



Invitational Research Symposium on
Technology Enhanced Assessments

Technology Enhanced Assessments in Mathematics and Beyond: Strengths, Challenges, and Future Directions

Gabrielle A. Cayton-Hodges, Elizabeth Marquez,
Madeleine Keehner, Cara Laitusis, Peter van Rijn,
Diego Zapata-Rivera, Malcolm I. Bauer, and Mark T. Hakkinen

May 7–8, 2012



Technology Enhanced Assessments in Mathematics and Beyond: Strengths, Challenges, and Future Directions

Gabrielle A. Cayton-Hodges, Elizabeth Marquez, Madeleine Keehner, Cara Laitusis, Peter van Rijn, Diego Zapata-Rivera, Malcolm I. Bauer, and Mark T. Hakkinen

Educational Testing Service

Executive Summary

In this paper, we focus on a technology enhanced prototype formative assessment task from the mathematics stream of the ETS research initiative Cognitively Based Assessment of, for, and as Learning (CBAL™; Bennett & Gitomer, 2009). The task, called *Proportional Punch*, is part of a family of assessments designed to capture evidence of students' understanding of, and skills in, proportional reasoning.

In mathematics, technology enhanced assessments (TEAs) allow us to evaluate aspects of student understanding that are not easily observable in paper-based assessments. One broad aim of CBAL assessments in this domain is to capture evidence of students' mathematical *thinking*. In this case, a key component of the task is an interactive simulation, the sweetness meter, which allows students to model the results of adding different ratios of cherry punch mix and water to a pitcher. The task also involves scaffolding of student understanding through multiple sequential subitems, as well as a range of typical and innovative item types. Together, these features allow us to capture evidence of students' thinking about proportions at different levels of sophistication, so that we can make fine-grained distinctions among student performances at subtly differing levels.

In this paper we provide an account of the development process for this task and for CBAL mathematics items in general. Here we explain the central role of insights from the cognitive and learning sciences in building theoretical models of mathematical competence and learning progressions for each component knowledge set (in this case, the learning progression for proportional reasoning, which aims to characterize how understanding of ratio and proportion develops through the middle school grade levels). We describe how these insights about student knowledge and understanding, in

conjunction with an evidence-centered design (ECD) approach to assessment development, feed into the design of complex, scenario-based tasks, as well as score reports and professional support materials for teachers.

Following this introduction to the task design process, we consider implications for instruction. Here we draw on our experiences of piloting CBAL mathematics tasks in schools, and we describe some of the valuable insights gained from working with teachers and students in situ. We also discuss some of the professional support materials that accompany formative CBAL mathematics tasks like this one, including teacher handbooks and innovative real-time reporting systems.

In the next section, on measurement and scoring, we discuss some of the challenges and opportunities inherent in TEAs. We consider the important but challenging question of how to characterize interactivity data and how to make use of the information it can potentially offer, including how these new data types might be used for building a validity argument as well as improving measurement precision and providing new kinds of evidence about student competency. We discuss implications for measurement models of these novel data types.

Next, we consider one of the most widely acknowledged strengths of TEAs: their ability to enhance engagement and motivation. Although the ability to engender high levels of student engagement is considered a central property of TEAs, it is important to consider how this plays off against other characteristics. To this end, we discuss the important issue of balancing engagement with cognitive load and with validity needs. We also consider another important strength of TEAs: their potential for capturing constructs that can otherwise be hard to measure.

Finally, we turn to the challenge of making TEAs accessible to all students. Here we focus particularly on mathematics assessments and the unique challenges they present. We propose a three-level framework for ensuring accessibility in these kinds of items, and we describe some of the prior and current research being carried out at ETS to explore the implications of technology enhancements for a range of student populations. We also present some early-stage work with accessible technologies that we believe have the potential to ameliorate some of the accessibility challenges of these assessment forms.

Finally, we end by identifying a set of key future research questions that touch on all of the themes in this paper. The varied nature of the many challenges and opportunities presented by TEAs makes clear the importance of taking a multi-disciplinary approach to research and development going forward. We view these as important issues to address for the future of TEAs.

As Bennett (1999) has observed, new technologies are changing many aspects of assessment, from overhead logistical and operational matters such as testing location and presentation platforms, to the critical minutiae of item design, evidence capture, and scoring methods. In the present paper, we are mostly focused on the latter types of issues, as we consider some of the exciting, yet at the same time challenging, implications of the shift towards technology enhanced assessment (TEA; cf. Baker, 2003).

TEAs and technology enhanced items (TEIs) are computerized test forms that typically incorporate properties such as digital media and multimodal representations of information; interactive components such as student-operated navigational tools, tutorials, and simulations; and the capability to capture digitally the student's test responses and other task-relevant interactive behaviors. The dynamic nature of these systems makes it possible to build extended, scenario-based assessments, which may involve multiple embedded items, sometimes capturing subtly differing levels of performance in a sequential set of steps that progress from one level to the next, and sometimes providing scaffolding to support student understanding or compensate for missing knowledge or errors of reasoning. Because they are interactive and process-oriented, they can also be designed explicitly to capture evidence of students' online *skills* as well as, or instead of, long-term *knowledge* (McFarlane, 2003). Many of the assessments currently in development or in use capitalize on this very rich potential (although some computerized discrete tasks also exist, which are essentially screen-based analogs of traditional test items).

There is a general recognition in the research literature that TEAs have the potential to shift the paradigm, not only for *how* we assess, but also for *what* we assess. As McFarlane (2003) has noted,

The advent of digital technologies has led to a shift in emphasis, or even a complete change, in the skill set that will serve pupils well when they enter the world of work. Key learning outcomes are seen as skills, not knowledge, and require a dynamic environment in which to capture the practice of these skills in action. (p. 261)

One key advantage of TEAs over traditional assessment media is their ability to offer a dynamic environment (McFarlane, 2003). This permits capture of new kinds of evidence, especially information about the *processes* students use to complete a task, and not just the final *outcome* of their response. However, this opportunity also presents challenges—although it is possible in theory to capture every observable action in the “click stream” of interactive behaviors, in order to obtain valid evidence of a student's competencies and cognitions, one needs to establish which of those actions are cognitively informative. Given the richness of the information that can be gathered, if the assessment has a formative aspect, one also needs to understand which aspects of the data provide the greatest educational value and how best to report that information to achieve this goal.

Dynamic systems also allow educators to be responsive to student actions as they occur. This is the foundational premise of computerized adaptive testing (CAT), in which student responses are analyzed automatically, on the fly, and the selection of subsequent test items is determined by their performance, with critical choice-points built into the system (see, for example, Sands, Waters, & McBride, 1997; van der Linden & Glas, 2000; Wainer, 2000). This flexibility also allows feedback to be provided to the student during formative assessments, according to their performance, but ensuring feedback is offered in effective ways requires substantial research efforts.

The aim of the present paper is to discuss some of the opportunities and challenges that arise from TEAs. We approach these issues from the perspective of a current ETS research initiative that has TEAs at its center. For the present paper, we take as our focal task an interactive, technology enhanced mathematics task designed to measure proportional reasoning skills. Where relevant to illustrate specific points we also refer to other TEAs.

The Focus of this Paper: A Technology Enhanced Assessment for Proportional Reasoning in CBAL Mathematics

In the domain of mathematics, TEAs allow educators to evaluate aspects of student understanding that are not readily observable in paper-based assessments. Technology can be incorporated, by design, into the development of online scenario-based tasks to enhance the constructs being assessed. For example, computer simulations can emulate the properties of real-world situations that map onto key mathematical constructs, allowing one to tap those constructs in the context of a rich scenario with authentic qualities. Complex tasks that incorporate multiple subitems can allow fine-grained distinctions among student performances at subtly differing levels. And interactive systems can allow one to study process as well as outcome, so that one can see how students engage with and make use of interactive tools for reasoning.

In the present paper, we focus on a technology enhanced task from the mathematics stream of the ETS research initiative Cognitively Based Assessment *of, for, and as Learning* (CBAL™; see Bennett & Gitomer, 2009). CBAL is described as “an innovative K–12 assessment system that documents what students have achieved (*of learning*); facilitates instructional planning (*for learning*); and is considered by students and teachers to be a worthwhile educational experience in and of itself (*as learning*)” (Bennett, 2010, pp. 71–72).

The CBAL theory of action incorporates four key components. The first element is composed of domain-specific *competency models*, which describe the capabilities that underpin competence in a given discipline. These include hypothesized *learning progressions* (the qualitative changes that occur in a learner’s level of sophistication for key concepts, processes, strategies, practices, or habits of mind). The competency models and learning progressions form the theoretical foundation of the CBAL approach. The remaining three key components are summative assessments, formative assessments, and professional support materials for teachers. Bennett (2010) argued that these four components can work together via a set of hypothesized action mechanisms to produce effects such as improved teacher understanding of the content domain and improved pedagogical skills generally, greater classroom focus on integrated and higher-order skills, routine use of formative assessment to inform instruction, and improved student engagement. Ultimately, it is envisaged that the enhanced information available to educational policy makers through this approach will permit more effective decisions and therefore better learning. In the present paper, we describe a TEA from the CBAL mathematics program.

Our focal TEA for the present paper is the CBAL formative task *Proportional Punch*. The primary assessment goals of this TEA are the Common Core State Standards (CCSS) content clusters *Understand ratio concepts and use ratio reasoning to solve problems* (Grade 6) and *Analyze proportional relationships and use them to solve real world and mathematical problems* (Grade 7). We selected this task for two reasons. First, the properties of this task enable illustration of many of the key strengths

and considerations of TEAs. Second, we view this area of mathematics as important to measure effectively—the family of skills and knowledge directly involved in and more broadly related to proportional reasoning is critical in middle-school mathematics, and mastery of this area has far-reaching implications for students’ readiness for college and careers.

In the sections that follow, we introduce the CBAL task design process and provide a detailed rationale for the task’s structure and purpose built on the CBAL philosophy, with special emphasis on the role that technology plays in relation to these. We look in detail at examples of several questions within this task to illustrate how student responses to those questions can begin to build evidence of understanding. We also discuss how that evidence relates to the CBAL-defined learning progression for proportional reasoning, which was used as a partial basis for the development of the task. In addition, we examine more generally questions of how the use of technology can help to provide a fuller and more engaging assessment experience, how sophisticated data capture and analysis methods can expand the notions of what it means to seek validity in different types of evidence, and how extended, componential, scenario-based tasks can broaden access to the assessment of key concepts for some students. Finally, we discuss some of the challenges for universal design that arise from TEAs and take a look at how ETS is attempting to address these challenges. We end with a call for future research on key questions raised by these still relatively new technology based approaches to assessment.

Technology Enhanced Task Development: An Illustration From CBAL Mathematics

The CBAL task design process begins with three major parts: (a) the content area to be assessed, (b) the CBAL model of mathematical competency, and (c) the library of learning progressions created or adapted for the CBAL project. In the case of our focal task, our aim was to create an assessment around proportional reasoning, which has two subcomponents, rates and ratios.

Once the content area is identified, the CBAL competency model¹ is examined, and the relevant competencies are identified (see Figure 1). Of course, no single task aims to measure all aspects of skills and knowledge for a given content area; rather, each CBAL task is situated within a family of tasks, and it is the total combination of tasks administered in the assessment that ensures full coverage of the range of capabilities to be measured.

Given the structure of the competency model (see Figure 1), we could approach the development of a proportional reasoning task in one of two possible ways: either through the *Represent* strand or through the *Argue* strand. For *Proportional Punch*, we chose to highlight *Argue*. In other words, in this task students have to justify and reason about proportions, but they would not be developing proportional representations (*Represent* would be assessed in another task within the assessment). Within *Argue* we focused on two subcompetencies: (a) *Use qualitative reasoning* and (b) *Justify*.

¹ For a detailed description of the CBAL competency model and its creation, see Graf (2009); see also Graf, Harris, Marquez, Fife, and Redman (2009).

Invitational Research Symposium on
Technology Enhanced Assessments



© 2009 Educational Testing Service. All Rights Reserved. 12/31/2009

Figure 1. CBAL middle-school mathematics competency model. Relevant nodes are outlined in red.

In the next step, we draw on the relevant CBAL learning progressions. Learning progressions are defined in several ways throughout various literatures. In CBAL, a learning progression is defined as “a description of qualitative change in a student’s level of sophistication for a key concept, process, strategy, practice, or habit of mind” (ETS, 2011, para. 1), with change potentially occurring due to a variety of factors including maturation and instruction, and each progression presumed to hold for most, but not all, students. As with all scientific research, the progressions are open to empirical verification and theoretical challenge (for more information, see ETS, 2011).

Proportional reasoning has been examined extensively in the literature on mathematical cognition (e.g., Noelling, 1980a; 1980b). Baxter and Junker (2001; cited in Weaver & Junker, 2004) conducted a comprehensive review of the research in this field and identified a series of developmental stages, shown in Table 1, which can be seen in students’ responses to problems involving these skills.

Baxter and Junker’s (2001; cited in Weaver & Junker, 2004) hypothesized learning progression identifies five levels of understanding in proportional reasoning: (a) intuitive, or qualitative, understanding; (b) early attempts to quantify; (c) recognizing multiplicative relationships; (d) accommodating covariance and invariance; and (e) functional and scalar relationships (see Table 1 for detailed descriptions). In particular, identified as key was “a shift from additive to multiplicative reasoning and from use of context-specific strategies to a generalized understanding of functions” (Baxter & Junker, 2001, p. 12).

Table 1. Development of Proportional Reasoning

Levels	Description
1. Qualitative	Young students generally possess a good deal of knowledge about quantity that permits them to answer questions about fairness or questions about more and less.
2. Early attempts at quantifying	Early attempts at quantifying often involve constant additive differences rather than multiplicative relationships. Students still rely on counting up or down.
Level 3. Recognition of multiplicative relationships	Students have the intuition that the differences change with the size of the numbers and that the change may be multiplicative in nature, but they do not necessarily realize that they need to consider the constantly increasing/decreasing difference between the related terms of each pair, that is, of each ratio. They rely on pattern matching or buildup strategies that are sometimes additive in nature. Problems with unit factors or sharing are easily solved, but others may lead to additive solution strategies.
Level 4. Accommodating covariance and invariance	Students begin to develop a change model that recognizes that while some quantities may be changing, relationships among the quantities remain invariant. However, they rely on coordinated buildup, unit factor, and other scalar approaches. Often strategies are context specific and not generalizable. Students may revert to additive strategies in unfamiliar or challenging situations.
Level 5. Functional and scalar relationships	Students recognize the invariant nature of relationships between pairs of changing quantities. They have a generalized model for solving proportionality problems and understand why it works. This is not to say that they use the same strategy in every context; rather they have a repertoire of strategies and use the most efficient for a given situation.

Note. Adapted from Baxter and Junker (2001), as cited in Weaver and Junker (2004), p. 8.

In the CBAL initiative, we use learning progressions in task development to examine the developmental stages at which students may be reasoning, based on their strategies, and we build on that information, in the context of the learning progressions, to determine next steps in instruction. Our focal task, *Proportional Punch*, is aimed at middle-school students, so its primary goal is to discriminate between Level 2 and 3 learning progressions. But the task's relatively rich design also allows us to identify characteristics of Levels 1 and 4 as well (Level 5 is quite advanced for middle-school students and thus the task need not reach for a discrimination of Level 5).

Once we have identified the relevant content and portions of the competency model and learning progression, we are ready to list the types of observable evidence that would allow us to make inferences about the underlying competencies. Types of evidence include interpreting a table of values, comparing two objects, perceiving patterns, communicating effectively, distinguishing between ratios and quantities, calculating with accuracy, and evaluating statements.

Using the content, competency model, learning progression, and evidence list, an initial prototype task is developed and iteratively refined. Once the prototype has stabilized, a task model is created. The task model captures the broad properties that are present in the prototype and which therefore should be represented in any equivalent or parallel tasks. The task model also helps to articulate how the content, competencies, and evidence are expected to interact in tasks generated from this framework. In *Proportional Punch*, the task model shows that at least one item in the task will have students comparing two objects (evidence) that differ in their proportions (competency) and that those differences will need to be quantified in order to be observable (learning progression Levels 1–2 distinction). The task model has several goals: (a) to map the evidence, competencies, and learning progression levels to be linked to ensure that all of the intended areas are targeted; (b) to determine which parts of the competency model, evidence list, or learning progression are not covered and will need to be targeted by other tasks in that family of assessments; and to fit new scenarios into the same task model to create multiple tasks, or *parallel forms*, which should have roughly equivalent assessment capabilities.

Finally, individual tasks are created by taking the task model and applying a real-world scenario and problem—in the case of our focal task, adding punch mix to water to make cherry punch. In *Proportional Punch*, the ratio between the mix and the water indicates how sweet the punch will be. An interactive tool that simulates making punch and determining its sweetness is available throughout the task (see Figure 2). Students are encouraged to try out the simulation themselves to become comfortable with it before the task begins. They are also given an additional constraint, which states that no more than two scoops of mix can dissolve in each cup of water, so if they add extra mix, it will settle on the bottom of the pitcher and the sweetness meter will stay at 2:1.

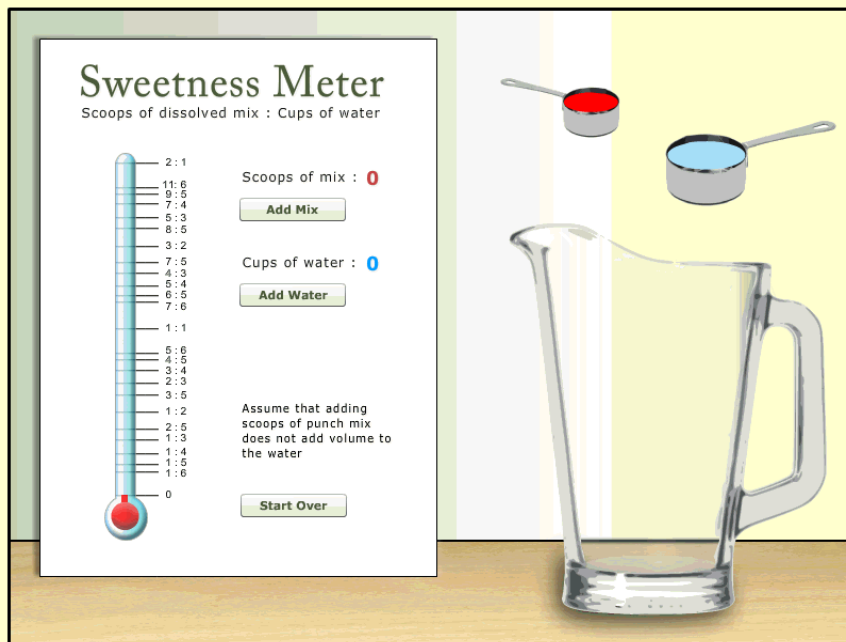



Figure 2. Sweetness-meter simulation.

Items in this task begin with relatively simple comparisons, such as one scoop of mix with three cups of water, versus two scoops of mix to three cups of water. These kinds of comparisons can be solved by applying intuitive (qualitative only) proportional reasoning skills, of the kind associated with students at Level 1 of the learning progression. The comparisons slowly become more challenging, so that students can no longer differentiate the sweetness by relying on intuitive reasoning alone but must begin quantifying the relationships (Level 2) and eventually reasoning multiplicatively (Level 3) and working with covariation (Level 4). Students are asked to reveal their strategies as they work through the task to help to discriminate between the levels of the proportional reasoning learning progression. Table 2 shows the Baxter and Junker (2001; cited in Weaver & Junker, 2004) proportional reasoning learning progression, but here the levels are illustrated with sample items and student responses from *Proportional Punch*.

Figure 3 provides an overview of the entire task design process. The next section covers how technology becomes incorporated into the design of a task and the implications of this for instruction.

Table 2. Proportional Reasoning Learning Progression With Examples

Description	Task items	Sample student responses
<p>Qualitative Young students generally possess a good deal of knowledge about quantity that permits them to answer questions about fairness or questions about more and less.</p>	 <p>If 1 scoop of punch mix is added, will the punch be sweeter? Yes/No Why?</p>	<p><i>If you add more mix, it will just get sweeter, and you aren't adding any more water.</i></p>
<p>Early attempts at quantifying Early attempts at quantifying often involve constant additive differences rather than multiplicative relationships. Students still rely on counting up or down.</p>	<p>Beth's punch recipe uses 5 scoops of mix for every 4 cups of water. Lula's recipe uses 3 scoops of mix for 4 cups of water. Which recipe makes the sweeter punch? Why?</p>	<p><i>Beth's recipe will be sweeter punch because there are more scoops of mix and the same amount of water as Lula's,</i> <i>BBBBB – Beth's punch mix</i> <i>BBBB – Beth's water</i> <i>LLL – punch mix – Lula's</i> <i>LLLL – Lula's water – more water than punch</i></p>
<p>Recognition of multiplicative relationships Students have the intuition that the differences change with the size of the numbers and that the change may be multiplicative in nature, but they do not necessarily realize that they need to consider the constantly increasing/decreasing difference between the related terms of each pair, that is, of each ratio. They rely on pattern matching or buildup strategies that are sometimes additive in nature. Problems with unit factors or sharing are easily solved, but others may lead to additive solution strategies.</p>	<p>Dean's punch recipe uses 4 scoops of mix and 5 cups of water. How many scoops of mix would Dean use for 50 cups of punch? Show your work.</p>	<p>40 $\frac{4}{5} \frac{8}{10} \frac{12}{15} \frac{16}{20} \frac{20}{25} \frac{24}{30} \frac{28}{35} \frac{32}{40} \frac{36}{45} \frac{40}{50}$ —</p>
<p>Accommodating covariance and invariance Students begin to develop a change model that recognizes that while some quantities may be changing, relationships among the quantities remain invariant. However, they rely on coordinated buildup, unit factor, and other scalar approaches. Often strategies are context specific and not generalizable. Students may revert to additive strategies in unfamiliar or challenging situations.</p>	<p>There are 3 students at each table that has 2 candy bars on it. The teacher wants to push 4 tables together. Would you rather get your share of candy before or after the tables are pushed together?</p>	<p><i>It doesn't matter because I will get 2/3 of a candy bar each way.</i></p> <p><i>2 bars 3 people</i> <i>8 bars 12 people</i></p>
<p>Functional and scalar relationships Students recognize the invariant nature of relationships between pairs of changing quantities. They have a generalized model for solving proportionality problems and understand why it works. This is not to say that they use the same strategy in every context; rather they have a repertoire of strategies and use the most efficient for a given situation.</p>	<p>Malcolm's recipe for making punch uses a ratio of 3 scoops of punch mix to 5 cups of water. How many scoops are needed to fill a 150 cup jug using Malcolm's recipe?</p>	<p>$\frac{x}{150} = \frac{3}{5}$ $5x = 450$ $x = 90$ works because $\frac{5x}{750} = \frac{450}{750}$</p>

Invitational Research Symposium on
Technology Enhanced Assessments

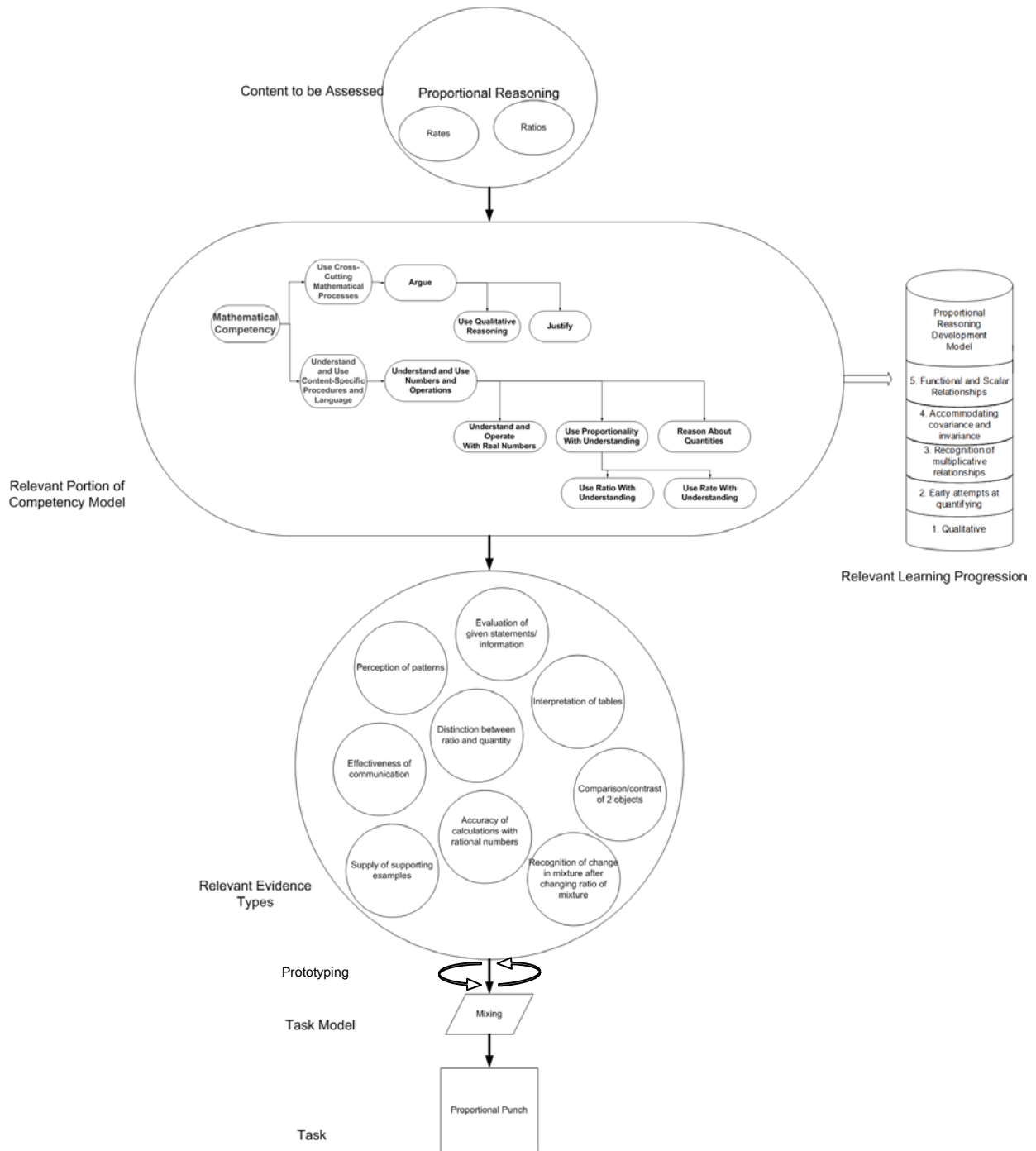


Figure 3. CBAL mathematics task design process.

Technology Integration and Implications for Instruction

Technology enhanced formative assessments allow one to assess, and drive instruction in, conceptual understanding and reasoning processes more effectively and authentically than traditional assessments. Such assessments can also simultaneously support and enhance instruction in the CCSS content and practice standards. The CBAL *Proportional Punch* task uses technology to assess a fundamental mathematical concept that is integral to the understanding of all mathematics. Moreover, as with all CBAL tasks, the integration of technology engages students in learning by allowing them to explore and express themselves as the digital natives they are. In this section of the present paper covers how the embedded technology within this task not only provides an accurate assessment of the student's level of understanding, but also allows one to identify opportunities for instructional intervention, founded on the hypothesized learning progressions and the level of attainment that an individual student has reached.

The focal task assesses the *Use Proportionality With Understanding* competency (see Figure 1) in the context of making punch and discriminates between levels of the proportional reasoning learning progression (see Table 2). Having a suggested level of performance can inform more focused and effective instructional interventions, and to that end, the task is presented in five sections. The task begins with students becoming acquainted with a computer simulation and answering qualitative questions about the relative sweetness (or strength) of different mixtures. Next, students answer simple quantitative questions involving specific ratios of scoops of mix to cups of water, and then students answer more difficult quantitative questions involving more complicated ratios. Finally, the task culminates in a generalization and an application of that generalization. In each section, students are asked to reveal their strategies so that interventions can be made before they move to the next section of the task, which targets a higher level of the learning progression for proportional reasoning.

The first section of the task is intended to familiarize students with a computer simulation composed of a punch-mixing tool and a sweetness meter that registers the sweetness of punch based on the ratio of scoops of dissolved mix to cups of water (see Figure 2). The meter allows students to discover equivalent ratios in an engaging and efficient way and is available to students throughout the task. However, the task is carefully scaffolded, so that students progress through increasingly challenging items and eventually tackle problems for which the simulation cannot be used, such as ratios composed of numbers that do not appear on the sweetness meter scale.

Figure 4 shows a question from Section 1, in which students practice using the simulation without an explicit identification of equivalent fractions.

Section 2 of the task assesses how well students understand proportional reasoning at the qualitative or intuitive level (Level 1 of the proportional reasoning learning progression). The literature cited earlier indicates that a strong *intuitive* understanding of proportion is essential to a *quantitative* understanding of proportion. Figure 5 shows a question from Section 2 in which students add an extra cup of water to see if the sweetness meter goes up or down. A response that indicates a student used the meter to help find the answer may mean that he or she lacks some relevant understanding.

Section 3 of this task moves from the qualitative level of the learning progression to early attempts at quantifying (Level 2). Numbers now appear in context and students are directed to answer in terms of numbers. Figure 6 shows a question from Section 3 that elicits evidence of rudimentary quantifying because there is no extension of pattern.

Proportional Punch Question # 1 Timer

CBAL MATH

Use the Mix Tool to help answer the questions.

Sweetness Meter
Scoops of dissolved mix : Cups of water

Answer the following questions in terms of the number of scoops of punch mix and cups of water.

- When does the meter read 0 ?
(Give supporting examples.)
- When does the meter read 1 : 1 ?
(Give supporting examples.)
- Give three specific examples that result in a meter reading of 1 : 2.

Figure 4. Proportional Punch Question 1.

Proportional Punch Section 2 Question # 2 Timer

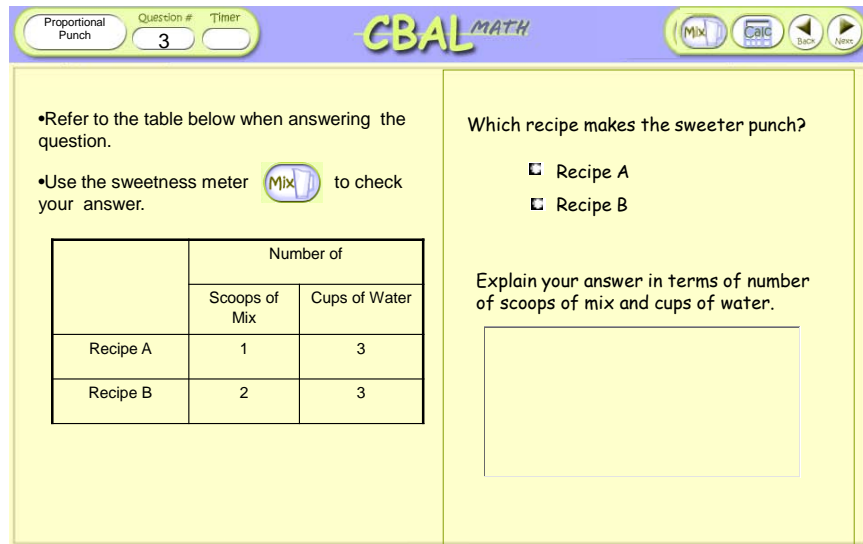
CBAL MATH

One cup of water is added to the punch in the pitcher, and no mix is added. The punch will be

- sweeter than it was.
- not as sweet as it was.
- just as sweet as it was.

Explain your answer.


Figure 5. Proportional Punch Question 2.



Proportional Punch Question # 3 Timer

CBAL MATH

•Refer to the table below when answering the question.

•Use the sweetness meter  to check your answer.

	Number of	
	Scoops of Mix	Cups of Water
Recipe A	1	3
Recipe B	2	3

Which recipe makes the sweeter punch?

Recipe A

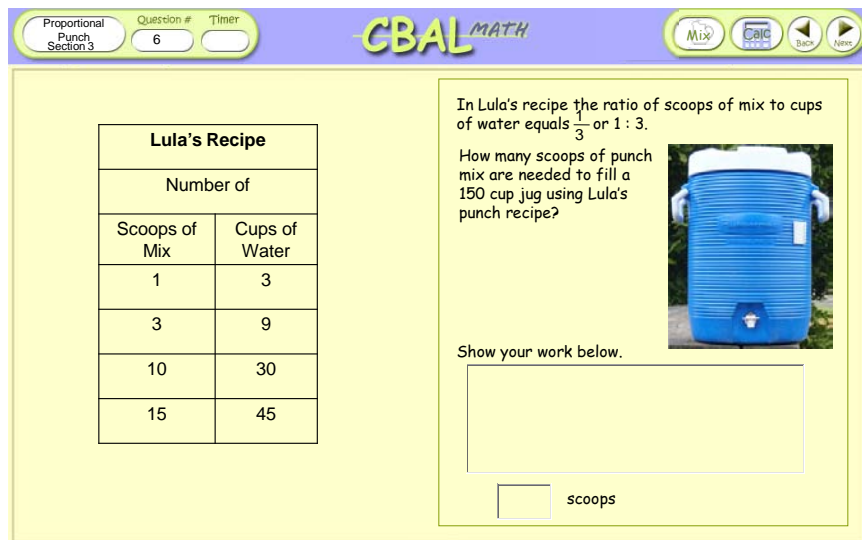
Recipe B

Explain your answer in terms of number of scoops of mix and cups of water.

Figure 6. Proportional Punch Question 3.

In this item in Section 3, students might correctly say that Recipe B makes sweeter punch, but it is important to examine their explanations. For example, students might say that A is sweeter because the difference between 2 and 3 is only 1, and the ‘smaller’ difference between numerator and denominator means a bigger fraction. That response could indicate a common misconception (a version of additive reasoning) or it might indicate students are still at Level 1 of the learning progression. In either case, an intervention is necessary. That intervention may simply be having students use the sweetness meter to check their belief that the smaller difference necessarily means a bigger fraction (e.g., it works for $\frac{2}{3}$ and $\frac{1}{3}$ but what about $\frac{3}{5}$ and $\frac{1}{2}$?).

In Section 4 of the task, student performance is assessed at Level 3 of the learning progression—recognition of the invariant nature of relationships between pairs of changing quantities and ability to use a generalized model based on multiplicative relationships. Figure 7 shows a question from Section Four that asks students to use a 1:3 ratio in order to make enough punch to fill a large jug.




Proportional Punch Section 3 Question # 6 Timer

CBAL MATH Mix Calc Back Next

Lula's Recipe	
Number of	
Scoops of Mix	Cups of Water
1	3
3	9
10	30
15	45

In Lula's recipe the ratio of scoops of mix to cups of water equals $\frac{1}{3}$ or 1 : 3.

How many scoops of punch mix are needed to fill a 150 cup jug using Lula's punch recipe?



Show your work below.

scoops

Figure 7. Proportional Punch Question 6.

Note that students who are still functioning at Level 2 of the learning progression may simply extend the table, rather than using a model that equates ratios and recognizes multiplicative relationships. It is possible that students may use $\frac{a}{b} = \frac{c}{d}$, in which case it might appear that students are at Level 5 of the proportional reasoning learning progression because they have a generalized model for solving the problem at hand. However, unless we have evidence showing that students understand the model works because they are finding numerators of equivalent fractions (i.e., $\frac{a}{b} = \frac{c}{d} \rightarrow \frac{ab}{bd} = \frac{bc}{bd}$), we cannot assert that they are truly at Level 5 of the learning progression. This important distinction is made explicit in the following excerpt from the *Proportional Reasoning Teacher Handbook*:

What to emphasize to students:

There are generalized methods for solving proportionality problems all of which are based on equivalent ratios. These generalized methods are efficient whereas extending a table is tedious. Demonstrating extension of table vs. solving a proportion may convince students of the more efficient method. Let the power of a model be revealed! Students often cross multiply to get a solution without knowing why it works so be sure to emphasize and remind them of the fact that when they cross multiply they are finding the numerators for fractions with common

denominators—e.g., if students solve the problem by cross multiplying, they should understand

$$\frac{x}{150} = \frac{3}{5}$$

that $5x = 450$

$$x = 90$$

works because $\frac{5x}{750} = \frac{450}{750}$

Anecdotally, we have found that when some middle-school teachers encounter this point, their reaction is akin to “Oh, *that’s* why it works!” Such responses suggest that the teacher handbooks are helping to improve teachers’ content knowledge, as well as improving instructional techniques:

In Section 5 of *Proportional Punch*, the objective is to see if the scaffolded questions leading to the final questions result in independence from the simulation and the ability to accommodate covariance and invariance, which would indicate that students are at Level 4 of the proportional reasoning learning progression. The penultimate question (Figure 8) asks students to generalize comparisons of sweetness, and the final question (Figure 9) asks students to demonstrate that their generalization works with numbers that do not lend themselves to the simulation.

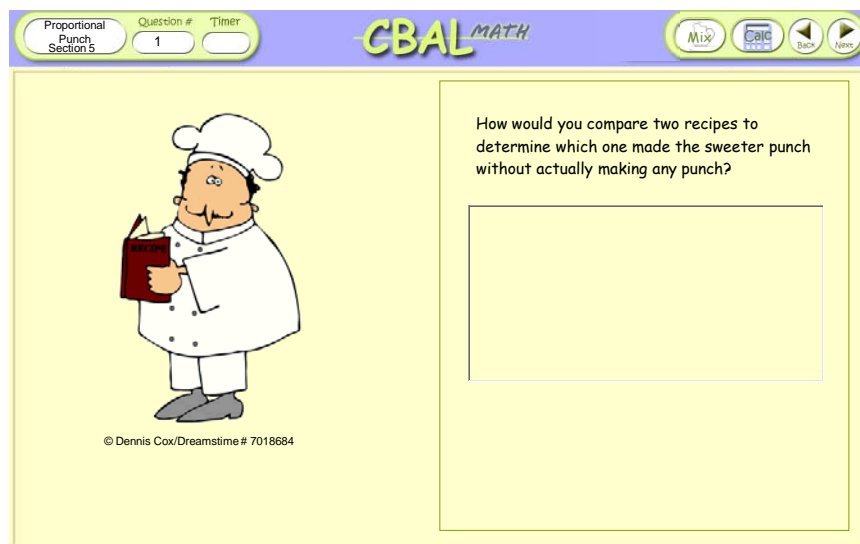
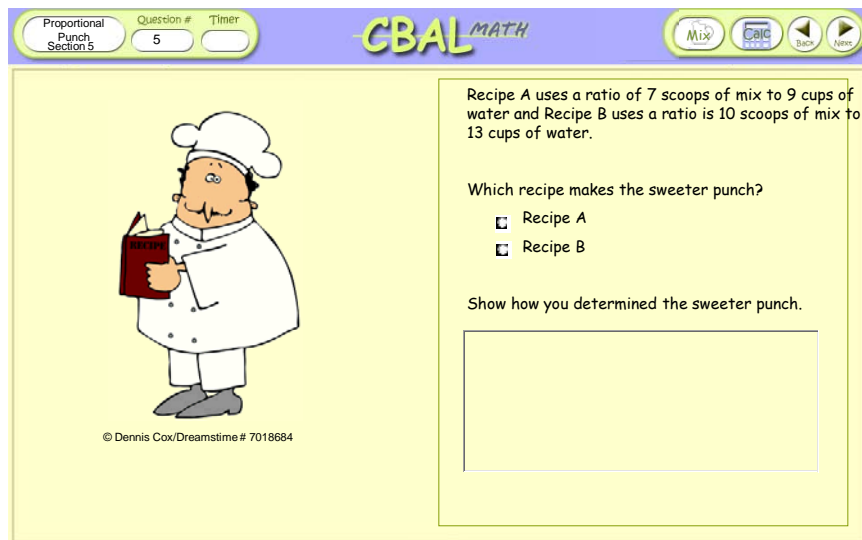


Figure 8. *Proportional Punch*, Section 5.



Proportional Punch Section 5 Question # 5 Timer

CBAL MATH Mix Calc Back Next

Recipe A uses a ratio of 7 scoops of mix to 9 cups of water and Recipe B uses a ratio is 10 scoops of mix to 13 cups of water.

Which recipe makes the sweeter punch?

Recipe A

Recipe B

Show how you determined the sweeter punch.

© Dennis Cox/Dreamstime # 7018684

Figure 9. Proportional Punch, final question.

One important result of embedding this task within a technology enhanced scenario was the opportunity to parse it into five discrete yet connected sections. This property of *Proportional Punch* gave rise to notable teacher creativity, both in terms of interventions and in preparations for subsequent sections. For example, the following is an excerpt from a lesson that we observed between Sections 2 and 3 of the task (recall that the sweetness meter registers sweetness based on the ratio of scoops of mix to cups of water):

Teacher: Yesterday we used the sweetness meter to discover different ratios for different combinations of scoops of mix compared to cups of water to make punch. Can anybody tell me a way to make a proportion of zero?

Student: Mix with no water.

Teacher: Can you prove it with the sweetness meter?

(Student comes up to teacher laptop and uses the sweetness meter to put a scoop of mix in the pitcher. Teacher directs students to notice any change in the meter.)

Teacher: Is there another way to get a proportion of zero?

Students: Yes....

Teacher: Prove it! (This became the class mantra for all the questions with this slide. Student uses the sweetness meter on the teacher laptop and just adds water to the pitcher.)

In addition to computer simulations, CBAL tasks feature computer functions that cannot be replicated in traditional testing. For example, in the *Dams and Droughts* algebra task students have a slider tool that helps them read complex graphs (Figure 10).

In the CBAL *Carnival Rides* algebra task is a tool that allows students to merge two graphs helping them identify the points that satisfy two constraints in this rudimentary linear programming task (Figure 11).

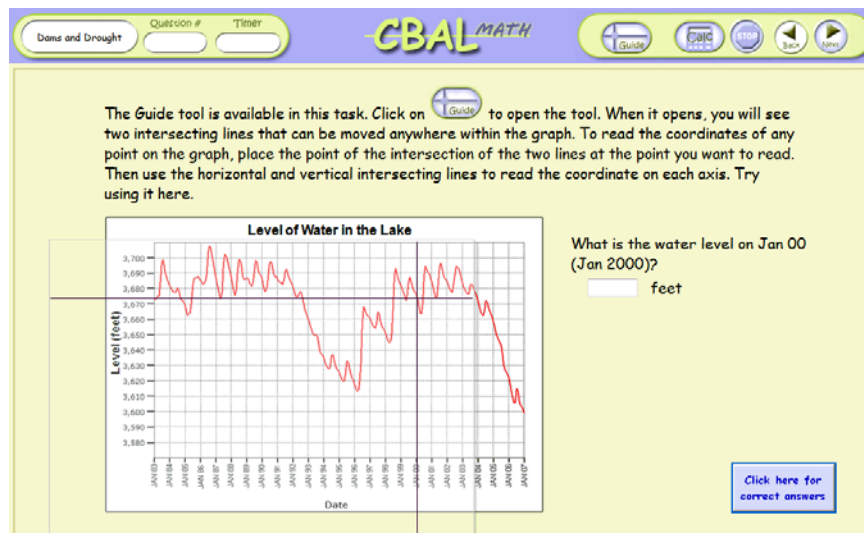


Figure 10. *Dams and Droughts* guide tool in use.

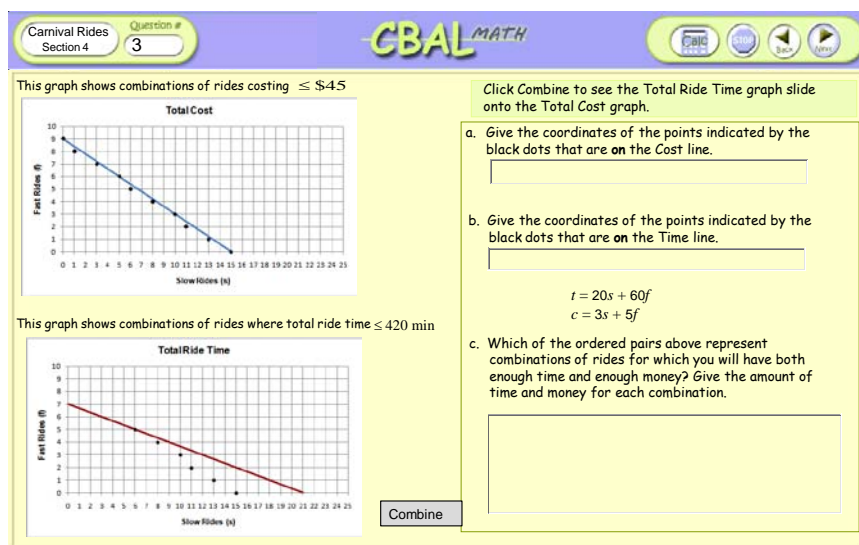


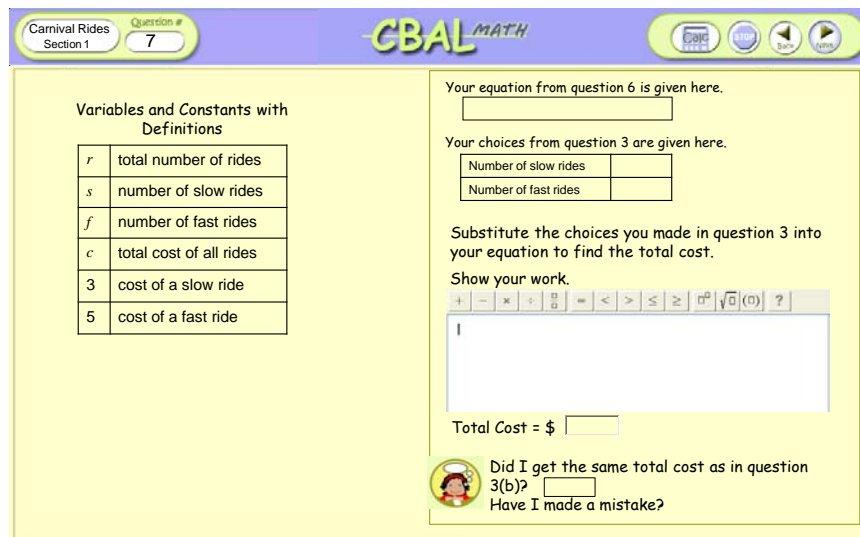
Figure 11. Graph merge, CBAL *Carnival Rides* task.

In the *Carnival Rides* task is also an example of a task in which technology is used to bring answers to earlier items forward to extend understanding. The item pictured in Figure 12 brings answers forward from two previous questions in order to have students test how well an equation they wrote works.

Overall, our pilot studies in CBAL mathematics have revealed that both students and teachers find these assessments engaging. Teachers also appreciate the real-time feedback they can obtain about student performance through the *VIA* online reporting system. It is clear that the formative CBAL TEAs in mathematics are valuable learning experiences for students and instructional enhancements for teachers, in and of themselves.

Measurement and Scoring: Potential New Insights From Technology Enhanced Assessments

Whether TEAs improve assessment in terms of validity and precision of measurement is an empirical question. In this section, we discuss some of the possibilities, and potential challenges, of TEAs for establishing validity and reliability from an educational measurement perspective (Baker, 2003). The focal task in this paper, *Proportional Punch*, is a formative assessment, but here we want to focus on some key issues that are relevant to summative assessment (assessment *of* learning). So throughout this section we will refer to a similar, but summative, task in the family of CBAL assessments on proportional reasoning called *Mix It Up*.



CBAL MATH

Carnival Rides Section 1 Question # 7

r	total number of rides
s	number of slow rides
f	number of fast rides
c	total cost of all rides
3	cost of a slow ride
5	cost of a fast ride

Your equation from question 6 is given here.

Your choices from question 3 are given here.

Number of slow rides	<input type="text"/>
Number of fast rides	<input type="text"/>

Substitute the choices you made in question 3 into your equation to find the total cost.

Show your work.

$+$ $-$ \times \div $\frac{\square}{\square}$ $=$ $<$ $>$ \leq \geq $^{\square}$ $\sqrt{\square}$ (\square) $?$

Total Cost = \$

Did I get the same total cost as in question 3(b)?

Have I made a mistake?

Figure 12. *Carnival Rides* task, reviewing previous answers.

The discussion in this section is divided into two parts: validity and reliability (for a detailed discussion of validity and reliability, see Brennan, 2006). Validity concerns the question: Are we measuring what we intended to measure? Reliability, which is directly related to precision of measurement, concerns the question: Would we get the same result if we were to repeat the measurement under identical conditions; in other words, can we get a consistent measurement?

Before we discuss validity and reliability for TEAs, it is important to stress that it is not easy to make sense of the overwhelming amount of data that these assessment types can produce (Bennett, Jenkins, Persky, & Weiss, 2003). A wide variety of behaviors can be captured by the computer, and these behaviors can show tremendous individual differences. Therefore, an important issue in designing TEAs is the question of what student behaviors to capture for the purpose of the assessment. What type of clicking behaviors are to be recorded? What sort of typing behaviors for constructed-response items should be recorded? Is only the final answer needed or are intermediate responses and editing behaviors also informative and important to capture? Should item-response times be captured and if so, how and when should these be captured when there are multiple items on one screen? How should response time be represented when students can access a simulation tool before answering the item or after answering the item but before clicking next, and what if students go back? All of these questions represent important considerations in the design of the task and the data capture and analysis protocols.

The computer makes it possible to track all such actions, but it is necessary to decide which ones are relevant for the purpose of the assessment. Some of these behaviors can be useful for building a validity argument, while others are useful for the actual measurement of the intended construct. The fact that students can be tracked as they proceed through an assessment presents some challenges and belies the classical notion of a test as a small, scientific, controlled experiment (Anastasi & Urbina, 1997).

Building a Case for Validity With Technology Enhanced Assessments

In terms of validity, TEAs offer many possibilities for capturing student actions that can be characterized as either *construct-relevant* or *construct-irrelevant* (e.g., Messick, 1989; for an example of a mathematics-specific empirical study, although for paper-based responses, see Fife, Graf, & Ohls, 2011). For example, if a TEA contains a simulation, observing that students use the simulation before answering the question can be counted as construct-relevant behavior and can be used for building a case for (construct) validity. In contrast, excessive mouse-clicking and navigating can be counted as construct-irrelevant behaviors. During item development, all kinds of student behaviors can be captured through piloting and used to identify and ameliorate construct-irrelevant lures or distractions, as well as difficulties in using the simulation or answering items with innovative response formats (e.g., sentence selection, drag-and-drop, click-and-click, copy-and-paste, numeric entry, equation, graph). Obviously, it should not be the case that such innovations lead to construct-irrelevant behaviors, and we need to provide at least some empirical support for that.

TEAs also offer the potential for gathering richer and more subtle data about a student; for example, different stages in a learning progression might be linked to different interactive behaviors and strategies for answering (e.g., Level 2 students often use the simulation, but answer the item incorrectly; Level 3 students often use the simulation and answer correctly; Level 4 students do not use the simulation and answer correctly). Capturing the variability in interactive behavior patterns such as

simulation use, responses, and response times can provide information on strategies, both successful and unsuccessful. Such evidence can be quite compelling in building a validity argument.

As an example, consider the task *Mix It Up*, from which a number of screenshots are shown in Figure 13, and the learning progression for proportional reasoning described in Table 1. Tracking the use of the paint-mixing tool in combination with the item responses could help researchers to obtain feedback on the stages in the learning progression, might help to fine-tune the descriptions of the stages, or might even result in changing them. Unfortunately, we do not yet have such data available.

Improving Measurement Precision With Technology Enhanced Assessments

In terms of reliability, TEAs with innovative response formats can benefit measurement precision. Typical selected response items always have the issue that students have a chance of guessing the correct answer. Changing to more innovative selected response formats that are available through interactive technologies, such as selecting elements in a table, ordering a list, or clicking points in a graph, circumvent this guessing issue because the number of possibilities quickly increases. When guessing behavior can be reduced, reliability and measurement precision benefit.

More elaborate selected-response items also allow for more elaborate scoring, so that more units of measurement can be created. For example, ordering a list allows for a wide range of partial credit scoring based on the (rank order) correlation between the observed order and the correct order. Another example can be found in a reading assessment in which students have to select sentences or paragraphs as support for their answers on a multiple-choice item. These two examples illustrate possible improvements that can be drawn from single items with innovative response format. Nevertheless, such items also require more time to answer, so that the overall improvement in precision can be limited.

Both challenges and opportunities exist in the use of TEAs as a window into student cognition, including the need for sophisticated measurement models and the issues associated with automated capture and characterization of both scorable and nonscorable observables. Nowadays it is also common that items have multiple classification both in terms of a content domain and a cognitive domain. For example, the Programme for International Student Assessment (PISA) uses such a double item classification (Organization for Economic Co-Operation and Development, 2012). These double classifications allow for multiple analyses of data: one for reporting results (e.g., a simple mathematics score) and one for finding support of a cognitive model (e.g., probabilistic relations between elements of the mathematics competency model; the sequencing of stages in a learning progression).


In *Mix It Up*, item responses can be modeled with a simple unidimensional item response theory (IRT) model to in order to produce a score. If needed, this score can be related to relevant background variables such as gender, socio-economic status, etc. Alternatively, item responses can also be modeled with a latent class model in which the classes are formed by the stages in the learning progression for proportional reasoning. A classification of the items in *Mix It Up* in terms of the stages of the progression can be used as a confirmatory model. The use of response times might be less suited for the first type of models that is to be used in reporting scores. However, for the purpose of finding support for a cognitive competency model, item response times can provide valuable information. There are a number of recent examples of the integration of cognitive psychology and psychometrics that make use of response times (Klein-Entink, Kuhn, Hornke, & Fox, 2009; Van der

Mix It Up Question # Timer
CBAL MATH

Mix It Up

Your class is helping fix up a teen center.

The class has received a donation of red paint and blue paint and has decided to mix them to get purple paint to paint the teen center.



Mix It Up Section 1 Question # 1 of 10 Timer
CBAL MATH

Paint Mixing Tool

Click on the **Add Red** and **Add Blue** buttons to experiment with mixing paint so that you can see how the ratio of red to blue affects the color of the paint.

Click on **Start Over** to empty the can and begin a new experiment.

Red:
 Blue:





Mix It Up Section 1 Question # 10 of 30 Timer
CBAL MATH

Refer to the tables below when answering this question.

Ratios of Shades of Purple

	Red Paint : Blue Paint	Red Paint Blue Paint
Rosie	2:3	$\frac{2}{3}$
Karen	5:6	$\frac{5}{6}$

Maximum Amount of Each Shade Using Donated Paint
(18 quarts of red and 24 quarts of blue were donated)

	Donated Red Used	Donated Blue Used	Total Amount of Paint Made	Donated Red Left	Donated Blue Left
Rosie	16	24	40	2	0
Karen	15	18	33	3	6

Keep these conditions in mind when answering the question below.

- All of the donated paint should be used.
- To save money, the least amount of additional paint will be purchased so that all the donated paint can be used.
- The class will paint the teen center in either Rosie's or Karen's shade.
- When red and blue are mixed together, only full quarts of paint are used.

Which ratio of red paint to blue paint do you recommend be chosen to meet all of the conditions above?

Rosie's
 Karen's

Explain your answer:

Figure 13. Screenshots from the CBAL mathematics summative task *Mix it Up*.

Maas, Molenaar, Maris, Kievit, & Borsboom, 2011). These examples make use of relatively straightforward selected-response item formats, and the challenges for TEAs in this respect are yet to be explored.

In addition, a problem with fitting the aforementioned psychometric models to tasks like *Proportional Punch* and *Mix It Up* is that a number of other problems can arise, including person-by-task interactions, local dependencies, and specific forms of multidimensionality (e.g., testlet or bi-factor structures; Rijmen, 2010). These problems typically arise when one is interested in aggregating a number of such tasks and making inferences about a student's overall performance. In addition, there are further problems when different assessments are made parallel or need to be equated. At present, ETS is addressing parts of these challenges, but many remain open as yet.

Strengths of Technology Enhanced Assessments: Engagement, Motivation, and Capturing Hard to Measure Constructs

In this section, we discuss some of the hypothesized strengths of TEAs. One of the most frequently cited advantages of these kinds of assessments is their ability to engage and motivate students. This is an important consideration—evidence from assessment of literacy suggests that less interesting tasks tend to underestimate student abilities (Guthrie & Humenick, 2004; Moje, Overby, Tysvaer, & Morris, 2008; Rickford, 2001). In the commercial world, video game designers have long been engaged in developing techniques for creating a continually engaging experience for users. But in the assessment world, too, the dynamic properties of TEAs can allow designers to introduce components that encourage student interaction and increase novelty. Studies have shown that novel tasks have the potential to increase the validity of an assessment. Wohlhueter (1966) speculated that novel tasks support mental stimulation and allow for the opportunity to recover from cognitive fatigue. In a meta-analysis, Guthrie and Humenick identified a number of specific factors that increase motivation and achievement in reading, including providing interesting text, providing clear conceptual goals, providing choice, and collaboration. Some of these properties are uniquely suited to, or can be more readily and effectively incorporated within, technology enhanced assessment formats.

User engagement is the connection that exists between a user and a resource. This connection may involve cognitive, emotional, and behavioral aspects. User engagement is associated with the user being absorbed by an artifact or a piece of technology and being motivated to use it (O'Brien & Toms, 2008). Attfield, Kazai, Lalmas and Piwowarski (2011) reviewed research on user engagement and identified several of its characteristics, which include: (a) *focused attention*, immediate feedback and interaction that could result in loss of self-consciousness, distortions in the perception of time and experiencing a state of flow (Csikszentmihalyi, 1990); (b) *positive affect*, or positive emotions experienced during interaction that could result in personal bond associated with the activity (Jennings, 2000); (c) *aesthetics*, or visual appeal of an interface (usability and aesthetics seem to be positively correlated; Tractinsky, Katz, & Ikar, 2000); (d) *endurability*, or the tendency of people to remember, want to repeat, and share experiences that they find enjoyable, useful, or engaging (O'Brien, Toms, Kelloway, & Kelly, 2010; Read, MacFarlane, & Casey, 2002); (e) *novelty*, or engagement that can be associated with new, surprising, unfamiliar, or unexpected experiences (O'Brien, et al., 2010; a balance between novelty and familiarity may help keep high levels of engagement; O'Brien & Toms, 2008); (f) *richness and control*—level of engagement can be explained in terms the richness of thoughts, actions, and

perceptions evoked by an activity and the level of control the users have to achieve their goals (Rozendaal, Keyson, & Ridder, 2009); (g) *reputation, trust and expectation*—trust is a necessary condition for engagement, and users will not enjoy interacting with systems or applications that are not trustworthy or that have a negative reputation; and (h) *user context*—level of user engagement in the lab may differ from that in the real context, and social norms, expectations, values, and trends influence how users engage (Harper, Rodden, Rogers, & Sellen, 2008).

TEAs can leverage many of these aspects to support user engagement. For example, feedback is a key component of formative assessments. Research shows that interactive computer applications that provide immediate, task-level feedback to students can positively contribute to student learning (e.g., Hattie & Timperley, 2007; Shute, 2008). Black and Wiliam (1998) reviewed the literature on classroom formative assessment and found evidence supporting the notion that providing frequent feedback to students improves student learning. Students in this approach are considered active participants working with teachers to implement formative strategies in the classroom. Immediate feedback that results from a direct manipulation of objects in a simulation can provide useful information to guide exploration or refine interaction strategies (e.g., interaction with the sweetness meter and punch-making simulation in the proportional punch task can help reveal student strategies as regarding proportional reasoning).

On the other hand, higher levels of interactivity can result in higher than expected cognitive processing demands, and this may reduce the measurement quality of the assessment. This type of demand, also called cognitive load, has been extensively explored (e.g. Mayer, 2005; Mayer & Moreno, 2003; Sweller, 1999; Wickens, 1980, 1991). For example, Mayer and Moreno (2003) defined three kinds of demands: (a) essential processing demands that are aimed at use of working memory for selecting, organizing, and integrating information related to the core learning goals or assessment needs (i.e., demands that are construct relevant); (b) incidental processing demands that are aimed at use of working memory for selecting, organizing, and integrating information not directly related to the core learning goals or assessment needs (i.e., demands that are not construct-relevant); and (c) representational holding, which describes the task demands for maintaining information in working memory (i.e., demands that may or may not be construct-relevant depending upon the nature of the construct). Principles of multimedia design and user interface design should be applied to improve users' experience with TEAs without imposing construct irrelevant high cognitive demands (see Clark & Mayer, 2011, for seven multimedia principles that apply to TEAs).

Similarly, highly interactive (e.g., immersive) TEAs may negatively affect student engagement, especially for students who are not used to interacting with these types of environments. And parallel environments that do not impose the same visual, motor, and auditory demands without changing the construct may need to be developed for particular groups of students (e.g., students with visual disabilities; see the following section of this paper for more discussion of this issue).

Prior work on applying principles of evidence-centered design (ECD; Mislevy, Steinberg, & Almond, 2003) in the development of new assessment tasks (cf. Shute, Ventura, Bauer, & Zapata-Rivera, 2009; Zapata-Rivera & Hansen, 2009; Zapata-Rivera, VanWinkle, Doyle, Buteux, & Bauer, 2009) shows that it is possible to design engaging TEAs that capture knowledge and skills traditionally considered hard to measure without negatively influencing the assessment quality of the task. Other researchers are applying the ECD principles in the creation of engaging simulation and gaming environments to

measure higher order thinking skills (also called 21stth century skills) such as problem-solving, modeling, and communication in a similar way (Behrens, Mislevy, DiCerbo, & Levy, 2011; Bennett, Persky, Weiss, & Jenkins, 2007; Shute & Torres, 2011; Zapata-Rivera & Bauer, 2011). Initial evidence seems promising, although more research is needed to establish effects on validity and reliability of including engagement-enhancing features in assessments.

We have employed ECD in the development of mathematics tasks for hard to measure constructs in our CBAL work (Graf et al., 2009; Harris, Bauer, & Redman, 2008). Using ECD, we have found that hard to measure constructs become much easier to measure the more clearly one defines the competencies, especially discriminations within competencies, and the types of evidence that can be used to make inferences about those competencies. This has led to the use of mathematical learning trajectories (Harris et al., 2008) that distinguish among levels of sophistication in reasoning about specific mathematical concepts. Developing well-defined measurement criteria and constraints enables the design of technology enhanced items that exist to serve an evidential purpose. Features can be included to allow for the collection of evidence to make inferences about specific competencies, and features that do not serve that purpose can be excluded. For example, in the Proportional Punch task, the sweetness meter and associated items were designed to collect evidence that distinguishes between, for example, (a) thinking in terms of additive differences (i.e., reasoning that is characteristic of Level 2 of the proportional reasoning learning trajectory) and (b) multiplicative reasoning. Making the jump to multiplicative reasoning is a significant hurdle for students, and so collecting evidence to distinguish students as each of the levels is important for both summative and formative purposes.

Enhancing user engagement or motivational aspects of an assessment should be done carefully, and including technology in tasks should be done with the intent of enhancing assessment quality. The purpose of the assessment provides guidance on the types of features that can be employed and allows for a clear justification for design features in terms of the overall validity argument for the assessment.

A Key Challenge for Technology Enhanced Assessments: How to Achieve Universal Design and Accessibility

Technology has the potential to provide unprecedented access to information and resources for students with disabilities, as many formerly expensive assistive technologies are now built-in features on off the shelf products. For example, mainstream products from Apple (e.g., iPad, iPhone, Apple TV) allow any user to turn on accessibility features such as the VoiceOver screen reader. In addition, other applications can be downloaded for free or low cost to allow a tablet or smartphone to be turned into an augmentative and alternative communication (AAC) device or magnifier. Finally, many newer technologies have the capability to interoperate with other assistive technology hardware (e.g., wireless refreshable braille displays).

During this time of blossoming technology solutions, much research has been dedicated to the development of accessible assessments. However, the primary focus of this research has tended to be on reading assessments, with limited research on mathematics-related TEAs. In fact, very little research has examined the accessibility of mathematics assessments for students with disabilities beyond the use of calculators or language supports (audio to read-aloud mathematics, glossary of terms, or simplified language; Bouck & Bouck, 2008; Elliott et al., 2010; Engelhard, Fincher, & Domaleski, 2010; Kettler et al., 2011; Wolf, Kim, Kao, & Rivera, 2009). In addition, research on the universal design (UD) of assessments

has resulted in broad principles for the development of assessments, but results have been mixed. For example, studies designed to implement the UD principle of “concise and readable text” (Johnstone, Altman, & Thurlow, 2006) have found that in some cases the item becomes easier for all test takers, while other studies have found that this feature closes the gap between ELLs and non-ELLs (Abedi, Lord, Hofstetter, & Baker, 2000; Sato, Rabinowitz, Gallagher, & Huang, 2010; Sireci, Li, & Scarpatti, 2003).

ETS researchers are continuing to refine and advance the accessibility of new TEAs through a multi-part approach that involves three stages of support, with the intention of meeting the needs of as many students as possible. In this approach, our base item development (first stage support) incorporates accessibility as an extension of ECD to meet the needs of most test takers (Hansen & Mislavy, 2008; Mislavy & Haertel, 2006). The second stage of support includes exploring additional features that can be added to TEA items to enhance accessibility for some students. Finally, the third stage of support includes the development of alternative assessment techniques for students who are not able to take the assessment under either of the previous stages of support. This approach allows the majority of students to be measured under standard conditions (base item), some students to be assessed with accommodations that are not appropriate for the majority of test takers, and a few students to be assessed on alternative administration of the same construct.

Our focal task, *Proportional Punch*, has several elements that allow the majority of test takers to be assessed under standard conditions. One element that is probably true for most TEA items is that they engage students who previously did not provide their best effort on “boring” paper-based multiple-choice tests. While most students benefit from increased engagement during assessment, some students with disabilities that impact attention, academic motivation, and learning are uniquely advantaged by highly engaging assessments. These subgroups may include students with Attention Deficit/Hyperactivity Disorder (ADHD), Autism Spectrum Disorder (ASD), Emotional Behavioral Disorders (EBD), and some students with Specific Learning Disabilities (LD) and mild to moderate Cognitive Impairments (CI).

It is important however that TEA items balance engagement without making excessive demands on working memory as many of the same disabilities identified above impact executive functioning and working memory. For this reason, TEAs should include a second element that is found in the Proportional Punch task: *Appropriate demands on working memory*. This is one of the principles of universal design for assessment (Thompson, Johnstone, & Thurlow, 2002), which requires that directions and procedures be intuitive, to allow students to focus on the construct being measured and not ancillary tasks (e.g., how to enter your answer or navigate). In addition, *Proportional Punch* reduces working memory demands by providing information to students in a thoughtful manner and requesting them to respond in sequential steps. These features allow students with attention or working memory deficits to process information without being overwhelmed. In addition, the sequential nature of the items lend themselves to rendering in alternate formats (e.g., audio) for students with print-based disabilities (e.g., dyslexia, low vision, blindness), because the structured nature of the items allows for easy referencing and review (via navigation) between (and within) the sequence using audio or braille presentation. Finally the Proportional Punch task allows for direct measurement of prerequisite or foundational skills and then supports students in higher level skills by providing the answers to previous steps. This is very important for the large numbers of students with disabilities who are performing well below grade level (Laitusis et al., 2010). In addition this challenge is likely to increase as states adopt the

more rigorous common core assessments and phase out alternate assessments based on modified achievement standards (AA-MAS).

Although these features of *Proportional Punch* allow students with disabilities that impact working memory, engagement, attention, and attainment of prerequisite skills to participate on the standard item (first-stage support in the accessibility model), some students with disabilities will require additional support. Next we will explain the features that we are investigating to allow students who are blind or visually impaired to test under the second stage of support in the accessibility approach.

The second stage approach to accessibility includes the use of audio presentation and haptics² that are integrated into the delivery of the task. While some may consider audio a first stage accessibility feature, there is a substantial amount of research on audio for mathematics that indicates this feature impacts item difficulty and item comparability and may not universally benefit all test takers, particularly at higher grade levels (Bolt & Thurlow, 2007; Bolt & Ysseldyke, 2006; Wolf et al., 2009). In addition, reading aloud mathematics symbols is not straightforward and some states may wish to regulate what is read aloud and monitor the use of audio presentation in mathematics (e.g., for inclusion in growth models or item difficulty statistics). That said, many students with print-based disabilities and some ELLs are likely to benefit from audio presentation of mathematics.

For sighted test takers, visual cues that may facilitate understanding or interpretation of information in a task can prove problematic when transforming into an audio presentation. For some mathematics items, it may be feasible to present graphical elements, such as graphs or geometric shapes, using tactile graphics (e.g., printed in hardcopy form using a tactile embosser). Such approaches, however, incur cost in terms of specialized hardware, production of specialized file formats, and potentially complex processes to support adaptive testing. Further, for dynamic and interactive tasks such as *Proportional Punch*, the use of hard copy, static, and tactile graphics is not suitable.

A variety of haptic technologies, ranging from specialized hardware (such as the Novint Falcon) to haptic feedback mechanisms built into off-the-shelf tablet devices (such as vibration) offer promising approaches to add nonvisual cues to structures that can be explored by students with visual impairments. Our research is examining the addition of nonspeech and haptic feedback to the *Proportional Punch* task. This can allow nonvisual exploration of the task as the student interacts with it, changing variables affecting the visual presentation, and in turn resulting in perceivable haptic and auditory cues.

A prototype of the *Proportional Punch* task has been implemented on a Google Android tablet, incorporating vibration-based haptic feedback and auditory cues (see Figure 14). The prototype is serving as a platform for exploring the effectiveness of haptic feedback as a mechanism to enhance the nonvisual presentation and navigation of interactive tasks with participants with visual impairments. Haptic technologies are rapidly evolving, with a variety of new technological approaches expected to appear in consumer products in the next several years. Understanding how these developments can enhance the accessibility of technology enhanced assessments is an exciting and promising area of research.

² Haptic displays add the modality of touch to computer-based systems, enabling the user to perceive shapes or textures through a variety of technological approaches.

TEA Item Adapted for Non-Visual Access

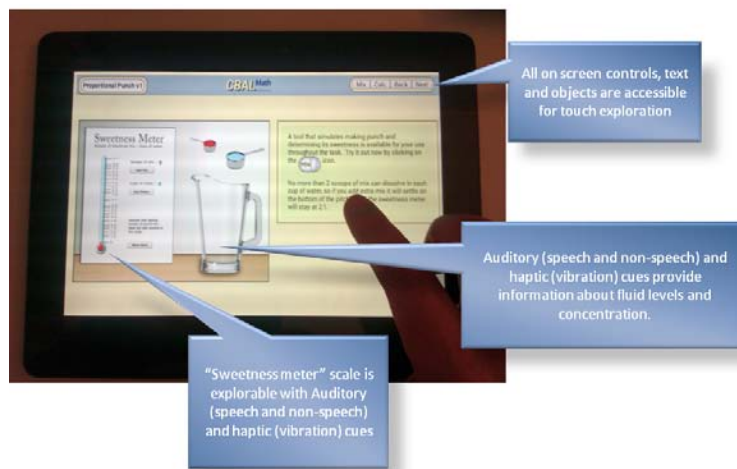


Figure 14. *Proportional Punch* adapted for accessibility by visually impaired users.

Finally, the third stage approach includes developing alternative assessments that measure the same construct as TEA items. While this approach is far from ideal, we recognize that regardless of the accessibility elements embedded in the first and second stages there will be some students who are unable to participate in the assessment. These groups may include a small number of students with seizure disorders who may be unable to use computers, students who are deaf/blind, or students with visual impairments who have not yet learned how to use assistive technologies such as screen readers. In addition, even when assistive technologies are embedded within the TEA tasks and students are proficient in using these assistive technologies, it is likely that some TEA items will become so dependent on technology skills that the item moves from an assessment of the intended knowledge, skills, and abilities (KSAs) to an assessment of a student’s working memory capacity or assistive technology proficiency. ETS recently completed a focus group with teachers of the visually impaired to evaluate several TEA items similar to *Proportional Punch* and concluded that even with the integration of assistive technologies such as those described in the second stage, it is likely that some TEA items will result in cumulatively excessive nonfocal requirements (Hansen, Laitusis, Frankel, & King, in press).

Conclusions and Future Directions

The present paper has highlighted some of the strengths and challenges of technology enhanced assessments. We have also identified some of the work being undertaken at ETS in relation to these. In closing, it seems appropriate to look to the future and identify some of the important questions that we feel should be addressed in research efforts focused on these relatively new forms of assessment.

1. Identifying and capturing informative process data: Now that we have the capability to track how students proceed through an assessment, an important research challenge is how to characterize these process data. As we have said, we could in theory capture every observable action made by the student, but it is important to establish which of these actions are cognitively informative and thus valuable for assessment purposes. Part of this process will be to establish which behaviors are useful for measurement of the intended construct and which may be more useful for building a validity argument.
2. Analyzing process data: Building on the above point, new analytical approaches may be required to allow us to extract useful information from the interactive behaviors captured. This may entail multiple classifications, to permit multiple analyses of data (e.g., content and cognitive domains). Response times may or may not be useful, depending on the aims of the analyses. We will need to decide whether to model responses with a simple unidimensional item response theory (IRT) approach to provide a score, or with a more sophisticated approach for a qualitatively different aim, such as a latent class model to confirm some theoretical structure such as stages in a learning progression.
3. Informing instruction: If the assessment has a formative aspect, we need to understand which aspects of the captured data provide the greatest educational value for providing feedback to the student and/or the teacher. We also need to consider what kinds of design framework and principles should guide the development of formative and professional support components intended to be used in conjunction with TEAs.
4. Implications for reporting: Given the potential richness of the available data, it will be important to ensure that reporting of results reflects the information accurately but in a way that end-users can benefit from. Empirical studies with stakeholders for summative assessments and with end-users of formative assessments will be important to achieve the right balance of richness and usability.
5. Adaptive testing: Dynamic systems can allow us to develop adaptive TEAs that respond flexibly to the student's level of performance. Careful research will be necessary to ensure this is done effectively and appropriately. At a more fundamental level, we need to understand the broad implications of introducing these kinds of features into any assessment situation.
6. Effects of engagement-enhancing features: More research is needed to establish the potential effects of including features to enhance engagement on the validity and reliability of TEAs. It will also be important to conduct research to ensure that TEAs balance engagement with working memory demands.
7. New assistive technologies and universal design characteristics: New presentation and control technologies such as dynamic haptic displays potentially provide great opportunities to enhance the accessibility of technology enhanced assessment. This is an exciting and promising area of research, but there is also a substantial amount of work to be done to understand the implications and challenges of incorporating these new technologies into assessment design, both for accessibility and in terms of validity.

These challenges are just a selection of the potential questions yet to be fully addressed by research in this field. It is therefore important that those of us involved in developing TEAs engage in rigorous and targeted empirical studies. This will help to ensure that these tasks not only possess TEA-specific strengths such as engagement and flexibility, but also can be demonstrated to be valid, informative, useful, and accessible forms of assessment.

References

- Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, 19(3), 16–26.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Attfield, S., Kazai, G., Lalmas, M., & Piwowarski, B. (2011). Towards a science of user engagement. In *UMWA 2011: Workshop on user modelling for web applications* (pp. 9–12). Retrieved from <http://research.yahoo.com/files/BarcelonaMarch01.pdf>
- Baker, E. L. (2003). Reflections on technology-enhanced assessment. *Assessment in Education: Principles, Policy, and Practice*, 10, 421–424.
- Baxter, G. P., & Junker, B. (2001, April). *Designing cognitive-developmental assessments: A case study in proportional reasoning*. Paper presented at the annual meeting of the National Council for Measurement in Education, Seattle, WA. Retrieved from <http://www.stat.cmu.edu/~brian/rpm/baxterjunkerncme.pdf>
- Behrens, J. T., Mislevy, R. J., DiCerbo, K. E., & Levy, R. (2011). An evidence centered design for learning and assessment in the digital world. In M.C. Mayrath, J. Clarke-Midura, & D. Robinson (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. Charlotte, NC: Information Age.
- Bennett, R. E. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and Practice*, 18(3), 5–12.
- Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives*, 8(2–3), 70–91.
- Bennett, R. E., & Gitomer, D. H. (2009). Transforming K-12 assessment: Integrating accountability testing, formative assessment, and professional support. In C. Wyatt-Smith & J. Cumming (Eds.), *Educational assessment in the 21st century* (pp. 43–61). New York, NY: Springer.
- Bennett, R. E., Jenkins, F., Persky, H., & Weiss, A. (2003). Assessing complex problem solving performances. *Assessment in Education: Principles, Policy, and Practice*, 10, 347–359.
- Bennett, R. E., Persky, H., Weiss, A., & Jenkins, F. (2007). *Problem-solving in technology rich environments: A report from the NAEP technology-based assessment project* (NCES 2007-466). Washington, DC: National Center for Educational Statistics.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.
- Bolt, S. E., & Thurlow, M. L. (2007). Item-level effects of the read-aloud accommodation for students with reading disabilities. *Assessment for Effective Intervention*, 33(1), 15–28.
- Bolt, S. E., & Ysseldyke, J. E. (2006). Comparing DIF across math and reading/language arts tests for students receiving a read-aloud accommodation. *Applied Measurement in Education*, 19, 329–355.
- Bouck, E. C., & Bouck, M. K. (2008). Does it add up? Calculators as accommodations for sixth grade students with disabilities. *Journal of Special Education Technology*, 23(2), 17–32.
- Brennan, R. L., (Ed.). (2006). *Educational measurement* (4th ed.). Westport, CT: Praeger.

- Clark, R. C., & Mayer, R. E. (2011). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (3rd ed.). San Francisco, CA: Wiley.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. London, England: Harper and Row.
- Elliott, S. N., Kettler, R. J., Beddow, P. A., Kurz, A., Compton, E., McGrath, D., Bruen, C.,... Roach, A. T. (2010). Effects of using modified items to test students with persistent academic difficulties. *Exceptional Children, 76*, 475–495.
- Engelhard, G., Fincher, M., & Domaleski, C. S. (2010). Mathematics performance of students with and without disabilities under accommodated conditions using resource guides and calculators on high stakes tests. *Applied Measurement in Education, 24*(1), 22–38.
- ETS. (2011). *Outline of provisional learning progressions*. Retrieved from <http://elalp.cbalwiki.ets.org/Outline+of+Provisional+Learning+Progressions>
- Fife, J. H., Graf, E. A., & Ohls, S. (2011). *Constructed-response mathematics tasks study* (ETS Research Report No. RR-11-35). Princeton, NJ: ETS.
- Graf, E. A. (2009). *Defining mathematics competency in the service of Cognitively Based Assessment for grades 6 through 8* (ETS Research Memorandum No. RM-09-42). Princeton, NJ: ETS.
- Graf, E. A., Harris, K., Marquez, E., Fife, J., & Redman, M. (2009). *Cognitively based assessment of, for, and as learning (CBAL) in mathematics: A design and first steps toward implementation* (ETS Research Memorandum No. RM-09-07). Princeton, NJ: ETS.
- Guthrie, J. T., & Humenick, N. M. (2004). Motivating students to read: Evidence for classroom practices that increase reading motivation and achievement. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 329–354). Baltimore, MA: Brookes.
- Hansen, E., Laitusis, C. C., Frankel, L., & King, T. (in press). Designing accessible technology-enabled reading assessments: Recommendations from teachers of students with visual impairments. *The Journal of Blindness Innovation and Research*.
- Hansen, E. G., & Mislevy, R. J. (2008). *Design patterns for improving accessibility for test takers with disabilities* (ETS Research Report No. RR-08-49). Princeton, NJ: ETS.
- Harper, R., Rodden, T., Rogers, Y., & Sellen, A. (2008) *Being human: Human-computer interaction in the Year 2020*. Cambridge, England: Microsoft.
- Harris, K., Bauer, M. I., & Redman, M. (2008). Cognitively-based developmental models used as a link between formative and summative assessment. In *Proceedings of the 34th International Association for Educational Assessment (IAEA) annual conference, Cambridge, England, September 7–12*.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.
- Jennings, M. (2000) Theory and models for creating engaging and immersive e-commerce websites. In *Proceedings of the 2000 ACM SIGCPR conference on computer personnel research* (pp. 77–85). New York, NY: ACM.
- Johnstone, C. J., Altman, J., & Thurlow, M. (2006). *A state guide to the development of universally designed assessments*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

- Kettler, R. J., Rodriguez, M. C., Bolt, D. M., Elliot, S. N., Beddow, P. A., & Kurz, A. (2011). Modified multiple-choice items for alternate assessments: Reliability, difficulty, and differential boost. *Applied Measurement in Education, 24*, 210–234.
- Klein-Entink, R. H., Kuhn, J.-T., Hornke, L. F., & Fox, J.-P. (2009). Evaluating cognitive theory: A joint modeling approach using responses and response times. *Psychological Methods, 14*, 54–75.
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 31–48). New York, NY: Cambridge University Press.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.
- McFarlane, A. E. (2003). Editorial: Assessment for the digital age. *Assessment in Education, 10*(3), 261–266.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: Macmillan.
- Mislevy, R., & Haertel, G. (2006). *Implications of evidence-centered design for educational testing* (Draft PADI Technical Report No. 17). Menlo Park, CA: SRI International.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessment (with discussion). *Measurement: Interdisciplinary Research and Perspective, 1*, 3–62.
- Moje, E. B., Overby, M., Tysvaer, N., & Morris, K. (2008). The complex world of adolescent literacy: Myths, motivations, and mysteries. *Harvard Educational Review, 78*, 107–154.
- Noelting, G. (1980a). The development of proportional reasoning and the ratio concept Part I: Differentiation of stages. *Educational Studies in Mathematics, 11*(2), 217–253.
- Noelting, G. (1980b). The development of proportional reasoning and the ratio concept Part II: Problem-structure at successive stages; problem-solving strategies and the mechanism of adaptive restructuring. *Educational Studies in Mathematics, 11*, 331–363.
- O’Brien, H., & Toms, E. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology, 59*(6), 938–995.
- O’Brien, H. L., Toms, E., Kelloway, K., & Kelly, E. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology, 61*(1), 50–69.
- Organization for Economic Co-Operation and Development. (2012). *PISA 2009 technical report*. Retrieved from <http://dx.doi.org/10.1787/9789264167872-en>
- Read, J., MacFarlane, S., & Casey, C. (2002). Endurability, engagement and expectations: Measuring children’s fun. *Interaction Design and Children, 2*, 1–23.
- Rickford, A. E. (2001). The effects of teacher education on reading improvement. *Reading Improvement, 38*(4), 147–169.
- Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement, 47*, 361–372.
- Rozendaal, M., Keyson, D. V., & Ridder, H. (2009) Product features and task effects on experienced richness, control and engagement in voicemail browsing. *Personal Ubiquitous Computing, 13*, 343–354.

- Sands, W. A., Waters, B. K., & McBride, J. R. (Eds.). (1997). *Computerized adaptive testing: From inquiry to operation*. Washington, DC: American Psychological Association.
- Sato, E., Rabinowitz, S., Gallagher, C., & Huang, C. (2010). *Accommodations for English language learner students: the effect of linguistic modification of math test item sets* (Final Report NCEE 2009-4079). Washington, DC: Department of Education, Institute of Education Sciences.
- Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.
- Shute, V. J., & Torres, R. (2011). Where streams converge: Using evidence-centered design to assess Quest to Learn. In M. Mayrath, J. Clarke-Midura, & D. H. Robinson (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (pp.). Charlotte, NC: Information Age Publishing.
- Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. J. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295–321). Philadelphia, PA: Routledge.
- Sireci, S. G., Li, S., & Scarpati, S. (2003). *The effects of test accommodation on test performance: A review of the literature* (Center for Educational Assessment Research Report No. 485). Amherst, MA: University of Massachusetts.
- Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report No. 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>
- Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting With Computers*, 13, 127–145.
- van der Linden, W. J., & Glas, C. A. W. (2000). *Computerized adaptive testing: Theory and practice*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Van der Maas, H. L. J., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory again: On the relation between process models for decision making and latent variables for individual differences. *Psychological Review*, 118, 339–356.
- Wainer, H. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Weaver, R., & Junker, B. W. (2004). *Model specification for cognitive assessment of proportional reasoning* (Department of Statistics Technical Report No. 777). Pittsburgh, PA: Carnegie Mellon University.
- Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 239–257). Hillsdale, NJ: Lawrence Erlbaum.
- Wickens, C. D. (1991). Processing resources and attention. In D. L. Damos (Ed.), *Multiple task performance* (pp. 3–35). Bristol, PA: Taylor & Francis.
- Wohlhueter, J. F., (1966). *Fatigue in testing and other mental tasks: A literature survey* (Research Memorandum No. RM-66-6). Princeton, NJ: ETS.
- Wolf, M. K., Kim, J., Kao, J. C., & Rivera, N. M. (2009). *Examining the effectiveness and validity of glossary and read-aloud accommodations for English language learners in a math assessment* (CRESST

- Report No. 766). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Zapata-Rivera, D., & Bauer, M. (2011). Exploring the role of games in educational assessment. In M.C. Mayrath, J. Clarke-Midura, & D. Robinson (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. Charlotte, NC: Information Age.
- Zapata-Rivera, D., & Hansen, E. (2009, April). *Analyzing the learning potential of existing games using evidence-centered design and cognitive task analysis*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Zapata-Rivera, D., VanWinkle, W., Doyle, B., Buteux, A., & Bauer, M. I. (2009). Combining Learning and assessment in assessment-based gaming environments: A case study from a New York City school. *Interactive Technology and Smart Education*, 6, 173–188.

Appendix

Table A1. Identification of Universal Design Issues: Proportional Punch Task Item 3

Step 1	Read the question: "Which recipe makes the sweeter punch?" and process the information in the table (number of scoops of mix and cups of water for Recipes A and B)	Understands construct-relevant concepts such as ratio and proportion.	Understands how to judge likely sweetness based on the ratios of two substances, and thus can compare two simple number-based ratios.	Can perceive the question and table headings and numbers	Understands unfamiliar construct-irrelevant English terms such as "punch"	Can read and process the wording in the question
				ELL V H M RLD	ELL V H M RLD	ELL V H M RLD
Step 2	Compare ratio of mix to water in Recipe A and Recipe B and decide which would be sweeter	Can reason qualitatively and quantitatively to compare 2:3 versus 1:3 (mix to water)	Can conclude that 2:3 is sweeter	Can perceive the question and table headings and numbers, and calculator if needed		
				ELL V H M RLD	ELL V H M RLD	ELL V H M RLD
Step 3	Select answer using radio button A or radio button B	Can conclude that 2:3 is sweeter		Can perceive the forced-choice selected response options in screen	Can manipulate mouse to click radio button to select answer	
				ELL V H M RLD	ELL V H M RLD	ELL V H M RLD
Step 4	Explain answer in constructed response box	Understands the relationship between amounts of mix, water, and concentration (sweetness)	Can explain these relationships and own thinking in a way that shows the reasoning behind their selection	Can perceive constructed response field on screen	Can use mouse and keyboard to enter and type into field	Can explain answer using unfamiliar English terms such as "punch" and "scoops"
				ELL V H M RLD	ELL V H M RLD	ELL V H M RLD
Step 5	Use sweetness meter to check answer and reflect on explanation of own reasoning	Understands how to generate ratios using the same proportions as given in the table	Can read off proportion and ratio information and make connections among different sources and types of data	Can perceive sweetness meter's visual properties and dynamic indicator that responds to changes made to ratios of mix and water	Can use mouse control to operate scoops and cups to precisely manipulate ratios and proportions	
				ELL V H M RLD	ELL V H M RLD	ELL V H M RLD
Step 6	Review answer and make any required changes	Understands how information gained from sweetness meter generation process relates to table	Can use information from sweetness meter to review and, if needed, amend prior explanation	Can read out visual representation of sweetness indicator change	Can use mouse and keyboard to enter and type into field	
				ELL V H M RLD	ELL V H M RLD	ELL V H M RLD

Table A2. Identifying Ways to Enhance Accessibility: Proportional Punch Task Item 3

Top row: identifying relevant populations and their needs. Middle row: specifying the difficulty. Bottom row: suggested solution.						Comments
STEP 1: Read the question: "Which recipe makes the sweeter punch?" and process the information in the table (# of scoops of mix and cups of water for Recipes A and B)	Population	Blind/Low Vision	English Language Learner	Reading-based learning disability	Motor/Dexterity Impairment	Audio reader and hyperlinks to translations would not be provided universally, as these features could distract or disadvantage some students
	Difficulty	Cannot perceive visually-presented text	Non-construct-relevant English terms (e.g., "Punch") may be unfamiliar in this context	Extracting meaning from text may be difficult		
	Solution	Ensure all visually presented text content can be rendered non-visually; provide audio reader	Provide hyperlinks to native language translations of unfamiliar English words	Provide audio button to hear the information read aloud		
STEP 2: Compare ratio of mix to water in Recipe A and Recipe B and decide which would be sweeter	Population	Blind/Low Vision				Audio reader would not be provided universally, as this feature could distract or disadvantage some students
	Difficulty	Numerical information in the table cannot be perceived				
	Solution	Ensure all visually presented numerical content can be rendered non-visually; provide audio reader				
STEP 3: Select answer using radio button A or radio button B	Population	Blind/Low Vision			Motor/Dexterity Impairment	Audio reader would not be provided universally, as this feature could distract or disadvantage some students
	Difficulty	Cannot perceive radio buttons; cannot accurately target mouse to relevant on-screen locations			Difficulty in activating on-screen controls that require mouse clicks	
	Solution	Provide non-visual cues (haptics, audio cues) that provide an equivalent, alternate modality for presenting the cue. Provide control mechanisms for easily navigating through on-screen response locations.			Provide multiple mechanisms for activating on screen controls, such as keyboard, touch, and switch inputs.	

Table A3. Identifying Ways to Enhance Accessibility: Proportional Punch Task Item 3

STEP 4: Explain answer in constructed response box	Population	Blind/Low Vision	English Language Learner	Reading-based learning disability	Motor/Dexterity Impairment	Audio and haptic cues, and hyperlinks to translations, would not be provided universally, as these features could distract or disadvantage some students
	Difficulty	Cannot perceive constructed response answer field	May have difficulty explaining answer using unfamiliar English terms such as “punch”, “scoops”		May have difficulty using mouse and keyboard to enter and type into field	
	Solution	Provide non-visual cues that provide an equivalent, alternate modality. Provide controls for easily navigating through on-screen response locations.	Provide hyperlinks to native language translations of unfamiliar English words.		Provide an alternate text input device and more time if needed	
STEP 5: Use sweetness meter to check answer and reflect on explanation of own reasoning	Population	Blind/Low Vision			Motor/Dexterity Impairment	Non-visual audio and haptic cues would not be provided universally, as this feature could distract or disadvantage some students
	Difficulty	Analog information (color cues) present in the “sweetness meter” or pitcher, numerical values in the scale, and control mechanisms cannot be perceived			Difficulty in activating on-screen controls that require mouse clicks	
	Solution	Provide non-visual cues (haptics, non-speech audio cues) that provide an equivalent, alternate modality for presenting the cue. Provide control mechanisms for easily navigating through on-screen data values. Provide contextual cues (labels) to facilitate interpretation of a numeric value			Provide multiple mechanisms for activating on screen controls, such as keyboard, touch, and switch inputs.	
STEP 6: Review answer and make any required changes	Population	Blind/Low Vision	English Language Learner		Motor/Dexterity Impairment	Audio labels and hyperlinks to translations would not be provided universally, as these features could distract or disadvantage some students
	Difficulty	Sequential nature of non-visual presentation of on-screen information places a burden on working memory	May have difficulty explaining answer using unfamiliar English terms		May have difficulty using mouse and keyboard to enter and type into field	
	Solution	Provide control mechanisms for easily navigating through on-screen data and contextual cues (audio labels) to facilitate interpretation of numeric values	Provide hyperlinks to native language translations of unfamiliar English words		Provide an alternate text input device and more time if needed	



Invitational Research Symposium on
Technology Enhanced Assessments

The Center for K–12 Assessment & Performance Management at ETS creates timely events where conversations regarding new assessment challenges can take place, and publishes and disseminates the best thinking and research on the range of measurement issues facing national, state and local decision makers.

Copyright 2012 by Educational Testing Service.

ETS is a registered trademark of Educational Testing Service (ETS).
CBAL is a trademark of ETS.