieving students along with students from the general population, if possible. Since all sessions were facilitated by Pearson and Educational Testing Service (ETS) staff, schools were informed that students requiring moderate accommodations could also be selected, though none were. Table 2 summarizes the demographic makeup of the cognitive lab participants.

Table 2. Demographic characteristics of participants in the May 2013 cognitive labs.

	Grade 4 (Social Studies)	Grade 5 (Science)	Grade 7 (Social Studies)	Grade 8 (Science)	HS (Social Studies)	HS (Science)
N	12	12	10	10	11	12
Male	4	6	5	5	7	7
Female	8	6	5	5	4	5
Caucasian	6	6	2	3	4	5
Hispanic	5	6	8	7	6	6
Asian	1	0	0	0	0	1
African American	0	0	0	0	1	0
LEP (Current/Exited)	2	3	3	3	6	5
SPED	2	3	1	2	1	2

As stated previously, the emphasis of the cognitive labs was on the usability of the testing system, items, tools, and help system. While all students will have the opportunity to take a practice test prior to taking the tests operationally, the students in the study were only shown the use of the system in the context of a demonstration of the “think aloud” protocol by one of the facilitators. Thus, the results of these labs form a baseline that shows what the students are able to do within the testing system using only the general technological skills that they have acquired during the course of their educational career without the benefit of training in or familiarity with TestNav. Areas where students had difficulties during the study

thus are a guide to the CDE and Pearson as they create the practice tests that will be available to students prior to taking the operational assessment.

Procedure

The procedure followed during the first and second sets of cognitive labs are broadly similar, but the procedures for the second set were refined based on the experience gained in the first set. Students were tested in the presence of one ETS content specialist who acted as the lead facilitator and one Pearson research scientist or content specialist who acted as observer and co-facilitator. Each student was tested separately, either in an individual room or in a common room, isolated from each other and otherwise free from distractions.

Each session began with the facilitators introducing themselves to the student and thanking him or her for participating. They explained that they were with the company (Pearson) that was developing new tests for the state in science and social studies, and that these tests were to be given on a computer, unlike previous tests which are administered in paper form. The student was told that he or she would, in essence, be “testing the test,” and that the test would not be graded nor would any feedback on the results be given to their parents, teachers, or anyone else. Students were also informed that the facilitators would be taking notes as they worked through the test and that their actions on the screen and audio (but not video) would also be captured so that they could be reviewed by the facilitators after the session was finished. They were told that these notes and recordings would be destroyed after the report documenting the labs was completed, and that they would not be shared with anyone besides the facilitators and the CDE.

After the introductions were completed and the forgoing information given to the student, one of the facilitators demonstrated the think aloud protocol to the student that he or she would be asked to use in working through the test. This protocol was demonstrated using a set of items from a grade and subject other than the one that the student would be addressing. For the first set of labs (high school), a grade 5 simulation was used, and the facilitator worked through several items associated with the simulation. For the second set of labs (elementary and middle school grades), the grade and subject of the facilitator’s sample items varied. During the first set of labs, students rarely noticed that there were tools available within TestNav. In order to alert students to the presence of the tools without explicitly directing the student to access them in the second set of labs, the facilitator drew their attention to their presence by using several during this initial demonstration of the think aloud protocol. In order to demonstrate the use of the “answer eliminator” tool, a multiple choice item was required. For this reason, the demonstration of the think aloud protocol during the second set of labs began with a multiple choice item.

At the conclusion of the demonstration of the think aloud protocol, the student was asked if he or she understood the task and if there were any questions. After verifying that the task was clear and that there were no questions, the facilitator then allowed the student to begin taking the test.

As the student worked through the test, the facilitators followed his or her progress, making note of any particular difficulties (or lack thereof) with the different item types present. More reticent students were prompted at regular intervals to verbalize their thought processes, and facilitators stopped them at various points throughout the test to further explore their interaction with particular aspects of TestNav, the items, and the tools. After the student finished the test (or after the allotted time period had elapsed), he or she

was asked to complete a survey that measured the student's perception of the usability of the system, items, and tools, and how well they liked taking a test on the computer. He or she also had an opportunity to give any comments on the experience that weren't otherwise covered in the survey. Finally, all students received a \$25 Visa or Amazon gift card to thank them for participating in the study.

Item Types

There were several types of items that were presented to students during the labs. Students were presented with traditional selected and constructed response items within the context of social studies performance events and science simulations. Both had a set of items that all referenced a common stimulus or set of stimuli. For the grade 4 performance event, this common set was composed of historical passages and addressed the history standard within social studies, while the grade 7 performance event targeted the geography standard and included three maps. Science simulations used an animation of a science experiment as their central stimulus, with all questions addressing various aspects of the depicted experiment. All selected and constructed response items used within the labs were associated with either a performance event or a simulation.

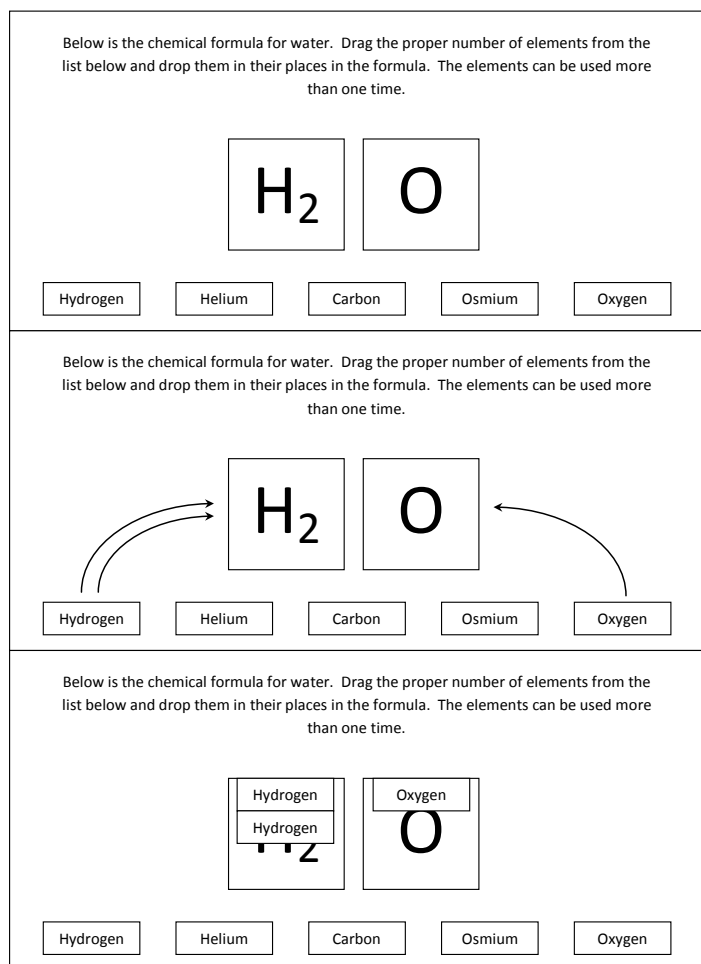


Figure 1. Drag-and-drop item (not an actual item on the test). The top pane shows the initial state of the item, the middle pane shows the actions required to answer the item correctly (drag two hydrogens and one oxygen and drop them into their respective boxes), and the bottom pane shows the final state.

Independent standalone items were all technology-enhanced and fell into two broad categories: drag-and-drop items and hotspot items. A sample drag-and-drop item is shown in Figure 1 on the previous page. The distinguishing characteristic of this item type is the manner in which the student indicates his or her answer—by clicking on an answer option (“dragger”) with the mouse, holding the mouse button down, dragging it over to a specific area on the screen and dropping it into a specific area (“landing box”) by releasing the mouse button. Both the dragger and landing box indicate the student’s response. In the sample item in Figure 1, to answer the item correctly, the student would drag the “Hydrogen” box over to the “H₂” box and drop it twice, and do the same with the “Oxygen” box, dropping it once.

Variations of this type of item exist and differ mainly in limitations on the draggers and/or landing boxes. Both draggers and landing boxes can be restricted in the number of times they can be used or items they can accept, respectively. In the sample item in Figure 1, the elements can be dragged any number of times to the H₂ and O landing boxes. Limitations can be placed upon the draggers—if each dragger can only be used once, it would no longer appear in its starting position after being used (visually cuing the student that it can only be used once). If the landing boxes are limited, then attempting to drop draggers above the limits would result in the attempted drop failing (the dragger moves back to its original position).

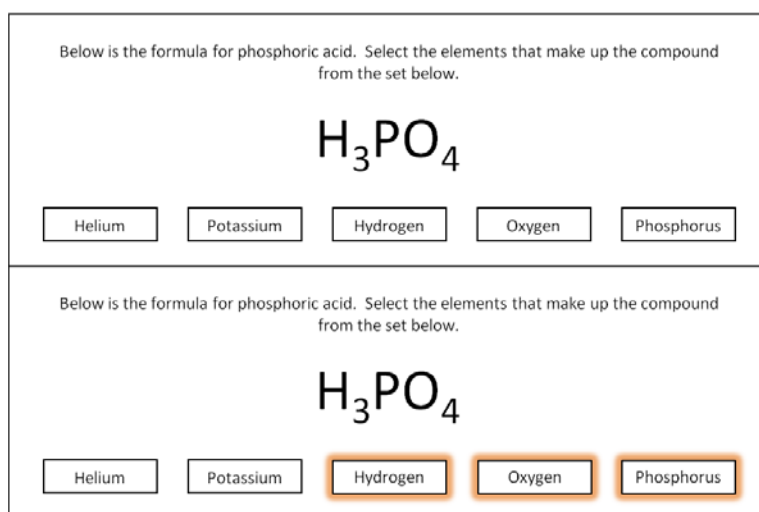


Figure 2. Hotspot item (not an actual item on the test). The top pane shows the initial state of the item and the bottom pane shows the final state of the item after the correct options ("hotspots") have been clicked on.

Figure 2 shows a sample hotspot item. This item is answered by clicking on the hotspot(s) that correspond to one’s desired response. The two variations of hotspot items depend on whether the student is limited in the number of hotspots that can be chosen. Typically, if the number of hotspots that can be selected is limited, the limit is one, and the student is cued to the limit by the first selected hotspot being deselected when a second hotspot is selected (much like the buttons on classic car radios).

For both drag-and-drop and hotspot items, the interactive elements were not limited to just rectangular boxes; draggers, landing pads, and hotspots can be any object on the screen. The cognitive labs included items with both rectangular and irregularly shaped objects. In addition, both unlimited items (i.e., students could drop any number of draggers or select any number of hotspots) and limited items (i.e., students could only drop one dragger per landing pad or could only select a single hotspot) were included.

Results

Performance Events and Simulations

Students were aware that there were no consequences for poor performance on the test items that they took during the labs, and this decreased level of motivation was most apparent in the lack of effort that was expended in answering the constructed response items—many of the students gave these items a perfunctory response and moved on to the next item. This made it difficult to gauge individual students' experience with the simulations or performance events as a coherent unit rather than as individual selected and constructed response items.

It is important to note the differences between the science simulations and the social studies performance events. Though both were comprised of sets of items that referenced a set of common stimuli and both were presented in “split-screen” format with the stimuli on one half of the screen and the current item being responded to on the other, there were important differences between the two. The simulations were animations that the student was expected to view before answering the related items. As such, each simulation item set was preceded by an introductory screen on the half of the screen occupied by the items that gave the student introductory information and prompted him or her to run the simulation. Thus the student was taken step by step through the actions necessary to run the simulation.

In contrast, the initial presentation of the social studies performance events to the student was with the exhibits and first item in the set in the split screen format. There were no introductory pages, just the first item in the associated set. Some students did not notice that the exhibits had multiple pages, despite there being navigation arrows and a label reading “Page 1 of 3” (for example) at the bottom of the first exhibit and attempted to answer the questions without paging through to the relevant exhibit. One suggestion from a student that there be thumbnails or reduced size representations of the exhibits to more prominently indicate the presence of multiple pages in the exhibits was prompted by this issue. In general, students who had not noticed that there were multiple pages quickly recognized that they were in fact present and were able to easily navigate to them when queried by a facilitator.

In most cases, students had little problem with either the simulation or the performance event. Because of its directive nature, the simulation presented the least problem for students to complete. Having to completely go through the simulation before being presented with the first question encouraged most students to examine the results of the simulation before answering the questions. In contrast, for the performance event, while most students were able to navigate through the stimuli for the performance event, some did not notice that there were multiple pages, and so attempted to answer the questions without having access to all of the information contained in the exhibits. This may be at least partially due to their experience with traditional paper based test items—the student is able to see the item and all associated stimuli simultaneously, and if further information is on multiple pages, then the physical pages are also immediately apparent to the student. In contrast, the prompts to the student that there are multiple exhibits in the online performance event are altogether more subtle and easier for students to miss, especially if they have limited experience with technology.

Since the performance events and simulations were all comprised of selected and constructed response items, in most cases the use of tools related to the item types was the same as for standalone items. However, the grade 7 performance event addressed geography, and included the explicit direction for

students to “use the Notepad tool to take notes” as they were studying the maps. A few students did access the Notepad tool after reading the directions; in every case the student briefly glanced at it and then dismissed it without using it in any way. Observation of the students as they were responding to the performance event items showed that for the most part the Notepad would have probably been of little use. Students were able to switch between the maps in the exhibits with no problem as they were composing their answer to the constructed response items or determining the best response to the selected response items.

Selected Response Items

With respect to the different item types on the test, selected response items presented the fewest problems for students. These items’ presentation on the screen was virtually identical to how they appear in a traditional paper-based test; consequently all students were very comfortable in responding to them. Any problems that students had with this item type were due to factors that were unrelated to the items being presented via computer (e.g., limited reading for understanding, limited reading comprehension, cognitive load). Selected response items were the only item type for which the Answer Eliminator tool was functional. In the initial labs with the high school students, tools were not explicitly referenced by the facilitator. Without this guidance, most students did not notice nor attempt to utilize the tools. In the second set of labs with the lower grades, the facilitator deliberately accessed the tools during the course of demonstrating the think aloud protocol to the students. With this indirect prompting, most students did access the tools and used one or more as they were answering the item.

Constructed Response Items

Generally, students also had little trouble answering the constructed response items, though most grade 4 students indicated that they were not familiar with responding to this item type on a computer. Those grade 4 students who were lower performers or who had limited proficiency in English had some difficulty in responding to the items. Finally, some grade 4 students experienced frustration when presented with consecutive constructed response items. Higher performing grade 4 students and most of the grade 5 and higher students were able to respond well to the item, and to use the embedded tools (cut, copy, paste, spell-check, undo, and redo) within the text editing box. These tools were more within the line of sight of the student and thus were more noticeable than those on the tool bar, and were the most accessed tools across all grades. Also, a misspelled word was immediately indicated with a wavy red underline, which tended to catch the student’s attention and clue them to the word-processing functionality built into the text editor. Since the functionality is virtually identical to that found in most word-processing programs, most students were familiar with and adept at using it.

Technology Enhanced (Drag-and-Drop and Hotspot) Items

For both the drag-and-drop and hotspot technology-enhanced item types, students had very few problems accessing and determining how to answer them. Since much of the regular interaction with a computer involves either clicking on or dragging objects, it is unsurprising that students were able to manipulate and answer the items with little difficulty. For students who did have some initial confusion about how to answer an item (usually this was with drag-and-drop items where the dragger was something other than a rectangular box or with hotspot items consisting of several paragraphs where the student responds by

clicking on a particular paragraph), context-sensitive highlighting of the dragger or hotspots when the mouse pointer rolls over them helped them to quickly determine how to indicate their response.

General Navigation and Interaction with the TestNav Interface

Figure 3 below displays a sample item and shows the elements that make up the TestNav interface. The very top bar (light blue) displays the examinee's name and the title of the test and also contains a button that allows exit from the test. Below that is the toolbar (darker blue) that contains all of the general tools for the test (tools specific to the constructed response text editor are embedded within the text editor box itself). Below that is a yellow banner that shows directions specific to the item. The white region that makes up most of the user interface contains the item itself. If the item were to extend beyond the confines of the screen, it would have scroll bars to allow access to the off-screen regions. Finally, the blue bar at the bottom of the screen contains navigation-related controls, including the "Next" and "Previous" buttons, an array of buttons that allow the student to jump to any question in the test, the button that sends the student to the "Section Review" screen, and information about where the student currently is within the test.

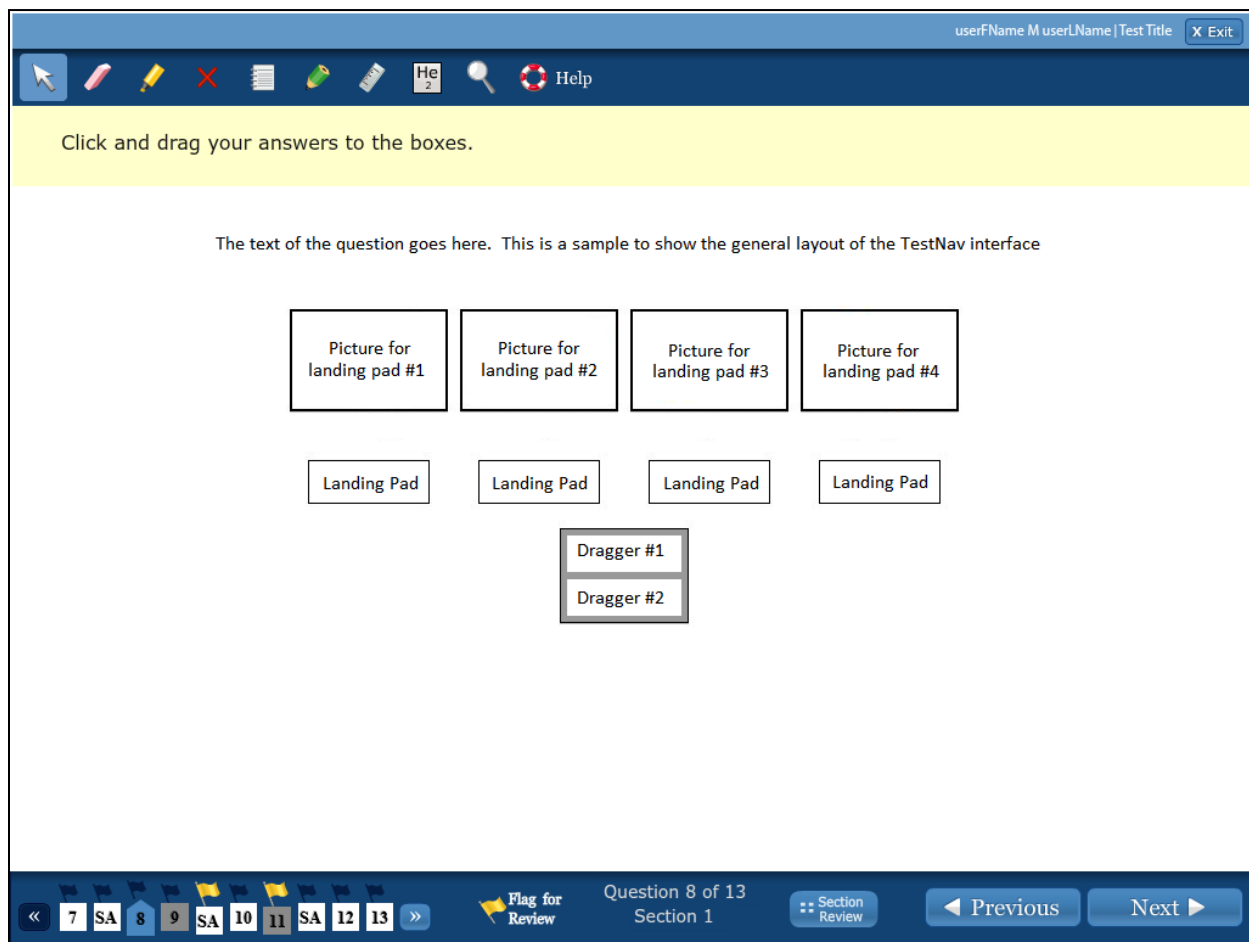


Figure 3. Sample item within the TestNav user interface.

In general, most students had little difficulty with basic navigation through the test. Moving from item to item was relatively straightforward—“Next” and “Previous” buttons are prominent within the TestNav interface, and are a conventional means of navigating through paginated content on the Internet. Most took little or no notice of the elements of the interface beyond the item itself and the “Previous” and “Next” buttons, especially in the first lab, where the facilitator did not access the tools in the course of demonstrating the think aloud protocol to the student.

One specific task that students were presented with when taking the test was meant to assess their ability to move from the current item to another non-consecutive item without having to move item by item using either the “Next” or “Previous” buttons and also to determine quickly whether any items remain unanswered. There are two ways to accomplish these within the TestNav interface. The first is to use the set of boxes in the lower-left corner of the screen that depict the states of ten items in the vicinity of the current item, Question 8 (see Figure 3). White boxes indicate unanswered items, grey boxes indicate answered items, and flags indicate items that the student has flagged for review. Clicking on any box takes the student to the corresponding item and the boxes at either end of the row scroll the boxes through the entire range of items in the current section.

The second way a student can navigate to an item in the test is to click on the “Section Review” button at the middle-right region of the lower tool bar. This brings up a dialog, shown in Figure 4 below.

Return to Test

Section 1 Review

Choose an item below or click *TEST OVERVIEW* to go to the Test Overview, or click *SUBMIT* to submit your test.

Test Overview

All Items	2 Flagged for Review	6 Answered	7 Unanswered
Question 1			! Unanswered
Question 2		✓ Answered	
Question 3		✓ Answered	
Question 4			! Unanswered
Question 5			! Unanswered
Question 6		✓ Answered	
Question 7			! Unanswered
Simulation Activity			
Question 8		✓ Answered	
Question 9		✓ Answered	
Simulation Activity	🚩 Flagged for Review		

SUBMIT

AND EXIT TEST

userFName M userLName | Test Title

Figure 4. Sample section review dialog within the TestNav user interface.

Each question displays the flag status and response state (answered or unanswered) of each item. In the middle of the lab session, the student was stopped and the facilitator asked, “Suppose that you wanted to find out how many questions you had answered and how many more remained. How do you think that you might find that information?”

If the student responded by asking if he should take a particular action, the facilitator responded, “Try doing that and see what happens.” If after a few minutes the student was unable to identify either of the two methods (either by using the boxes or the “Section Review” button), or if the student was obviously frustrated, the facilitator would then ask the student whether he thought that the “Section Review” button would take him to where he might be able to review the status of the items on the test, and would then encourage the student to try it.

Once the student had identified one of the two approaches, or had been guided to the section review dialog by the facilitator, he or she generally had little trouble moving around the test, understanding how answered and unanswered items were represented, and what purposes flags might be used for. One fourth grade student initially thought that the green checkmarks represented items that he had answered correctly, but quickly realized that they only indicated that the items had been answered after carefully reviewing the dialog.

Summary

In summary, virtually all students within the United States today are very familiar with technology, using it at home and at school. It is ubiquitous in their lives, and they generally can extend their current experience to new technology and user interfaces. All of the students sampled in these labs were completely unfamiliar with the TestNav interface, although some had had experience taking other online tests prior to the labs. All were able to successfully access and answer at least a portion of the items, and in general, problems that they did experience were due not to the interface or to the technology, but were instead due to other factors, including LEP students’ difficulty in understanding the language, and other students being confused by the wording of the questions. Students who did experience initial difficulties with particular aspects of the technology-enhanced items were usually able to resolve them in a short amount of time without help from the facilitator.

The low frequency of usability issues in a sample that received only a minimal amount of training prior to their taking the test bodes well for operational use of the TestNav interface starting in the spring of 2014. Recommendations from this study are limited to allowing students to become familiar with the TestNav interface through exposure to practice tests prior to the administration of the test. While the test is being administered, proctors trained in the capabilities of the TestNav system should be available to assist students should they encounter difficulties during the course of testing. The students in the sample showed a high degree of facility in responding to the questions, and only a small amount of supplemental training is needed to acquaint them with the tools and navigation of the TestNav interface. The surveys given to the students after completion of the test included a question that asked them to indicate whether they preferred paper or computer-based tests. All except two students indicated that they preferred the computer-based version, and many commented that it had been an enjoyable experience.