Student Engagement Framework for Smarter Balanced Assessment Consortium

Prepared by Catherine Taylor (CTB McGraw-Hill),
Kari Kokka, Linda Darling-Hammond, and Jack Dieckmann
Stanford Center for Assessment, Learning, and Equity (SCALE), and
Vivian Santana Pacheco and Susan Sandler, Sandler Foundation

Introduction

Assessment consortia partnerships designing assessments aligned with the Common Core State Standards have promised innovative approaches that better reflect the full range of standards -- including higher-order thinking and performance skills -- and that are more fair and accessible to the full range of students in our diverse nation.

Engaging students in meaningful applications of their knowledge is a key aspect of both addressing the standards and providing greater access. Not only do the standards emphasize the importance of meaningful engagement in contextualized tasks, but evidence suggests that engagement is strongly related to student performance on assessment tasks, especially for students who have been traditionally less advantaged in school settings. In the traditional assessment paradigm, however, engagement has not been a goal of test development, and concerns about equity have focused on issues of bias and accessibility. The traditional approach to avoid bias has resulted in creating highly decontextualized items. Unfortunately, this has come at the cost of decreasing students' opportunities to *create meaning* in the task as well as their motivation to *cognitively invest* in the task, thereby undermining students' opportunities to adequately demonstrate their knowledge and skills.

Assessment designs that include performance tasks, technology-enhanced items, and constructed-response items offer a ripe opportunity to develop tasks that engage all students of diverse backgrounds. Performance tasks and constructed-response items can allow students to demonstrate their evaluation, synthesis, analysis and application skills by providing multiple solution pathways to address the task, in contrast to multiple choice tests that privilege single solution pathways or one "right" interpretation of text.

The Design Specifications for the Smarter Balanced Assessment Consortium Performance Tasks, for both ELA and Mathematics, focus heavily on student engagement. This document adds to the information in the Design Specifications to describe design elements that can increase student engagement, with a special eye toward engagement considerations for traditionally underserved students. This document also adds to our thinking about ways to increase student engagement with the Smarter Balanced traditional interim and summative tests as well as classroom formative assessments.

In what follows, we begin with an overview of research on student engagement. We then present a set of design challenges, recommendations, and guiding questions for item and performance task writers' (see Table 1), followed by design considerations for each of the following elements: Pre-Assessment Classroom Activities (part of the Smarter Balanced Performance Tasks), Design of Items and Tasks, Design of Standardized Assessments as a whole, and Scoring of the Items and Tasks. Summaries and resources of student engagement literature are also presented in Appendix C.

Student Engagement: A Brief Review of Research

Student engagement in classroom activities and assessments is acknowledged to be "a highly desirable goal with positive outcomes for all parties" (Bryson and Hand 2007, p. 354). It is a complex of internal and external behaviors that are necessary for effective '[mental] interaction with content' (Moore, 1989). School engagement describes students' feelings, behaviors, and thoughts about their school experiences. It is an important predictor of academic outcomes such as achievement, standardized test scores, and high school completion. Bodovski and Farkas (2007) followed students from K through grade 3 and found that student engagement was a stronger predictor of achievement test scores than either initial test scores or time spent studying.

Engagement has been shown to be malleable and responsive to variations in the learning environment (Fredricks, Blumenfeld, & Paris, 2004). For example, engagement can be improved through changes in teachers' relationships with students, instructional strategies, and the nature of tasks and assessments (Dotterer & Lowe, 2011). Features of the classroom and tasks that matter include relevance, personalization, autonomy, sense of belonging, authenticity, and collaboration, as we discuss below.

Relevance and *personalization* of assessment items are of heightened importance for engagement considerations of students of traditionally underserved groups. Students of privileged backgrounds have been found to be more able to compartmentalize and carry out irrelevant, decontextualized, obscure tasks (Darling-Hammond, L., Barron, B., Pearson, D. P., et al. 2008) than traditionally underserved students. Historically privileged students also tend to be more 'test wise' or savvy with the unspoken skills and strategies to successfully tackle a test (Arbuthnot, 2011). Although the causes of these differences are unclear, they have implications for the design features of large scale assessments.

Students' abilities to connect scenarios that are relevant and/or personalized to their own lives appear to improve engagement and performance (Meier, 2008). For example, Walkington and Petrosino (2012) found that personalization increased student performance on algebraic word problems. They studied 145 9^{th} graders of three Algebra classes where teachers utilized Cognitive Tutor Algebra, a computer based tutoring system that individualizes instruction

through adaptive problem selection, hints, and feedback. Students in the experimental group solved algebraic word problems matched to their self-reported interests (e.g. sports, music, art, games) in prior student surveys and interviews. Personalization increased student performance, especially with more cognitively challenging problems and for struggling students (For a more complete report of the study, see Appendix A).

Equally important in this study was the relevance of problems to students' lives. Problems *relevant* to students' lives were easier for students to solve than those not connected to their experience, even when they were personalized to an expressed interest. Research on the dimensions of mathematics test performance suggests that, when contexts are relevant to students' lives, struggling readers perform much better than expected on test items despite the fact that substantial reading is involved (Taylor & Lee, 2004). These findings suggests the importance of relevance in choosing assessment topics, and the possibilities for improving student performance through personalization when relevance is lacking.

Student engagement may also be improved through contexts that support *autonomy* (Connell & Wellborn, 1991) and *sense of belonging*. When autonomy is supported in the classroom, self-regulation and behavioral engagement increase. The structure of the school context should strengthen students' sense of the efficacy of their own learning strategies, their perceptions of their own abilities to learn, and their sense that achievement is under their own control. When students develop faith in their own competence, subsequent *cognitive* engagement increases. If they develop a *sense of belonging* in the school context, their *emotional* engagement increases. Affective engagement is the strongest predictor of achievement for historically underserved students; while affective and behavioral engagement in addition to prior achievement, are the strongest predictors of achievement for more affluent students (Connell, Spencer, & Abner, 1994).

Student collaboration has also been found to greatly increase student engagement (Chan & Leijten, 2012; Cooper & Speece, 1990; Greenwood, 1996, 1991; Greenwood, Delquadri & Hall, 1989; McCormick, Kinzie, & Korkmaz, 2011). This may be both a function of the social benefits of collaboration -- that is, the way in which it may enhance a sense of belonging and participation -- and because it boosts student interest and gives them more pathways to understanding the material through discussing their thinking and hearing others' ideas.

Authenticity also matters to student engagement and performance. Research by Newman et al. found that when students in elementary and middle school classrooms engage in *authentic work*, the quality of their academic performance increased (Newman & Associates, 1996; Newman, Bryk, & Nagaoka, 2001). Authentic intellectual work (Newman and colleagues, 1996, 2001) involves construction of knowledge, disciplined inquiry, and value beyond the classroom.

As good teachers know, these ways of engaging students can trade off against each other. If students need to study something that is far from their experience, they can be engaged by the format of the tasks or the way they are drawn into the work or the opportunity they have to collaborate with others.

Engagement Considerations for Traditionally Underserved Students

Additional engagement considerations for students of historically underserved groups include concerns of *stereotype threat*. Students of some subgroups (e.g., females, African Americans) often underachieve when faced with stereotype threat. Stereotype threat occurs when students perform below their actual ability because of concerns about gender or racial stereotypes (Steele, 2003) or when presented with abstract tasks irrelevant to their own lives. Steele and his colleagues have tested specific ways of framing and presenting tasks and tests that have demonstrated significantly higher performance from groups that experience stereotype threat (girls in math, African Americans on various kinds of tests, etc.) These approaches should be considered as classroom activities are developed and as test instructions are framed.

Interpretation of text may be of particular concern for students of different socioeconomic and cultural backgrounds as well as English Language Learners. Student interpretation of test items is often mediated by socioeconomic and cultural factors, reducing the validity of the assessment. For example, Solano-Flores and Trumbull (2003) analyzed the different ways students interpreted 1996 4th grade NAEP mathematics items. The "lunch money" item read as follows.

> Sam can purchase his lunch at school. Each day he wants to have juice that costs 50¢, a sandwich that costs 90¢, and fruit that costs 35¢. His mother has only $1.00 bills. What is the least number of $1.00 bills that his mother should give him so that he will have enough money to buy lunch for 5 days?

This item is intended to measure proficiency with addition, multiplication, and rounding. However, interviews of students' interpretation of the question of three student subgroups (high SES suburban white, low SES inner city African American, and low SES rural Native American) revealed great variation. 84% of white students read the question as intended, whereas only 56% and 52%, respectively, of Native American and African American students read the sentence as intended. Solano-Flores and Trumbull also found that 10% and 18%, respectively, of the Native American and African American students interpreted the word *only* as restricting the number of dollars (i.e., "His mother has only one dollar"); however, this interpretation was not observed with white students in the study.

Taylor and Lee (2011) found that, when reading test items ask students to interpret complex text, girls, Latino Americans, African-Americans, and Asian-Americans perform better than expected on constructed-response items whereas whites and boys tend to perform better than expected on

multiple-choice items. These results suggest that readers' interpretations of text are influenced by background experiences – interpretations that may not be represented in the answer choices for multiple-choice items.

For English Language Learners, inclusion of unnecessary *language complexity* can also provide inaccurate measures of student understanding. However, assessments that use clear language embedded in realistic contexts have been found to provide ELL students greater opportunities to demonstrate understanding than traditional discrete-item assessments (Abedi, 2010). To address concerns of reading load for ELL students we provide an example of a linguistically modified performance task that is clear without compromising its rich context in Appendix B (Abedi, 2010). In the following section and Table 1 we offer recommendations for text presentation without compromising richness and context of tasks for SBAC assessments.

Scoring rubrics, while not tied explicitly to engagement, are of further consideration for students of traditionally underserved backgrounds. For example, Smitherman (2000) found that, in student writing responses on NAEP, high school responses that included African American vernacular English were more likely to receive a lower score when scorers used a holistic rubric (without specified weights for each element). In mathematics Moschkovich (2008) found multiple meanings behind students' reasoning expressions when examining third grade English Language Learners' problem solving discourse, highlighting the importance of evaluators' interpretation of students' mathematical reasoning.

We address concerns highlighted in this literature review for SBAC assessment design in the table and discussion below.

Assessment Tensions and Smarter Balanced Assessment Design Recommendations

Several tensions present challenges to assessment design. For example, efforts to personalize assessments are met with questions about standardization. Similarly student collaboration in the assessment context can raise concerns about measuring individual performance. Below we list tensions in assessment design between concerns for engagement and concerns associated with the ways in which fairness has often been sought in traditional test design through standardization and decontextualization. We also present design recommendations that address these tensions in Table 1. Each design recommendation is also accompanied with questions for item writers' consideration. Following Table 1 we offer a bulleted list of Design Recommendations by each of the following Task Components: Classroom Activity, Design of the Task, and Scoring of the Tasks.

Among the tensions that emerge as assessment designers seek to balance engagement with a range of other goals for assessment, are the following.

1. How can assessments offer relevant themes, when students from majority and non-majority groups have often had different experiences?
2. Can assessments be personalized without threatening key aspects of standardization?
3. Can assessments include opportunities for collaboration when the outcome is intended to describe individual performance?
4. How can assessments balance the demands of authentic intellectual work, which results in open-ended responses that must be interpreted, and the desire for ease of scoring?
5. How can assessments balance the rich context needed for authentic performance tasks with concerns for minimizing linguistic complexity?
6. How can assessments stimulate engagement and encourage students to understand and persevere in the task when test objectivity has traditionally been defined as precluding engagement with peers or adults?
7. How can assessments be scored in ways that allow diversity of responses without sacrificing reliability and ease of scoring?

These tensions are discussed further, with suggestions for resolution in Table 1 below. We note that there ways to balance different pathways to engagement in performance tasks in particular, such that task developers can aim to purposefully engage students on one or more levels, by:

- connecting to young people's interests and experiences,
- provoking their curiosity with lively stimuli
- communicating through technologically current media, or
- drawing them into a dramatic scenario in which they take personal agency to complete a purposeful task.

Table 1. SBAC Design Recommendations to Increase Student Engagement

Challenge	The Traditional Testing Approach	Recommended SBAC Approach (in task and through classroom interaction activity)	Questions for task developers to consider
1. Familiarity and relevance matter to student engagement and performance, yet students have had different experiences. Different contexts are more familiar and relevant to some students than others.	Design tasks for a 'mainstream' student audience. Avoid contexts or examples that may be unfamiliar to subgroups of students, such as sports, hobbies, activities or social/political issues.	Assessment items and contexts should be ***familiar*** and ***relevant*** to many students at the age group being tested and should be sensitive to the experiences of non-dominant and traditionally underserved students of diverse backgrounds. Include a mix of items and tasks that draw on the experiences and cultural contexts of both dominant and non-dominant groups. As a group, test items and performance tasks should represent a variety of diverse backgrounds. In ELA, for example, tasks might present students with texts discussing the Harlem Renaissance, Chinese Exclusion Act, Delano Grape Strike, and texts from a variety of authors (e.g. Chinua Achebe, Toni Morrison, Gabriel Garcia Marquez, Jhumpa Lahiri, James Baldwin.) Use sports, hobbies, and other contexts as interesting activities, but choose widely known sports or activities and avoid those that are known primarily to specific subgroups. Be sure to provide sufficient background knowledge about activities that are part of an item or task context for students to perform the task. (e.g. Don't assume that students know how a particular sport is scored or played.) Use classroom activities to "level the playing field" by familiarizing students with the context of the task in an engaging way that reduces anxiety about non-familiarity.	• What prior knowledge, familiarity, or experience is expected, implied, assumed, and/or required of the task (e.g. how basketball games are scored, how the stock market works, how sleds, surfboards, or boats operate etc.)? Many test items and all performance tasks will have a context that is differentially known to different students by geography, interests or activities, etc. As much as possible, background knowledge associated with the context should be made explicit to the students and incorporated into the item or task itself. For Performance Tasks in particular, sufficient background knowledge should be provided in explanations in pre-assessment Classroom Interaction Activities or in the tasks themselves so that students with no prior exposure to the context are not disadvantaged. • For Smarter Balanced Performance Tasks, in what ways will classroom-based interaction adequately familiarize students with the given context or setting to level the playing field when they lack prior knowledge or experience? • Why would a student care about this topic? Consider a diversity of students in addressing this question from different backgrounds (e.g. such as but not limited to different socioeconomic, cultural, age, geographic, sexual orientation, religious, and language backgrounds). • Does the context privilege certain subgroups? For example, if a context is primarily known to upper-income students (e.g. traveling to Europe on vacation; sailing in a regatta) or to students with specific life experiences (e.g., making sense of urban bus schedules), the context is not appropriate for the purpose of this assessment). • Is this topic or context timely and is its relevance up to date?

Challenge	The Traditional Testing Approach	Recommended SBAC Approach (in items, tasks and through pre-assessment classroom activities)	Questions for task developers to consider
2. Personalization, choice, and student agency increase engagement. Standardization has often been interpreted as precluding personalization or choice.	Little to no choices offered in an effort to preserve standardization and equity. Little to no consideration of the student's role in tasks; items posed in the third person (e.g. Mary went to the store).	To engage students, Smarter Balanced items and tasks, while standardized, should present opportunities for ***personalization*** and choice with opportunities for ***autonomy*** and agency. Tasks can be personalized by allowing students to play a role in solving a problem, taking agency and engaging in choices about how to approach the problem and what decisions to make. • Allow students to personalize their response to the task to their own lives, settings, and contexts. • Invite students of diverse backgrounds to envision themselves in the role they are placed in through a performance task. • Offer opportunities for student agency and choice in classroom formative assessments • Give students an opportunity to provide their own interpretations of text Where possible, allow choice of tasks and prompts as is common in IB, AP, and other countries' assessments. This allows students to choose tasks that connect with student interest, relevance to their lives, and/or familiarity.	• What is the student's role in the task? Is this role purposeful, meaningful, appropriate, and accessible? For example, a performance task that asks students to create a business plan might seem realistic, engaging, and authentic it is not accessible to students because at their age they have never done this nor have had any exposure to this. • How are students invited to enact agency and/or choice through the task? Elementary aged students may engage in a task to write a letter to a parent or family member to advocate and argue for something they desire and care about while a high school aged student may prepare a speech to student government (although not all schools have student government programs), principal, or teachers in their school to suggest changes in school policies.
3. Student collaboration improves engagement. A focus on standardization and individual performance has typically eliminated opportunities for collaboration.	Proctor reads script with minimal instructions when administering the exam.	Include collaboration with peers in pre-assessment classroom activities that set the background for formative assessments and summative performance tasks. Classroom activities can be standardized in terms of the activities they include. Have students work independently on assessment items and tasks that follow classroom activities, minimizing concerns about so-called 'cheating' or not measuring students' individual performance. Classroom activities should be facilitated by students' regular classroom teachers to increase familiarity, reduce anxiety, and enhance students' sense of belonging.	• During pre-assessment classroom activities, are students afforded opportunities to engage with peers and ask questions of their teacher? • Does the work provide a standard set of stimuli with teacher facilitation, information/guidelines that are clear and common, but not a lockstep script? • How will the teacher facilitator's role contribute to increasing engagement? • How will the classroom-based interaction include a variety of modalities (e.g. partner/group work, listening to audio, watching video, etc. where possible)? • What modes can be used to increase engagement? (e.g. partner or group work)

Challenge	The Traditional Testing Approach	Recommended SBAC Approach (in task and through classroom interaction activity)	Questions for task developers to consider
4. Students invest in authentic and purposeful intellectual work, which requires reasoning. Such work is more complex and less easily scored by machine.	Use items that are multiple-choice, discrete, decontextualized and lend themselves to computer based scoring.	Invite students to engage with ***authentic intellectual*** tasks that involve construction of knowledge, disciplined inquiry, and value beyond the classroom. • Structure tasks so that they can be scored reliably with analytical rubrics. • Build rubrics and scoring criteria to account for multiple answers that are within a known range of acceptable answers. • Select anchor papers that represent a wide range of high quality but diverse responses/performances to guide scorer training. • Be alert to opportunities to move from hand scoring to eventual AI scoring (e.g. items that may be propositionally scored).	• Is the context rich and authentic to students of diverse backgrounds? • To what extent does the task address the central ideas and modes of inquiry in the discipline? • To what degree does instruction involve students in manipulating information and ideas to arrive at conclusions that solve an open-ended problem? • How is the task structured to provide a range of acceptable right answers that can be analytically scored?
5. Tasks with rich contexts invite engagement. Linguistically dense items may disadvantage English learners or struggling readers.	Create unnecessary text complexity (that is not construct-relevant) Or Reduce text and reduce context in an effort to offer manageable text (however, this can result in decontextualized, inauthentic items not relevant to students' lives)	Smarter Balanced assessment items and tasks should present ***manageable text*** (both for English Language Learners and for assessments where reading comprehension is accessible to most students) without stripping the task of its ***rich context.*** • Use clear text, familiar words, and removing construct-irrelevant complex language without removing construct-relevant language. See Appendix B for a linguistically modified mathematics task from the New Jersey SRA, High School Proficiency Assessment. • Use a classroom activity to be sure that key concepts and vocabulary have been introduced and to familiarize students with the context. • Where appropriate present stimuli in more than one mode and/or format. These may include graphic, tabular, audio-visual, and pictorial as well as text-based stimuli (though not all at once to avoid overwhelming students). Allow students ready access to glossaries to decode text. • Invite students to demonstrate their understanding through multiple modes (i.e. using graphing tools, creation of tables, equations, and figures to communicate solution strategies in mathematics).	• Does the task or test as a whole offer a variety of information sources? (text, visuals, figures, tables, charts, etc.) • Should the text be linguistically modified to be readily accessible? (See Appendix B for an example.) • What additional formats may be used to present stimuli that may engage more students? • In what multiple ways are students invited to demonstrate their understanding? • Do student response capabilities also allow for multiple solution pathways and modes of representation? • Is text comprehension and understanding of the context/scenario supported through the classroom activities?

Challenge	The Traditional Testing Approach	Recommended SBAC Approach (in task and through classroom interaction activity)	Questions for task developers to consider
6. Clear and encouraging expectations engage students in the intended task.	Procedural directions are provided to students in an authoritative voice. Tests assume that students understand the format of the test and the types of responses expected.	Test and performance task expectations should be clear and engaging to encourage students to persevere with Smarter Balanced assessments without compromising assessment objectivity. • Structure assessments in such a way that students feel a personal stake in the task and that their success is valued and expected. • Minimize stress by providing adequate time, access to resources, and focus on concepts and skills students perceive to be important. • Minimize messages regarding competitive performance. • Minimize 'stereotype threat' by communicating that the purpose of the assessment is to support students' learning	• In what ways do the Classroom Interaction Activity and Performance Task communicate test taking strategies and high expectations for students' success? • Are expectations for student responses (including conventions) clear to the student? For example, the audience and format expected in responses should be clear, e.g. "Write a letter to your school principal with your recommendation."
7. Specificity of rubrics and diverse anchor papers can increase fairness	Multiple-choice items are scored by computer.	Scoring rubrics should be analytic and should articulate the diversity and range of appropriate student responses to ensure fairness. • Performance Tasks should be hand scored, with capability for students to respond in some languages other than English. • Specify weights should be given to each element in the rubric. • A diversity of possible student responses should be included in anchor papers or scoring guides • Provide opportunities for students to earn partial credit for partial understanding or partial completion.	• Does the rubric anticipate the full range of possible student responses? • Is the rubric clear enough for scorers to fully understand how to award points? • Is the rubric specific enough to increase inter-rater reliability and norming of scores? • Is the rubric susceptible to bias (unintentional or implicit)? • Does the rubric include specific weights for each element? • Does the rubric include opportunities for partial credit for partial understanding?

Design Recommendations by Assessment Component

Student engagement shows positive outcomes for achievement, standardized test scores, and connection with the academic discipline. Positive effects of engagement range from primary school through college. Although prior achievement is the best predictor of later achievement for high performing students, engagement is the best predictor of achievement for low performing students. Design recommendations that improve student engagement are listed below by assessment type and performance task component.

Pre-Assessment Classroom Activities (Part of the Smarter Balanced Performance Tasks)

- The Classroom Activity should ***activate students' prior knowledge*** and ***connect the topic*** to students' own lives.

- The Classroom Activity can enhance engagement and access by creating ***familiarity*** with the context and introducing key concepts and vocabulary so that students are ready to engage the substance of the task.

- The Classroom Activity may provide an opportunity to offer ***personalization, student choice, and agency*** through the type of activities it uses. For example, students may choose which side of an argument they wish to represent in a brief class debate. Students might vote on their preferences or share their opinions about a topic.

- Inclusion of video, audio, and use of manipulatives during the Classroom Activity may further invite engagement through the addition of ***multiple communication modalities***.

- The Classroom Activity can help create a supportive learning environment focused on all students' success, through ***active and collaborative learning*** (such as hands-on learning in small groups where students work together to make sense of concepts and solve problems).

- The classroom activity should incorporate ***visualization*** of important ideas and concepts. The use of visual models accompanied by discussion to support an understanding of the context is likely to increase engagement with the Performance Task.

- The Classroom Activity provides an opportunity to ***decrease stress*** by creating a climate of care, focusing on big ideas rather than discrete bits of knowledge, and communicating directions and expectations for the subsequent Performance Task. Improving socio-emotional context (a sense of belonging, social support, positive relationships with teachers) through the Classroom Activity may engage more students in the subsequent Performance Task. Stress can also be minimized by providing adequate time, access to resources, and focus on concepts and skills students perceive to be important.

- Assessment events should be introduced and structured in such a way that students feel their ***success is valued and expected***. Messages regarding competitive performance should be

minimized. Teacher facilitators can be prompted to convey expectations to students such as, "On this assessment, there are many different ways to accomplish the task. This task has several parts, and there are several 'right' ways to solve the problem. Look over the entire task to understand the overarching goal. Use what you know, think creatively. If you get stuck review the questions and information provided to see what you have already answered. If possible, skip to the following prompt and return to the one that stumped you later." Messages can also be constructed to ***minimize stereotype threat***, such as affirmations about the fact that, while difficult, the task is one at which students are expected to be able to succeed.

Design of the Formative, Interim, or Summative Assessment Event

- Assessment events should be introduced and structured in such a way that students feel their ***success is valued and expected***. Messages regarding competitive performance should be minimized.

- Teacher facilitators can be prompted to convey expectations to students such as, "On this assessment, there are many different ways to accomplish the task. This task has several parts, and there are several 'right' ways to solve the problem. Look over the entire task to understand the overarching goal. Use what you know, think creatively. If you get stuck review the questions and information provided to see what you have already answered. If possible, skip to the following prompt and return to the one that stumped you later."

- Messages can also be constructed to ***minimize stereotype threat***, such as affirmations about the fact that, while difficult, the task is one at which students are expected to be able to succeed.

Design of the Items and Tasks

- Items and performance tasks should engage students in ***authentic, hands-on work*** that involves inquiry, use of authentic disciplinary thinking and processes, and the construction of new knowledge. Tasks should involve inquiry-based learning (questioning, investigating, drawing conclusions, reasoning from evidence), higher-order reasoning, and 'sense-making.'

- Tasks should be ***relevant*** to students' own lives and experiences and have value beyond school. Whenever possible, the work should be anchored in ***real world contexts***. When their work is placed in academic/theoretical contexts, it should be work worth doing – work that the student feels is purposeful, interesting, and challenging yet achievable for the student.

- The focus of assessments should be on ***important, core ideas*** rather than discrete, abstract bits of knowledge and/or skill.

- Whenever possible, assessment events should provide opportunities for ***autonomy, agency, and choice***. Students should be able to take a role in solving a meaningful problem or developing meaningful interpretations of text in a fashion that allows them to choose how they will approach the problem and that allows them to make decisions in the process.

- Where possible, assessments should provide ***auto-feedback*** (i.e. clues to the causes of difficulties as well as opportunities for attacking the task in a new, more informed way -- assessment as learning).

- Where possible performance tasks should ***integrate subject matter***, demonstrating how concepts and skills relate across disciplines; and how students can learn in context.

- Assessment items and contexts should be ***familiar*** and ***relevant*** to the majority of students at the age group being tested and should be sensitive to the experiences of non-dominant and traditionally underserved students of diverse backgrounds. Sports, hobbies, and other activities may be used as contexts, but developers should choose widely known sports or activities and avoid those that are known primarily to privileged groups. Be sure to provide sufficient background knowledge about activities that are part of a task context for students to perform the task. (i.e. Don't assume that students know how a game is scored or played.)

- Smarter Balanced assessment items and performance tasks should present ***manageable text*** (both for English Language Learners and for struggling readers in assessments where reading comprehension is not construct relevant) without stripping the task of its ***rich context.*** Provide clear text, use familiar words, without removing construct-relevant language. For example, we present an example in Appendix B of a linguistically modified mathematics test item from the 2002-2003 New Jersey SRA, High School Proficiency Assessment. The linguistic modification retains the context in which the task is set.

- Stimuli should be presented in ***multiple modes and formats*** (use graphs, charts, figures, etc.). Where possible invite students to demonstrate their understanding through multiple modes (i.e. using graphing tools, creation of tables, equations, and figures to communicate solution strategies in mathematics).

- Smarter Balanced formative assessments and performance tasks can offer ***choices*** for students. Student choice of task completion from a menu of tasks is successfully used in other exams such as the A-Level British Council and International Baccalaureate assessments. While SBAC does not (yet) offer student choice of tasks we suggest that SBAC move in this direction as SBAC develops to meet the needs of 21st century thinking.

- Smarter Balanced might consider as a future option that students be surveyed about their interest (from a fixed number of choices) prior to presenting a performance task. A question may ask students, "From the following topics which one are you most interested in? a) music, b) sports, c) art, or d) games." The subsequent performance items and tasks may be

set using the context chosen by the student. (For example, if the performance task aims to capture student understanding of linear functions and lines of best fit and a student chooses music as their topic of interest (from the limited four choices), data can represent record sales trends over time and ask students to compare and contrast CD sales and iTunes sales patterns. Alternatively if a student selects sports, a performance task may present Oakland A's and Los Angeles Angel's ticket sales over time, ask for the lines of best fit, and other questions related to prediction and comparison to capture student thinking and facility with using linear functions to make real world predictions and decisions. If data show that average performance is similar across contexts, performance tasks can measure the same mathematics content but through the context that the student may be most interested in.

Scoring of Items and Tasks

- Scoring rubrics should be ***analytically constructed***, and should include specific weights for each element, emphasizing meaning and central understandings and minimizing aspects of communication that may be associated with culture and language background, where these are not the focus of the construct.

- Rubrics or scoring criteria should anticipate the ***diversity and range*** of possible student responses to ensure fairness in evaluation, particularly for students of traditionally underserved backgrounds. Scoring criteria should include the range of plausible anticipated student answers to alert scorers to award credit and partial credit when they see a novel or untraditional way of answering a question that they themselves might not anticipate as a 'correct' or 'partially correct' response.

- Smarter Balanced Performance Tasks should maintain ***open-ended responses*** that are, at least initially, human scored, in order to assess complex modeling, reasoning, and communication skills.

- Smarter Balanced Performance Tasks should, in some cases, allow students to respond in ***languages other than English***.

Conclusion

Smarter Balanced summative, interim, and formative assessments intend to measure 21st century skills and higher-order thinking. By definition a performance task "is an item type designed to provide students with an opportunity to demonstrate their ability to apply their knowledge and higher-order thinking skills to explore and analyze a complex, real-world scenario" (Design Specifications for Mathematics Performance Tasks for Smarter Balanced Assessment Consortium, July 2013). However, strategies can be used to ensure that the Smarter Balanced summative, interim, and formative assessments encourage student engagement through classroom activities that prepare students for assessments, allowing for student autonomy and

choice in assessment events, providing items and tasks that are worthwhile and relevant to the lives of students, and communicating a belief in the value of each student. The recommendations provided in this document offer a framework for considerations of engagement for all students of diverse backgrounds (e.g. such as but not limited to different socioeconomic, cultural, age, geographic, sexual orientation, religious, and language backgrounds).

Engagement considerations are of ongoing concern and importance. We see this as an important first step and look forward to continued dialogue and actions to continually address concerns of student engagement in Smarter Balanced assessments.

References

Abedi J. (2010). Performance Assessment for English Language Learners. Stanford, CA: Stanford University. Stanford Center for Opportunity Policy in Education.

Arbuthnot, K. (2011). *Filling in the Blanks: Understanding Standardized Testing and the Black-White Achievement Gap.* Charlotte, NC: Information Age Publishing.

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *Elementary School Journal, 108,* 115-130.

Bryson, C., and L. Hand. 2007. The role of engagement in inspiring teaching and learning. *Innovations in Education and Teaching International 44,* 349–62.

Carini, R. M., Kuh, G. D., & Klein, S. P. Student engagement and student learning: Testing the linkages. *Research in Higher Education, 47*, 1-32.

Connell, J. P., Spencer, M. B., & Aber, J. L. (1994). Educational risk and resilience in African American youth: Context, self, action, and outcomes in school. *Child Development, 65,* 493–506.

Darling-Hammond, L., Barron, B., Pearson, D. P., et al. (2008). P*owerful Learning: What We Know About Teaching for Understanding,* pp. 74-76. San Francisco: Jossey-Bass.

Dotterer, A. M., & Lowe, K. (2011). Classroom context, school engagement, and academic achievement in early adolescence. *Journal of Youth and Adolescence, 40*, 1649-1660.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, *59*, 117–142.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59–109.

Furlong, M. J., & Christenson, S. L. (2008). Engaging students at school and with learning: A relevant construct for all students. *Psychology in the Schools*, *45*, 365–368.

Gehlbach, H., Brinkworth, M. E., and Harris, A. D. (2011). Changes in teacher–student relationships. *British Journal of Educational Psychology*.

Howard, T. C. (2010). *Why Race and Culture Matter in Schools: Closing the Achievement Gap in America's Classrooms. Multicultural Education Series*. Teachers College Press. 1234 Amsterdam Avenue, New York, NY 10027.

Jackson, Y. (2011). *The Pedagogy of Confidence: Inspiring High Intellectual Performance in*

Urban Schools, pp.77-84. New York: Teachers College Press.
Jimerson, S. R., Campos, E., & Greif, J. L. (2003). Toward an understanding of definitions and measures of school engagement and related terms. *The California School Psychologist, 8*, 7–27.
Kuh, G. D. (2001). Assessing what really matters to student learning: Inside the national survey of student engagement. *Change, 33*(3), 10–17. 66.
Kuh, G. D. (2003). What we're learning about student engagement from NSSE. *Change, 35*(2), 24–31.
Kuh, G. D. (2004). *The national survey of student engagement: Conceptual framework and overview of psychometric properties.* Bloomington: Indiana University Center for Postsecondary Research and Planning.
Kuh, G. D. (2008). High-impact educational practices: What they are, who has access to them, and why they matter. Washington: Association of American Colleges and Universities.
Kuh, G. D. (2009). The national survey of student engagement: Conceptual and empirical foundations. New Directions for Institutional Research, 141, 5–21.
Lau, S., & Roeser, R. W. (2002). Cognitive abilities and mental processes in high school students' situational engagement and achievement in science, Educational Assessment, 8, 139-162.
Lubienski, S. T. (2002). A closer look at Black-White mathematics gaps: Intersections of race and SES in NAEP achievement and instructional practices data. *Journal of Negro Education*, 269-287.
Martiniello, M. (2008). Language and the Performance of English-Language Learners in Math Word Problems. *Harvard Educational Review,* 78:2.
Meier, T. (2008). *Black Communications and Learning to Read: Building on Children's Linguistic and Cultural Strengths.* New York: Lawrence Erlbaum.
Moore, M. 1989. Three types of interaction. *American Journal of Distance Education* 2: 1–6.
National Research Council and the Institute of Medicine (NRCIM). (2004). *Engaging schools: Fostering high school students' motivation to learn*. Washington, DC: The National Academies Press.
Newmann, F. M., Wehlage, G. G., & Lamborn, S. D. (1992). The significance and sources of student engagement. In F. M. Newmann (Ed.), *Student engagement and achievement in American secondary schools*. (pp. 11–39). New York: Teachers College Press.
Park, S. Y. (2005). Student engagement and classroom variables in improving mathematics achievement. *Asia Pacific Education Review, 6*, 87-97.
Solano-Flores, G. The Cultural Validity of Assessment Practices. In Basterra, M. del R., Trumbull, E. & Solano-Flores, G. (Eds.). (2011). *Cultural Validity in Assessment: Addressing Linguistic and Cultural Diversity,* pp.3-21. New York: Routledge.
Solano-Flores, G., & Li, M. (2009). Generalizability of Cognitive Interview□Based Measures Across Cultural Groups. *Educational Measurement: Issues and Practice*, *28*(2), 9-18.

Solano-Flores, G., & Nelson□Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, *38*(5), 553-573.

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher*, *32*(2), 3-13.

Steele, C. (2003). Stereotype threat and African-American student achievement. In T. Perry, C. Steele & A. G. Hilliard (Eds.), *Young, gifted, and black: Promoting high achievement among African-American students*. Boston: Beacon Press, pp. 109-130.

Taylor, C.S. & Lee, Y. (2012). Gender DIF in Tests with Mixed Item Formats. *Applied Measurement in Education, 25*, 246-280.

Taylor, C.S. & Lee, Y. (2011). Ethnic DIF and DBF in Reading Tests with Mixed Item Formats. *Educational Assessment, 16,* 35-68.

Taylor, C.S. & Lee, Y. (2004, December). Is Reading a Dimension in the WASL Mathematics Test? Using differential item functioning to examine the dimensions of WASL mathematics. Presented at the Washington State Assessment Conference, Seattle, WA.

Strutchens, M. E., & Silver, E. A. (2000). NAEP findings regarding race/ethnicity: Students' performance, school experiences, and attitudes and beliefs. *Results from the seventh mathematics assessment of the National Assessment of Educational Progress*, 45-72.

Walkington, C., & Sherman, M. (2012). Using adaptive learning technologies to personalize instruction: The impact of interest-based scenarios on performance in algebra. In *Proceedings of the 10th International Conference of the Learning Sciences. Sydney, Australia*.

Wigfield, A., & Wentzel, K. R. (2007). Introduction to motivation at school: Interventions that work. *Educational Psychologist*, *42*, 191–196.

Appendix A

Example of four personalized variations on original algebra story problem

Walkington, C., & Sherman, M. (2012). Using adaptive learning technologies to personalize instruction: The impact of interest-based scenarios on performance in algebra. In *Proceedings of the 10th International Conference of the Learning Sciences. Sydney, Australia*.

Original Problem	One method for estimating the cost of new home construction is based on the proposed square footage of the home. Locally, the average cost per square foot is estimated to be $46.50.
Sports	You are working at the ticket office for a college football team. Each ticket to the first home football game costs $46.50.
Music	You are helping to organize a concert where some local R&B artists will be performing. Each ticket to the concert costs $46.50.
Art	You have been working for the school yearbook, taking pictures and designing pages, and now it's time for the school to sell the yearbooks for $46.50 each.
Games	You work for a Best Buy store that is selling the newest Rock Band game for $46.50.

Appendix B

Linguistically Modified Test Item

New Jersey Department of Education, 2002–03 SRA Mathematics Performance Assessment Task

ORIGINAL ITEM: Dorothy is running for president of the student body and wants to create campaign posters to hang throughout the school. She has determined that there are four main hallways that need six posters each. A single poster takes one person 30 minutes to create and costs a total of $1.50.

What would be the total cost for Dorothy to create all the needed posters? Show your work.
If two people working together can create a poster in 20 minutes, how much total time would Dorothy save by getting a friend to help her? Show your work.

If Dorothy works alone for 3 hours and is then joined by her friend, calculate exactly how much total time it will take to create all the necessary posters. Show your work.
Omar, Dorothy's opponent, decided to create his posters on a Saturday and get his friends Janice and Beth to help. He knows that he can create 24 posters in 12 hours if he works alone. He also knows that Janice can create 24 posters in 10 hours and Beth can create 24 posters in 9 hours. How long will it take them, if all three of them work together, to create the 24 posters? Round all decimals to the nearest hundredth. Show your work.

When Omar went to purchase his posters, he discovered that the cost of creating a poster had increased by 20%. How many posters will he be able to create if he wants to spend the same amount of money on his posters as Dorothy? Justify your answer.

LINGUISTICALLY MODIFIED ITEM: You want to plant 6 roses in each of four large pots. Planting a single rose takes you 30 minutes and costs $1.50.
What is the total cost to plant all the roses? Show your work.

With a friend's help, you can plant a rose in 20 minutes. How much total time do you save by getting a friend to help? Show your work.

You work alone for 3 hours, and then a friend joins you. Now how much total time will it take to plant all the roses? Show your work.

You can plant 24 roses in 12 hours. Your friend Al can plant 24 in 10 hours and your friend Kim can plant 24 in 9 hours. How long does it take the three of you to plant 24 roses together? Round all decimals to the nearest hundredth. Show your work.

You just discovered that the cost of purchasing a rose increased by 20%. How many roses can you plant with the same amount of money that you spent when a rose cost $1.50? Justify your answer.

Source: Abedi J. (2010). Performance Assessment for English Language Learners. Stanford, CA: Stanford University. Stanford Center for Opportunity Policy in Education.

Appendix C

Summaries and Resources of Student Engagement Research

Factor	Brief Summary of Research	Sources
Work that is relevant to the lives of students	Student engagement increased from 68 to 91 percent when students did "…activities that encourage students to draw on their previous knowledge and experiences, engage in critical thinking, and apply what they learn to their own lives." Type of instruction (lecture versus authentic work) explained 15% of the variance in 'student engagement' scores.	Cawthon, S. W., Dawson, K., & Ihorn, S. (2011). Activating student engagement through drama-based instruction. *Journal for Learning through the Arts, 7* (1), 1-29.
Authentic intellectual work	Authentic intellectual work (Newman and colleagues, 1996, 2001) involves construction of knowledge, disciplined inquiry, and value beyond the classroom. Authentic intellectual work can be shown in three places: instruction, assessment, and student work. (To what degree does instruction involve students in manipulating information and ideas to arrive at conclusions that produce new meanings for the student? To what extent does instruction address the central ideas of the discipline? To what degree are students involved in exchanges with the teacher and/or their peers about disciplinary ideas? To what extent do the assessment and instruction provide connections to the larger context in which students live?) When students in elementary and middle school classrooms engaged in authentic work, quality of academic performance increased regardless of SES, ethnicity, gender or prior achievement level.	Newman, F. M., & Associates (1996). *Authentic achievement: Restructuring schools for intellectual quality.* San Francisco: Jossey-Bass. Newman, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago: Consortium on Chicago School Research.
Student self assessment	National Survey of Student Engagement (NSSE), an annual survey of student engagement at the college level, was developed to measure student engagement so that colleges and universities could self-evaluate and improve the quality of instruction in order to increase student engagement (Kuh, 2001, 2003, 2004, 2008, 2009). NSEE assesses students' involvement with activities and conditions likely to generate learning'' (Coates 2006). It also looks at "the policies and practices that institutions use to induce students to take part in these activities'' (Kuh 2003). Using longitudinal and cross sectional data, NSEE found that level of engagement was a good predictor of undergraduate GPA.	Kuh, G. D. (2001). Assessing what really matters to student learning: Inside the national survey of student engagement. *Change, 33*(3), 10–17. 66. Kuh, G. D. (2003). What we're learning about student engagement from NSSE. *Change, 35*(2), 24–31. Kuh, G. D. (2004). *The national survey of student engagement: Conceptual framework and overview of psychometric properties.* Bloomington: Indiana University Center for Postsecondary Research and Planning.

Factor	Brief Summary of Research	Sources
Social-emotional context of the classroom	Social, instructional, and organizational climate of schools influences both students' engagement and their academic achievement (e.g., Eccles et al. 1998; Patrick et al. 2007; Ryan and Patrick 2001)…. Children who feel a sense of belonging and social support are more likely to be engaged and participate in school (Deci and Ryan 1985; Wentzel 1997). Children who are in conflict with their teachers may not feel connected or supported and may disengage from classroom activities (Connell 1990). Ladd and Burgess (2001) found that when teacher-child conflict was greater, students were less engaged in the classroom, were less likely to enjoy school, and were at increased risk for poor academic performance. Further, Baker (2006) found that teacher-child conflict was associated with lower report card grades and standardized test scores. Classrooms rated as having a positive climate were associated with children being more engaged in classroom activities and higher achievement.	Baker, J. A. (2006). Contributions of teacher–child relationships to positive adjustment during elementary school. *Journal of School Psychology, 44*, 211–229. Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum. Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3: Social, Emotional, and personality development* (5th ed., pp. 1017–1095). Hoboken, NJ: Wiley. Ladd, G. W., & Burgess, K. B. (2001). Do relational risks and protective factors moderate the linkages between childhood aggression and early psychological and school adjustment? *Child Development, 72,* 1579–1601. NICHD Early Child Care Research Network. (2005). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. The *Elementary School Journal, 105*, 305–323. Patrick, H., Ryan, A. M., & Kaplan, A. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. *Journal of Educational Psychology, 99*, 83–98. Ryan, A. M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal, 38*, 437–460. Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology, 89*, 411–419.

Factor	Brief Summary of Research	Sources
Integration of subject matter content	Connecting language arts, math, science, and arts through integrated instructional units focused on real world problems increased student engagement and achievement. In authentic problems, students are presented with a scenario in their first session. They are required to identify issues, research the principles underlying the issues, and learn material in context; scenarios are intended to motivate the students to engage in learning and understand the issues that underlie the problem.	Racknor, W., & Drake, S. M., (2011, Fall). *Curriculum integration: One school district's journey*. Toronto, Canada: Education Canada, Canadian Education Association.
Active learning, collaboration, and instruction focused on students' success	Study of trends in NSSE over four years for freshmen and seniors suggest that high engagement scores were associated with more focus on active and collaborative learning experiences, better student-faculty interactions, and a supportive learning environment. Supportive environmental factors included focus on student learning, helping students set personal goals, giving students an opportunity to close the gap between themselves and those who were more advanced.	McCormick, A. C., Kinzie, J., & Korkmaz, A. (2011, April). *Understanding evidence-based improvement in higher education: The case of student engagement.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
Visualizing	Students' ability to develop and use visual models of mathematics and science concepts improves engagement and performance.	Cifuentes, L. (2004). Visualization for middle school students' engagement in science learning. *Journal of Computers in Mathematics and Science Teaching, 23,* 109-137.
Decreasing stress	High achieving students in constant high stress contexts have decreased motivation and engagement and may resort to illegal drugs to manage the demands; decreases in stress increase motivation. Students' engagement increases when they believe that the educators care about their success. Strategies for decreasing stress: block schedules, climate of care, and focusing on understanding of the big ideas rather than many discrete bits of knowledge	Pope, D. (2010). Beyond doing school: From "stressed out: to "engaged in learning". Toronto, Canada: Canadian Education Association, 4-8.

Factor	Brief Summary of Research	Sources
Inquiry-based learning	Introducing enquiry-based learning in a first-year seminar course a university level significantly affected the learning behaviors of students, led to greater motivation to succeed, and to enhanced reasoning and processing skills that were transferred to other courses throughout undergraduate experience; all students benefited from EBL in terms of achievement, engagement, and ability to access and use resources to support learning (Murray & Summerlee, 2007; Summerlee & Murray, 2010). Studies of inquiry based classrooms in science showed that students developed deeper conceptual understand as well as increased engagement, interest and positive attitudes toward science. The 'five stages' include framing research questions, designing investigations, conducting investigations, collecting data, and drawing conclusions (reasoning from evidence) (Chang & Mao, 1999; Schwartz, Lederman, & Crawford, 2004). However, focus on teaching to standardized tests tends to undermine focus on student learning, trivialize the targets of learning, and discourage inquiry teaching. Studies of the classroom assessments found that, assessments in low income schools mirrored the types of test questions found on standardized tests and focused on low level skills while assessments in middle to upper income schools focused on reasoning and inquiry (Madaus, 1999; Madaus, et al, 1992).	Chang, C., & Mao, S. (1999). Comparison of Taiwan science students' outcomes with inquiry-group versus traditional instruction. *The Journal of Educational Research, 92*, 340–346. Madaus, G. F. (1999). The influence of testing on the curriculum. In M. J. Early & K. J. Rehage (Eds.), *Issues in curriculum: A selection of chapters from past NSSE yearbooks.* Chicago: The University of Chicago Press. Madaus, G.F., West, M. M., Harmon, M.C., Lomax, R. G., & Viator, K. A. (1992). *The influence of testing on teaching math and science in grades 4-12.* Chestnut Hill, MA: Boston College; Funded by National Science Foundation and Center for the Study of Testing, Evaluation, and Educational Policy. Murray, J., & Summerlee, A. J. S. (2007). The impact of problem-based learning in an interdisciplinary first-year program on student learning behaviour. *Canadian Journal of Higher Education 37*(3), 87-107. Schwartz, R., Lederman, N., & Crawford, B. (2004). Developing views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry. *Science Education, 88*, 610–645. Summerlee, A. & Murray, J. (2010). Impact of enquiry-based learning on academic performance and student engagement. *Canadian Journal of Higher Education, 40*, 78-94.

Factor	Brief Summary of Research	Sources
Assessment as Learning: Student engagement during testing	Students' engagement during testing increased when given an opportunity to use a 'cheat sheet' or to consult with peers to clarify problems.	Skidmore, R. L., & Aagaard, L. (2004). The relationship between testing condition and student test scores. *Journal of Instructional Psychology, 31*, 304-313.
Assessment tasks that allow higher order reasoning and 'sense-making'	Tasks that are well designed and support meaningful conjectures contribute to students' motivations to participate and their dispositions toward mathematics.	Mueller, M., & Maher, C. A. (2009). Learning to reason in an informal math after-school program. *Mathematics Education Research Journal*, *21*(3), 7–35. Mueller, M., Yankelewitz, D., & Maher, C. (2010). Promoting student reasoning through careful task design: A comparison of three studies. *International Journal for Studies in Mathematics Education, 3*(1), 135–156. Yankelewitz, D. (2009). *The development of mathematical reasoning in elementary school students' exploration of fraction ideas.* Unpublished doctoral dissertation, Rutgers, The State University of New Jersey, New Brunswick. Yankelewitz, D., Mueller, M., & Maher, C. (2010). Tasks that elicit reasoning: A dual analysis. *Journal of Mathematical Behavior*, *29*, 76–85.
Peer learning (learning together) and feedback	Boud, Cohen and Sampson (1999) define peer learning as 'the use of teaching and learning strategies in which students learn with and from each other without immediate intervention from the teacher' (p. 413). In their study, Chan and Leijten (2012) found that students involved with peer learning were deeply engrossed in examining other student's approaches to the welding tasks and then evaluating the process just undertaken. "The feedback strategies proposed are not difficult to introduce to learners and teachers but lead to improved student engagement, improved student meta-cognition and enhanced skill practice and learning." (p. 23)	Boud, D, Cohen, R. and Sampson, J (1999) Peer learning and assessment, Assessment and Evaluation in Higher Education, 24(4), 413-126. Chan, S., & Leijten, F. (2012). Using feedback strategies to improve peer learning in welding. International Journal of Training Research, 10 (1), 23-29.

Factor	Brief Summary of Research	Sources
Hands-on learning in small groups	Hands on learning experiences in small groups – even with worksheets and textbooks – with high quality tasks increase engagement while lectures and whole class instruction decrease student engagement ((Cooper & Speece, 1990; Greenwood, 1996, 1991; Greenwood, Delquadri & Hall, 1989). Best fitting statistical model shows engagement as a mediating variable between instruction and test scores.	Cooper, D.H., & Speece, D. L. (1990). Instructional correlates of students' academic responses: Comparisons between at risk and control students. *Early Education and Development, 3,* 270-300. Greenwood, C. S. (1996). The case for performance-based instructional models. *School Psychology Quarterly, 11*, 283-296. Greenwood, C. S. (1991). A longitudinal analysis of time to learn, engagement, and academic achievement in urban versus suburban schools. *Exceptional Children, 57,* 521-535. Greenwood, C. R., Delquadri, J., & Hall, R. V. (1989). Longitudinal effects of peer tutoring. *Journal of Educational Psychology, 81*, 371-383.

Factor	Brief Summary of Research	Sources
Timely, usable feedback`	If our interactions with students are to be pedagogically effective, students must engage with them. This is particularly the case with formative assessment feedback which relies for its effectiveness on being applied at some point in the future (Handley & Williams, 2011). Three factors influence students' engagement with feedback. First, students must be able to make sense of the feedback. If it is written in disciplinary language that is not familiar to students or using terms that differ across disciplines, understanding is hampered. Second, if feedback has no bearing on what students do next, they are unlikely to attend to it. Finally, students cannot effectively use feedback if they don't understand the criteria (expectations) for the work. Feedback should help students close the gap between current performance and the desired performance level. Two effective ways to give feedback so that students engage with it are to give feedback on drafts and to give students exemplary work so that they understand what the ultimate goals of instruction area. For feedback to be effective, it has to be used in a structured and meaningful manner. Hattie and Timperley (2007) advocate the use of three forms of feedback: • Feed up – which is to answer the question of whether the learning objectives are being met. • Feedback – providing the learner with the actual performance level on the learning activity. • Feed forward – to the learner on what needs to be done to improve learning or move to the next objective. The feedback cycle requires all three forms of feedback to be present for effective feedback to be completed.	Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessments. *Phi Delta Kappan, 92* (1), 81-90. Black, P., & Wiliam, D. (2009). Developing a theory of formative assessment. *Educational Assessment, Evaluation, and Accountability, 21*, 5-31. Boud, D. 1995. *Enhancing learning through self-assessment.* London: Routledge. Handley, K., & Williams, L. (2011). From copying to learning: Using exemplars to engage students with assessment criteria and feedback. *Assessment & Evaluation in Higher Education, 36*, 95-108. Hattie, J., and Timperley, H (2007) The power of feedback, *Review of Educational Research, 77*(1), 81-112. Sadler, D. 1989. Formative assessment and the design of instructional systems. *Instructional Science 18,* 119–44.

Factor	Brief Summary of Research	Sources
Content that influences Differential Item Functioning	In her chapter on differential item functioning (DIF), Arbuthnot reviews research on math and verbal test DIF and ability. Among studies that examine the items that showed performance differences between Black and white students, one study found that particular item content genres favored Black students over white students. In math, purely quantitative math problems and those involving symbols favored Black students. While on verbal subject tests, reading items that featured humanities, human relations, and Black history-topics favored Black students. In the same chapter, Arbuthnot attempts to answer the question of *why* one group of students performs differently than another. Arbuthnot examines the explanations that are not typically considered. The first is cultural differences. The cultural familiarity hypothesis tests the assumption that culture influences the performance on assessments. Researchers found that Black students performed differentially worse than white students on items involving words typically identified as high frequency . This may be explained by cultural differences because there may be an assumption that all students are exposed to the high frequency words, but cultural differences may lead students to be exposed to different set of words. The second explanation Arbuthnot suggests is the degree to which a test item is engaging and interesting to a student or groups of students. Specifically, the more interested a particular group of students is in the topic of the test item, the more likely that item will differentially favor that particular group (56). Arbuthnot examines student perceptions and behaviors within the testing environment. "Test-wiseness" refers to a student's ability to use their test knowledge to their advantage – for example, knowing test-taking strategies (e.g. most SAT preparation courses give student insight to the structure of the test, how to eliminate wrong answers) to boost one's test score. She found that white students are more likely than Black students to have this test-wiseness. In particular, Black students are more likely to skip questions on the SAT, where white students will guess because there is a penalty of skipped questions and no penalty for guessing.	Arbuthnot, K. (2011). *Filling in the Blanks: Understanding Standardized Testing and the Black-White Achievement Gap,* pp. 50-58. Charlotte, NC: Information Age Publishing.

Factor	Brief Summary of Research	Sources
Text Complexity for ELL's in Math	Martiniello explores the literature on text and lexicon complexity to understand DIF between ELL students and non-ELL student performance. The literature reveals the importance of not having unnecessarily complex words or sentence structure in math word problems to give ELL student test-takers proper accessibility. She chose to analyze six items that had a DIF between ELL and non-ELL students. Part of the analysis uses a think-aloud protocol where ELL Spanish-speaking students deconstruct their process of reading and answering each item. The article includes how students interpreted items differently from how they were intended to be read and how this affected their performance on the item.	Martiniello, M. (2008). Language and the Performance of English-Language Learners in Math Word Problems. In *Harvard Educational Review,* 78:2, 2008.
Cultural Validity	Solano-Flores explains that the way "students interpret test items and respond to them are mediated by cultural factors" and may not relate to the knowledge or skills being assessed. In order to be valid, student scores should only reflect the knowledge being assessed and not other factors. This chapter describes four aspects of cultural validity that help examine testing practices from a cultural perspective: theoretical foundations, population sampling, item views and test review. Solano-Flores and Trumbull (2003) analyzed the different ways students interpreted 1996 a 4th grade NAEP mathematics item. Interviews of students' interpretation of the question of three student subgroups (high SES suburban white, low SES inner city African American, and low SES rural Native American) revealed great variation. 84% of white students read the question as intended, whereas only 56% and 52%, respectively, of Native American and African American students read the sentence as intended. Solano-Flores and Trumbull also found that 10% and 18%, respectively, of the Native American and African American students interpreted the word *only* as restricting the number of dollars ("His mother has only one dollar") however, this interpretation was not observed by white students in the study.	Solano-Flores, G. The Cultural Validity of Assessment Practices. In Basterra, M. del R., Trumbull, E. & Solano-Flores, G. (Eds.). (2011). *Cultural Validity in Assessment: Addressing Linguistic and Cultural Diversity,* pp.3-21. New York: Routledge. Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher*, *32*(2), 3-13.

Factor	Brief Summary of Research	Sources
Equity and themes	The content of the themes that an assessment item touches on can make a difference in a student's performance. When a student is more engaged in a reading passage, he or she is more likely to comprehend it than a passage of similar complexity that he or she finds less engaging. Not only is comprehension easier, but students are more likely to persist in the face of difficulty when engaged. Similarly, writing prompts that are engaging generate more elaborated and developed writing than others. While engagement can boost the ability of all students to demonstrate what they know on an assessment, engagement is particularly important for the students least well-served in schools. More affluent students often find it easier to compartmentalize and to carry out academic tasks related to themes that they find dry and trivial. Some of the racial and cultural groups that currently do the worst in school also have a cultural orientation to learning through tasks that are meaningful and authentic and have difficulty with pursuing tasks on themes that seem irrelevant. The more barriers a student faces, the more the student needs and benefits from the boost of finding the themes that an assessment item addresses to be interesting and meaningful. Not all student populations will be equally engaged by all themes. When thinking about the issue of engagement in large-scale assessments, it is important to find themes that would not compound disadvantage Themes should either be: a.) engaging to all student subgroups; or b.) engaging to groups least well-served in schools while not hurting the performance of other student populations. The issue of engaging themes is relevant to four types of items that will be included in the English language arts and math assessments aligned to the common core standards: 1.) reading passages, 2.) writing prompts, 3.) projects and tasks involving research, and 4.) math word problems.	Darling-Hammond, L., Barron, B., Pearson, D. P., et al. (2008). P*owerful Learning: What We Know About Teaching for Understanding,* pp. 74-76. San Francisco: Jossey-Bass.

Factor	Brief Summary of Research	Sources
Linguistic and Cultural Strengths	Meier establishes that in the literature about reading comprehension, understanding is facilitated by a connection that a student has with their own experiences; this could be a text-to-self-connection, a text-to-text connection, or a text-to-world connection. In this section, Meier provides teachers with strategies to help students make these connections so that students can then draw their own connections when they read materials on their own in other settings. Meier makes a point about the content of reading materials in classrooms: it is important for available reading books in a classroom have content that reflect the student demographics in that class. It is also the duty for a teacher to expose their students to literary texts that feature many cultural backgrounds. In this section, Meier models themes from African American literature for elementary school-aged children and outlines the ways that these cultural books have underlying themes that speak to most children's experiences. Meier reviews studies and literature that illustrate the cultural aspects of being raised in predominately African American communities and the linguistic skills that are developed from an early age. This review highlights vivid examples of non-academic contexts that are where young African American children learn how to use cognitively complex language behavior like building arguments, taking context into account, understanding multiple meanings of words, and using metaphorical and figurative language. These are the examples that are particularly applicable to reading cognitively complex texts.	Meier, T. (2008). *Black Communications and Learning to Read: Building on Children's Linguistic and Cultural Strengths,* pp. 119-121. New York: Lawrence Erlbaum.

Factor	Brief Summary of Research	Sources
Interpreting Student Performance	Smitherman presents her findings of a study where she analyzes 17-year old African American student NAEP writing samples and the scores they received between 1969 and 1989. The purpose of the analysis was to test a hypothesis of a decrease in the use of Black language in student writing over time and understand how the use of Black language in student writing affected test scores. The study found the use of Black language had decreased in a 10 year period, but only in narrative essays. The study also found that between 1969 and 1979, student writing that included Black language was more likely to receive a higher score with the primary trait scoring rubric (a rubric that analytically scored writing with minimal weights based on grammar and syntax). In 1984 and 1988, student writing that included Black language was more likely to receive a lower score when according to a holistic rubric (a rubric that includes an assessment of grammar, mechanics, and syntax without specified weights for each element).	Smitherman, Geneva. (2000). African American Student Writers in the NAEP 1969-88/89[1992] and 'The Blacker the Berry, the Sweeter the Juice' [1994]. In *Talkin that Talk: Language, Culture and Education in African America,* pp. 163-191. New York: Routledge Press.
Interpreting student writing responses	Kathy Escamilla and María Coady find that understanding writing conventions and pattern of ELL students' native language can be used in the assessment of writing samples written in English to fully understand student mastery of writing and ELA standards. Additionally, if ELL student writing is assessed in both English and their native language, or through "contrastive analysis," the evaluator can analyze the writing skills that a student has mastered that cannot yet be expressed in English because of the student's level of language acquisition. The article provides samples of student writing and the approach to this method for writing assessment.	Escamilla, K. & Coady, M. (2001). Assessing the writing of Spanish speaking students: issues and suggestions. In J. Tinajero and S. Hurley (Eds.). *Handbook for Literacy Assessment for Bilingual Learners*. Boston: Allyn & Bacon.

Factor	Brief Summary of Research	Sources
Interpreting ELL Student Performance in Math	Moschkovich examines third grade English-language learner group problem-solving discourse during a math class. Her data, recorded conversations among students and their teacher, reveals the verbalization of student math thought processes and reasoning patterns. She finds that multiple meanings exist behind reasoning expressions. Moschkovich illustrates the way that recognizing these multiple meanings helped a teacher bridge student understanding to formal vocabulary and concepts. This recognition also helped the teacher not dismiss student understanding and grasp of math concepts.	Moschkovich, J. (2008). "I Went by Twos, He Went by One": Multiple Interpretations of Inscriptions as Resources for Mathematical Discussion. *Journal of the Learning Sciences*, 17:4, 551-587.