

Rightsizing the Smarter Balanced English Language Arts (ELA) Assessment for Connecticut

March 2016



Connecticut State Department of Education

Contents

Introduction 3

The Purpose of the State Summative Assessment 4

Structure and Content 4

Reliability..... 7

Time 9

From Standards to Curriculum and Instruction 10

Conclusion..... 11

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 that all students 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 that is aligned to these standards.

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. In spring 2015, all Connecticut districts successfully administered the Smarter Balanced operational assessment. The CSDE released results from the spring 2015 assessment in August 2015.

Subsequently, on October 7, 2015, the Connecticut State Board of Education adopted the SAT as the statewide assessment in lieu of the Smarter Balanced Assessment for Grade 11 public school students. This change was made to minimize duplicative testing for Grade 11 students while ensuring that the assessment aligns with the state standards and provides a reliable measure of student achievement. This change was initiated by Governor Malloy, recommended by education advocates, supported by the Connecticut General Assembly, and approved by the U.S. Department of Education as part of Connecticut's ESEA Flexibility Request.

While this change was occurring for students in Grade 11, the CSDE simultaneously began taking a closer look at the Smarter Balanced assessment in other grades to see if the assessment burden could be reduced while maintaining high reliability and continued alignment to state academic standards. The CSDE had encouraged and supported districts to take a critical look at the array of assessments being used locally to ensure that redundant testing was reduced or eliminated; it was time for the CSDE to do the same. On February 25, 2016, Governor Malloy and Commissioner Wentzell announced that the Smarter Balanced English Language Arts (ELA) assessment would be trimmed "to be smart about testing, limit anxiety, and boost learning time." This report provides a detailed explanation of the many analyses and considerations that guided the CSDE's recommendation and ultimate decision.

The Purpose of the State Summative Assessment

An assessment must be conducted for a clear and stated purpose. The assessment instrument that is used must provide the information necessary to support the decisions for that purpose.

The state summative academic assessment is an important component for ensuring that we – the state, districts, and schools – are fulfilling on the promise of a high quality education for all students that prepares them for college, careers, and life. This summative assessment is akin to an annual physical. It can inform district and school accountability. It can help local leaders to evaluate the broad effects of curriculum and instructional interventions. It can also measure student growth over time. However, it is not meant as a diagnostic measure to directly inform a teacher’s classroom instruction on a daily/weekly basis. Such determinations are left to those who are better attuned to the academic needs of our students i.e., our teachers and our instructional leaders. Consequently, the state summative assessment should remain focused on those elements that provide the essential information to fulfill the purposes outlined above without unduly burdening our teachers, students and families.

Connecticut’s legacy assessments in ELA and Mathematics (the Connecticut Mastery Test – CMT – or the Connecticut Academic Performance Test – CAPT) were also used for the purposes stated above. However, the past hyper-focus on test scores and the detailed reporting practices created a climate that led to an unfortunate practice of heavy test preparation.

With the adoption of the Connecticut Core Standards and the aligned Smarter Balanced assessments, Connecticut has been strongly encouraging teaching to the standards and not to a test. The CSDE firmly believes that the art and science of teaching and learning should be guided by the creative application of the rigorous standards. We are confident that the teachers and local leaders will make the right curricular and instructional decisions that will best meet the academic needs of our students. While the state summative assessment is an important lever for positive change, it is not the only one.

Structure and Content

The Smarter Balanced ELA assessment in each grade is comprised of a computer-adaptive test (CAT) and an extended reading/writing assignment which is termed as a performance task (PT). This extended reading/writing assignment is also administered on the computer; however, unlike the computer-

adaptive section of the assessment, this extended reading/writing assignment is a fixed form test i.e., the test questions do not *adapt* to the students' responses.

To evaluate the content of the overall ELA assessment, the CSDE reviewed the [Smarter Balanced assessment blueprint](#). A summary of the ELA blueprint for grades 3-5 and grades 6-8 is presented below.

Smarter Balanced Assessment Blueprint – ELA/Literacy Grades 3 - 5						
Claim	Content Category	Stimuli		Items		Total Items by Claim
		CAT	PT	CAT	PT	
1. Reading	Literary	2	0	7-8	0	14-16
	Informational	2	0	7-8		
2. Writing	Organization/Purpose	0	1a	5	1	11
	Evidence/Elaboration	0		5		
	Conventions	0		5		
3. Speaking/Listening	Listening	3-4	0	8-9	0	8-9
4. Research/Inquiry	Research	0	1b	6	2-3	8-9
Smarter Balanced Assessment Blueprint – ELA/Literacy Grades 6 – 8						
1. Reading	Literary	1-2	0	4-7	0	13-17
	Informational	2-3	0	9-10		
2. Writing	Organization/Purpose	0	1a	5	1	11
	Evidence/Elaboration	0		5		
	Conventions	0		5		
3. Speaking/Listening	Listening	3-4	0	8-9	0	8-9
4. Research/Inquiry	Research	0	1b	6	2-3	8-9

The CAT contains the vast majority of the test items for both grade ranges. It incorporates not only traditional multiple choice items but also multi-select (i.e., multiple answer options must be selected to get the full score for that item), technology enhanced (e.g., highlight, click and drag, matching), and constructed response items (i.e., short-text paragraph writing). There are two to three short-text items in the CAT. The PT includes two or three items in Claim 4, one or two of which are short text items while the other is a selected response item. Each PT also has one full write (i.e., essay) in Claim 2 that is scored across three traits: Organization/Purpose, Evidence/Elaboration, and Conventions.

The test uses “assessment targets” to align test items to one or more Connecticut Core Standards. The assessment targets for each claim are listed below:

Claim 1: Reading	
Literary Texts	Informational Texts
1 - Key Details 2 - Central Ideas 3 - Word Meaning 4 - Reasoning and Evidence 5 - Analysis Within and Across Texts 6 - Text Structures and Features 7 - Language Use	8 - Key Details 9 - Central Ideas 10 - Word Meaning 11 - Reasoning and Evidence 12 - Analysis Within and Across Texts 13 - Text Structures and Features 14 - Language Use
Claim 2: Writing	Claim 3: Speaking and Listening
1 - Write Brief Texts* 2 - Revise Brief Texts* 3 - Compose Full Texts* 4 - Use Text Features 5 - Language and Vocabulary Use 6 - Edit 7 - Technology <i>* items may be either narrative or information/explanatory or opinion/argument</i>	Listening 4— Listen/Interpret
Claim 4: Research/Inquiry	
Grades 3-5	Grades 6-8
1— Plan/Research** 2— Interpret and Integrate Information 3— Analyze Information/Sources 4— Use Evidence <i>** This assessment target is not assessed in the summative assessment. The core standard expects students to conduct short research projects in order to: (i) answer multi-step questions; (ii) present an opinion; or (iii) investigate different aspects (subtopics) of a broader topic or concepts using multiple sources.</i>	1— Plan/Research*** 2— Analyze/Integrate Information 3— Evaluate Information/Sources 4— Use Evidence <i>*** This assessment target is not assessed in the summative assessment. The core standard expects students to conduct short research projects in order to: (i) explore a topic, issue, or problem; and (ii) analyze interrelationships among concepts or perspectives.</i>

With the exception of the essay writing component (#3 “compose full texts” under the Writing claim), all other assessment targets in all claims are assessed in the CAT. Also, all targets in the Reading and Listening claims are assessed *only* in the CAT.

Reliability

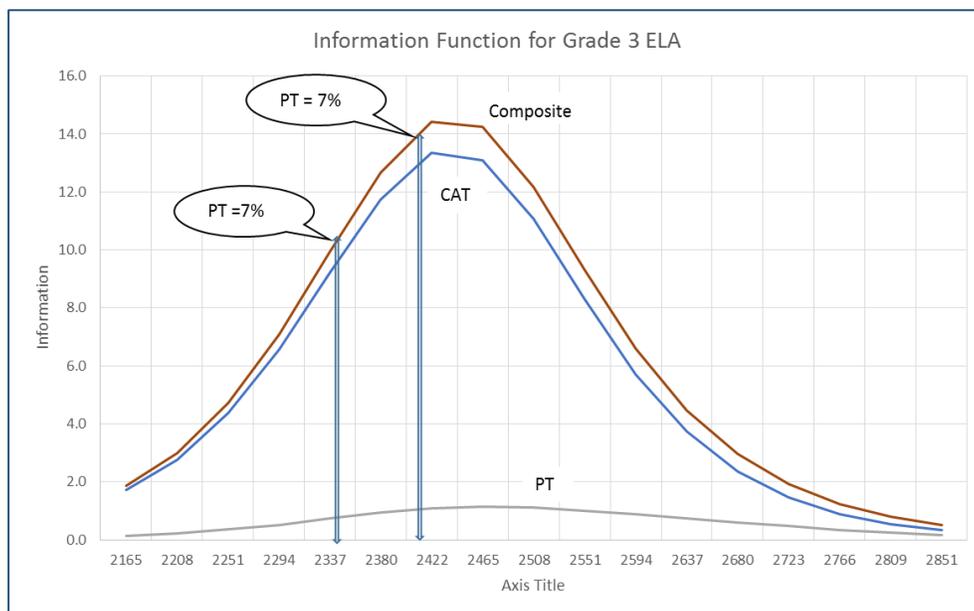
Reliability refers to the precision, stability, and consistency of test scores. The reliability of the Smarter Balanced ELA assessment for Connecticut students was ascertained using two approaches:

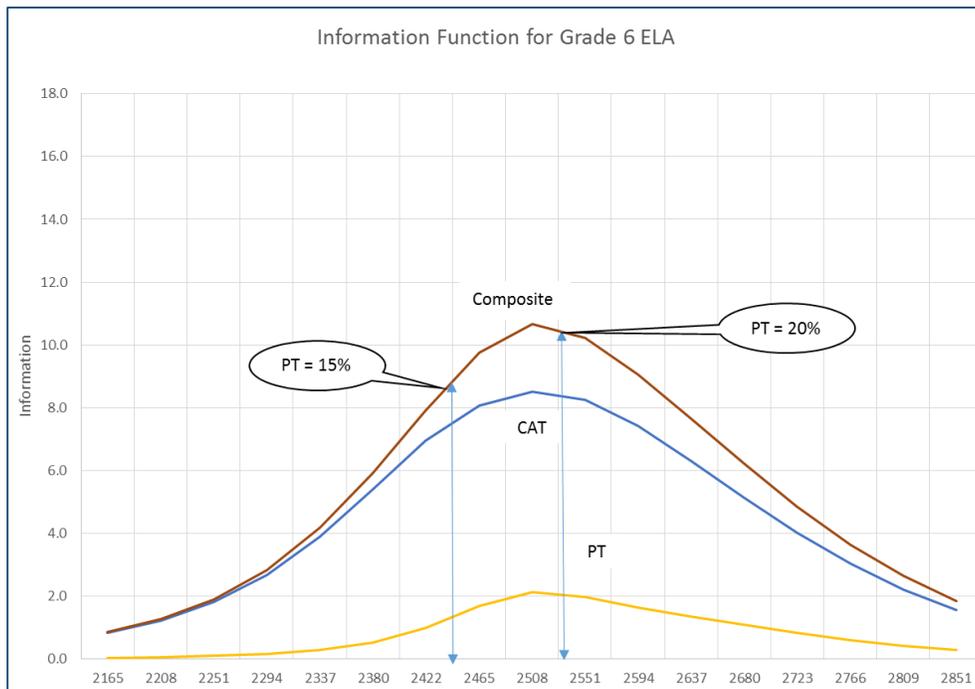
- Test Information; and
- Marginal reliability.

“Test information” is a psychometric statistic that quantifies the amount of information that can be derived from a test. It is the sum of the “item information” from all items on a test. The amount of information an item will yield is dependent on the *precision* with which it estimates the ability of a wide range of students.

To determine the “test information” for an adaptive test where different students are exposed to different test items, the CSDE constructed two sample fixed-length assessments, one for ELA Grade 3 and another for ELA Grade 6. These tests were constructed using items from the item pool that were administered to a large number of Connecticut students in the spring of 2015.

Once the test items were assembled, information functions were constructed for: (i) all items; (ii) CAT items only; and (iii) PT items only. They are presented below.





The CAT items (blue curve) provide much more information than the PT items (yellow curve) toward the total composite score (red curve). The additional percentages of information from the PT at the achievement level cut scores are listed. Overall, 80 percent to 93 percent of test information is derived from the CAT.

Another statistic for expressing the reliability of a computer adaptive assessment that is comparable to reliability statistics generated in classical test theory is called *marginal reliability*. In item response theory (IRT – the psychometric theory behind the Smarter Balanced assessment as well as CMT and CAPT) unlike in classical test theory, the measurement error varies across the spectrum of student ability. Therefore, to generate a single test reliability statistic, these measurement errors are averaged. The resulting statistic ranges from 0 to 1 with higher values representing greater test reliability.

The marginal reliabilities for the CAT only test as compared to the overall test (CAT + PT) are presented below. The CAT only test remains a highly reliable assessment.

Grade	CAT Only	Overall (CAT + PT)
3	0.905	0.914

6	0.876	0.896
---	-------	-------

Time

Data from the 2014-15 assessment reveal that on average, Connecticut students spent more time on the ELA PT than on the ELA CAT (see table below). These times are based on the actual hours and minutes from login to logout in the online testing system.

GRADE	ELA CAT	ELA PT	Total
3	1:33	1:46	3:19
4	1:35	1:48	3:23
5	1:30	1:47	3:17
6	1:33	1:38	3:11
7	1:21	1:30	2:51
8	1:19	1:26	2:45

By contrast, in Mathematics, the PT takes less time on average than the CAT. As a result, the total *estimated* average testing times based on the *actual* average testing times from the 2014-15 operational assessment without the ELA PT are presented below. Overall, students on average should complete both the ELA and Mathematics state summative assessment in around 3.0 to 3.5 hours. These are substantially lower than Smarter Balanced estimates with the ELA PT (6.0 to 6.5 hours) or CMT actual testing times (5.25 to 6.5 hours).

Grade	ELA (CAT Only)	Mathematics (CAT and PT)	Total
3	1:33	1:49	3:22
4	1:35	1:45	3:20
5	1:30	2:07	3:37
6	1:33	1:55	3:28
7	1:21	1:37	2:58
8	1:19	1:43	3:02

The elimination of the PT is expected to provide annual cost savings for the State of Connecticut because two people will no longer be required to manually score the approximately 235,000 essays. Moreover, at the local level, it will repurpose testing time for instruction. The ELA PT was administered in two sittings; therefore students had to be scheduled to go through a computer lab (or other testing situation) twice in order to complete the ELA PT. With this reduction, about 800 schools (excludes high schools) will no longer have to accommodate two PT testing events into their calendar. The instructional time that is redeemed for local schools is expected to exceed the actual times between login and logout.

From Standards to Curriculum and Instruction

The decision to eliminate the ELA PT from the state summative assessment should not be viewed as an indication that teachers and instructional leaders must now modify what they teach or how they teach it. The Connecticut Core Standards – not assessments – should inform important choices regarding curriculum and instruction.

While the state summative assessment is an important lever for positive change, it is not the only one. The Smarter Balanced assessment system is comprised of more than just the summative assessment. Another critical component of the system is the optional interim assessments. These interims are not standardized assessments but: (i) are designed to inform teaching and learning; (ii) can allow teachers to monitor progress and improve instruction throughout the year; (iii) are aligned to the standards; and (iv) offer some of the following important features:

- The **test items** in those interims are of high quality as they are developed under the same conditions, protocols, and review procedures as those used in the summative assessments. They assess the same Connecticut Core Standards, adhere to the same principles of Universal Design in order to be accessible to all students, and provide evidence to support claims in mathematics and ELA/literacy.
- Interim **assessment blocks** focus on smaller sets of related concepts and provide more detailed information for instructional purposes. While there is no better place to learn these skills than as part of a dynamic and interactive classroom setting, teachers can use the interim assessments throughout the year to inform instruction by measuring student progress, identifying areas of strengths and weaknesses, and assessing the skills of incoming students.

- The scope, sequence, number, and timing of the interim assessments are locally determined. Moreover, a teacher may administer an interim assessment that is above or below the student’s current grade level to best target the instructional needs of the student.
- Though the interims are not “standardized”, they can give teachers near real-time information about how their students are progressing because the tests are scored locally (either machine-scored by the test software or scored by teachers using specified rubrics). By contrast, on the summative assessment, there is no separate score for the ELA PT (it is subsumed into overall score) and results are provided *after* the end of the school year.
- The CSDE continues to present in-depth workshops about the purpose, uses, and scoring of the interim assessments to help teachers and instructional leaders to maximize the benefits of the system.

Writing Portfolio Project: The CSDE will also be creating a collection of K-12 tools to support the instruction and assessment of writing utilizing research-based, Connecticut Core Standards (CCS) aligned strategies. The purpose of this portfolio is support the development of student writing across elementary through secondary education in preparation for postsecondary schooling or career. These optional tools will be available for use to all districts and will include existing, vetted CCS aligned instructional and professional development resources (Smarter Balanced writing rubrics and student exemplars, Achieve writing materials, CT Core Standards website materials, etc.) as well as resources that will be created in collaboration with Connecticut educators (student writing models, K-2 grade-level rubrics, student-friendly learning targets and rubrics). The K-12 writing tools will be conveniently organized by grade band (K-2, 3-5, 6-8, and 9-12) and will support the instruction and assessment of all purposes of writing across the various content areas.

- Project Timeline:
 - Spring 2016: Plan and research
 - Summer 2016: Organize and create professional learning activities
 - Fall 2016: Develop materials, evaluate resources, and gather additional input
 - Winter 2017: Statewide launch and training

Conclusion

Connecticut has been and will continue to be a leader in administering high quality assessments to support teaching and learning. However, since the passage of the No Child Left Behind Act in 2001, poor practices developed in some places where curriculum and instruction decisions were driven by the test and not the standards. Since 2010, the field of public education has been working methodically to correct this. After adopting rigorous standards, Connecticut practitioners collaborated with other states to develop assessments that are aligned to those standards. Alongside these changes, the state also modified its accountability system to look beyond test scores and provide a more holistic, multifactor perspective of district and school performance. These changes were further supported from the federal level by ESEA Flexibility (i.e., waiver) and the new Every Student Succeeds Act (ESSA).

This rightsizing of the Smarter Balanced ELA assessment is one more step in this effort toward finding the right place for the state summative assessment in the state's educational system. The CSDE will continue to have the highest expectations for each and every student and demand the highest levels of accountability for schools and districts; it will, however, leave curriculum and instructional decisions to the experts: our teachers and our instructional leaders. We are confident that those who are closest to the students are best positioned to make decisions that will most benefit student learning.