Smarter Balanced in Connecticut: A Progress Report

INTRODUCTION

The Connecticut State Board of Education adopted the Common Core State Standards in 2010. These standards, referred to in our state as the Connecticut Core Standards, are designed to promote equity by ensuring all students, no matter where they live, are well prepared with the skills and knowledge necessary to collaborate and compete with their peers in the United States and abroad.

Also in 2010, Connecticut joined the Smarter Balanced Assessment Consortium and partnered with other states to develop the Smarter Balanced Assessment System. As a governing state in the consortium, Connecticut educators have been actively involved in many aspects of this new assessment system, including item development, standard setting, report development, and creation of professional learning resources for teachers.

In Spring 2014, educators in nearly 90 percent of Connecticut's school districts successfully implemented a large-scale field test of new assessment items and the accompanying technology. Over the past year, thousands of Connecticut educators have also participated in important professional learning activities. These efforts have positioned Connecticut well for the first operational administration of the Smarter Balanced assessments occurring between March and June 2015.

NEW STANDARDS. NEW TESTS. NEW RESULTS.

In November 2014, Connecticut educators, including CSDE staff, participated in setting initial achievement levels for the Smarter Balanced assessments in mathematics and ELA/literacy. The achievement level setting process required a careful review of assessment items and field test data to identify threshold scores for each of the levels. The achievement levels approved by Connecticut and the other governing states in the consortium can inform and support discussions about student performance.

The achievement-level setting process yielded initial estimates of the percentages of students across the consortium states achieving at each of the four levels in mathematics and ELA/literacy, based on consortium-level field test data. This information cannot be generated from the field test separately for Connecticut students, as the sampling plan for achievement-level setting was developed to be representative of the consortium as a whole, and is not representative of each participating state. However, there are other data sources that can provide reasonable estimates for Connecticut.

The Common Core State Standards and the Smarter Balanced assessments were informed by the National Assessment of Educational Progress (NAEP). Known as the "nation's report card," NAEP is the largest nationally representative and continuing assessment of students' knowledge in reading and math on a
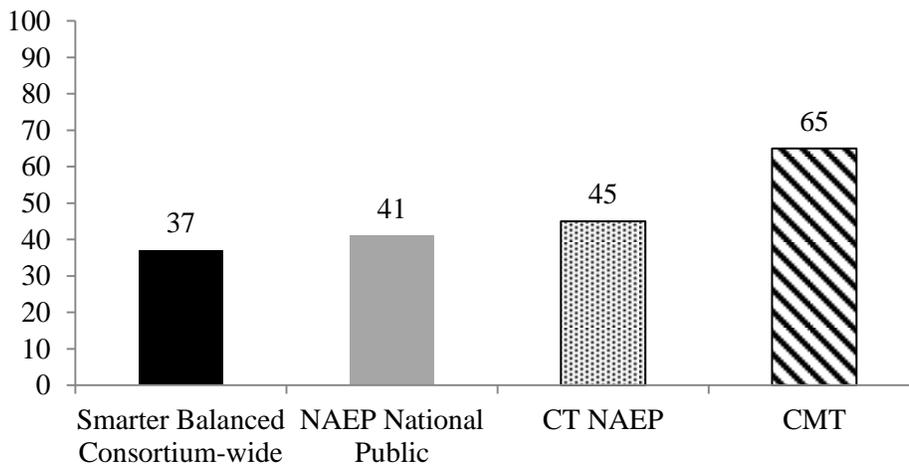
state-by-state basis. Proficient performance (level 3 of 4) on the NAEP is indicative of "competency over challenging subject matter."

Figures 1 through 4 present the best available data in Grades 4 and 8 in mathematics and ELA/literacy across several assessments[1]. In addition to the percentage of students across the consortium states performing at or above level 3 on the Smarter Balanced field test, the figures also include:

- the percentage of students in national public schools scoring at or above NAEP "proficient" in 2012-13;
- the percentage of Connecticut public school students scoring at or above NAEP "proficient" in 2012-13; and
- the percentage of students scoring at or above the CMT Goal level in 2012-13.

The data reveal similarities between Smarter Balanced performance at or above level 3 and NAEP performance at or above proficient. This is not surprising because Smarter Balanced was informed by the NAEP; NAEP items were included in the Smarter Balanced field test; and student responses to the NAEP items were viewed alongside performance on field test items as part of achievement level setting.

Figure 1. Grade 4 mathematics.

Smarter Balanced Consortium-wide = percentage of students performing at Smarter Balanced Level 3 and above in 2013-14
NAEP National Public = percentage of students nationwide performing at or above the NAEP "proficient" level in 2012-13
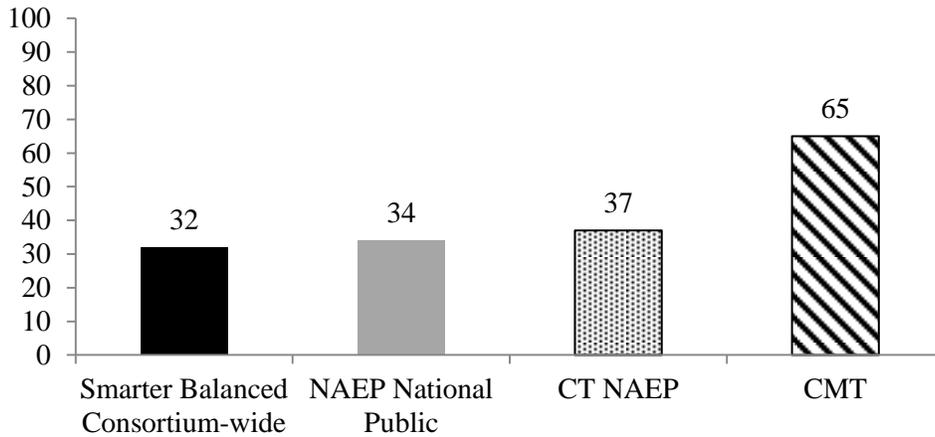*CT NAEP = percentage of students in Connecticut performing at or above the NAEP "proficient" level in 2012-13*
CMT = percentage of students performing at or above CMT Goal in 2012-13

This graph illustrates the similarities between- Smarter Balanced performance at or above level 3 and NAEP performance at or above proficient in Grade 4 mathematics.

---

[1] The comparative information provided here is purely informational; it would be inappropriate to draw definitive conclusions from these data. Moreover, given that these assessments were developed separately, are based on different content standards, are administered differently, and have dramatically different stakes for students and schools, strict point-by-point comparisons should not be made.
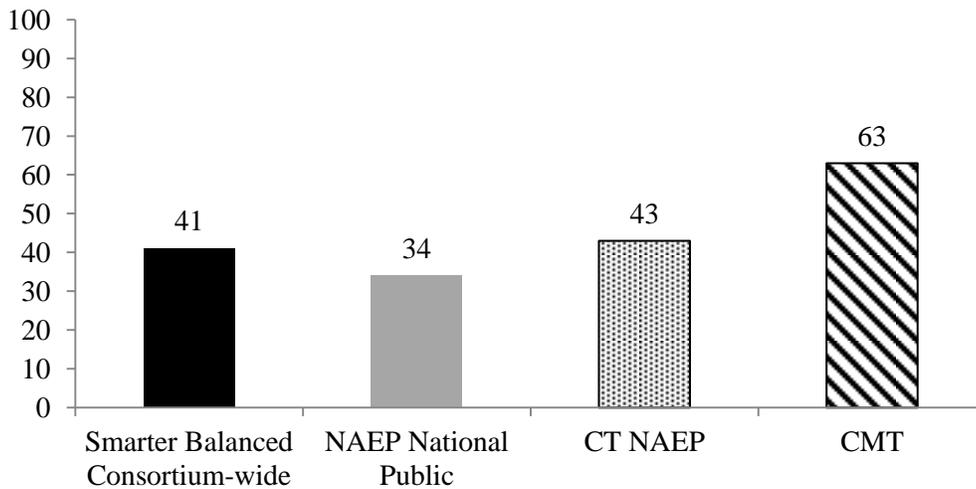
Figure 2. Grade 8 mathematics.

Smarter Balanced Consortium-wide = percentage of students performing at Smarter Balanced Level 3 and above in 2013-14
NAEP National Public = percentage of students nationwide performing at or above the NAEP "proficient" level in 2012-13
CT NAEP = percentage of students in Connecticut performing at or above the NAEP "proficient" level in 2012-13
CMT = percentage of students performing at or above CMT Goal in 2012-13

This graph illustrates the similarities between Smarter Balanced performance at or above level 3 and NAEP performance at or above proficient in Grade 8 mathematics.
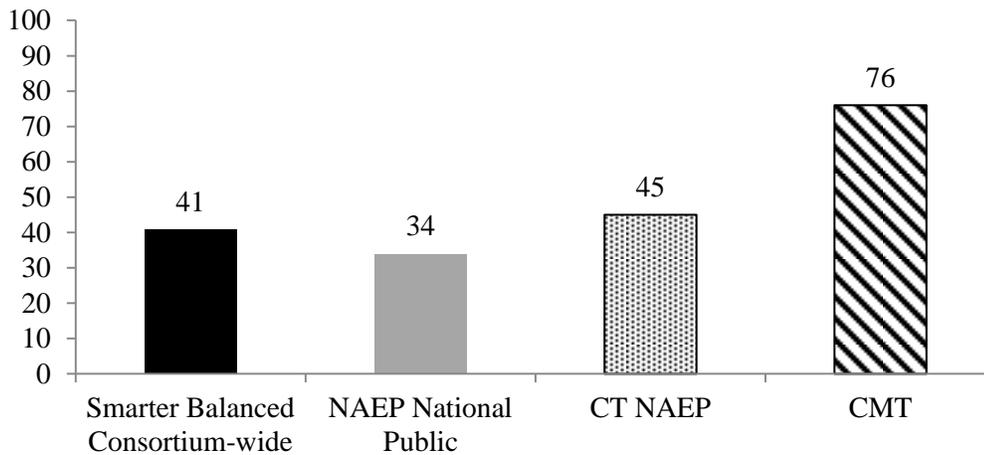
Figure 3. Grade 4 reading.

Smarter Balanced Consortium-wide = percentage of students performing at Smarter Balanced Level 3 and above in 2013-14
NAEP National Public = percentage of students nationwide performing at or above the NAEP "proficient" level in 2012-13
CT NAEP = percentage of students in Connecticut performing at or above the NAEP "proficient" level in 2012-13
CMT = percentage of students performing at or above CMT Goal in 2012-13

This graph illustrates the similarities between Smarter Balanced performance at or above level 3 and NAEP performance at or above proficient in Grade 4 reading.

Figure 4. Grade 8 reading.

Smarter Balanced Consortium-wide = percentage of students performing at Smarter Balanced Level 3 and above in 2013-14
NAEP National Public = percentage of students nationwide performing at or above the NAEP "proficient" level in 2012-13
CT NAEP = percentage of students in Connecticut performing at or above the NAEP "proficient" level in 2012-13
CMT = percentage of students performing at or above CMT Goal in 2012-13

This graph illustrates the similarities between Smarter Balanced performance at or above level 3 and NAEP performance at or above proficient in Grade 8 reading.

LOOKING FORWARD

As we administer the Smarter Balanced operational assessment, it is critical to recognize that Connecticut's ELA/literacy and mathematics achievement results in 2014-15 will represent a new baseline for performance in those subject areas. The reason is that the results will be based on the first administration of a completely new state assessment that is aligned to new, more rigorous standards. The data presented in this report should provide all stakeholders with some perspective as the state prepares to administer the new assessment and receive the results. The achievement results for 2014-15 will reflect a change in expectations for students, not a change in their abilities.

Beyond 2014-15, in addition to measuring achievement annually, it will be equally if not more important to measure growth in student achievement from this new baseline. Together, measures of annual achievement and growth will more fully reflect students' progress toward the fulfillment of the promise of college and career readiness. These measures will also acknowledge the efforts of Connecticut educators who have been working tirelessly over the past few years to implement the Connecticut Core Standards.

Lastly, it is important to underscore that the achievement levels should be used to inform meaningful conversations regarding the performance of students and groups of students. Relying solely on the achievement levels alone to characterize student performance would be an oversimplification. Therefore, in the years ahead, the CSDE will increasingly report achievement and growth results using more continuous measures of performance (e.g., vertical scale scores).

A TEST OF THE TEST

In January 2014, the Connecticut State Department of Education (CSDE) received approval from the United States Department of Education (USED) to allow districts the flexibility to choose to administer the Smarter Balanced mathematics and English language arts (ELA) field tests in 2013-14 in lieu of the Connecticut Mastery Test (CMT) and the Connecticut Academic Performance Test (CAPT) – CSDE's legacy assessments. The CSDE sought this flexibility to:

- support districts as they continued to implement the Connecticut Core Standards;
- avoid "double-testing" students so that teachers can (a) reclaim instructional time that would have otherwise been lost to redundant standardized testing and (b) focus on teaching the new standards;
- provide districts the opportunity to test their hardware and network capacity for the new computer-based state assessment in 2014-15 (i.e., Smarter Balanced) in a low-stakes environment; and
- enable students to experience current assessment content, get familiarized with the new computer-based testing system, and meaningfully inform item selection for the 2014-15 state assessment.

Nearly 90 percent of districts participated in the field test. In these districts, more than 250,000 students were tested. The CSDE received considerable feedback from schools indicating that they benefited from the experience and believe they are well prepared for the upcoming 2014-15 administration.

To provide high-level summary information related to the large-scale field test administration, the CSDE analyzed Connecticut's item-level student responses for machine-scored items and provides the following information.

CONTENT CLAIMS: "BIG PICTURE" LEARNING OUTCOMES

The Smarter Balanced operational assessments will cover the full range of college- and career-ready knowledge and skills contained in the Connecticut Core Standards. To do this, each test item is associated with assessment targets and overall content claims. Content claims such as Communicating Reasoning (mathematics claim 3) or Research/Inquiry (ELA/literacy claim 3) are the broad statements of the learning outcomes that the assessment system is measuring. Claims describe what students should be able to do to be college and career ready in mathematics and ELA.

Tables 1 and 2 below provide the mathematics and ELA/literacy claims for reference. Please note that when 2014-15 mathematics results are reported, claims 2 and 4, Problem Solving and Modeling and Data Analysis respectively, will be combined.

Table 1. Claims for the Mathematics Summative Assessment.

Mathematics Claim	Brief Explanation of Mathematics Claim
Claim 1— Concepts and Procedures	"Students can explain and apply mathematical concepts and interpret and carry out mathematical procedures with precision and fluency."
Claim 2— Problem Solving	"Students can solve a range of complex well-posed problems in pure and applied mathematics, making productive use of knowledge and problem solving strategies."
Claim 3— Communicating Reasoning	"Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others."
Claim 4— Modeling and Data Analysis	"Students can analyze complex, real-world scenarios and can construct and use mathematical models to interpret and solve problems."

Table 2. Claims for the English Language Arts/Literacy Summative Assessment.

ELA/Literacy Claim	Brief Explanation of ELA/Literacy Claim
Claim 1— Reading	"Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts."
Claim 2— Writing	"Students can produce effective and well-grounded writing for a range of purposes and audiences."
Claim 3— Speaking* and Listening	"Students can employ effective speaking and listening skills for a range of purposes and audiences."
Claim 4— Research/Inquiry	"Students can engage in research and inquiry to investigate topics, and to analyze, integrate, and present information."

* At this time, speaking is not assessed on the Smarter Balanced summative assessment.

CSDE staff analyzed student responses to more than 4,200 machine-scored mathematics items across all seven tested grades. The majority of these items (73 percent) assessed Concepts and Procedures (claim 1), 15 percent assessed Communicating Reasoning (claim 3), and 12 percent assessed the combined claims 2 and 4 designed to measure Problem Solving and Modeling and Data Analysis.

Overall, in most grades, mathematics items measuring Concepts and Procedures (claim 1) were more likely to be answered correctly while items measuring Problem Solving and Modeling and Data Analysis (claims 2 and 4) were less likely to be answered correctly. This finding is consistent with student performance on the Connecticut Mastery Test and other assessments where students were more successful with items that required them to apply standard procedures and less successful when the items required complex thinking and application of knowledge.

The field-test data analyses also included a review of responses to approximately 3,800 machine-scored ELA/literacy items associated with the four ELA/literacy claims. Thirty-six percent of the items assessed Reading (claim 1), 26 percent assessed Writing (claim 2), 23 percent assessed Listening (claim 3), and 15 percent assessed Research/Inquiry (claim 4). The items aligned to Writing (claim 2) were more likely to

be answered correctly by students in the lower grades. In the higher grades, items aligned to Research/Inquiry (claim 4) were less likely to be answered correctly.

DEPTH OF KNOWLEDGE LEVELS: A MEASURE OF RIGOR AND COMPLEXITY

Every assessment item developed for the Smarter Balanced assessments is tagged during the item writing and review process with a Depth of Knowledge (DOK) level. The DOK levels range from DOK Level 1 to DOK Level 4. The DOK levels reflect the cognitive rigor associated with each item. Many educators may be familiar with this type of rating system having studied and applied Bloom's Taxonomy of Educational Objectives in their classrooms. Smarter Balanced items classified at higher DOK levels are considered more complex. Different DOK systems are used in each Smarter Balanced content area. The systems are included for reference in the appendixes to this report.

In mathematics, the analysis of field test items by DOK level was limited to items measuring Recall and Reproduction (DOK Level 1), Basic Skills and Concepts (DOK Level 2), and Strategic Thinking and Reasoning (DOK Level 3). Since the analysis was conducted using machine-scored items exclusively, it was not possible to evaluate the performance of items written to measure Extended Thinking (DOK Level 4) since that level of complexity was only measured through constructed-response items.

Overall, in most grades, mathematics items at DOK Level 1 were more likely to be answered correctly than items at DOK Level 2; items at DOK Level 3 were least likely to be answered correctly. As items increase in complexity, they appear to be more challenging to students.

As with mathematics items, ELA/literacy items written at the lowest DOK level requiring students to recite basic facts or demonstrate basic understanding were most likely to be answered correctly by Connecticut students, with DOK Level 2 items requiring some mental processing beyond simple recall following close behind. The data show that items written at DOK Level 3 and DOK Level 4 were more challenging for students. However, as the item pool transitioned from elementary grades to middle school grades, there was less dramatic differentiation in the average percentage of correct responses across the DOK levels. It is difficult to make strong statements about these patterns, but the data may be indicating that as students advance from elementary school to middle school, responding to items of increasing complexity becomes less difficult.

MOVING STATE ASSESSMENTS FORWARD THROUGH TECHNOLOGY

The Smarter Balanced assessments are computer-based assessments. The use of technology for the administration of the assessment allows for assessing content and skills using a variety of item types. While CMT and CAPT items were limited to multiple-choice (i.e., selected response) and constructed response items, Smarter Balanced incorporates additional types of items (e.g. multi-selection, matching item, equation response). For an explanation of the Smarter Balanced item types and to view a tutorial specific to each type, please visit http://ct.portal.airast.org/item-type-tutorials/.

Based on a review of the field test data, multi-select items (items with more than one correct response) consistently presented the greatest challenge to students, while multiple-choice items, a format that is quite familiar to students, received on average the greatest percentage of correct responses. As educators continue using the new standards and delivering instruction that demands greater cognitive complexity, students will increase their familiarity of the varied item types and their ability to demonstrate what they know and can do.

Besides innovative test question formats, a computer-delivered assessment allows for customization at the student level. The operational Smarter Balanced assessments will have a computer-adaptive component as well as a performance task component. When a student begins to respond to items in the computer-adaptive component, the test delivery system will adjust to the student's skill level by changing the difficulty of the questions presented. This approach avoids providing students with many items that are too easy or items that are too difficult causing unnecessary frustration. This is a major shift in test design that will mean a more efficient test-taking experience for the student and more precise information about achievement for educators, students, and families.

Appendix A
Cognitive Rigor Matrix/Depth of Knowledge (DOK) Levels for Mathematics

The depth(s) of knowledge (DOK) that students need to bring to the Smarter Balanced Mathematics items/tasks have been identified using the Cognitive Rigor Matrix developed by Hess, Carlock, Jones, and Walkup ("snapshot" shown below). This matrix draws from two widely accepted measures to describe cognitive rigor: Bloom's (revised) Taxonomy of Educational Objectives and Webb's Depth-of-Knowledge Levels.

Depth of Thinking (Webb) + Type of Thinking (Revised Bloom)	DOK Level 1 Recall and Reproduction	DOK Level 2 Basic Skills and Concepts	DOK Level 3 Strategic Thinking and Reasoning	DOK Level 4 Extended Thinking
Remember	- Recall conversions, terms, facts			
Understand	-Evaluate an expression -Locate points on a grid or number on number line -Solve a one-step problem -Represent math relationships in words, pictures, or symbols	- Specify, explain relationships -Make basic inferences or logical predictions from data/observations -Use models/diagrams to explain concepts -Make and explain estimate	-Use concepts to solve non-routine problems -Use supporting evidence to justify conjectures, generalize, or connect ideas -Explain reasoning when more than one response is possible -Explain phenomena in terms of concepts	-Relate mathematical concepts to other content areas, other domains -Develop generalizations of the results obtained and the strategies used and apply them to new problem situations
Apply	-Follow simple procedures -Calculate, measure, apply a rule (e.g., rounding) -Apply algorithm or formula -Solve linear equations -Make conversions	-Select a procedure and perform it -Solve routine problem applying multiple concepts or decision points -Retrieve information to solve a problem -Translate between representations	-Design investigation for a specific purpose or research question - Use reasoning, planning, and supporting evidence -Translate between problem and symbolic notation when not a direct translation	-Initiate, design, and conduct a project that specifies a problem, identifies solution paths, solves the problem, and reports results
Analyze	-Retrieve information from a table or graph to answer a question -Identify a pattern/trend	-Categorize data, figures -Organize, order data -Select appropriate graph and organize and display data -Interpret data from a simple graph -Extend a pattern	-Compare information within or across data sets or texts -Analyze and draw conclusions from data, citing evidence -Generalize a pattern -Interpret data from complex graph	-Analyze multiple sources of evidence or data sets
Evaluate			-Cite evidence and develop a logical argument -Compare/contrast solution methods -Verify reasonableness	-Apply understanding in a novel way, provide argument or justification for the new application
Create	- Brainstorm ideas, concepts, problems, or perspectives related to a topic or concept	-Generate conjectures or hypotheses based on observations or prior knowledge and experience	-Develop an alternative solution -Synthesize information within one data set	-Synthesize information across multiple sources or data sets -Design a model to inform and solve a practical or abstract situation

Appendix B

Depth of Knowledge Levels for English Language Arts/Literacy
(Adapted from Appendix C of the Webb Depth of Knowledge Levels)*

DOK Level	Brief Explanation of DOK Level for Reading
Reading Level 1	Requires students to receive or recite facts or to use simple skills or abilities to demonstrate basic comprehension of a text. Items require a shallow understanding of text and often consist of verbatim recall from text, slight paraphrasing of specific details, or simple understanding.
Reading Level 2	Requires the engagement of some mental processing beyond recalling or reproducing a response; it requires both comprehension and subsequent processing of text or portions of text to demonstrate a closer understanding of the text. Inter-sentence analysis of information is required.
Reading Level 3	Requires students to show a deeper knowledge and understanding of the more abstract ideas in a text (or between texts) by explaining, generalizing, or connecting ideas, identifying themes, making inferences, and analyzing connections. Students must support their reasoning with details or examples from the text(s).
Reading Level 4	Requires higher-order thinking in which students examine information from at least one passage (often dual texts) to develop hypotheses or perform complex analyses of the connections, perspectives, and themes across texts.

DOK Level	Brief Explanation of DOK Level for Writing
Writing Level 1	Requires students to write or recite simple facts to demonstrate a basic understanding (i.e., to list basic ideas or words, engage in simple spelling, or write simple sentences). Students are expected to use appropriate grammar, punctuation, capitalization, and spelling.
Writing Level 2	Requires some mental processing in which students are engaged in developing first-draft writing for a number of purposes and audiences. Students begin to connect ideas using simple organizational structure.
Writing Level 3	Requires some higher-level mental processing in which students develop multi-paragraph compositions that may demonstrate the use of complex sentence structure, synthesis, and analysis and, when appropriate, should support their ideas with details or examples from one or more sources. Students are engaged in editing and revising to improve the quality of the composition.
Writing Level 4	Requires higher-level thinking in which students develop a multi-paragraph composition that demonstrates the ability to synthesize and analyze complex ideas or themes supported by evidence from a sources or sources. Students must demonstrate an awareness of audience, purpose, and voice.

*The above descriptions of Webb's depth of knowledge (DOK) levels are excerpted from the *Web Alignment Tool (WAT) Training Manual, Draft Version 1.1*. (Webb, 2005, pp.45-46 and 70-75) and do not include the full breadth of the DOK descriptions for reading and writing. Please refer to the manual for a complete description of the Depth of Knowledge Levels.