K-12 Center at ETS

Invitational Research Symposium on
Technology Enhanced Assessments

Design and Development of Technology Enhanced Assessment Tasks: Integrating Evidence-Centered Design and Universal Design for Learning Frameworks to Assess Hard-to-Measure Science Constructs and Increase Student Accessibility

Geneva D. Haertel, Britte Haugan Cheng, Renee Cameto, Reina Fujii, Christopher Sanford, Daisy Rutstein, and Kathryn Morrison

May 7–8, 2012

K-12 Center at ETS
Invitational Research Symposium on
Technology Enhanced Assessments

Design and Development of Technology Enhanced Assessment Tasks: Integrating Evidence-Centered Design and Universal Design for Learning Frameworks to Assess Hard to Measure Science Constructs and Increase Student Accessibility

Geneva D. Haertel, Britte Haugan Cheng, Renee Cameto, Reina Fujii, Christopher Sanford, Daisy Rutstein, and Kathryn Morrison

SRI International

Executive Summary

Along with the promise and potential embodied in technology enhanced assessment come many challenges. Some challenges are foundational and faced by all assessment designers, such as establishing the validity, reliability, precision of measurement, and fairness of their new measures. Other challenges are less familiar such as broader inclusion of students in accountability measures. In all cases, assessment designers require appropriate tools to conceptualize and tackle design challenges.

This paper describes a design methodology for improving the validity of inferences about the performance of students on large-scale science assessment tasks. We present (a) an introduction to validity and its instantiation in an evidence-centered design (ECD) framework; (b) an overview of the ECD and universal design for learning (UDL) frameworks and a description of how these frameworks were integrated within a working web-based assessment design system; (c) how the integration was accomplished using three approaches; (d) examples of science assessment items and tasks developed using an integrated ECD/UDL approach and the design documentation that is generated during the process and (e) a review of these three assessment tasks in terms of their accessibility, technology enhanced features, implementation obstacles, scoring and validation methodologies, and efficiencies conferred in terms of costs and time. Six recommendations provide guidance about how an ECD process, including principles of UDL, improves validity of assessments and enhances their fairness in the context of technology enhanced assessments.

Foundational Challenges of Assessment: Validity, Reliability, Precision of Measurement, and Fairness

The Challenge of Validity

In this section of this paper we will provide a brief chronology of how validity, as a technical quality of assessments, has evolved over the past century. We will describe how we integrated two conceptual frameworks, one for evidence-centered design (ECD) and one for universal design for learnding (UDL), and instantiated them in an-line assessment design system (referred to as PADI). Our goal in developing the PADI assessment design system was to support assessment designers in developing technically sound measures and associated design documentation that is re-usable and can guide the design of future items and tasks.

The validity of an assessment is its most important quality. Over the years, many scholarly articles have captured the evolution of the term *validity*, including definitions of the term, conceptual frameworks specifying different types of validity, and guidelines for implementing validity studies. Initial approaches to validity focused on criterion-related definitions. Assessment specialists assumed that the criterion measure provided the "real" value of the variable of interest and the validity of new measure was determined by how well it predicted the criterion scores. The newly developed measure was considered valid for any criterion for which it provided accurate estimates (Thorndike, 1918). In the 1950s, conceptions of validity began to change. Cureton (1950) argued that the essential question of validity was "how well a test does the job it is employed to do" (p. 621). Cronbach and Meehl's (1955) seminal article turned the assessment community's attention toward construct validity. Thus, by the mid-1950s, definitions of validity were re-conceptualized and based on criterion *and* construct-based models.

In time, construct validity was regarded as a unified framework for the study of an assessment's validity. According to Kane (1996), construct-based models of validity included many types of evidence of validity, including "content and criterion-related validity, reliability and the wide range of methods associated with theory testing." (p. 324).

In the past 20 years, the definition of validity has been further transformed. This time, assessment experts, including Kane (1992, 1994, 1996), Kane, Crooks, and Cohen (1999), Mislevy, Steinberg, and Almond (2003), Messick (1989, 1994) began to describe assessment as a process of systematic arguments and chains of inferences. Mislevy, in his article *Validity by Design* (2007), makes the point that Toulmin diagrams, introduced by British philosopher Stephen Toulmin (1958) provided terminology that sets forth the structure of a simple argument, composed of claims, data, warrants, backing, and alternative explanations. Henry Wigmore (1937) developed a system for charting the structure of arguments with multiple propositions, chains of reasoning, dependent claims, and various data in the context of a judicial analysis. Both Toulmin and Wigmore advanced assessment design, as well as other disciplines, by providing argument structures in the context of evidentiary reasoning (Bachman, 2005; Kane, 1992; Messick, 1989).

Validity came to be seen as an explanation rather than a prediction or correlation. Most importantly, validity was understood as an extended analysis of many kinds of evidence that could be collected and studied during the assessment design, development, and validation phases. Validation of an assessment took its place in the well-established tradition of proposing an interpretation of a particular phenomenon and subjecting that phenomenon to empirical and conceptual examination.

The conceptualization of validity as an argument and chain of reasoning is operationalized in the ECD process on which we are reporting and is further reflected in the assessment argument specified in the student, evidence and task models described by Messick (1994).

Assessment as an evidentiary argument. More than half a century ago, Cronbach and Meehl, (1955) noted that an assessment could be structured in terms of argument for the purpose of validation. In fact, although assessments might look quite different or be used for different purposes or in different contexts, they share the property that assessment, by its nature, is always a process of reasoning from limited evidence of what students say, do, and make in particular settings, to claims about what they know and can do more broadly (Messick, 1994).

Operationalizing the assessment argument in ECD. The view of assessment as an evidentiary argument is the foundation of the ECD framework. The goal of ECD is to develop a coordinated and coherent assessment or assessment system by fleshing out an assessment argument across five layers of work that begin with the analysis and organization of the conceptual domain to be assessed and culminate in the delivery, scoring and reporting of the assessment results to stakeholders. The design pattern tool was designed to fit within the ECD framework and to support assessment designers in the second layer of work, referred to as domain modeling. To provide the theoretical grounding of the design pattern tool and its interplay with the other elements in the assessment development process, we briefly introduce the five layers of work entailed when ECD is used to design an assessment.

Layers in the design process. Evidence-centered assessment design was first proposed systematically by Mislevy et al. (2003). Over the past decade, the principles, patterns, examples, common language and knowledge representations for designing, implementing and delivering educational assessment using the processes of ECD have been further elaborated (Mislevy & Haertel, 2006).

Figure 1 and Table 1 present each of the five layers of ECD as applied to assessment. Reading Figure 1 from the top down, we can see the successive refinement and reorganization of knowledge about the content domain and the purpose of the assessment being implemented—from a general substantive argument to an increasingly specific argument that identifies the elements and processes needed in its operation. Different experts may carry out the work at different layers of the design process. The ECD framework provides a common language that facilitates efficient communication among these layers. Table 1 characterizes each layer in terms of its role in the assessment design process, the key concepts and entities used, and the knowledge representations and tools that are used to achieve each layer's purpose. A brief introduction to each layer follows.

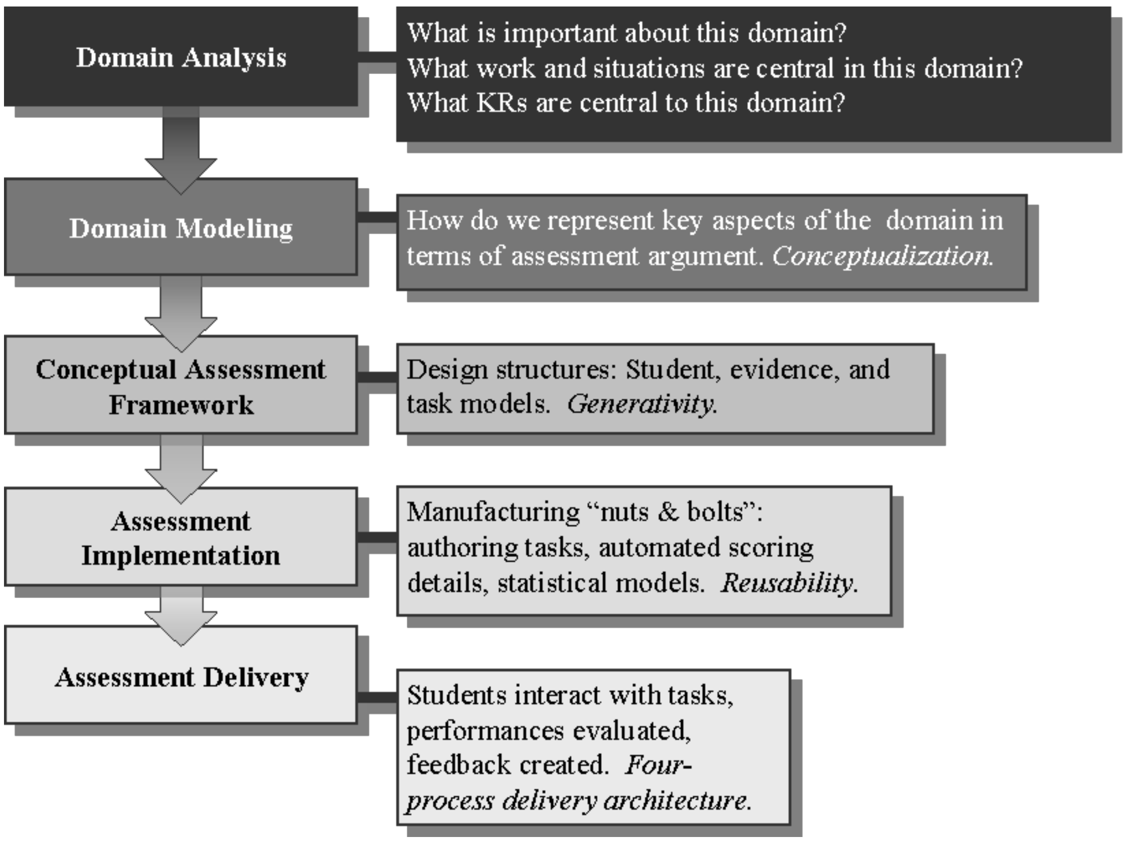


Figure 1. Layers of Evidence-Centered Design for educational assessment.

Table 1. Layers of Evidence-Centered Design for Educational Assessments

Layer	Role	Key entities	Selected knowledge representations
Domain analysis	Gather substantive information about the domain of interest that has implications for assessment; how knowledge is constructed, acquired, used, communicated.	Domain concepts, terminology, tools, knowledge representations, analyses, situations of use, patterns of interaction.	Representational forms and symbol systems used in domain (e.g., algebraic notation, Punnett squares, maps, computer program interfaces, content standards, concept maps).
Domain modeling	Express assessment argument in narrative form based on information from Domain Analysis.	Specifications of knowledge, skills, or other attributes to be assessed; features of situations that can evoke evidence; kinds of performances that convey evidence.	Design patterns; "big ideas", Toulmin and Wigmore diagrams for assessment arguments; assessment blueprints, ontologies, generic rubrics.
Conceptual assessment framework	Express assessment argument in structures and specifications for tasks and tests, evaluation procedures, measurement models.	Student, evidence, and task models; student, observable, and task variables; rubrics; measurement models; test assembly specifications; task templates and task specifications.	Algebraic and graphical representations of measurement models; task templates and task specifications; item generation models; generic rubrics; algorithms for automated scoring.
Assessment Implementation	Implement assessment, including presentation-ready tasks and calibrated measurement models	Task materials (including all materials, tools, affordances); pilot test data to hone evaluation procedures and fit measurement models.	Coded algorithms for rendering tasks, interacting with examinees and evaluating work products; tasks as displayed; IMS/QTI representation of materials; ASCII files of item parameters.
Assessment Delivery	Coordinate interactions of students and tasks: task-and test-level scoring; reporting.	Tasks as presented; work products as created; scores as evaluated.	Renderings of materials; numerical and graphical summaries for individual and groups; specifications for results files.

Domain analysis. As the first level, *domain analysis* marshals substantive information about the content domain. Assessment designers use this substantive information to understand the knowledge, skills, and abilities people use in a domain of interest, the representational forms they use, characteristics of good work, and key features of situations that commonly occur in the domain of interest. All of this information has important implications for assessment design, although usually most of the sources were neither originally created to support assessment nor presented in the structure of an assessment argument. For example, the National Science Education Standards (NRC, 1996), and state science standards provide key information at the domain analysis layer of work conducted in the design of a large-scale statewide, science assessment. For more specific examples of the work conducted at the domain analysis layer see the discussions of the development of a design pattern for observational investigation (Mislevy et al., 2009)[1] and a design pattern for experimental investigation (Colker et al., 2010)[2]. A thorough analysis of the content domain of interest, is prerequisite for generating a design pattern which is the product of the work conducted in the next layer of ECD called domain modeling.

Domain modeling. In the *domain modeling* layer, information identified in *domain analysis* is organized along the lines of the assessment argument. Without getting tangled in the technical details of assessment, this layer directs assessment designers to clarify what is meant to be assessed, and how and why to do so. Design patterns (DPs), as a tool, were developed as part of the original Principled Assessment Designs for Inquiry (PADI) project (see Mislevy et al., 2003[3]) to support work at the *domain modeling* layer of ECD. DPs help the assessment designer think through the key elements of an assessment argument in narrative form. Key attributes of DPs are provided in a subsequent section of this paper.

The three subsequent layers of the ECD framework involve the specification, implementation, and delivery of assessment tasks that build on the assessment arguments first sketched in domain modeling and represented in DPs. We review the three subsequent layers (i.e., conceptual assessment framework, assessment implementation, and assessment delivery) briefly to show the connection between the narrative form of assessment arguments that *design patterns* represent. See Almond, Steinberg, and Mislevy (2002) and Mislevy and Riconscente (2006) for further discussion of these three layers.

Conceptual assessment framework (CAF). The CAF concerns technical specifications for operational elements of the assessment. An assessment argument laid out in narrative form at the domain modeling layer is here expressed in terms of coordinated technical specifications, such as measurement models, scoring methods, and delivery requirements. The commonality of data structures and reusability of the CAF models offer opportunities to bring down the costs of task design, which is especially important for technology-enabled tasks. At the CAF layer, the PADI project produced another ECD-associated design tool, task templates, to guide the creation of families of tasks at a more detailed level than that of a DP (Baxter & Mislevy, 2005[4]; Mislevy & Riconscente, 2005[5]).

---

[1] http://ecd.sri.com/downloads/ECD_TR2_DesignPattern_for_ObservationalInvestFL.pdf
[2] http://ecd.sri.com/downloads/ECD_TR8_Experimental_Invest_FL.pd
[3] http://padi.sri.com/downloads/TR1_Design_Patterns.pdf
[4] http://padi.sri.com/downloads/TR5_IDFramework.pdf

Assessment implementation. The fourth layer, assessment implementation, includes activities carried out to prepare for the operational administration for testing examinees. These activities include authoring tasks, calibrating items into psychometric models, piloting and finalizing scoring rubrics, producing assessment materials and presentation environments, and training interviewers and scorers. Such activities are all in accordance with the assessment arguments foreshadowed in DPs at the domain modeling layer and specified in the CAF.

Assessment delivery. The final ECD layer, assessment delivery, includes presenting tasks to examinees, evaluating performances to assign scores, and reporting the results to provide feedback or support decision making. See Mislevy and Haertel (2006) for more details about kinds of tools produced by other research projects for the final two layers.

Instantiation of ECD in PADI assessment design system. In 2001, SRI International and the University of Maryland, University of California at Berkeley, the BioKIDS Project at the University of Michigan, and the FOSS Project at the Lawrence Hall of Science collaborated on the design and implementation of an online assessment design system referred to as PADI. PADI drew on new understandings in cognitive psychology and recent advances in measurement theory and technology to create a conceptual framework and supporting software tools for use in the design of assessments. The PADI project developed a set of online assessment resources that support the design of evidence-centered assessments, which were initially applied in the domain of science inquiry. The online assessment system was designed, however, to support the development of assessments in any content domain, based on any theory of learning (e.g., cognitivist, behaviorist, socio-cultural) and serving any assessment purpose (e.g., large scale, statewide assessments, diagnosis of gaps in learning, benchmark examinations, summative examinations). The conceptual framework that underlies the PADI assessment system uses a template series that needs to be completed by the assessment designer. The designer is prompted to articulate the student, evidence, and task models which undergird the assessment argument that is central to the ECD process. When articulated, these models can help establish the claims and warrants that are required in the ECD process.

A DP*,* which is intentionally broad and not technical, enables designers to fill in a template that implicitly contains the assessment argument. Centered around the constructs of interest (referred to as knowledge, skills, and abilities, or KSAs), a DP is organized in a way that leads toward the more technical work of designing particular tasks or task templates. A task template is a more complex and hierarchical data structure populated with definitions of student model variables, work products, evaluation procedures, task model variables, and the like, thereby rendering a general blueprint for a family of assessment tasks. Using template structures makes it possible to create assessment elements and processes that can be reused for different assessment purposes. The PADI assessment design system, which is focused on the design of items and tasks, also is supported by an assessment delivery system that includes a calibration and a scoring engine. In this paper, we focus on the highest level template supported in PADI: the DP.

Attributes of PADI DPs. PADI assessment DPs (analogous to those in architecture and software engineering) capture design rationale in a reusable and generative form in the domain modeling layer of

---

[5] http://padi.sri.com/downloads/TR9_ECD.pdf

ECD. They help designers think through substantive aspects of an assessment argument in a structure that spans specific content domains, grade levels, and purposes (Mislevy et al., 2003). Assessment designers working with the PADI design system use the web-based interface shown in Figure 2 below.

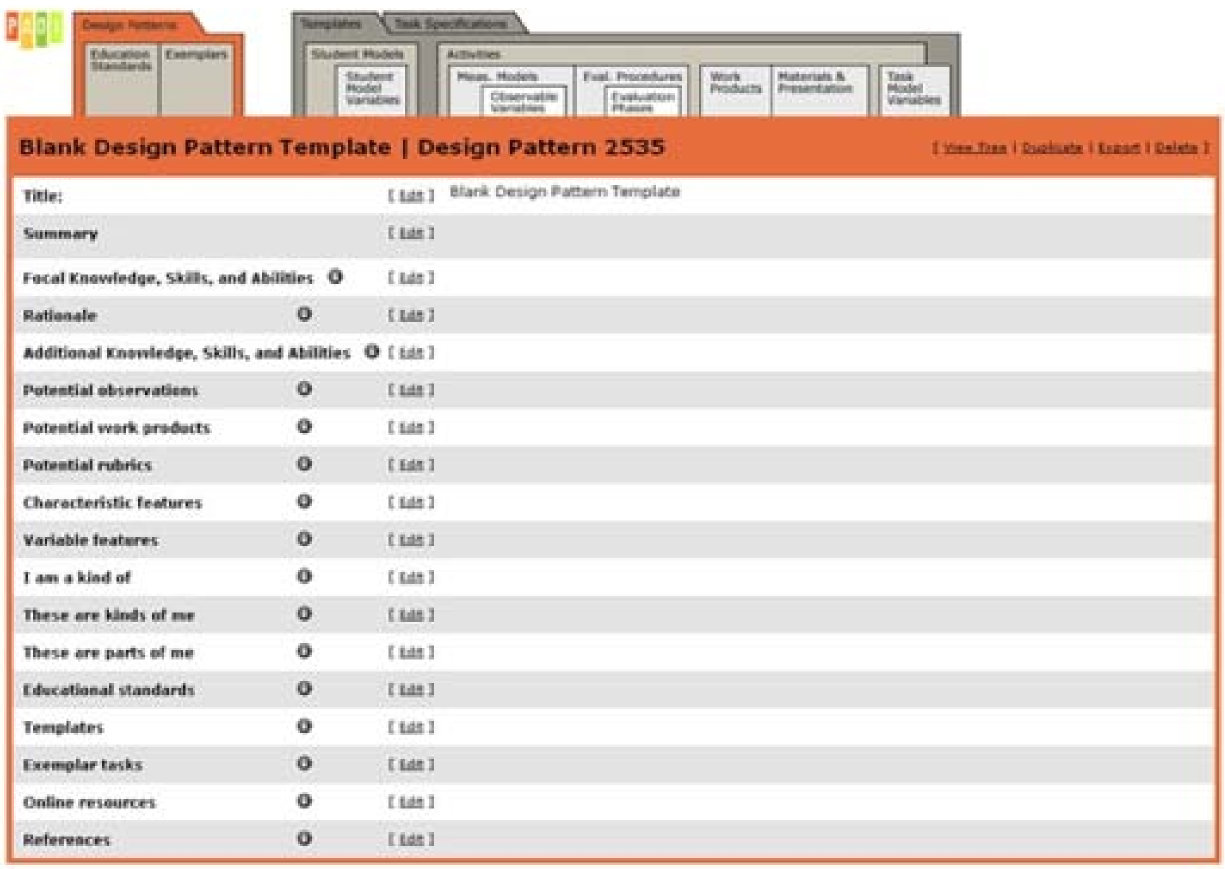


Figure 2. PADI design pattern template (blank).

The key attributes of a DP are described below:

Focal knowledge, skills, and abilities (KSAs). Focal KSAs are the primary knowledge/skills/abilities targeted by the DP. These are the competencies of interest that will ultimately be the focus of an assessment.

Rationale. The rationale describes the nature of the KSAs of interest and how they are manifested and articulates the theoretical connection between data to be collected and the claims to be made.

Additional knowledge, skills, and abilities (KSAs). These are the other knowledge/skill/abilities that may be required in a task (Mislevy et al., 2003). AKSAs may include declarative knowledge and prerequisite skills in a content domain. Or they may include non-construct relevant knowledge and skills that are needed for success on the item/task but not the target of the assessment. For tests of academic subjects, the abilities to "see" and "hear" are typically additional KSAs. On the other hand, for assessments of sight and hearing, respectively, sight and hearing are likely to be defined as focal KSAs. There are many disabilities that involve impairments of sight, hearing, or both (e.g., blind, low vision, color-blind, deaf, hard to hear, deaf-blind). Cognitive issues such as dyslexia, attention deficit, and executive processing limitations can also be addressed using additional KSAs. Deficits in such additional KSAs can cause unduly low scores among test takers with disabilities. In order to address the needs of students with disabilities in the assessment design process, we integrated a UDL framework with the PADI assessment design system. A brief description of the UDL framework is presented in the next section.

Potential observations. Features of the things students say, do, or make that constitute the evidence.

Potential work products. Some possible things one could see students doing that would give evidence about the KSAs.

Potential rubrics. Scoring schemes that turn students' work products into observable variables (scores).

Characteristic features. Characteristic features of the assessment are the features that must be present in a situation in order to evoke the desired evidence about the focal KSAs (Mislevy et al., 2003).

Variable features. Variable features are described as features that can be varied to shift the difficulty or focus of tasks (Mislevy et al., 2003). Variable features have a particularly significant role with respect to test takers with disabilities and other sub-populations (e.g., speakers of minority language). Much of our attention will be on manipulating variable features to reduce or eliminate demands for additional KSAs in which there may be a deficit while making sure (to the extent possible) that demands for focal KSAs have not been changed.

Instantiation of UDL framework in the PADI assessment design system. UDL helps to meet the challenge of diversity by suggesting flexible assessment materials, techniques, and strategies (Dolan, Rose, Burling, Harris, & Way, 2007). The flexibility of UDL empowers assessors to meet the varied needs of students and to accurately measure student progress. The UDL framework includes three guiding principles that address three critical aspects of any learning activity, including its assessment. The first principle, multiple means of representation, addresses the ways in which information is presented. The second principle is multiple means of action and expression. This principle focuses on the ways in which students can interact with content and express what they are learning. Multiple means of engagement is the third principle, addressing the ways in which students are engaged in learning (Meyer & Rose, 2006;

Rose & Meyer, 2002; Rose, Meyer, & Hitchcock, 2005). These principles provide structure for the infusion of UDL into assessment design.

Principle I. Provide multiple means of representation (the "what" of learning). Students differ in the ways that they perceive and comprehend information that is presented to them. For example, those with sensory disabilities (e.g., blindness or deafness), learning disabilities (e.g., dyslexia), language or cultural differences, and so forth, may all require different ways of approaching content. Others may simply grasp information better through visual or auditory means rather than printed text.

Principle II: Provide multiple means of action and expression (the "how" of learning). Students differ in the ways that they can interact with materials and express what they know. For example, individuals with significant motor disabilities (e.g. cerebral palsy), those who struggle with strategic and organizational abilities (executive function disorders, ADHD), those who have language barriers, and so forth, approach learning tasks very differently and will demonstrate their mastery very differently. Some may be able to express themselves well in writing text but not oral speech, and vice versa.

Principle III: Provide multiple means of engagement (the "why" of learning). Affect represents a crucial component to learning. Students differ markedly in the ways in which they can be engaged or motivated to learn. Some students enjoy spontaneity and novelty, while others do not, preferring strict routine. Some will persist with highly challenging tasks while others will give up quickly.

In reality, there is no one means of representation, expression, or engagement that will be optimal for all students in all assessment situations; providing multiple options for students is essential. In addition to the three principles of UDL expressed above which provide general guidance on the infusion of UDL into the assessment, we identify particular categories of student needs (perceptual, expressive, language and symbols, cognitive, executive functioning, and affective) that are required for successful performance on assessment tasks, but are not the targets of interest. Using PADI, we can select assessment task features that provide non-construct relevant supports to address student needs.

Three approaches to UDL integration. Based on our work across multiple design projects, we have derived three approaches to the integration of UDL in an ECD process, shown in Table 2 below. SRI takes the perspective that the selection of an approach to infusing UDL must be considered in terms of the assessment's goals, context, target population, existing design assets and assumptions of the design and development process. The three approaches represent three degrees of implementation of UDL.

- In the first approach, the principles of UDL are used to guide analysis of task demands or AKSAs that can challenge students but are not the target of the assessment. Six categories of AKSAs are specified, based on the UDL principles. The assessment designer considers each category of AKSAs to identify possible sources of construct-irrelevant variance and then generates possible design solutions to mitigate the effects of the construct-irrelevant variance. This first approach is conducted using PADI design patterns.
- In the second approach, each of the six categories of AKSAs are linked to a set of variable features that can be used in the task to support the identified AKSAs. The assessment designer can select from among the set of variable features presented those they wish to implement. Designers can also generate variable features to be used in task design. The second approach is also conducted using PADI design patterns.

- In the third approach, detailed AKSAs are identified within each of the six categories. These AKSAs offer even more specific guidance to designers. In addition each AKSA is linked to a set of variable features that can be instantiated in the task to mitigate the influence of AKSAs. Using PADI design patterns, the assessment designer is required to consider each possible AKSA; they select those that are appropriate to their task. This selection dynamically generates a PADI task template that provides the menu of variable features to be considered in design and documents the designer's choices. The third approach is conducted using both PADI design patterns and PADI task templates.

Each approach addresses issues of fairness through the UDL lens. It is the constraints, assets, and assumptions underlying each design and development process that will determine the degree to which analysis of AKSAs and their links to variable features can be conducted. *Designers can apply any of these approaches to a design process that forward engineers new items or to the modification of existing items.*

The prior section of this paper was devoted to validity and how it is instantiated in the ECD process. In the following section, we will address the technical qualities of reliability, precision of measurement, and fairness.

The Challenge of Reliability

This section of the paper describes how the use of ECD is able to enhance the reliability of an assessment. The articulation of the assessment argument, which is the essence of the ECD process, *should* result in items and tasks in which the knowledge and skills being assessed, the evidence being collected, and the features of the items and tasks that elicit that evidence are well aligned. This coherence is represented in the student, evidence, and task models. Thus, a careful implementation of ECD *should* result in items/tasks that share construct-relevant content and features; thereby, increasing the reliability and validity of the score interpretations. Although such an argument is plausible, the reliability of assessment items and tasks remains unpredictable. Even when assessment items are aligned to the target content and sources of construct-irrelevant variance have been identified and reduced, there are still Inconsistencies in task administration and scoring, as well as other sources of error, that contribute to unreliable measurement.

Generalizability studies are well-suited to estimating the reliability of hard-to-assess technology enhanced assessment tasks. Many hard-to-assess science constructs are multi-dimensional (e.g., the need to assess both science content and practices). When these assessment items and tasks are delivered via technology, the number and kinds of additional KSAs that are required for successful performance increases. The internal structure of these complex assessment situations is best estimated using a "g" study to determine the extent to which the assessment scores reliably vary across the different facets of the measurement situation, such as, raters, items, and tasks (American Educational Research Association, American Psychological Association, & National Council for Measurement in Education [AERA, APA, & NCME], 1999). We have not conducted a generalizability study to date, but we have recently proposed several such studies to examine the quality of assessments being produced using the ECD approach.

The Challenge of Precise Measurement

Because ECD is a construct-centered design approach, Kane's writings on precision of measurement are particularly germane to our work (Kane, 1996, 2010, 2011). When constructs are defined narrowly, indicating the kinds of observations used to produce the observed scores, we can be confident about the accuracy of our limited conclusions. For example, we can define a construct for planning an experiment in terms of the particular science domain and context. If, however, the constructs are defined broadly, such as proficiency in life sciences, there is more inferential risk and uncertainty. Even domains that are very narrowly defined may require the design and development of several tasks, scoring of assessment responses by many raters, and several testing administrations. Thus, even in narrow domains, there are many possible sources of error.

ECD allows the definition of the construct of interest to be as narrow or broad as the assessment designer would like. Attention to the specification of the student, task, and evidence models draws the designer to consider issues of precision of measurement from the outset of the design process. In implementing the ECD process substantial effort is invested in identifying, estimating, and controlling sources of measurement error. Some of this is accomplished in specifying the additional KSAs that serve as counterfactuals and are able to explain students' poor performances. Use of technology, as well as rich item and task contexts and UDL features, need to be addressed in terms of additional KSAs in order to mitigate measurement error.

Challenge of Fairness

Achieving fairness in assessment through the integration of ECD and UDL has been a key goal in our work. The 1999 edition of the *Standards for Educational and Psychological Testing* (APA, AERA, NCME) recognized fairness as a fundamental issue of test validity. Our goal to build fair assessments is expressed in thoughtfully applying the discipline of ECD in order to provide *all* students with an opportunity to perform at their best in assessment situations. The infusion of UDL into the assessment design from the very beginning is critical to removing barriers to accessibility. The *Standards* specifically address the incorporation of UDL as a means for developing tests that are fair to all examinees.

Much of the practice of ECD is focused on the identification of sources of construct-irrelevant variance that can result in faulty interpretations of scores. Assessment design choices that are not carefully examined can contribute to the development of test items that employ unfamiliar language and syntax, poorly understood social and cultural item contexts and task stimuli, as well as modes of representations (visual, oral, behavioral) that be systematically biased against sub-groups with limited access to those modalities. Fairness in the assessment situation requires that task contexts be equally familiar, appropriate, and accessible to all students. Articulation of task models from the beginning of the assessment design process reduces the likelihood that items and tasks will be developed that are biased against particular groups.

More recently, with the advent of technology enhanced assessment delivery systems, students who are unfamiliar with particular hardware and software are at disadvantage in some computer-based testing situations. In particular, those from diverse socio-economic and cultural groups, diverse language backgrounds, and individuals with disabilities need to be considered when technology-based items and tasks are presented.

How does ECD guard against the design of unfair tests? The practice of ECD makes the assessment designer aware of the many kinds of additional KSAs that can contribute to faulty inferences about students' assessment performances. In our work, we consider three broad types of additional KSAs: (a) cognitive background (sometimes referred to as prerequisite knowledge), (b) student needs (perceptual, expressive, language and symbols, cognitive, executive processing, and affective) and (c) technology-related knowledge and skills. As mentioned earlier in this paper, the student's needs are identified based on principles of UDL. These needs, if not addressed in the testing situation, can result in a student's poor performance even though she may possess the knowledge and skills of interest.

In applying the ECD process, we identify the focal KSAs that compose the construct we are assessing. Next, the knowledge and skills that are required to successfully complete an item, but are not the target of the assessment, are identified and labeled as additional KSAs. Then, we reduce the influence of these additional KSAs on a student's assessment performance by identifying variable features that can be designed into the assessment. These variable features are used to provide non-construct relevant supports. This process of linking the additional KSAs to variable features that support performance without compromising the measurement of the construct of interest guards against inappropriate interpretations of the test score.

During the ECD process, we also identify the potential observations needed to provide evidence of whether a student has acquired the knowledge and skills of interest. In articulating these observations, the assessment designer considers whether all students have an adequate opportunity to acquire the knowledge and skills required to perform the focal KSAs. Thus, the role of "opportunity to learn" is prominently considered during the design and development process.

New Assessment Challenges

In the following section of this paper, we will address three challenges that are not foundational to technical qualities of assessment, but rather reflect the context in which current assessments are being designed and delivered to today's students.

The Challenge of Engaging Today's Students

The national dialogue around student assessment now encompasses all students as compared to prior compartmentalization that excluded students with disabilities from accountability metrics. From a policy perspective, the definition of "today's students" now includes students who might have previously been exempt from state-level assessments given their special education designation. Beginning with the No Child Left Behind Act in 2001, states must include students with disabilities in reports of performance and progress. Developing assessment design frameworks that can produce assessment tasks appropriate and accessible for a wide range of students requires new tools and approaches, including those that can interface with frameworks underlying instructional and assessment materials (i.e., UDL) that *are* specifically designed to meet the needs of students with disabilities.

Moreover, not only has the range of students being tested increased, state-of-the-art of assessment design now includes the use of context-rich, situated tasks often presented in online or computer-based testing environments. These tasks often involve story narratives to increase student engagement and motivation and, theoretically, present students with conceptual links previously unavailable in paper-and-pencil testing to support student's cognitive engagement. Technology enhanced tasks also support the use of open-ended, interactive contexts that focus on student reasoning processes, permit multiple solution paths, and present varied stimuli and concepts that were impossible in paper-pencil assessment (e.g., students can fold proteins to create new chemicals to eradicate diseases (Williams, 2009).

The same characteristics of technology enhanced tasks that are desirable in terms of assessing students' extended reasoning may present accessibility barriers to students with disabilities. Students with cognitive disabilities, for example, may be overwhelmed with extended reasoning tasks by virtue of their cognitive load, memory demands, or executive functioning demands. Research has shown that some combinations of stimuli can overwhelm students' working memory. Chandler and Sweller (1992) documented the split attention effect where students' learning was hampered from the combination of animation, narration and on-screen text as compared to just animation and narration.

An ECD process can guide designers in the application of UDL principles as they consider ways to recruit interest, sustain effort, and provide options for self-regulation. For example, designers might consider ways that students can monitor their progress as they work through a task. Variable features that could be implemented to help students monitor their progress could include a progress bar, intermittent messages to the student about their progress, or interactive navigation to support students' working through an extended task. Illustrations of design choices made in the development of three science assessment tasks based on principles of UDL are presented in the next section of the paper including those aimed at limiting unnecessary sources of cognitive load, but also content-related design decisions (e.g., those involved in the judicious inclusion of relevant subject matter appropriate for different cultural, socioeconomic, ethnic, disability, and gender groups) that were made to maximize student engagement.

The Challenge of designing Assessment Tasks That Link to Day-To-Day Instructional Practices Aligned With Student Needs

Within domain modeling of the ECD process, designers articulate design elements that reflect the assessment of that domain but also reflect aspects of instruction in that domain. Within a domain, designers specify the KSAs including the canonical knowledge representations used in that domain. These are also, in instructional terms, intended learning goals (Krajcik, McNeill, & Reiser, 2008). Designers identify the work products that students would be expected to produce to demonstrate proficiency in a domain. In addition, designers identify qualities of those work products that provide evidence of student understanding, and thereby define the kinds of activities in students would engage in an instructional context.

Equally important to domain modeling is the identification of additional KSAs that may be required of students when learning about a domain, but are not the long-term outcomes of interest. For example, as students work with scientific data, related math skills will be drawn upon. These additional KSAs are often intertwined with the focal or target KSAs of an assessment of that domain and these interactions must be well understood so that a student's performance can be linked to an additional KSA and supported by the classroom teacher. AKSAs provide teachers with information that a student may have a gap in related knowledge that can be ameliorated. For example, additional KSAs reflecting a need for technology skills may indicate that the teacher needs to provide additional instruction in the use of software and hardware.

Variable features, articulated in domain modeling, can take the form of the very same scaffolds that are the critical feature of instruction, used to ensure that instructional content is accessible to students. For example, use of multiple representations in instruction can help make instructional concepts salient (Ainsworth, 2006) and might also be used in an assessment design to ensure that focal or target KSAs are the primary focus of a task, rather than additional KSAs. Similarly, vocabulary support, demonstrations of processes, and contrasting cases might be used in both instructional and assessment contexts. Taken together, the set of variable features defined in domain modeling represent the wide range of needs present in classrooms and, ultimately, in the assessment context.

ECD provides a set of tools and vocabulary to model the domain of interest, effectively modeling many aspects of the instruction that would be used in a domain. As will be illustrated later in this paper, by combining the ECD and UDL frameworks, assessment designs can be linked to, if not embody, the day-to-day instructional contexts of students and address the range of student needs present in classrooms.

The Challenge of Financial Feasibility

There are trade-offs between the use of traditional assessment design methods and using an ECD approach. SRI has substantial experience implementing ECD for small and large scale assessment design and development. We have created about 250 design patterns over the past decade and developed around 500 associated items. We have used ECD to build small assessments that served the purposes of individual research projects and evaluations, as well as large scale state assessments. We have applied ECD in a variety of domains—English/language arts, mathematics, science, economics, computer science, and IT workforce development. We have produced assessments for K–12, community college, and University students. In building these assessments, we have developed selected and constructed response items, suites of items, scenario-based items, and short essays. Thus, we are well aware that ECD has a fair amount of implementation costs in the initial design phase. These costs reflect the development of the design patterns and template series. Over the course of the assessment development, however, these initial costs should reduce the number of items that need to be developed as fewer items will be expected to be rejected during the review process. It is our experience that the ECD design and development process results in the development of items that meet the criteria of internal and stakeholder review committees and are better aligned with the content standards and design frameworks that they are intended to address.

As development proceeds, other returns on investment occur. Design patterns and templates can be used to create clones of discrete items, by changing their surface features. Having ample documentation on the design decisions that were used to create these items (design patterns) provides value information that can be readily applied to the creation of clones, thereby reducing the cost of the development.

Scenario-based items can be developed from one or more design patterns. Variants, which are tasks that have been altered significantly in terms of their context or the particular knowledge and skills they measure, can be more quickly produced by using the same design pattern or task template but applying the attributes within the design pattern/template to: (a) different domain (reading charts and graphs in mathematics vs. reading charts and graphs in science), (b) subdomain (understanding complex systems in life science vs. physical science), (c) topical areas within a domain (understanding how to plot points on a graph versus interpreting the intercept of a graph), or (d) significantly increasing the complexity of the evidence being collected (interpreting results of a simple experiment with one independent variable vs. interpreting results from an experiment with two independent variables and two dependent variables). The development of these variants is made more efficient by the reusability at the design layer as well as at the development layer. Attributes from different design patterns can be combined to enhance existing tasks to include additional skills or to modify the focus of the tasks. The process of combining attributes can also be used to create new tasks. The modeling of the assessment that was conceptualized and documented at the domain modeling layer of ECD now can be re-used to facilitate the design and development of future items and tasks. The hard won understandings and relationships that assessment designers poured over in the early stages of ECD now confer benefits in the quality and rigor of future items and tasks.

ECD is an iterative process in which multiple experts are providing input into the different layers of task development. The work done on at a particular layer of ECD is not performed in isolation and several different types of experts may be working on integrated at multiple layers of the process simultaneously. (See Figure 1 for a depiction of the multiple layers.) This process helps ensure the coherence of the assessment argument. The alignment of the student, task and evidence models that are being articulated help ensure that an item/task is measuring what it is supposed to be measuring, as there are multiple checks at each layer of the process and the decisions that went into the item design process are documented clearly and can be refined, if needed. The application of an ECD process among experts from different backgrounds establishes a common language of assessment design and development that individuals from different fields such as content specialists, teachers, special educators and assessment specialists can acquire and use. This ability to communicate provides support to ensure that the final item/task conforms to the vision of the group of experts who contributed to the development of the design patterns, task templates, items and assessment which is eventually produced (Baxter & Mislevy, 2005).

The ECD process as depicted in Figure 1 not only supports item development, but the whole process by which an entire assessment is designed, administered and scored. Documentation of the domain content, item stimulus materials, technology requirements, and evidence models supports the versioning of the different components that are created en route to the development of an operational assessment. ECD supports the documentation and compilation of a library of design patterns, tasks templates and associated items.

Finally, ECD confers cost benefits in that it specifies the information required to deliver the assessment. For a thorough discuss of this layer of development see the paper by Almond, Steinberg, and Mislevy (2002) on the four-process delivery system. The four processes specified include: presentation process, response processing, summary scoring process, and activity selection process. ECD calls for detailed specificity of technical information required by software engineers for the purposes of task presentation, delivery and scoring (e.g., response capture and asset management). As technological capabilities advance in terms of scoring, delivery and adaptivity, ECD is able to provide a framework for the systematic incorporation of these new approaches.

Exemplar Science Task I: Scenario-Based, Technology Enhanced—Pinball Car Race

Overview

The Pinball Car Race is a middle-school science assessment task that was designed to test a student's knowledge of both science content and practices. The science content being assessed is knowledge of forms of energy in the physical sciences. In particular, knowledge of potential and kinetic energy and that objects in motion possess kinetic energy. In the assessment task, students observe the compression of a spring attached to a plunger, the same type of mechanism as those used to put a ball "in play" in a pinball machine. The student observes that when the plunger is released, it pushes a toy car forward on a racing track. The potential energy in the compressed spring is transformed, on the release of the plunger, into kinetic energy which moves the toy car along the racing track. The student is then asked to plan an investigation to examine how the properties of the compression springs influence the distance the toy car travels down the race track.

In the following section, we identify the cross-cutting concepts, core disciplinary ideas and science practices, drawn from the new framework for K–12 science education (National Research Council, 2012) to which the task is aligned. We describe the assessment scenario, including the number and format of the questions posed to students and the types of technology employed in the task. Finally, we present the design pattern, Experimental Investigation, which guided the task development.

Cross-cutting concepts, core disciplinary ideas, and science practices. Using the newly developed *Framework for K–12 Science Education* (National Research Council, 2012), we identified the cross-cutting concepts to which the task was related. In this case, the cross-cutting concepts were "energy and matter: flows, cycles and conservation." These concepts are introduced to students by providing them with opportunities to track changes in energy and matter into, out of, and within systems. There were three core disciplinary ideas that align with the pinball task: PS3 A, B, and C. These ideas are, respectively, definitions of energy, conservation of energy and energy transfer, and relationship between energy and force.

The pinball task aligns to two science practices as set forth in the K–12 science framework (National Research Council, 2012). These practices are: planning an investigation (Practice 3) and analyzing and interpreting data (Practice 4). The following paragraphs describe how Science Practices 3 and 4 are incorporated in the assessment task.

In the pinball task, a student poses a hypothesis that can be investigated using the simulation presented in the task. The student selects three of nine compression springs to be used in the pinball plunger and initiates a simulation, which generates a table of data that illustrates how far the race car travelled on the race track using the particular compression springs that were selected. Data representing three trial runs are presented each time the simulation is initiated. The student runs the simulation twice for a total of six trials of data for each of the three springs selected.

The properties of the compression springs used in the simulation vary along two dimensions: number of coils and thickness of the wire. The student poses a hypothesis about how these properties might influence the distance the race car travels after the spring plunger is released. The experiment requires that students vary or control each of the properties of the spring. The student decides whether one or both of the properties of the spring will serve as independent variables and whether one or more of the variables will serve as control variables. In establishing the role of the properties of the spring as independent and control variables, a student fulfills one of the essential elements required in planning an experimental investigation. In completing the assessment task, the student also decides how many trials of data are needed to produce reliable measurements and whether the properties of the springs need to be varied and additional data collected before the hypothesis can be confirmed or disconfirmed.

Once a student has decided on the levels of the properties of the spring to be tested, the simulation produces a table of data and the student must graph the data and analyze the results. Based on the results, the student may revise their hypothesis and run the experiment again; changing the settings of the variables (e.g., select springs with more or fewer coils or springs with wider or thinner coils) to reflect a revision of their model of how the properties of the springs influence the distance the toy car travels. Finally, the student interprets their data in terms of the relationship of the properties of the spring and the distance the toy car travels. The student then indicates whether their hypothesis is confirmed or disconfirmed.

Description of assessment scenario. Eleven scenes and 14 questions comprise the pinball scenario assessment task. Of these 14 questions, three are content-focused and 11 are focused on science practices. Our intent was to build a task that would provide evidence of both science content and practice, but students' proficiency on the science practices was of greater interest; thus, there were a greater number of questions associated with science practices. We might want to use the scenario to identify particular aspects of planning an investigation that might cause difficulties for a student. Thus, the task includes questions about aspects of the investigation, including: posing a hypothesis, setting up an experiment, and graphing, interpreting, and explaining data displayed on a graph.

Five types of technology are implemented in the assessment task. The scenario begins with an animation that illustrates how the spring-loaded plunger is pulled back and released which causes the movement of the toy car down the race track. There are several open-ended questions that require the student to enter text into a text box, including a justification of the student's hypothesis, an explanation of the spring settings chose for the experiment and interpretations of the data presented in the tables and graphs. The student also is asked to make selections of the level of the properties of the spring to be varied or controlled using a drop down menu. When the student selects the levels of the properties of the spring, a table of data is generated with values for the springs selected and each of these springs is tested three times. Each student runs the experiment twice in order to complete the assessment task.

Design pattern used: Experimental investigation. The pinball assessment task was designed using the Experimental Investigation design pattern. Appendix A contains a copy of this design pattern, which was created as part of the NSF-funded project, *An Application of Evidence-Centered Design to States' Large-Scale Science Assessment.* See the project's technical report, Colker et al. (2010), for a detailed description of the attributes of the design pattern. See Mislevy et al. (2003) for an introduction to the design pattern tool.

Design and development process: Pinball Car Race. Two constructs are addressed in the pinball task. The first, focused on physical science content, is the broad construct of "forms of energy," in particular, potential and kinetic energy. The construct of planning and conducting an experimental investigation, however, is a science practice that cuts across all of the science disciplines. Most of the design and development effort associated with pinball has been focused on the science practice construct. If the pinball task were to be used to measure both physical science content and the science practices, additional content items would need to be integrated into the scenario.

Appendix A presents the experimental investigations design pattern. It supports the writing of storyboards and items that address scientific reasoning and process skills in planning and conducting experimental investigations. In experimental investigations, it is necessary to manipulate one or more of the variables of interest and to control others while testing a prediction or hypothesis. This contrasts with observational investigations, where variables typically cannot be manipulated. This design pattern may be used to generate groups of tasks for science content strands amenable to experimentation. In order for students to have a well-rounded understanding of the scientific method, they need to be familiar with the context and methods of experimental investigations.

Focal knowledge, skills, and abilities. The relevant focal KSAs are as follows:

- Ability to identify, generate, or evaluate a prediction/hypothesis that is testable with a simple experiment
- Ability to plan and conduct a simple experiment step-by-step given a prediction or hypothesis
- Ability to recognize that at a basic level, an experiment involves manipulating one variable and measuring the effect on (or value of) another variable
- Ability to identify variables of the scientific situation (other than the ones being manipulated or treated as an outcome) that should be controlled (i.e., kept the same) in order to prevent misleading information about the nature of the causal relationship
- Ability to interpret or appropriately generalize the results of a simple experiment or to formulate conclusions or create models from the results

The student behaviors or performances/products that will be accepted as evidence of the KSAs in the pinball assessment task are specified as potential observations and works products.

Potential observations. The relevant potential observations are as follows:

- Generate a prediction/hypothesis that is testable with a simple experiment
- Plausibility(Explanation) of plan for repeating an experiment
- Correct identification of independent and dependent variables
- Accuracy in identifying variables (other than the treatment variables of interest) that should be controlled (held constant) or made equivalent (e.g., through random assignment)
- Plausibility(Explanation) of design for a simple experiment
- Accuracy in critiquing the experimental design, methods, results, and conclusions of others
- Correctness of recognized data patterns from experimental data

Work products. The relevant work products are as follows:

- Select, identify, or evaluate an investigable question
- Complete some phases of experimentation with given information, such as selection levels or determining steps
- Identify or differentiate variables that do and do not need to be controlled in a given scientific situation

- Generate an interpretation/explanation/conclusion from a set of experimental results

Below are descriptions of the pinball tasks and stimuli that are intended to elicit the student performances and products specified above (see screenshots in Appendix A for Exemplar Task 1—Pinball). The characteristic and variable features that guided the development of the pinball task are presented below.

Characteristic features. The relevant characteristics are as follows:

- Presentation of situation of scientific interest where variables can be (or have been) practically altered to address a causal prediction
- Presentation of situation requiring the design or conduct of a controlled experiment
- Presentation or representation of an experimental design
- Presentation of observed result from an experiment requiring the development of explanations, conclusions, or models

Variable features. Variable features intended to influence the difficulty of the pinball task are specified below. Some of these variable features are UDL supports.

- Content (strand) context: Science topics within Earth/Space science, physical science, life science
- Which one of multiple phases of experimental investigation will be addressed: posing a hypothesis; identifying independent, dependent controlled variables; systematic collection of data; analyzing data; presenting results; interpreting results
- Qualitative or quantitative investigation or a combination
- Ease or difficulty with which the treatment (independent) variable can be manipulated
- Are manipulated variables given or to be determined
- The number of variables investigated and the complexity of their interrelationships
- Number of variables that need to be controlled to unambiguously study the relationship between the manipulated variable and the outcome variable
- Data representations used: tables; charts; graphs

Accessibility for individual students Table 2 below shows three approaches that our project teams at SRI International used to infuse UDL features into assessment tasks/items. SRI takes the perspective that the selection of an approach to infusing UDL must be considered in terms of the assessment's goals, context, target population, existing design assets and assumptions of the design and development process. In the experimental investigation design pattern, we would consider features of UDL to be considered under the attribute additional knowledge, skills, and abilities (AKSAs). While the Experimental Investigation design pattern does not specify individual additional KSAs suggested by the UDL framework, we did review the task and infuse UDL features during the design process.

Approach I was used to develop Task Exemplar I: Pinball Car Race. Table 3 illustrates how UDL Principles were infused in this task.

Table 2. Three Approaches to the Infusion of UDL in Assessment Items/Tasks Within a Particular Item/Task Context

UDL approach	The method of UDL infusion	Item/task exemplar
I	Attend to UDL Principles	Task Exemplar I: Pinball Car Race
II	1. Identify sources of construct irrelevant variance for each focal KSA 2. Link to Task Model Variables as means to mitigate construct irrelevant variance	Item Exemplar II: Bicycle Rider
III	Use task templates to prompt for links between additional KSAs and task model variables including support for background (cognitive) additional KSAs and UDL supports	Task Exemplar III: Recycling

Table 3. UDL Principles (Categories of Students' Needs) Supported by Variable Features

UDL principle (category of student need)	Potential task model variables addressing UDL principle
Perception	Screen presentation will include Variable font size Option for altering screen contrast Option for magnification or zoom Optional text-to-speech
Expression	Range of response options required (radio buttons, drop down menu, text input) Range of student support for producing response (speech-to-text)
Language and symbols	Provision of multiple representations of symbols (linguistic labels for symbols, define abbreviations, illustrations for key variables, etc.). Provision of definitions of nonconstruct relevant terminology Use of students' dominant language
Cognition	Use of a response template (cloze item format) Use of context to heighten salience Highlighting key terms and ideas (bold and underline) Use of multiple representations (data in a table, graph, and text; illustrations of variables Support for memory transfer (Automatic transfer of student response to new problem situations) Alternative conceptualization (pinball animation of potential and kinetic energy)
Executive functioning	Breaking task into manageable units Icons to encourage thinking and reflection (once you click next you cannot go back) Onscreen progress monitoring
Affective	Use of scenario or real-world context to heighten engagement Age-appropriate materials Interactive narrative (selection of parameters for experiment)

Appendix D contains the completed accessibility forms provided by the symposium leaders. We completed these forms for the Pinball Race Car task and found that while they helped identify the need for accommodations to support students with particular disabilities, the grain size of the solutions we identified were at a broader and more general level than the solutions we typically apply in our work. For example, using the ECD-based approach illustrated in this paper, we often infuse UDL practices at the level of individual items, thereby specifying changes in the assessment prompts, stems, distracters, and graphics. When we completed the accessibility forms, we found that most of the solutions we identified were more typical of accommodations such as large text size, translation of text in dominant language, or use of progress monitoring bars. These solutions apply to the overall assessment in general and apply less frequently at the item level; nor do these changes reflect the need for more thoughtful design of the cognitive demands placed on students, such as the presentation logic of items, metacognitive processes, and use of advance organizers.

Specification of technology options. The options for uses of technology within pinball and the rationale for the choices made are reflected in task templates and in task model variables. In designing and developing pinball the templates series was constrained to one template—the experimental investigation design pattern. The design pattern attributes that are encompassed within the task model, include variable features and may include technology requirements and enhancements. Below we address several technology considerations that were addressed during the design of the pinball task.

Programming environment. The programming environment that will likely be used when the task is fully implemented is a Java web application on the back end with HTML 5, and Javascript, and CSS on the front end. HTML was chosen for the universal availability of web browsers at schools.

Task logic and presentation (item elements and order). The logic and structure of pinball was determined by the narrative structure of the scenario—planning and conducting an investigation. The number of scenes allowed was constrained by the client as was the layout of screens. Background text and graphics were placed on the left-hand side of the screen and the prompt and response capture on the right-hand side of the screen. This layout was used consistently throughout the scenario. Principles of UDL, as well as good assessment practice argue for the presentation of item/task stimuli in proximity to relevant text and item prompt.

Task stimuli. Options for task stimuli included animations, simulations, drag and drop, gridded items and interactive graphs. The final selection of stimuli included was driven by the standards to be covered, the desire to minimize extraneous information, attention to the principles of UDL and the familiarity of students with technology enhanced tasks of a similar nature. The amount of text presented for any item was minimized. Particular requirements of task content also determined appropriate use of stimuli—for example, multiple representations of a single phenomena might be required. For purposes of UDL, we might present data from an experiment in a prose narrative, in a tabular form and in a graph.

Response capture. Several forms of response capture were considered, including: drag and drop, text, radio buttons, interactive graphs (possibly linked to equations constructed by student, data points, or simulations), speech, and so on. For the pinball task, the assessment designers chose text capture, radio buttons, interactive graphs, a simple simulation, and cloze procedures to align to the focal KSAs of the Experimental Investigation design pattern. For example, to allow students to show evidence of their ability to plan an experimental investigation, in this case, an experiment about the properties of compression springs and their influence on the distance a toy car travels on a race track, students use drop-down menus to select levels of independent and control variables, set parameters and run a simulation that produces a table of data, use interactive graphs to display their data, and provide written explanations of scientific phenomena.

Adaptivity. The assessment context in which pinball was embedded was not adaptive. Scoring and item selection are possible technologies to be employed at this level of design.

Practical obstacles to implementation (constraints on item presentation or student response format). The pinball task has not been implemented in classrooms, though it has been implemented in cognitive lab settings with a small number of middle school students. See Exemplars II (Bicycle Rider) and III (Recycling) for discussions of constraints of implementation.

Scoring methodologies, measurement models, and reporting. The pinball task was developed as a prototype. It demonstrates the use of design patterns as a tool to inform the development of technology enhanced, scenario-based tasks of hard-to-assess science concepts. The task could be scored several ways. It could be scored by summing those items aligned primarily to content standards and by summing those items aligned primarily to practice standards. Thus, the task would produce two subscores—one for science content and one for science practice. Or, the task could generate an overall score based on the aggregation of all items. In which case, the assessment designer would consider each item to be aligned to both content and practice standards. An argument would be made that the items aligned primarily to content (knowledge of potential and kinetic energy) are also aligned to identifying scientific principles (application is a science practice); in addition, the items aligned primarily to practice also require the use of knowledge of physical sciences. This latter scoring scheme is in keeping with the recently released K–12 Science Framework (National Research Council, 2012). The National Research Council proposes that the integration of science and practices in instruction and assessment is essential.

The constructed response items (e.g., provide an explanation for selection of independent variables, plot points on a graph, justify whether the hypothesis is confirmed or disconfirmed) would be scored using partial credit rubrics; whereas, the multiple choice items could be scored dichotomously (accurate/inaccurate). To score the partial credit items, we would have to train the scorers to use the rubrics in a consistent way in order to attain a reasonably high-level of percent agreement with other scorers. Scoring sessions would have to be conducted where scorer consistency and drift were monitored periodically and scorers could be retrained, if necessary.

Pilot testing, field testing and validation. To establish the technical qualities of the pinball task, several kinds of studies of validity and reliability would need to be conducted. To establish the validity of the scenario-based task, we would recommend the use of cognitive labs or interview (Ericsson & Simon 1980; 1993), small-scale pilot testing in several classrooms to collect at least 100 student responses per item in the scenario, expert panels to review the alignment of items to standards, sensitivity and bias reviews to guard against inadvertent biases in the item contexts, reviews of grade-level appropriateness of items by teachers, an instructional sensitivity study to determine whether the individual items are sensitive to instruction, and field-testing with several hundred students for the purposes of scaling the data using an appropriate measurement model, determining the internal structure of the scenario-based task, and conducting DIFF analyses of relevant subgroups of students. Bayes nets may be used to acquire probabilistic estimates of students' ability to answer related items correctly. Below we highlight some of the features of these studies.

Cognitive labs could be conducted with a variety of students to ensure that the desired evidence is being elicited by the items and tasks as planned. It is particularly useful to conduct cognitive labs with students at high, medium and low levels of prior science achievement. By examining these different groups of students, we acquire evidence about how the tasks perform, both in terms of cognitive difficulty and in terms of engagement. While grade-level appropriateness of tasks is often gauged through expert panels, during which the judgments of experienced teachers are solicited on an item-by-item basis and for the scenario; grade-level appropriateness also can also be judged by the level of engagement of students during cognitive labs and interviews. During cognitive labs, students can be observed, one-on-one, as the administration is conducted and levels of interest, anxiety, enjoyment and frustration are documented.

It is also desirable to conduct a small scale pilot where student performances can be collected from four to five classrooms (about 100 students) in order to calculate some basic item analysis data—item difficulties, percent correct, distributions of performances, and so on. This information can be used to revise the directions provided to students, to determine if a tutorial to familiarize students with the delivery system and different types of response capture. Items that prove too difficult or too easy can be re-evaluated and, if necessary, the structure of the scenario narrative can be addressed. When such revisions are complete, the scenario-based assessment task can undergo large-scale field testing.

After the scenario has been reviewed and revised based on the pilot data, field testing is necessary. A large-scale field test should be used to collect basic item statistics on the individual items in the scenario and on the scenario as a whole. DIFF analyses could be conducted on sub-groups of students to establish how the scenario performs when taken by students in high incidence disability categories vs. the general education population, by males versus females, by students participating in free lunch vs. no-free lunch programs or by students with high, medium and low levels of prior science achievement. An opportunity to learn or instructional sensitivity student can be conducted as part of the large scale field test to determine whether students, whose teachers report have been taught about forms of energy and how to conduct an experiment, outperform their counterparts who have not been exposed to the same content and practices.

Efficiencies (cost, design). The use of ECD as an approach to the design and development of high-stakes, assessments has received attention from both the assessment and education policy communities alike. There is much hope among educators that the use of this design approach will lead to the development of assessments of constructs that are more important than the recall of facts and declarative knowledge, which has often been the province of traditional, multiple choice assessments. Some educational researchers believe that the use of ECD is especially well-suited to the design of hard-to-assess constructs and is able to integrate UDL principles and technology requirements in the design process in a systematic, manageable way. Like all innovations, there are trade-offs in cost and design. Among the most obvious efficiency that deserve attention is the use of an approach that permits work to go be done at multiple layers of the design process simultaneously. template series and the advantages of creating variants from existing templates.

The use of a template series to specify the attributes of the assessment argument ensures that the assessment designer remains “in touch” with the fundamental elements required to design coherent assessments—these elements include, identification of the knowledge skills and abilities to be measured, identification of the types of evidence needed to measure those abilities, and specification of the types of features that tasks must include to elicit the desired behaviors. There are efficiencies conferred on item writers who use design patterns and task templates. They are able to take advantage of the careful specification of the assessment argument and use the intellectual resources of the design pattern and task templates to produce assessment tasks and items that are coherent and embody the assessment argument described by Messick (1989), Mislevy and Haertel (1996), and Kane (1992, 2006). In addition, to increased coherence among the elements of the assessment task, the use of the ECD process, if systematically implemented, contributes substantially to the content validity of the assessment being designed. Even though item writers may be able to write valid items/tasks more quickly using templates produced through the ECD process, the production of those templates can be time-consuming and labor intensive. Therefore, the use of the ECD process is particularly beneficial when tests are being designed and developed that are for large-scale purposes and will require the generation of new items to replace those that have been used in "live" test settings. The re-usability of the templates is a return on the investment of time and effort required to create the templates.

Another benefit of the template series re-usability is expressed in the variants that can be created from existing templates. Task variants, scenarios that focus on the use of experimental design in other areas of science, including (e.g., life science, earth science) or use of other variables in the same scientific area (e.g., replace properties of compression springs in a physical science scenario with changes in the mass of objects in a scenario focused on force in motion) can be efficiently produced using design patterns and task templates.

Exemplar Science Task II: Discrete Multiple Choice Item, UDL-Infused, Online — Bicycle Rider Item

Overview

The bicycle rider item is a middle-school assessment item designed to test both an area of science content and an inquiry skill. The science content being assessed is the student's knowledge of forces and motion in the physical sciences. The inquiry skill concerns the student's ability to use appropriate tools and technologies to gather, analyze, and interpret data. The item itself describes how a person rides a bike at changing or constant speeds over time. The item then asks the respondent to choose which one of four graphs, each illustrating a different relationship between speed and time, best characterizes the bicycle rider's travel.

The bicycle rider item was taken from a practice test from one state's large scale middle school science assessment. As part of work for the *Principled Science Assessment Design for Students with Disabilities* project at SRI, this item was aligned with a specific design pattern. This alignment allowed for an analysis of the item's features in terms of both ECD and UDL attributes of the design pattern. This, in turn, led to the identification of ways in which this item could be revised to potentially minimize sources of construct-irrelevant variance, thereby increasing accessibility for all students. As the item is discussed in sections below, reference is made to both the original and revised versions of the bicycle rider item.

In the next section, we identify the relevant cross-cutting concepts, core disciplinary ideas, and science practices from the new framework for K–12 science education (National Research Council, 2012) for the bike rider item. We then describe the original version of the bike rider item and how it was presented in its online platform. Next we describe the design pattern that was aligned to the item and how this design pattern guided the development of modifications to the item and its presentation in the online platform. Also discussed are the lab and field testing that compared the original and revised versions of the bicycle rider item and the implications for efficiencies in revising other existing assessment tasks.

Cross-cutting concepts, core disciplinary ideas, and science practices. Using the recently released *Framework for K–12 Science Education* (National Research Council, 2012), we identified one cross-cutting concept to which the original (and revised) bicycle item was related. This cross-cutting concepts was “Patterns,” which is described as “Observed patterns of forms and events guide organization and classification, and they prompt questions about relationships and the factors that influence them” (p. 84). In the bicycle rider item, the student is asked to think about the pattern of the relationship between the speed with which the bicycle travels and the time interval over which it travels.

The bicycle rider item also aligned with one core disciplinary idea set forth in the new K–12 framework, namely PS2: Motion and Stability: Forces and Interaction. Within this core idea, a sub-strand designated as PS2.A — Physical Science — Forces and Motion — was applicable. In the bicycle rider item, the student is presented with a description of a rider exerting force through pedaling, to start and accelerate the motion (speed) of the bicycle. At a certain point in time, the rider chooses to keep the speed of the bicycle constant (which may or may not involve using lesser or greater amounts of force). Throughout the entire ride, the force of the bike pressing down on the pavement is met with an equal force of the pavement pushing up on the tires. The student is then presented with four different graphic representations of the described ride (speed over time) and asked to choose which one of the four representations most closely fits that description.

Last, the bicycle rider item aligns to Practice 4 set forth in the K–12 science framework — namely, Analyzing and Interpreting Data. This Practice references the notion of using “spreadsheets, databases, tables, charts, graphs, statistics, mathematics, and information and computer technology to collate, summarize, and display data and to explore relationship between variables . . .” (p. 63). In the bicycle rider item, the student is asked to interpret the information about a bicycle ride in terms of speed and time, and to either anticipate how that information would be represented graphically or to look at four different graphs and analyze how the time and speed relationships in each one differ, thus leading to the selection of the best fitting graph.

Description of bicycle rider item. The bicycle rider item is one of 21 discrete, multiple-choice items that were revised, UDL-infused, and field tested in the *Principled Science Assessment Design for Students with Disabilities* project. The original versions of these 21 items, in turn, were taken from a larger pool of items used for a formative assessment (i.e., practice test) in one state’s middle school large-scale science assessment program.

The state that developed the original bicycle rider item employed the technology of online assessment for its middle school science assessments, and this specific technology was carried over and used for the item development and field testing in this project. Specific features of the online testing platform that were available were as follows:

- Progress monitoring on the screen (breadcrumbs across top of screen)
- Variable font size, magnifier, contrast
- Text to speech
- Radio buttons for multiple choice response capture
- Testing environment tools: highlighter, striker, eraser, ruler, calculator

Design pattern for bicycle rider item: Interpreting data in tables, charts, and graphs. The original version of the bicycle rider item was aligned with a design pattern titled "interpreting data in tables, charts, and graphs." This design pattern was developed in collaboration with one state department of education for the *Principled Science Assessment Design for Students with Disabilities* project. Also, this design pattern infused principles of universal design for learning (UDL) into specific design pattern attributes. Appendix B presents the complete design pattern. See Haertel, DeBarger, Villaba, Hamel, and Colker (2010) for a discussion of the integration of UDL into design patterns.

Design and development process: Bicycle rider item. As developed in its original version for a state-wide assessment, the bicycle rider item is designed to measure two constructs. The first construct is in the physical science content area of forces in motion. The second construct is a science practice that is applicable to all science content areas — namely, the science practice of understanding relationships among data as represented in canonical science and mathematical forms (i.e., tables, charts and graphs). For this single item, then, both the science content and science practice constructs are integrated. Because the item is multiple-choice and scored dichotomously (correct vs. incorrect), the single score can be interpreted to reflect a both a student's abilities in the science content and science practice areas. The science practice construct of the bicycle rider item closely aligns to the design pattern because the item asks the student to interpret how given data would be represented graphically.

Appendix B presents the interpreting data in tables, charts, and graphs design pattern. This design pattern supports the writing of storyboards or items that involve understanding and interpreting data and data variable relationships as represented in table, chart, or graphic forms. Given that every science content area has the potential to involve data, this design pattern can be used to generate groups of tasks in all science content areas. Thus, it can be easily used to generate variant assessment items that reduce future design and development costs.

The primary or **focal KSAs** to be assessed in the bicycle rider item include:

- Ability to compare and /or contrast multiple representations and the data represented therein.
- Ability to describe simple mathematical relationships or trends among data.
- Ability to draw conclusions or make predictions based on data.

The student behaviors or performances/products that will be accepted as evidence of the KSAs in the bicycle rider item are specified as potential observations and work products.

Potential observations. The relevant potential observations are:

- Identification of representational forms of data that communicate the same mathematical relationships among data (or trends in data).

Work products. The relevant work products are:

- Selection of an inference or prediction (selected response)

The features of tasks or stimuli that should elicit those cognitive behaviors and performances specified above are presented in the design pattern as characteristic features or variable features. For the bicycle rider item, these include:

Characteristic features. The relevant characteristic features are:

- The presentation contains numeric data
- The presentation includes at least one representational form
- The presented data are in a scientific context

Variable features. The variable features intended to influence difficulty of the task are given below. Some of these variable features are UDL supports and will be discussed in further detail in the next sections.

- Number or representations presented
- Types of representations
- Amount of data
- Complexity or representational form(s)
- Number of variables represented in the table, graph, or chart
- Amount of content knowledge required
- Data source (student collected vs. provided)
- Perceptual features: Representational format
- Language and symbols: Supports for vocabulary and symbols
- Cognitive features: Supports for background knowledge
- Cognitive features: Options that guide information processing
- Executive features: Supports for managing information
- Affect features: supports for intrinsic motivation

Accessibility for individual students. Table 4 below shows three different approaches for infusing of UDL into assessment tasks or items that have been employed in various SRI project contexts. The selection of an approach should be considered in terms the assessment goals, context, population, existing design assets, and assumptions of the design and development process. In the Interpreting data in tables, charts, and graphs design pattern, we would consider the potential sources of construct irrelevant variance to be identified under the attribute of AKSAs. Closely linked to the AKSAs is the attribute of variable features, which suggests features of the task that can be manipulated in an effort to minimize construct irrelevant variance, thereby increasing accessibility for individual students.

Table 4. Approaches to Infusion of UDL in Assessment Tasks Within Particular Item/Task Context

UDL approach	Infusion of UDL	Task exemplar
Approach One	Attend to UDL Principles	Exemplar I: Pinball
Approach Two	1. Identify sources of construct irrelevant variance for each focal KSA 2. Link to task model variables as means to mitigate construct irrelevant variance	Task Exemplar II: Cyclist
Approach Three	Use task templates to prompt for links between AKSAs and task model variables including support for background (cognitive) additional KSAs	Task Exemplar III: Recycling

The revision of the original bicycle rider item illustrates Approach Two. By taking the original item and analyzing it in terms of the aligned design pattern, it was possible to identify possible sources of construct irrelevant variance related to individual students' learning needs in terms of perception, expression, language and symbols, cognition, executive functioning, and engagement (affective). These sources were listed as AKSAs in this design pattern. Next, it was important to link the AKSAs to the types of task model variables (in this case, listed as variable features in the design pattern) that could be used to support students' non-construct relevant needs. Through reviewing the variable features, it was possible to identify a manageable set of modifications to the bicycle rider item that potentially could reduce the construct irrelevant variance. These modifications led to a revised version of the original item and the field testing of both items for comparison purposes. See Appendix B for screenshots of both the original and revised bicycle rider item (Original Cyclist Item and Revised Cyclist Item).

The specific UDL principles implemented in the revision of the bicycle rider item are described in Table 5 below. See the interpreting data in tables, charts, and graphs design pattern in Appendix B to see how these UDL features are represented in the design pattern template.

Table 5. UDL Principles (Categories of Students' Needs) Supported by Variable Features in Bicycle Rider Task

UDL principle (category of student need)	Task model variables implemented to address UDL principles in cyclist task
Perceptual features	Flexible size of text and images Flexible amplitude of speech and sound Adjustable contrast Flexible layout Visual graphics Verbal descriptors (spoken equivalents for text and images) Automatic text to speech
Skill and fluency	Alternative to written response (radio buttons)
Language and symbols	Embedded support for key terms Alternate syntactic levels (simplified text) Support for decoding (digital text and automatic text to speed)
Cognitive features	Using explicit examples to emphasize critical concept (minutes cyclist accelerating and at constant speed) Presentation of graphical representation simultaneously as compared to one at a time (reduce cognitive load)
Executive features	Reduced working memory Locate items near relevant text-on-screen Progress monitoring
Affect features	Real-world context to heighten engagement Age-appropriate materials

In comparing the original and revised (see Appendix B) versions of the bicycle rider item, two examples of modifications serve to illustrate the application of UDL via the design pattern attributes. First, note that in the wording of the prompt of the original item, no context is given for the ride or its amount of time. In the revised version of the item, the wording of the prompt references who is riding the bike and the amount of time the ride takes (i.e. adding up to six minutes). This revision was guided by attending to the cognitive and affective UDL categories, whereby a real-world context and explicit example of time is added. Second, note that in the presentation of the original item, the four graph options are presented within an image that needs to be enlarged to be viewed well, and furthermore, the graphs are given a letter (A through D) that much be referenced in order to make the radio button answer choice. In the revision, each of the four graphs is already enlarged (eliminating the enlargement step) and the radio buttons appear directly adjacent to the graphs (eliminating the letter choice translation). This minimizing of extra steps speaks directly to the expression, cognitive, and executive functioning UDL categories. Several of the revision choices were facilitated by the technology platform of the item, which is discussed in more detail below.

Specification of technology options. The options for uses of technology within the item format for the bicycle rider item and the rationale for the choices made are reflected in the aligned design pattern (interpreting data in tables, charts, and graphs) and its task model variables. The design pattern included an extensive list of AKSAs (cognitive background and UDL) and variable features which could be drawn on during the revision of the item, many of which could be operationalized through technology requirements and enhancements.

Programming environment. The original bicycle rider item already was available as an item on an existing online assessment delivery system presented in web browsers. Revisions to the item involved coordinating with the managers of the online delivery system and creating an additional set of assessment forms that contained both the original and revised versions of the bicycle rider item. Making modifications for the revised version of the item involved HTML coding.

Task logic and presentation (item elements and order). The assessment context from which the bicycle item was analyzed and revised was a multiple choice, online practice test that students used to prepare for their statewide science test. All items were discrete multiple choice items with a single question, involving simple, straightforward prompts and four distractors. Many of the items contained graphics. These included simple line diagrams, illustrations of scientific phenomena, and occasional charts, tables and graphs. Each item had to stand-alone and be presented on a single screen. The item stimuli had to be proximal to relevant text, item prompt, and distractors. The task logic of the bicycle rider item centered on presenting students with four possible graphs and asking them to identify the graph that represented the text description specified in the prompt. Students had to be able to see all four of the graphs depicted in the distractors in order to compare them with each other and with the written description provided in the prompt. To manage the cognitive load of the item, all of the information had to be simultaneously available to the student.

Task stimuli. Text and pictorial stimuli were selected to reduce cognitive demands (load) on students by making sure that each of the four graphs and the prompt were visible at the same time. We did not want to use roll-overs or any presentation form that would increase the students' cognitive load. Both written and visual depictions of the bicycle rider's speed over time had to be presented.

Response capture. Options include only radio buttons. No text capture, interactive graphing, drag and drop or checkboxes were typically used.

Adaptivity. Scoring and item selection are possible technologies to be employed at this level of design. In this case, the assessment context was not adaptive.

Practical obstacles to implementation (constraints on item presentation or student response format). This original bicycle rider item was presented as part of student's yearly large-scale science assessment practice in one state. For purposes of field testing the original and revised versions of this item (along with 20 other items), tasks were designed, delivered, and scored according to the state's procedures, which included providing accommodations for students who would normally receive them during testing, as per their Individualized Education Plan (IEP). The field-testing implementation process included a system tutorial to familiarize students with testing procedures and tools as well as a short set of practice items. Furthermore, school computers and servers in the participating states were set up to deliver items via the online assessment system and were updated regularly to ensure software compatibility. Local IT staff were involved throughout the planning and administrative process of field testing. In sum, important staff and technology resources were required for successful implementation of field testing (or any future testing) of the item.

Scoring methodologies, measurement models, and reporting. In the context of its original assessment, the bicycle rider item was scored as a single score point contributing to an overall proficiency score in science. It could also have been used to contribute to a subscore in physical sciences if desired. As a multiple-choice item, it was scored as a dichotomous item using a scoring key. For purposes of field testing the original and revised versions of the item, the same scoring methodology was employed. The main emphasis of the field testing was to examine comparisons for individual items (original versus revised) rather than to generate overall proficiency scores.

Pilot testing, field testing, and validation. The bicycle rider item illustrates a process whereby an existing item was analyzed for its ECD and UDL features and revised in an attempt to minimize potential sources of construct irrelevant variance. When any item revisions are made, it is important that they undergo a series of piloting and field testing before their implementation.

Cognitive labs. Cognitive lab testing should be the first step conducted to ensure that the construct is being elicited as planned across a range of students. To assure grade-level appropriateness, lab testing should include students in grades both above and below target grade. Furthermore, it may be advisable to conduct cognitive labs with students represent a range of prior science achievement (e.g., high, average, and low).

For purposes of the limited validity study in the *Principled Science Assessment Design for Students with Disabilities* project at SRI, the bicycle item was tested in a cognitive lab with two students. Results indicated that students understood the directions and how to use the technology of the online assessment system. A refinement of the item prompt resulted in response to the students' feedback in the lab.

Field testing. In order to determine the quality and usability of revised items, a large-scale field test should be conducted to collect basic item statistics (e.g., item difficulty, percent correct, distributions of performance). Furthermore, DIFF analyses could be conducted on sub-groups of students to establish how the items perform when taken by students in high incidence disability categories vs. the general education population, by males vs. females, or by students with high, medium and low levels of prior science achievement. An opportunity to learn or instructional sensitivity study also can be conducted as part of a large-scale field test to determine whether students whose teachers report having taught the item constructs outperform their counterparts who have not been exposed to the same content and practices.

The field testing of the original and revised bicycle rider item in the SRI project took place in five middle schools (two schools from State 1 and three schools from State 2) during the winters of 2011 and 2012, respectively. Pre-identification of students ensured that both general education and students with a high-incidence disability would be adequately represented (a priority for the UDL focus of the project). Each student participated in two testing sessions (approximately one month apart) where he or she completed two testing sessions online. Sessions were designed so that the student would receive both the revised and original version of the items balanced for order of presentation (e.g., for any given item, some students received the revised item in the first (or second) session and the original item in the second (or first) session).

Descriptive statistics for students with complete data on the bicycle rider item are presented below (where an incorrect answer was scored 0 and a correct answer was scored 1).

Tables 6 and 7 present sample sizes, means, and standard deviations for all students who completed both the original and revised versions of the bicycle rider item, broken down by state. For State 1, where 163 students completed both items, the mean for the revised item was .822 compared to a mean of .761 for the original. For State 2, where 61 students completed both items, the mean for the revised item was .836 compared to a mean of .787 for the original. The standard deviations decreased for the revised item for both states.

Table 6. State 1 Basic Item Statistics for Bicycle Rider Item

	N	Mean	SD
Bicycle rider revised	163	.822	.384
Bicycle rider original	163	.761	.428

Table 7. State 2 Basic Item Statistics for Bicycle Rider Item

	N	Mean	SD
Bicycle rider revised	61	.836	.373
Bicycle rider original	61	.787	.413

Table 8 presents *N*s, means, and standard deviations for general education and students with disabilities in each of the two states for the bicycle rider item. In State 1, for the 125 general education students, the mean for the revised item was .872 compared to a mean of .848 for the original. In State 1, for the students with disabilities, the mean for the revised item was .658 compared to a mean of .474 for the original. In State2, for the 36 general education students, the mean for the revised item was .917 compared to a mean of .861 for the original. In State 2, for the students with disabilities, the mean for the revised item was .720 compared to a mean of .680 for the original. It's worth noting that for both states and both student populations, the means are consistently in a direction favoring the revised UDL-infused item over the original item. Whether these differences are significant has yet to be tested.

Table 8. Paired Basic Item Statistics

States	Student	Item	*N*	Mean	SD
1	Gen Ed	Bicycle rider - original	125	.848	.3604656
		Bicycle rider - revised	125	.872	.3354342
		Difference (R-O)	125	.024	
	Disability	Bicycle rider - original	38	.474	.5060094
		Bicycle rider - revised	38	.658	.4807829
		Difference (R-O)	38	.184	
2	Gen Ed	Bicycle rider - original	36	.861	.351
		Bicycle rider - revised	36	.917	.280
		Difference (R-O)	36	.056	
	Disability	Bicycle rider - original	25	.680	.476
		Bicycle rider - revised	25	.720	.458
		Difference (R-O)	25	.040	

Validation. Over the next year, we will conduct the following analyses to understand the effects of the UDL-based item revisions on the item's performance: (a) a logistic DIF analysis to determine if there is a differential pattern of responses by students with disabilities and by students in different states, (b) a logistic analysis of the moderation effect of prior statewide math and reading achievement on the effect of item revision, (c) a Rasch IRT model to simultaneously determine the science functioning level for each student and the estimated difficulty levels for the original and redesigned items, (d) a logistic analysis of whether accommodations (such as more time on test) reduces the difference between the probability of getting the original and redesigned items correct, (e) an examination of the distribution of score differentials to identify a possible subpopulation of students who perform particularly better or worse on the redesigned items, (f) a logistic analysis to examine whether gain scores are greater for items that measure more complex scientific reasoning than items that measure declarative and procedural knowledge structures, (g) a logistic analysis to identify which, if any, design alternatives and item features predict the gains (or losses) in performance, and (h) comparison of factor structures of original and redesigned items by various subgrouping of students (for example, general education vs. students with disabilities, or high academic achievement vs. low academic achievement).

We also believe that it would be useful to conduct an IRT latent class modeling analysis to identify subsets of students who have similar responses to the items (which of course, would include any types of differential responses to the original and revised items) and then to further examine the characteristics of those students. However, given the limited set of characteristics that we have available to differentiate the groups (i.e., students with disabilities or general education, and math and science achievement scores for a subset of the students) we will conduct the logistic analyses specified earlier rather than the latent class modeling to determine the effects of those characteristics on item response.

In addition to the analyses described above, we believe that a rigorous study of the validity of the bicycle rider item would require gathering three additional sources of evidence. These include: (a) determining the correspondence of the original and revised items' content to the assessment argument implicit in the Interpreting Data in Tables, Charts, and Graphs design pattern; (b) correlating scores on the original bicycle rider item with the total scores for the other 20 original items in the assessment as compared to correlating the scores for the revised bicycle rider item with the total scores for the other 20 revised items in the assessment; (c) examining student response processes via cognitive labs for each of the original and revised bicycle rider items. We will use results from these comparisons of original and revised items to make recommendations for revising the assessment argument in our design documents and assessment items.

Efficiencies (cost, design). The general efficiencies to be gained in using ECD and UDL to generate large-scale assessment items were discussed in some depth with regards to the pinball car race assessment task. The bicycle rider item is an example of how the features of an existing item can be analyzed in terms of the ECD and UDL attributes of a broadly applicable design pattern (as compared with reverse-engineering an item into its own unique design pattern). The revision of the bicycle rider item created an efficiently produced item clone, whereby some surface features of an individual multiple choice item were altered. Specifically, the test designers were able to alter particular parameters in the item prompt and present four new improved graphs as distractors while using the same benefits of the available online testing technology platform. Such clones for a wide range of multiple-choice items could be produced by using existing design patterns and task templates that aligned with the items and by altering the settings of the variable features to attend to the accessibility needs of all students. This is a comfortable, relatively easy process that we could expect to be quicker than the reverse engineering of items.

Exemplar Science Task 3: Scenario-Based, Technology Enhanced—Recycling Suite

Overview

Recycling is a high-school mathematics assessment task that was designed to test students with significant cognitive disabilities who take a state's alternate assessment based on alternate achievement standards (AA-AAS). While the achievement standards on AA-AAS may reflect less depth and breadth for the educational standards being assessed, the tasks must be aligned with the general education standards for the grade to which the student is assigned. Students with significant cognitive disabilities challenge conventions with respect to the teaching and assessing of academic content. Assessment has been instrumental in changing the learning expectations of these students which in turn is beginning to influence classroom instructional practices. Assessment designers are challenged to develop assessments that adequately and reliably show what these students know and can do. The sheer variability in this target population, the assumptions about measuring their achievement, and the variability of design implementation procedures make traditional assessment design approaches inapplicable without some reformulation (Gong & Marion, 2006; Government Accountability Office, GAO, 2009; Ryan, Quenemoen, & Thurlow, 2004). The methods used to date in designing alternate assessments and selecting their content are varied but typically do not match the technical rigor used for designing general education assessments (Bechard, 2005).

The U.S. Department of Education's Enhanced Assessment Grant funded Alternate Assessment Design–Mathematics (AAD-M, Cameto, Haertel, Debarger, and Morrison, 2010) project was the first to address systematically the specification of grade-level academic content for alternate assessments of students with significant cognitive disabilities through the application of ECD and the principles of UDL. The grade level standards used to guide the development of assessment tasks for this project were the National Council of Teachers of Mathematics (NCTM, 2005). ECD was used to develop the design pattern and to systematically consider the application of principles of UDL. Although the suite of items (four items graduated in complexity) focuses on mathematics, the content being assessed is also often covered in science instruction and tested as part of science assessments.

The tasks were developed for a paper-pencil assessment context, but for the purposes of this paper and presentation, we will show a task model template developed with technology specifications included.

Cross-cutting concepts, core disciplinary ideas, and science practices. The construct to be measured for the suite of assessment tasks presented here, analysis and description of univariate data, comes from the NCTM content area of data and probability and the NCTM expectation: For univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics.

Typical students in grades 9-12 must learn how to use appropriate methods to analyze data. Being able to display, describe, and calculate summary statistics for univariate data is an important skill in understanding how data can be used in everyday applications. It is also a precursor in understanding and using bivariate data. Grade-level standards from NCTM addressed under the specified expectation included being able to describe central tendency, spread, and shape of univariate and bivariate data; use a variety of summary statistics and graphical displays to analyze the characteristics of the data; construct histograms, dotplots, stem and leaf plots, box plots, or scatter plots and select from among them to assist in understanding data; ability to comment on overall shape of the plots and on points that do not fit the general shape; explain differences in measures of central tendency (e.g., mean or median) and spread (e.g., standard deviation or interquartile range); recognize the sample mean and median can differ greatly for a skewed distribution; and understand that for data that are identified by categories (e.g., gender, favorite color) bar graphs, pie charts, and summary tables are often used to report the relative frequency or percent in each category (NCTM, 2005).

While standards for AA-AAS must be aligned to grade level, they are typically reduced in breadth and depth (Nagle, Cameto, Haertel, Debarger, & Morrison, 2011). First, the EAG project conducted a cross-walk of the NCTM expectations with three participating states' mathematics content standards to select a set of standards that were common across the states and the NCTM expectations (Fitton & Cameto, 2011). Secondly, the states' standards that addressed the NCTM expectations guided the breadth and depth of the design patterns and suites of tasks that were developed.

Description of assessment item suite. The EAG project created a suite of four items for a selected focal knowledge skill and ability (KSA) associated with a design pattern. The items in the suite were graduated in complexity from grade level interaction with the concepts of central tendency to simple attention to the stimulus materials. This approach was deemed appropriate to address the very heterogeneous nature of learning characteristics of students with significant cognitive disabilities. The items were designed to be individually administered to a student by their teacher or an administrator familiar with the student's specific learning needs. Each student begins the testing at Item A1. Item A1 begins the assessment process with a focus on an additional KSA or prerequisite skill associated with the focal KSA in the design pattern. If he or she answers A1 correctly, the student is administered the next most difficult item, B followed by the most difficult item, C. Both of these items assess the focal KSA. If, however, the student answered the initial item, A1, incorrectly, they were then administered Item A2 which addressed only whether the student could successfully attend to the stimulus materials (See Appendix C).

The suite of items focuses on the interpretation of tabular and graphical displays of data based on a scenario in which students participate in a recycling contest. In Item A1 of the suite the student is asked to read a histogram presenting the frequency with which three types of prizes were awarded and determine which prize was given to the most students. Those that answer incorrectly are administered Item A2 that asks the student to attend to the item stimulus materials. In Item B the student is presented with a frequency table of the number of students who received each type of prize and are asked to create a histogram based on the data and they are asked to identify the mode of the data. In Item C the student is presented with a table displaying data for ten students and the number of aluminum cans they collected. They are asked to select a type of data display (pie chart, histogram, or box plot) to show the number of students who received each of three prizes, create the data display based on the data, and answer questions about the measures of central tendency (i.e., mean, mode) based on the data. The item directives, instructions for administrators, stimulus materials, and correct answers are provided in Appendix C.

Design pattern for recycling item suite: Data analysis and probability. The Recycling assessment task was designed using the data analysis and probability (for univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics) design pattern. Appendix C contains the design pattern. See Mislevy et al. (2003) for an introduction to the design pattern tool.

Design and development process: Recycling item suite. The focal knowledge, skills, and abilities (FKSA) selected from the design pattern to use to design the suite of items is: the ability to answer a question about data by identifying, creating, and using a graphical display, and calculating and using a summary statistic. Only one FKSA was selected to create the recycling suite. If a more comprehensive assessment of data analysis and probability were being designed other FKSAs could be employed. See the complete set of FKSAs identified in the design pattern in Appendix C.

The student behaviors or performances/products that will be accepted as evidence of the FKSA (potential observations) were specified as: Given data and a question, student correctly determines the type of display to be created; correctly creates the display; correctly interprets the display; and correctly calculates or selects the summary statistic that answers the question [order of steps can vary depending on the question and type of representation to be created (e.g., to create a box plot, one needs to first calculate the median).

The potential work products were specified as: selection/expression of the type of display to be created; creation of data display; expression of relationships among variables represented in the data display; and selection/calculation of the summary statistic. The item stimuli used in the Recycling task to elicit the specified cognitive behaviors and performances for the suite can be found in Appendix C.

The characteristic features likely to evoke the desired evidence include: a data set about which questions can be answered through graphic displays and/or summary statistics; the presentation of a framing question that incorporates familiar context; and the use of only univariate measurement data.

Accessibility for individual students. Table 9 below shows the three approaches taken to infuse UDL in various SRI project contexts. We take the perspective that the selection of an approach must consider the assessment goals, context, population, existing design assets, and assumptions of the design and development process. Approach Three was used in development of the Recycling assessment suite of items.

Table 9. Approaches to Infusion of UDL in Assessment Tasks Within Particular Assessment Context

UDL approach	Infusion of UDL	Task Exemplar
Approach One	Attend to UDL principles	Exemplar I: Pinball Race Car
Approach Two	Identify sources of construct irrelevant variance for each focal KSA Link to task model variables as means to mitigate construct irrelevant variance	Task Exemplar II: Bicycle Rider
Approach Three	Use task templates to prompt for links between AKSAs and task model variables including support for background (cognitive) additional KSAs	Task Exemplar III: Recycling

Additional knowledge skills, and abilities (AKSAs). AKSAs are the other knowledge/skill/abilities that may be required in a task (Mislevy et al., 2003). AKSAs may include declarative knowledge and prerequisite skills in a content domain (i.e., cognitive back ground knowledge). Variable features linked to cognitive background AKSAs provide support for any prerequisite skill necessary for success on the item by reducing the cognitive load required to engage the skill. The cognitive background AKSAs and their associated variable features included in this assessment to provide support for prerequisite skills in a content domain are identified in Table 10.

Table 10. Additional Knowledge, Skills, and Abilities (AKSAs) in the Cognitive Background Domain

AKSA	Variable feature
Knowledge of what histograms, dot plots, stem-and-leaf, and box plots are	Provide student with familiar materials that support student's understanding of histograms, dot plots, stem-and-leaf and box plots, pretaught vocabulary and symbols Provide student with definitions and examples of various types of data and their corresponding types of graphical representations (e.g., histograms, dot plots, stem-and-leaf and box plots)
Knowledge of prerequisite vocabulary and symbols, and basic understanding of concepts (e.g., median, outliers, spread, minimum and maximum, range, box plot, stem-and-leaf, mode)	Provide student with a glossary (including illustrations) of pretaught vocabulary and symbols
Knowledge of what data are (e.g. a number that represents a property of some item)	Provide student with a glossary (including illustrations) of pre-taught vocabulary and symbols Provide student with a mechanism to highlight key features of a graph to support a consideration or an understanding of the appropriateness of different representations Provide student with a mechanism to highlight or mask parts of the data set
Ability to add, subtract, multiply and divide	Provide student with a calculator

AKSAs may also include non-construct relevant knowledge and skills needed for success on the task but not the target of the assessment (i.e., UDL variable features - perceptual, skill and fluency, language and symbols, cognitive, executive, and affective needs). Specifically, these non-construct relevant skills are necessary to make an item accessible to a student (see Table 10). For example, the ability to perceive images in an item is one AKSA from the perceptual domain. Depending on the needs of individual students, UDL variable features may be implemented to reduce limitations in perceiving images presented in a given item prompt or stimulus material. By implementing these UDL variable features, the focus of the assessment is aimed at the focal construct of an item rather than issues of accessibility. A crosswalk of all potential additional KSAs and their links to potential variable features are included in the general alignment between AKSAs and variable features document in Appendix C.

Variablefeatures. As an example of the process, the variable features intended to support non-construct relevant knowledge and skills needed for the executive functioning domain are identified in Figure 3. Potential variable features are linked to AKSAs that were selected as the design pattern was being created and deemed likely to be important in the subsequent development of items. During the item development phase, the potential variable features are pruned to a set that are most appropriate for use in the final item set. For example, the potential variable features for executive functioning are displayed in Figure 3. By designating a potential variable feature as "Implemented" or "Yes", it then becomes a selected variable feature. Selected variable features identified as "Implemented" have been incorporated into the actual item; whereas, those identified as "Yes" may be implemented by the administrator to meet the individual needs of students at the time of administration.

Selected Variable Features: Executive	• Representations of progress eg: **No** • Prompts, scaffolds, and questions to monitor progress, to "stop and think", and for categorizing and systematizing: **Yes** • Prompts and scaffolds to estimate effort, resources, and difficulty: **No** • Guides, checklists, graphic organizers, and/or templates for goal setting, prioritizing, breaking long-term objectives into reachable short-term goals, self-reflection, and self-assessment: **Yes** • Adjust levels of challenge and support eg: **Implemented**	• Representations of progress eg: **No** • Prompts, scaffolds, and questions to monitor progress, to "stop and think", and for categorizing and systematizing: **Yes** • Prompts and scaffolds to estimate effort, resources, and difficulty: **No** • Guides, checklists, graphic organizers, and/or templates for goal setting, prioritizing, breaking long-term objectives into reachable short-term goals, self-reflection, and self-assessment: **Yes** • Adjust levels of challenge and support eg: **Implemented**	• Representations of progress eg: **No** • Prompts, scaffolds, and questions to monitor progress, to "stop and think", and for categorizing and systematizing: **No** • Prompts and scaffolds to estimate effort, resources, and difficulty: **No** • Guides, checklists, graphic organizers, and/or templates for goal setting, prioritizing, breaking long-term objectives into reachable short-term goals, self-reflection, and self-assessment: **Yes** • Adjust levels of challenge and support eg: **Implemented**	• Representations of progress eg: **No** • Prompts, scaffolds, and questions to monitor progress, to "stop and think", and for categorizing and systematizing: **No** • Prompts and scaffolds to estimate effort, resources, and difficulty: **No** • Guides, checklists, graphic organizers, and/or templates for goal setting, prioritizing, breaking long-term objectives into reachable short-term goals, self-reflection, and self-assessment: **No** • Adjust levels of challenge and support eg: **Implemented**

Figure 3. Selected variable features: Executive.

Note: The "eg" notation indicates that additional examples for the variable feature are described in the general alignment between AKSAs and variable features document in Appendix C.

The variable features intended to influence difficulty of an item are further specified in the Recycling Suite's item complexity notes (see Appendix C) and are described using the following categories:

- Depth of knowledge being assessed

- Complexity of data distribution (number of data points)
- Complexity of data (e.g., single digit, double digit)
- Type of data presented in table (raw, ordered, categorized)
- Number and complexity of categories in the display
- Degree of contextualized information (high, medium, low– Variation in the presence of contextual information of several types; types of contextual information such as a description of where the numbers come from; the kinds of objects being quantified)
- Presence or absence of a framing question that incorporates familiar context (framing question is unifying theme of all sub-questions and it's not going to require a response – the sub-questions will require responses)
- Use of conceptual or procedural sub-questions [questions that fall under framing question]
- Type of data display to be created (e.g., pie chart vs. histogram);
- Type of summary statistic to select or calculate (e.g., mean, median, mode)
- Amount and type of scaffolding
- The aspects of the task that may be varied to improve accessibility for individual students, and how those variable features would be applied

Specification of technology options. The Recycling Suite of Items was originally designed to be administered via paper and pencil. However, in order to be able to deliver the suite of items via computer, options for the use of technology were specified in the Technology requirements task template (see Appendix C). The technology requirements task template indicates the kinds of technology that may be included in the task presentation and delivery of the Recycling Suite of Items.

The Recycling Suite was built from a design pattern with an extensive list of AKSAs (cognitive background and UDL) and variable features. The identification of AKSAs and associated variable features is necessary to be able to provide an assessment environment that is accessible to the widest range of needs and abilities associated with this extremely heterogeneous population of students.

Programming environment. The programming environment that will be used for the Recycling Suite is a Java web application on the back end with HTML5, and Javascript and CSS on the front end. HTML will be used due its universal availability at schools.

Task logic and presentation (item elements and order). The Recycling Suite of Items was developed to comprise four items aligned to the same content standard but of graduated complexity: from "near grade level" to requiring basic attention. Items at varying levels of complexity are required to assess the wide range of cognitive abilities of students taking alternate assessments (see Table 11). A description of the order in which items are presented to students is provided earlier in this paper, in the section titled Description of Assessment Item Suite. It is also graphically represented below in Figure 4.

Table 11. Recycling Suite: Item Descriptions

Item	Complexity
C	Near grade level content standard. Most complex item of the suite
B	Next-closest to content standard and presumed to be less-complex than item C
A1	Presumed to be less-complex than item B, often times aligned to a pre-requisite skill
A2	Less complex than item A1; basic attention required only

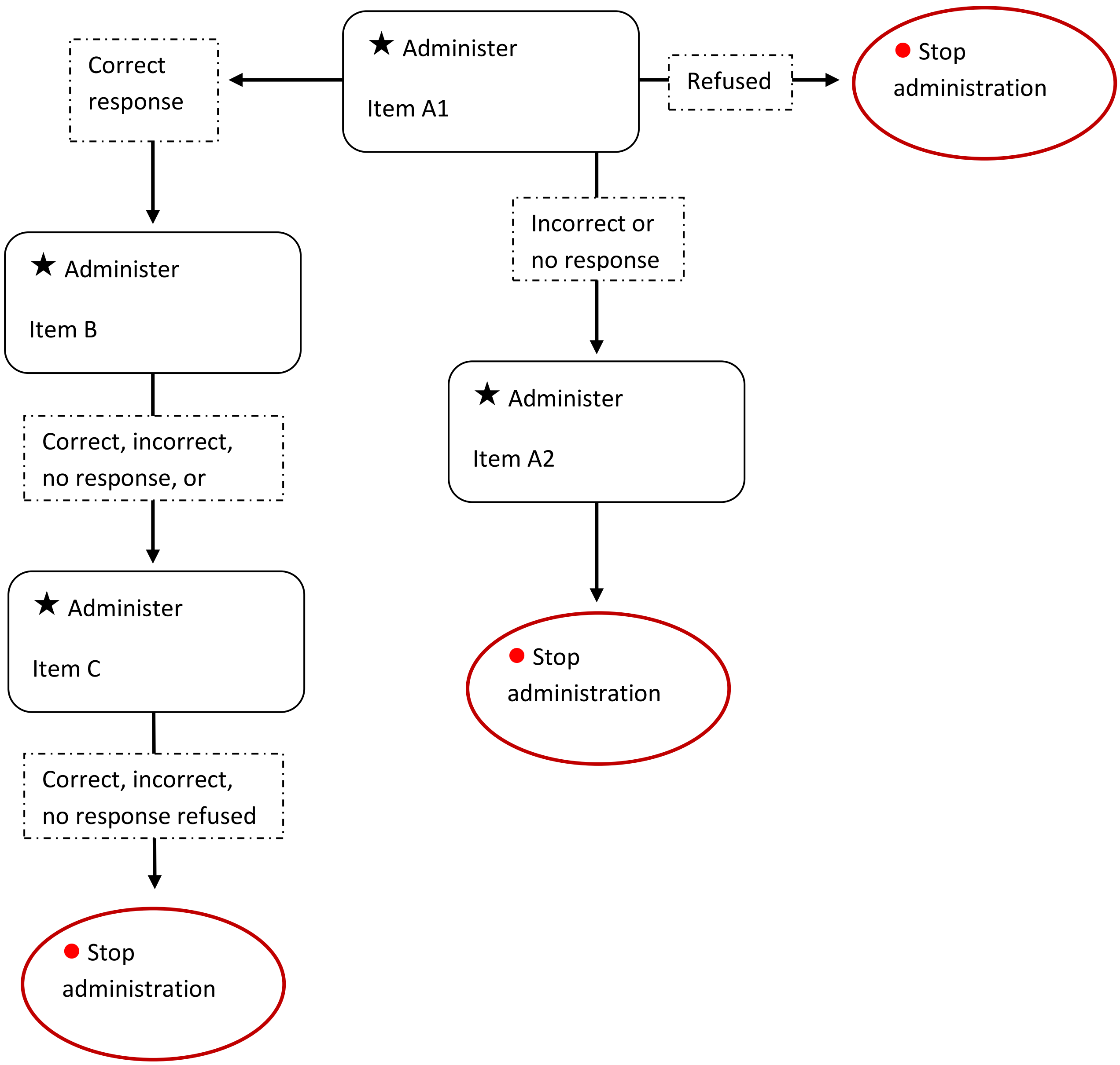


Figure 4: Flowchart of suite administration.

The items are individually administered by the student's teacher or a trained administrator. In the Recycling Suite of Items, assessment designers employed a range of item formats (e.g., open-ended, selected response). Many of the items contained graphics, including simple charts and tables.

Generally, segmented presentation of items was implemented to reduce cognitive load. As a result, some items contain multiple parts that may be dependent on the answer to a previous part. While this type of item structure creates statistical interdependencies among the item parts, each associated item part is re-set so that a student who answers a prior item part incorrectly is provided the correct answer and has an opportunity to answer subsequent items correctly. Item parts successively reveal information needed to complete subsequent item parts. For example, Item C contains 4 subquestions. Question C1 asks for the calculation of the mean and Question C2 asks the student to use the mean to answer another question. If the student completes Question C1 correctly, he/she proceeds directly to Question C2. If the student completes Question C1 incorrectly, the correct answer is provided to the student prior to proceeding to Question C2.

Task stimuli. Graphical displays were selected for inclusion based on the data analysis and probability content to be assessed and the cognitive demands (load) that students would encounter. Both written and visual depictions relevant to the everyday context of the items (recycling event) were presented. If these items were technology enhanced there could be dynamic graphs generated, simulations of the recycling activity to support student prediction, and other visual representations of data to support data analysis and calculation.

Response capture. Options include selected response, open-ended construction of a graphical display, open-ended interpretation of a graphical display, and calculation of summary statistics. If the task were technology-based we could employ interactive graphing including multiple representations of data and data relationships to illustrate key concepts, drag and drop of "data" into dynamic graphs, text capture or checkboxes.

Adaptivity. A limited adaptive model was developed based on item complexity. A simple item selection rule was established to guide item presentation.

Practical obstacles to implementation (constraints on item presentation or student response format). To date, there is limited information about performance of students with significant cognitive disabilities on technology enhanced tasks (Wehmeyer, Palmer, Smith, Davies, & Stock, 2008). While there has been some limited use of online presentation of discrete assessment items, little research has been conducted on interactive, scenario-based tasks. Concerns that some students with disabilities will have particular challenges with regard to using computers are prevalent (Bechard et al., 2010). For example, students' with limited fine-motor abilities may be hampered in their use of keyboards (touchpads have been suggested for item delivery in these cases). The expense of technology solutions to cover the range of assistive technologies in use with students with disabilities (i.e., replacing head wands with eye tracking technologies) may be prohibitive. In addition, the efficiency of administration of technology-based assessment in general education classrooms may not apply to students with significant disabilities as they may require a teacher or aide to administer the assessment. Also, familiarizing students with significant disabilities to the assessment delivery system will be required as students need to be familiar with the testing environment. Finally, item development is complicated by students' with cognitive disabilities need for multiple representations of information including visual representations where possible (Copley & Ziviani, 2004; Wehmeyer, Smith, Palmer, & Davies, 2004).

Specific instructions to the assessment administrator for individual, face-to-face, pencil and paper administration are provided in Appendix C.

Scoring methodologies, measurement models, and reporting. The item types in the Recycling Suite are, with one exception, constructed response. The selected response item would be scored dichotomously as correct or incorrect. The constructed response items would utilize a partial credit scoring model. Many of the constructed response items of the Recycling Suite that currently require human scoring could be automated if a technology enhanced assessment were to be implemented.

Pilot testing, field testing, and validation. As part of the AAD-M project, task tryouts were conducted which allowed researchers to examine implementation of the 30 exemplar suites of items that were developed. The focus of the task tryouts was the identification of refinements that could be applied to the 30 design patterns and exemplar item suites developed by the project. The suites of items were administered by trained teachers to students with significant cognitive disabilities eligible for AA-AAS in school settings. A data collection instrument was used to collect information useful to identify any difficulties that warranted refinement to the item suites. After refinements to the items, a logical next step would be a larger field test with a representative sample of AA-AAS eligible students.

Three states recruited students (*N* = 186) to participate in the task tryouts. Nearly half (48%) of participating students were classified as having an intellectual disability, 20% having autism, 20% having multiple disabilities, and the remaining 10% were distributed across the remaining federal disability categories. Participating students were enrolled across the range of assessment and accountability grades (3 through 8 and high school): elementary (39%), middle school (28%), and high school (29%, Seeratan, Nagle, Cameto, Haertel, Debarger, &Morrison, 2011). Participating students were classified at three communication levels: pre-symbolic (23%), emerging symbolic (29%), and symbolic (48%) (Browder, Flowers, & Wakeman, 2008).

Each item suite was administered to a subset of the total population depending on the assignment of the item suite to a particular state, the grade level of the item suite and the number of students recruited. Each suite was administered to a minimum of 6 students. The Recycling Suite was designed for a high school NCTM expectation and administered to 14 high school students.

A data collection instrument was designed to collect demographic, performance data on students, and students' exposure to instruction in the content of each item. Teachers also provided demographics and information about their teaching experience, as well as feedback on the characteristics of the item directives and stimulus materials, level of student engagement, and an evaluation of whether they considered the item appropriate for the student being assessed and for students with significant cognitive disabilities in general. A subset of item administrations were videotaped to ensure administration fidelity. Quantitative analysis of teacher administration data revealed that 97% of teachers administered the items in the correct sequence.

Below we present some basic statistical information describing the performance of students on the Recycling Suite of Items. (Wagner, Cometo, & Haertel, 2011)

Fourteen students were administered items from the Recycling Suite. Item A1 begins the assessment process with a focus on an additional KSA or prerequisite skill associated with the focal KSA in the design pattern. If he or she answers A1 correctly, the student is administered the next most difficult item, B followed by the most difficult item, C. Both of these items assess the focal KSA. If, however, the student answered the initial item, A1, incorrectly, they were then administered Item A2, which addressed only whether the student could successfully attend to the stimulus materials.

Information was provided to teachers/administrators about how to categorize the communication levels of students who were administered the Recycling Suite of Items. Teachers rated individual students into one of three categories, that is low (four students), moderate (four students), and high (six students). Each level of communication is described below:

Low. The student has not yet acquired the skills to discriminate between pictures or other symbols (and does not use symbols to communicate). The student may or may not use objects to communicate; may or may not use idiosyncratic gestures, sounds/vocalizations, and movements/touch to communicate with others. A direct and immediate relationship between a routine activity and the student's response may or may not be apparent. The student may have the capacity to sort very different objects, may use trial and error. Mouthing and manipulation of objects leads to knowledge of how objects are used. The student may combine objects (e.g., place one block on another).

Medium.The student may use some symbols to communicate (e.g., pictures, logos, objects). The student is beginning to acquire symbols as part of a communication system. The student may have limited emerging functional academic skills. Representations probably need to be related to the student's immediate environment and needs.

High. The student communicates with symbols (e.g., pictures) or words (e.g., spoken words, assistive technology, ASL, home signs). The student may have emerging or basic functional academic skills, including emerging writing or graphic representation for the purpose of conveying meaning through writing, drawing, or computer keying.

Students with higher levels of communication generally performed more successfully than students with lower levels of communication on the items designed for the EAG AAD-Mathematics project. Results for Recycling Suite of Items are described below and are presented in Table 12. This is one type of analysis that is being conducted on the performance of individual items.

Of students in the low communication category, three of the four (75%) either did not respond to or refused to participate in Item A1. One student responded incorrectly to the item. None of the students given Item A2 responded to the item prompt. Based on the response to Item A1, none of the students in this category were given items B and C.

All of the students with a moderate level of communication (four students) responded correctly to Item A1. Based on the response to Item A1, none of the students in the moderate category were given Item A2. The majority of students with a moderate level of communication responded incorrectly to Item B (three students, 75%). One student responded, "I don't know." For Item C, the majority of students in the moderate category responded "I don't know" (three students). One student responded incorrectly to the item.

The majority of students with a high level of communication (six students) responded correctly to Item A1 (5 students, 83%). One student in this category responded incorrectly to the item. Based on the response to Item A1, one student was given Item A2. Although the student's response to A2 was scored as incorrect, it was noted that the student pointed to one of the histogram categories when asked to "Look at/touch the histogram." Of the students with a high level of communication given item B, one student (20%) responded correctly; and four students (80%) responded incorrectly or "I don't know." For Item C, one student (20%) responded correctly; and four students (80%) either responded "I don't know" (three students) or did not respond (one student).

Table 12. Analysis of Recycling Suite by Communication Level and Response Category

		Correct	Incorrect	Don't know	No response	Refused
Low (*n* = 4)	A1	0	1	0	2	1
	A2	0	0	0	3	0
	B	0	0	0	0	0
	C	0	0	0	0	0
Moderate (*n* = 4)	A1	4	0	0	0	0
	A2	0	0	0	0	0
	B	0	3	1	0	0
	C	0	1	3	0	0
High (*n* = 6)	A1	5	1	0	0	0
	A2	0	1	0	0	0
	B	1	2	2	0	0
	C	1	0	3	1	0

Efficiencies (cost, design). Design documents such as the design pattern and template of item features provide resources to guide the development of item clones and variants. Changing surface features of discrete items and selected-response options can efficiently produce clones. Changing the features of the item context or presentation and/or employing a different set of the variable features identified in the design pattern can produce item variants. For example, in the Recycling Suite of Items, the test designer could alter the parameters in the prompt and present new data and graphs.

Novice item writers with little background in UDL can make use of established links creating additional efficiencies in item development. However, it is important that item writers have experience with the instructional needs of students with significant cognitive disabilities to effectively create items that allow these students to show what they know and can do.

Recommendations

1. **Keep in mind the first principle of ECD: The observations of student performances that you collect must be clearly aligned with the claims to be made about students' performances.**

   1a. **Rich documentation of the assessment argument and its components is essential.** This includes, for example, decisions made at the domain modeling layer, how content identified at the domain analysis layer was narrowed and expressed in design patterns. Justifications for these and all design choices must be clearly and fully explicated such that all team members, including item writers who may not have participated in early design activities, can revisit and understand design rationales.

1b. **To achieve alignment among observations and claims, the assessment design process must include several types of expertise:** instructional and disciplinary content expertise at the appropriate age/grade levels, knowledge of the population of students being assessed, and knowledge of ECD and how it should be implemented. Content and instructional experts, for example, must contribute to the articulation of the student, task and evidence models. As part of the student model, content experts are required to identify research-based learning progressions and grade level expectations that align with Focal KSAs. In the evidence model, these experts must provide direction around the construction of rubrics and scoring based on task designs and proficiencies identified in the student model. Content and instructional experts further ensure alignment by guiding selection of task variable features, identifying grade appropriate stimuli and response formats as well as familiar task contexts that will elicit desired performances.

2. **Increase the fairness of assessments for all students through ECD.** The APA's Code of Fair Testing Practices in Education described an "obligation to provide and use tests that are fair to all test takers regardless of age, gender, disability, race, ethnicity, national origin, religion, sexual orientation, linguistic background, or other personal characteristics." ECD is particularly well-suited to support designers in the identification of construct-irrelevant task demands that introduce error into student scores. For students with disabilities, in particular, the range of task demands that may influence student performance, but are not the target of the assessment, are many. UDL can address the range of challenges faced by students with disabilities, some of which may also challenge students without disabilities (i.e., reading difficulties). Drawing from the UDL framework, construct-irrelevant task demands related to perception, language and symbols, skill and fluency, cognition, executive functioning and affect have been incorporated in the design pattern structure. Likewise, a range of variable features that may be used to reduce these construct-irrelevant demands have also been incorporated in the design pattern. Tasks designed with UDL considerations should support students as they access, interact with and respond to stimuli to better represent their true knowledge and skills. Infusing UDL principles throughout the design process, made possible by ECD, should increase the fairness of assessments for all students.

3. **Use ECD to design measures of hard-to-assess constructs that require complex, multi-step tasks.** Imagine large-scale assessments of the future where an immersive assessment task opens with an avatar presenting students with a problem in their local community. Students enter a virtual lab, select parameters to design an experiment to gather evidence to solve the problem. In designing their own experiment, the student selects independent variables and numbers of trials to be conducted. In conducting the experiment, the student operates simulations of laboratory equipment, generates interactive graphs and tables that reflect the choices he or she made. Finally, the student generates text explaining the results of the experiment that he or she designed and how their experiment suggests community action. With these technological affordances, we can anticipate increased engagement and enthusiasm on the part of students, but the articulation of student and evidence models becomes challenging. The multiple dependencies among student

responses as they move through such an elaborate scenario need to be addressed in an evidence model. In addition, the number of additional KSAs requisite to successful performance that might produce construct-irrelevant variance need to be identified and mitigated, if possible. As illustrated in this paper, ECD is a powerful tool to help us in the careful design of rich, immersive assessment tasks, supporting designers as they think through the coherence of the student, evidence and task models.

It is essential to examine the pervasive assumption that the introduction of technology or the incorporation of the latest technological advances is always beneficial for students. In addition to inequitable availability of resources to both assess or instruct using technology, considerations of accessibility suggest that each use of technology must be carefully considered in terms of fairness to ensure that construct-irrelevant variance is minimized. As has been illustrated in this paper, ECD can be used to analyze the intersection of students' knowledge and skills and the demands conferred by technology, identify sources of construct-irrelevant variance, and provide information about the tradeoffs of using a particular technology in the assessment situation.

4. **Customize the design process to the purpose and context of your assessment situation.** Assessments should be designed to meet the particular context and purpose for which they are intended. The design considerations implicit in the development of formative versus summative assessments, for example, will have significant impact on the assembly and evidence models underlying the assessment. ECD structures and design activities support designers in considering the assessment purpose, context, and consequences. As a comprehensive design process, ECD can be applied to all content domains, using all types of item/task formats, for all assessment purposes. It can be used to design assessments that are entirely multiple-choice items as well as assessments those that incorporate a range of student response formats. More importantly, an ECD process *makes explicit* the design choices that customization to a particular context entails and prompts designers to justify and document these design choices.

   4a. **Identify an appropriate approach to the integration of UDL.** The resources available in any particular design project may, unfortunately, not allow for the comprehensive process we have derived across the body of our work to be systematically implemented. We have concluded that three approaches, or levels of implementation can each confer significant benefit to the design process, especially in terms of accessibility and fairness. The three approaches, all supported by ECD, are described in the Challenges of this paper and exemplified in the three task illustrations. Identifying priorities and leverage points in a particular design context may suggest that one of the three approaches is most appropriate for that context. For example, using the second approach, SRI made use of existing design assets (items), revising them based on principles of UDL through an ECD process. This approach is illustrated in this paper using the bicycle rider item. *Designers can apply any of these approaches to a design process that forward engineers new items or to the modification of existing items.*

5. **Articulate evidence models that integrate observable variables from multiple sources.** As technology-enabled tasks become more ubiquitous, the use of evidence models will permit assessment designers to gather evidence of student knowledge, skills and abilities from a variety of sources, including: discrete and scenario-based items on accountability tests, from simulations and

modeling activities, and from performances of students in serious games. The wealth of data produced in technology-enabled tasks can be gathered from a multitude of sources, providing multiple opportunities to observe student performance and increasing the robustness of student proficiency estimates.

6. **Gather empirical evidence to establish validity.** Although an ECD approach can support the construction of a valid assessment argument, developers are obligated to gather evidence of validity through empirical studies. Application of the ECD can increase the alignment of assessment content to educational standards and/or constructs of interest. However, the ECD process alone does not ensure that the test has inferential validity, that it is instructionally sensitive, or that it distinguishes between novice and expert performances. Therefore, developers must collect additional data to support the chain of reasoning resulting in inferences about student performance.

Acknowledgments

Research findings and assessment tasks described in this paper were supported by the following projects: Principled Assessment Design in Inquiry [National Science Foundation, REC-0089122 and REC-0129331]; An Application of Evidence-Centered Design to a State's Large Scale Science Assessment [National Science Foundation, DRK-12 initiative, DRL-0733172]; Principled Assessment Science Assessment Designs for Students with Disabilities [Institute of Education Sciences, US Department of Education, R324A070035]; Applying Evidence-Centered Design to Alternate Assessments in Mathematics for Students with Significant Cognitive Disabilities [US Department of Education, Contract to State of Utah, 09679]; Alternate Assessment Design Reading (AAD-R): Evidence-Centered Design for Alternate Assessment [US Department of Education, Contract to State of Idaho, S368A090032]. In addition, SRI International provided Strategic Business Thrust (SBT) funds. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183–198.

Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment, 1*(5). Retrieved from http://www.bc.edu/research/intasc/jtla/journal/v1n5.shtml

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Bachman, L. F. (2005). Building and supporting a case for assessment use. *Language Assessment Quarterly, 2*, 1-34.

Baxter, G., & Mislevy, R. (2005). T*he case for an integrated design framework for assessing science inquiry* (PADI Technical Report No. 5). Menlo Park, CA: SRI International.

Bechard, S. (2005). *Developing alternate assessments using expanded benchmarks from a nine state consensus framework in reading, writing, mathematics, and science.* Paper presented at the 2005 National Council on Measurement in Education (NCME) Annual Conference, Montreal, Canada.

Bechard, S. , Sheinker, J. , Abell, R. , Barton, K. , Burling, K. , et al. (2010). Measuring cognition ofstudents with disabilities using technology-enabled assessments: Recommendations for anational research agenda. *Journal of Technology, Learning, and Assessment*, *10*(4).

Brennan, R. L. (2001). *Generalizability theory*. New York, NY: Springer-Verlag.

Browder, D. M., Flowers, C., & Wakeman, S. Y. (2008). Facilitating participation in assessments and the general curriculum: Level of symbolic communication classification for students with significant cognitive disabilities. *Assessment in Education: Principles, Policy, and Practice, 15*(2), 137–151.

Cameto, R., Haertel, G., DeBarger, A., Morrison, K. (2010). *Alternate assessment design—mathematics technical report 1: Project overview: applying evidence-centered design to alternate assessments in mathematics for students with significant cognitive disabilities.* Menlo Park, CA: SRI International.

Chandler, P. & Sweller, J, (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*, 233–246. doi: 10.1111/j.2044-8279.1992.tb01017.x

Colker, A. M., Liu, M., Mislevy, R., Haertel, G., Fried, R., & Zalles, D. (2010). *A design pattern for experimental investigation* (Large-Scale Assessment Technical Report No. 8). Menlo Park, CA: SRI International.

Copley, J., & Ziviani, J. (2004). Barriers to the use of assistive technology for children with multiple disabilities. *Occupational Therapy International*, *11*(4), 229.

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Cureton , E. E. (1950). Validity. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 621–695). Washington, DC: American Council on Education.

Dolan, R. P., Rose, D. H., Burling, K., Harms, M., & Way, D. (April, 2007). *The universal design for computer-based testing framework: A structure for developing guidelines for constructing* *innovative computer-administered tests.* Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Ericsson, K., & Simon, H. (1980). Verbal reports as data. *Psychological Review*, *8*(3), 215–251.

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data*. MIT Press, Cambridge, MA.

Fitton, R., & Cameto, R. (2011). *Alternate assessment design mathematics. Technical report 3: Crosswalk: Domain analysis aligning National Council of Mathematics (NCTM) expectations with state extended mathematics standards.* Menlo Park, CA: SRI International.

Gong, B., & Marion, S. (2006). *Dealing with flexibility in assessments for students with significant cognitive disabilities.* Minneapolis: University of Minnesota, National Center On Educational Outcomes. Retrieved from http://cehd.umn.edu/NCEO/onlinepubs/synthesis60. html

Government Accountability Office. (2009). *No Child Left Behind: Enhancements in the Department Of Education's review process could improve state academic assessments* [GAO 09-911]. Retrieved from http://www. gao. gov/new. items/D09911. pdf

Haertel, G., Haydel DeBarger, A., Villalba, S., Hamel, L., & Mitman Colker, A. (2010). *Integration of evidence-centered design and univeral design principles using PADI, an online assessment design system* (Assessment for Students with Disabilities Technical Report No. 3). Menlo Park, CA: SRI International.

Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, *3*, 527–535.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*(3), 527–535.

Kane, M. T. (1994). Validating interpretive arguments for licensure and certification examinations. *Evaluation and the Health Professions, 17*, 133–159.

Kane, M. T. (1996). The precision of measurements. *Applied Measurement in Education, 9*(4), 355–379

Kane, M.T. (2006). Validation. In R.L. Brennan (Ed.), *Educational measurement* (4th ed.,pp. 17–64). Westport, CT: American Council on Education.

Kane, M. T. (2010). *Errors of measurement, theory, and public policy.* Princeton , NJ : Educational Testing Service.

Kane, M. T., Crooks, T. J., & Cohen, A. S. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice, 18*(2), 5–17.

Krajcik, J., McNeill, K. L., & Reiser, B. J. (2008). Learning-goals-driven design model: Developing curriculum materials that align with national standards and incorporate project-based pedagogy. *Science Education, 82*(1), 32.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: Macmillan.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13–23.

Mislevy, R., Liu, M., Cho, Y., Fulkerson, D., Nichols, P., Zalles, D., Fried, R., … Hamel, L., (2009). *A design pattern for Observational Investigation Assessment Tasks* (Large-Scale Assessment Technical Report No. 2). Menlo Park, CA: SRI International.

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design: Layers, structures, and terminology. In S. Downing & T. Haladyna (Eds.), *Handbook of test development* (pp. 61–90). Mahwah, NJ: Erlbaum.

Mislevy, R.J., & Riconscente, M. M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology. PADI | principled assessment designs for inquiry.* Menlo Park, CA: SRI International.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). *On the structure of educational assessments* (CSE technical report). Washington, DC, and University of California at Los Angeles: Center for the Study of Evaluation; National Center for Research on Evaluation, Standards, and Student Testing

Mislevy, R., Hamel, L., Fried, R., G., Gaffney, T., Haertel, G., Hafter, A., Murphy, R., Quellmalz, E., Rosenquist, A., Schank, P., Draney, K., Kennedy, C., Long, K., Wilson, M., Chudowsky, N., Morrison, A., Pena, P., Songer, N., Wenk, A. (2003). *Design patterns for assessing science inquiry (PADI Technical Report 1).* Menlo Park, CA: SRI International.

Mislevy, R., & Haertel, G. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 24*(4), 6–20.

Mislevy, R., Hamel, L., Fried, R., G., Gaffney, T., Haertel, G., Hafter, A.… Wenk, A. (2003). *Design patterns for assessing science inquiry* (PADI Technical Report No. 1)*.* Menlo Park, CA: SRI International.

National Council of Teachers of Mathematics. (2005). *Principles and standards for school mathematics.* Reston, VA: Author.

Nagle, K. Cameto, R., Haertel, G., DeBarger, A., & Morrison, K. (2011). *Alternate assessment design—mathematics technical report 2: Current state of mathematics assessment in alternate assessment.* Menlo Park, CA: SRI International.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

National Research Council. (2012). *A framework for K–12 science education: Practices, cross-cutting concepts, and core ideas.* Washington DC: National Academies Press.

Rose, D. H., & Meyer, A. (2002). *Teaching every student in the digital age: Universal design for learning.* Alexandria, VA: ASCD.

Rose, D. H., & Meyer, A. (2006). *A practical reader in universal design for learning.* Cambridge, MA: Harvard Educational Press.

Rose, D. H., Meyer, A., & Hitchcock, C. (2005). *The universally designed classroom: Accessible curriculum and digital technologies.* Cambridge, MA: Harvard Education Press.

Ryan, J. M., Quenemoen, R. F., & Thurlow, M. L. (2004). *I say potato, you say potato: The assessment-speak gap between general and alternate assessment experts.* Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Seeratan, K.L., Nagle, K., Cameto, R., Haertel, G., DeBarger, A., Morrison, K., & Fujii, R. (2011). *Alternate assessment design: Pilot task tryouts design and analysis* (Mathematics Technical Report No. 7). Menlo Park, CA: SRI International.

Thorndike, E. L. (1918). The nature, purposes, and general methods of measurement of educational products. In S. A. Courtis (Ed.), *The measurement of educational products (*17th yearbook of the National Society for the Study of Education, Pt. 2. pp. 16–24). Bloomington, IL: Public School.

Toulmin, S. E. (1958). *The use of argument*. Cambridge, UK: Cambridge University Press.

Wagner, S., Cameto, R., & Haertel, G. (2011). *Alternate assessment design: Summary of task and item try-outs* (Mathematics Technical Report No. 10). Menlo Park, CA: SRI International.

Wehmeyer, M. L., Smith, S. J., Palmer, S. B., & Davies, D. K. (2004). Technology use by students with intellectual disabilities: An overview. *Journal of Special Education Technology*, *19*(4), 7–21.

Wehmeyer, M., Palmer, S., Smith, S., Davies, D., & Stock, S. (2008). The efficacy of technology use by people with intellectual disability: A single-subject design meta-analysis. *Journal of Special Education Technology*, *23*(3), 21.
Wigmore, J. H. (1937). *The science of judicial proof* (3rd ed.). Boston, MA: Little, Brown, & Co.
Williams, V. (2008). educational gaming as an instructional strategy. In C. Bonk et al. (Eds.), *Proceedings of world conference on e-learning in corporate, government, healthcare, and higher education 2008* (pp. 215–220). Chesapeake, VA: AACE.

Appendix A. Exemplar I. Science Task: Scenario-Based, Technology Enhanced—Pinball Car Race

Design Pattern Used for Pinball Car Race

PADI | Design Patterns | Education Standards | Exemplars | Templates | Task Specifications | Student Models | Student Model Variables | Activities | Meas. Models | Observable Variables | Eval. Procedures | Evaluation Phases | Work Products | Materials & Presentation | Task Model Variables

Hello csanford | Account Settings | Logout | Edit Model | Export to PDF

Experimental Investigation TEAS Version | Design Pattern 2249 [Duplicate | Permit | Delete | View: View (vertical)]

Title [Edit]
Experimental Investigation TEAS Version

Overview [Edit]
This design pattern supports the writing of storyboards and items that address scientific reasoning and process skills in experimental investigations. In experimental investigations, it is necessary to manipulate one or more of the variables of interest and to control others while testing a prediction or hypothesis. This contrasts with observational investigations, where variables typically cannot be manipulated. This design pattern may be used to generate groups of tasks for science content strands amenable to experimentation. details

Use [Edit]
U1. This design pattern supports the construction of tasks that address experimental investigations - that is, investigations where experimental methods are appropriate (as compared with investigations where only observations of phenomena are possible). In order for students to have a well-rounded understanding of the scientific method, they need to be familiar with the context and methods of experimental investigations.

Focal knowledge, skills, and abilities [Edit]
Fk1. Ability to distinguish between experimental and observational methodology
Fk2. Ability to recognize that when a situation of scientific interest includes aspects that can be altered or manipulated practically, it is suitable for experimental investigation details
Fk3. Ability to recognize that the purpose of an experiment is to test a prediction/hypothesis about a causal relationship details

http://design-drk.padi.sri.com/padi/index.jsp

Fk4. Ability to identify, generate, or evaluate a prediction/hypothesis that is testable with a simple experiment
Fk5. Ability to plan and conduct a simple experiment step-by-step given a prediction or hypothesis
Fk6. Ability to recognize that at a basic level, an experiment involves manipulating one variable and measuring the effect on (or value of) another variable details
Fk7. Ability to identify variables of the scientific situation (other than the ones being manipulated or treated as an outcome) that should be controlled (i.e. kept the same) in order to prevent misleading information about the nature of the causal relationship details
Fk8. Ability to recognize variables that are inconsequential in the design of an experiment details
Fk9. Ability to recognize that steps in an experiment must be repeatable to dependably predict future results
Fk10. Ability to recognize that random assignment to treatment conditions (i.e. levels of the independent variable) is an important way to rule out alternative explanations for a causal relationship details
Fk11. Ability to interpret or appropriately generalize the results of a simple experiment or to formulate conclusions or create models from the results

Additional knowledge, skills, and abilities [Edit]
Ak1. Content knowledge (may be construct relevant) details
Ak2. Prerequisite knowledge from earlier grades details
Ak3. Prerequisite experience assessing or conducting component steps of an investigation details
Ak4. Ability to collect, organize, analyze, and present data details
Ak5. Familiarity with representational forms (e.g., graphs, maps) details
Ak6. Student needs based on UDL categories may be included (Perceptive, Expressive, Language and Symbols, Cognitive, Executive Functioning, Affective)

Potential observations [Edit]	Po1.	Accuracy in identifying situation suitable for experimental investigation
	Po2.	Plausibility of a measurable research question being raised
	Po3.	Plausibility of hypothesis as being testable by a simple experiment
	Po4.	Plausibility/correctness of design for a simple experiment
	Po5.	Correct identification of independent and dependent variables
	Po6.	Accuracy in identifying variables (other than the treatment variables of interest) that should be controlled (held constant) or made equivalent (e.g., through random assignment).
	Po7.	Plausability/correctness of steps to take in the conduct of an experiment
	Po8.	Plausibility of plan for repeating an experiment
	Po9.	Correctness of recognized data patterns from experimental data
	Po10.	Plausibility/correctness of interpretation/explanation of experimental results
	Po11.	Accuracy in critiquing the experimental design, methods, results, and conclusions of others
	Po12.	Generate a prediction/hypothesis that is testable with a simple experiment
Potential work products [Edit]	Pw1.	Select, identify, or evaluate an investigable question details
	Pw2.	Identify or differentiate independent and dependent variables in a given scientific situation
	Pw3.	Identify or differentiate variables that do and do not need to be controlled in a given scientific situation
	Pw4.	Complete some phases of experimentation with given information, such as selection levels or determining steps
	Pw5.	Generate or identify data pattern from results in a simple experiment
	Pw6.	Generate an interpretation/explanation/conclusion from a set of experimental results
	Pw7.	Critiques of peers on their choice of experimental procedures or explanations of experimental results details
	Pw8.	Given an experiment with unexpected or confusing results, identify possible reasons details
Potential rubrics [Edit]		
Characteristic features [Edit]	Cf1.	Focus on Nature of Science (Strand I in MCA) benchmarks that relate to experimental investigations at the appropriate grade level
	Cf2.	Presentation of situation of scientific interest where variables can be (or have been) practically altered to address a causal prediction details
	Cf3.	Presentation of situation requiring the design or conduct of a controlled experiment details
	Cf4.	Presentation or representation of an experimental design
	Cf5.	Presentation of observed result from an experiment requiring the development of explanations, conclusions, or models details
Variable features [Edit]	Vf1.	Content (strand) context details
	Vf2.	Which one of multiple phases of experimental investigation will be addressed
	Vf3.	Qualitative or quantitative investigation or a combination
	Vf4.	Ease or difficulty with which the treatment (independent) variable can be manipulated
	Vf5.	Are manipulated variables given or to be determined?
	Vf6.	The number of variables investigated and the complexity of their interrelationships details
	Vf7.	Number of variables that need to be controlled to unambiguously study the relationship between the manipulated variable and the outcome variable details
	Vf8.	Length of time over which the experiment must be conducted in order to study the potential impact of the treatment variable
	Vf9.	Data representations details
	Vf10.	Variable features may be added to support student needs associated with UDL categories (Perceptual - Screen presentation will include variable font size, Option for altering screen contrast, Option for magnification or zoom, Optional text-to-speech; Expressive - Range of response options required (radio buttons, drag and drop), Range of student support for producing response (speech-to-text); Language and Symbols - Provision of multiple representations of symbols (linguistic labels for symbols, define abbreviations, etc.), Provide definitions of non-construct relevant terminology, Use of studentsâ?? dominant language; Cognitive - Use of a concept map, Use of a response template, Use of context to heighten salience, Highlighting key terms and ideas; Executive Functioning - Breaking task into manageable units, Icons to encourage thinking and reflection, On-screen progress monitoring; Affective - Use of scenario or real-world context to heighten engagement, Age-appropriate materials -Interactive narrative (gaming), Affirmation of participation

Narrative structure	[Edit]	Cause and effect. An event, phenomenon, or system is altered by internal or external factors. Investigation. A student or scientist completes an investigation in which one or more variables may be observed or manipulated and data are collected Change over time. A sequence of events is presented to highlight sequential or cyclical change in a system.
National educational standards	[Edit]	NSES 8ASI1.1. Identify questions that can be answered through scientific investigations. Students should develop the ability to refine and refocus broad and ill-defined questions. An important aspect of this ability consists of students' ability to clarify questions and inquiries and direct them toward objects and phenomena that can be described, explained, or predicted by scientific investigations. Students should develop the ability to identify their questions with scientific ideas, concepts, and quantitative relationships that guide investigation. NSES 8ASI1.2. Design and conduct a scientific investigation. Students should develop general abilities, such as systematic observation, making accurate measurements, and identifying and controlling variables. They should also develop the ability to clarify their ideas that are influencing and guiding the inquiry, and to understand how those ideas compare with current scientific knowledge. Students can learn to formulate questions, design investigations, execute investigations, interpret data, use evidence to generate explanations, propose alternative explanations, and critique explanations and procedures. NSES 8ASI1.3. Use appropriate tools and techniques to gather, analyze, and interpret data. The use of tools and techniques, including mathematics, will be guided by the question asked and the investigations students design. The use of computers for the collection, summary, and display of evidence is part of this standard. Students should be able to access, gather, store, retrieve, and organize data, using hardware and software designed for these purposes. NSES 8ASI1.4. Develop descriptions, explanations, predictions, and models using evidence. Students should base their explanation on what they observed, and as they develop cognitive skills, they should be able to differentiate explanation from description, providing causes for effects and establishing relationships based on evidence and logical argument. This standards requires a subject knowledge base so the students can effectively conduct investigations, because developing explanations establishes connections between the content of science and the contexts within which students develop new knowledge.
		NSES 8ASI1.5. Think critically and logically to make the relationships between evidence and explanations. Thinking critically about evidence includes deciding what evidence should be used and accounting for anomalous data. Specifically, students should be able to review data from a simple experiment, summarize the data, and form a logical argument about the cause-and-effect relationships in the experiment. Students should begin to state some explanations in terms of the relationship between two or more variables. NSES 8ASI1.6. Recognize and analyze alternative explanations and predictions. Students should develop the ability to listen and to respect the explanations proposed by other students. They should remain open to and acknowledge different ideas and explanations, be able to accept the skepticism of others, and consider alternative explanations. NSES 8ASI1.7. Communicate scientific procedures and explanations. With practice, students should become competent at communicating experimental methods, following instructions, describing observations, summarizing the results of other groups, and telling other students about investigations and explanations.

Description of Pinball Car Race Prototype Task

This is an interactive computer task for middle school science focusing on physical science content and scientific investigation. Before the assessment begins, students are instructed that words and phrases presented in bold and underlined have a roll over feature that shows their definition and that images have a similar roll over feature that provides a description of the image.

In the first scene, the examinee is introduced to the pinball car race. This race is designed to demonstrate the difference between potential and kinetic energy. Images and animation are used to increase student engagement. The examinee is provided with some background information regarding the experiment that will be the focus of this task.

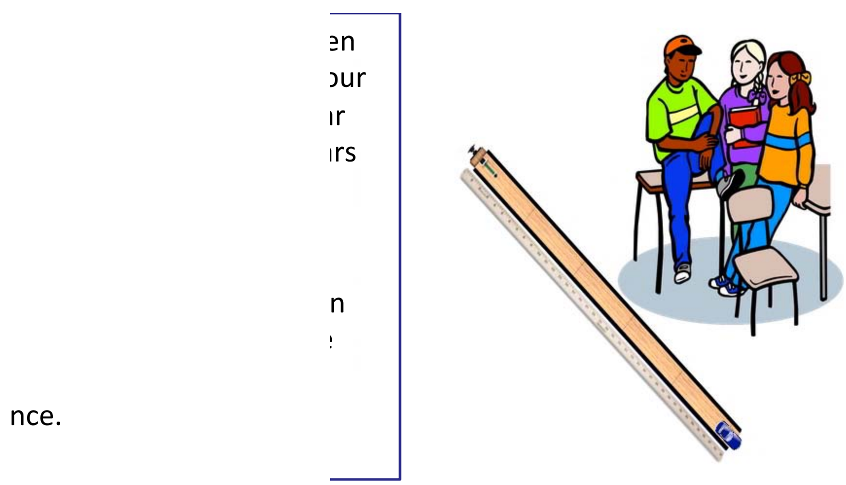


Figure A1. Scene 1 showing the start of the task.

In Scene 2, the examinee is instructed to play the animation that shows the operation of the spring in a Pinball Car race. The examinee must refer to the timer associated with the animation to answer the questions about the time segment in which the spring has the most potential energy, and the point in which the spring has the most kinetic energy.

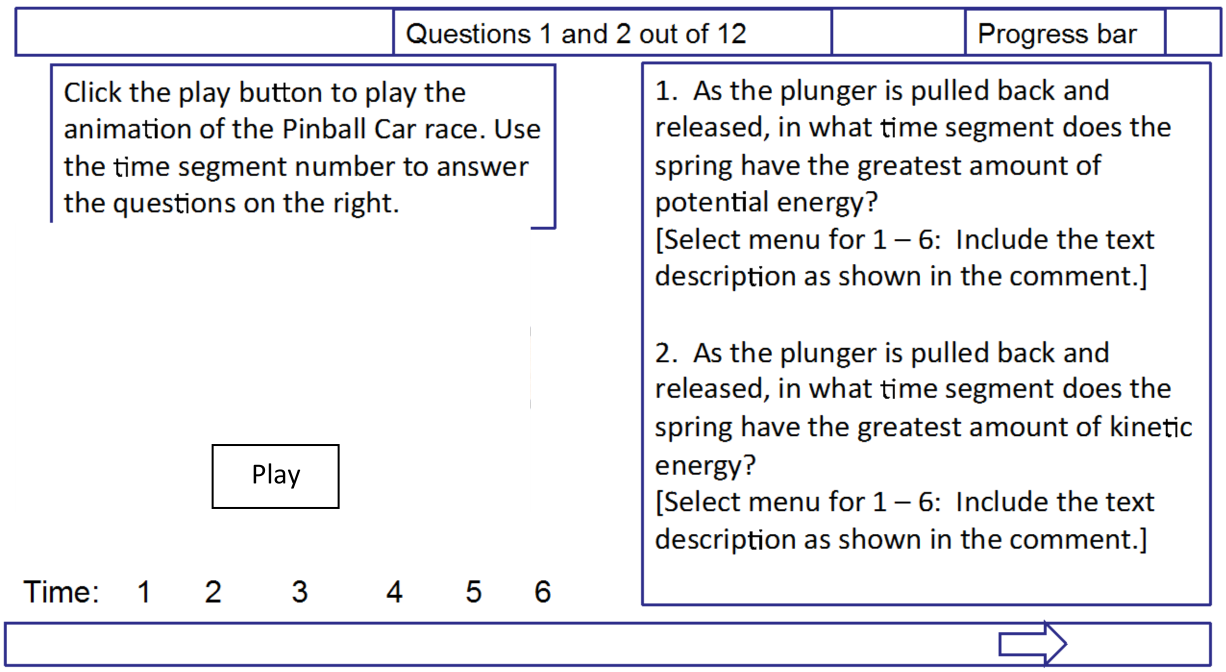


Figure A2. Scene 2 – Questions 1 and 2 out of 12.

This scene introduces the experiment the examinee will be asked to design. Two properties of springs are described, the number of coils and the thickness of the wire. The examinee is informed that these properties will be used as variables in the experiment. The opportunity to see visual representations of springs that use these properties is provided.

	Background Information		Progress bar	

Design an experiment to determine which spring will make the car go the longest distance. For the experiment you will do trial races using different springs.

The springs differ on two properties, the number of coils and the thickness of the wire.

These two properties might affect the amount of potential energy each spring can store which would then affect the distance the car could travel. These two properties will be used as the variables in the experiment.

The properties of the springs are provided below. You may click options from each column to see examples of the springs.

y
- Some
- Very few

Thickness of the Wire

- Very thick
- Moderately thick
- Thin

Figure A3. Scene 3 – Background information.

In Scene 4, the examinee is again presented with a visual representation of the variables that will be used in the experiment. The examinee selects a hypothesis for the experiment that uses either the number of coils or the thickness of the wire. Once the hypothesis is selected, the examinee is asked to explain how the variable selected affects the potential and kinetic energy of the spring.

	Questions 3 and 4 out of 12		Progress bar	

3. Pick a hypothesis to test by filling in the blanks in the statement below:

I hypothesize that the distance the car travels is affected by the __(number of coils/thickness of the wire)**.**

I believe that the car would travel farther with springs that have __(more coils/ less coils, thicker wire/thinner wire)**.**

4. Explain your hypothesis: What effect does the ___(FILL in answer from above: number of coils or thickness of the wire)____ have on the potential and kinetic energy of the spring?

Figure A4. Scene 4 – Questions 3 and 4.

The hypothesis the examinee selected in Scene 4 appears at the top of Scene 5. In this scene, the examinee designs an experiment to test this hypothesis by selecting settings for the number of coils and the thickness of the wire of three springs. The examinee is asked to explain why the spring settings chosen are appropriate for testing the hypothesis.

Questions 5 and 6 out of 12 | Progress bar

Your hypothesis is: [FILL IN PREVIOUS ANSWER] Springs with_(more coils/less coils/ thicker wire/thinner wire) will make the car go farther.

Now conduct an experiment to test your hypothesis. You can test three springs to find out which makes the car go the longest distance.

Use the table on the right to pick the settings for the three springs you will test. The settings for thickness and number of coils can be different or the same for each spring.

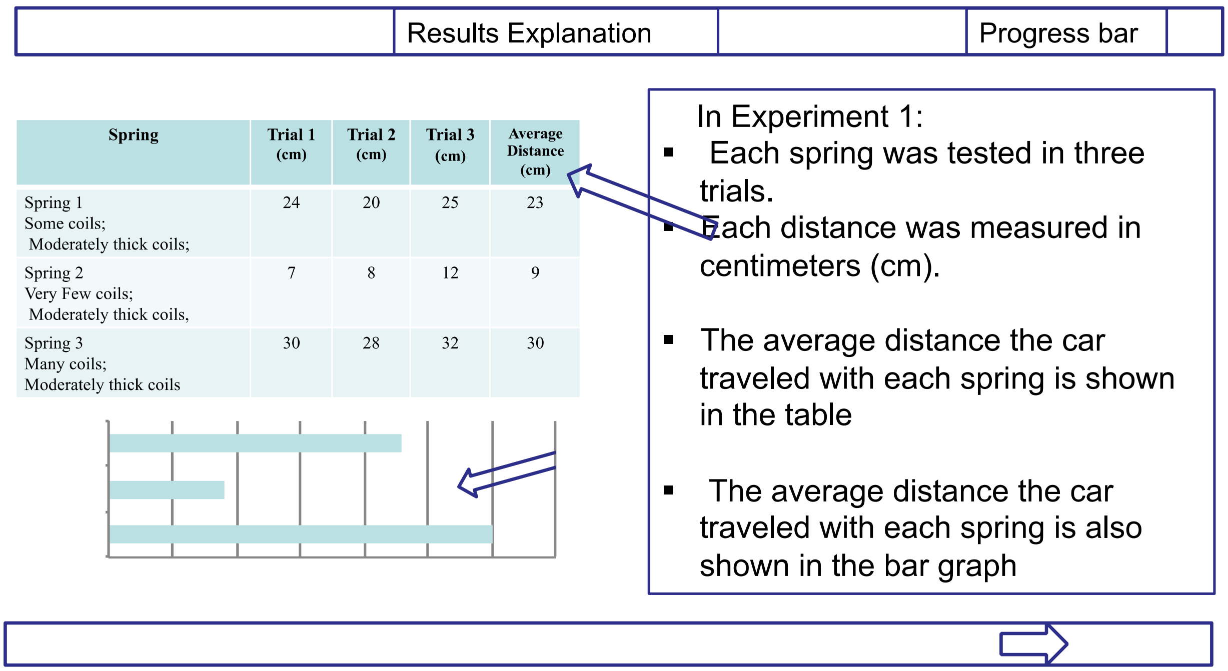
The setting you choose for each of these variables should be based on your hypothesis.

5.	Spring 1	Spring 2	Spring 3
Number of Coils	o Many o Some o Very few	o Many o Some o Very few	o Many o Some o Very few
Thickness of the wire	o Very Thick o Moderately Thick o Thin	o Very Thick o Moderately Thick o Thin	o Very Thick o Moderately Thick o Thin

6. Explain why the settings you chose are appropriate to test your hypothesis.

Figure A5. Scene 5 – Questions 5 and 6.

In Scene 6, results from the experiment are presented in a table and in a bar graph. This scene provides students with assistance reading the table and bar graph since these are not focal KSAs for this task.



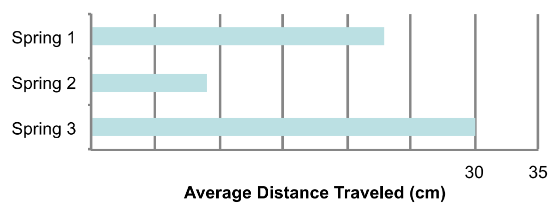
Spring	Trial 1 (cm)	Trial 2 (cm)	Trial 3 (cm)	Average Distance (cm)
Spring 1 Some coils; Moderately thick coils;	24	20	25	23
Spring 2 Very Few coils; Moderately thick coils,	7	8	12	9
Spring 3 Many coils; Moderately thick coils	30	28	32	30

Figure A6. Scene 6 – Results explanation.

In this scene, the examinee/user is asked how the data relates to the hypothesis and to explain their reasoning.

Question 7 out of 12 | Progress bar

Spring	Trial 1 (cm)	Trial 2 (cm)	Trial 3 (cm)	Average Distance (cm)
Spring 1 Some coils; Moderately thick coils;	24	20	25	23
Spring 2 Very Few coils; Moderately thick coils,	7	8	12	9
Spring 3 Many coils; Moderately thick coils	30	28	32	30



Your hypothesis is: [FILL IN PREVIOUS ANSWER] Springs with_(more coils/less coils/thicker wire/thinner wire) will make the car go farther.

7a. Based on the results of Experiment 1, which of the following is true?

- ❑ These results **support** my hypothesis
- ❑ These results **contradict** my hypothesis
- ❑ These results **do not provide information** about my hypothesis

7b. Explain your answer:

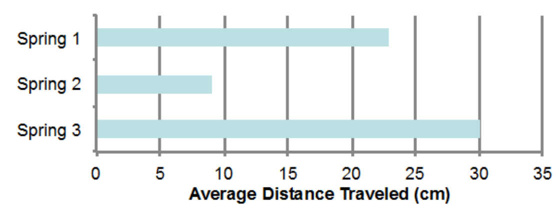
Figure A7. Scene 7 – Question 7.

In Scene 8, the examinee is again presented the experiment 1 results and their hypothesis and is asked if the design of the experiment was appropriate for the hypothesis. If the examinee answers, "Yes," then he or she is asked what changes could be made to the settings to provide additional information about the hypothesis. If the examinee answers, "No," then he or she is asked what changes could be made to the settings to better test the hypothesis.

Question 8 out of 12 | Progress bar

Experiment 1 Results

Spring	Trial 1 (cm)	Trial 2 (cm)	Trial 3 (cm)	Average Distance (cm)
Spring 1 Some coils; Moderately thick coils;	24	20	25	23
Spring 2 Very Few coils; Moderately thick coils,	7	8	12	9
Spring 3 Many coils; Moderately thick coils	30	28	32	30



Your hypothesis is: [FILL IN PREVIOUS ANSWER] Springs with_(more coils/less coils/thicker wire/ thinner wire) will make the car go farther.

8a. Were the settings you chose for your experiment appropriate for your hypothesis?
_______ Yes _______ No

8b. (If Yes) What changes could you make to the settings of the springs that would provide additional information about your hypothesis?

(If No) What changes could you make to the settings of the springs that would better test your hypothesis?

Figure A8. Scene 8 – Question 8.

In this scene, the examinee is provided with the opportunity to carry out the changes they described in the previous scene. The examinee is again asked to choose settings for the two variables and to explain why these settings are appropriate for the hypothesis.

Questions 9 and 10 out of 12 | Progress bar

Your hypothesis is: [FILL IN PREVIOUS ANSWER] Springs with_(more coils/less coils/ thicker wire/thinner wire) will make the car go farther.

You can rerun your experiment to obtain more information about your hypothesis.

You can either change the settings of your springs or leave them the same.

Remember that you are still testing the same hypothesis so your settings for the springs must reflect this hypothesis.

9.	Spring 4 	Spring 5 	Spring 6
Number of Coils	• Many ○ Some ○ Very few	○ Many • Some ○ Very few	○ Many ○ Some • Very few
Thickness of the wire	○ Very Thick • Moderately Thick ○ Thin	○ Very Thick • Moderately Thick ○ Thin	○ Very Thick • Moderately Thick ○ Thin

10. Explain why the choices you made for the settings of the springs are appropriate for your hypothesis.

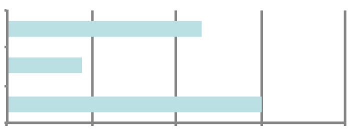
Figure A9. Scene 9 – Questions 9 and 10.

Results from Experiment 2 are provided in this scene. Examinees are asked how the results of the second experiment relate to the hypothesis, and to explain their reasoning.

Question 11 out of 12 | Progress bar

Experiment 2 Results

Spring	Trial 1 (cm)	Trial 2 (cm)	Trial 3 (cm)	Average Distance (cm)
Spring 4 Some coils; Moderately thick coils;	24	20	25	23
Spring 5 Very Few coils; Moderately thick coils,	7	8	12	9
Spring 6 Many coils; Moderately thick coils	30	28	32	30



Your hypothesis is: [FILL IN PREVIOUS ANSWER] Springs with_(more coils/less coils/thicker wire/thinner wire) will make the car go farther.

11a. Based on the results of Experiment 2, which of the following is true?

- ❑ These results **support** my hypothesis
- ❑ These results **contradict** my hypothesis
- ❑ These results **do not provide information** about my hypothesis

11b. Explain your answer:

Figure A10. Scene 10 – Question 11.

The bar graph results from both experiments are shown in Scene 11. Examinees are asked to use these results to pick a spring that makes the car goes the longest.

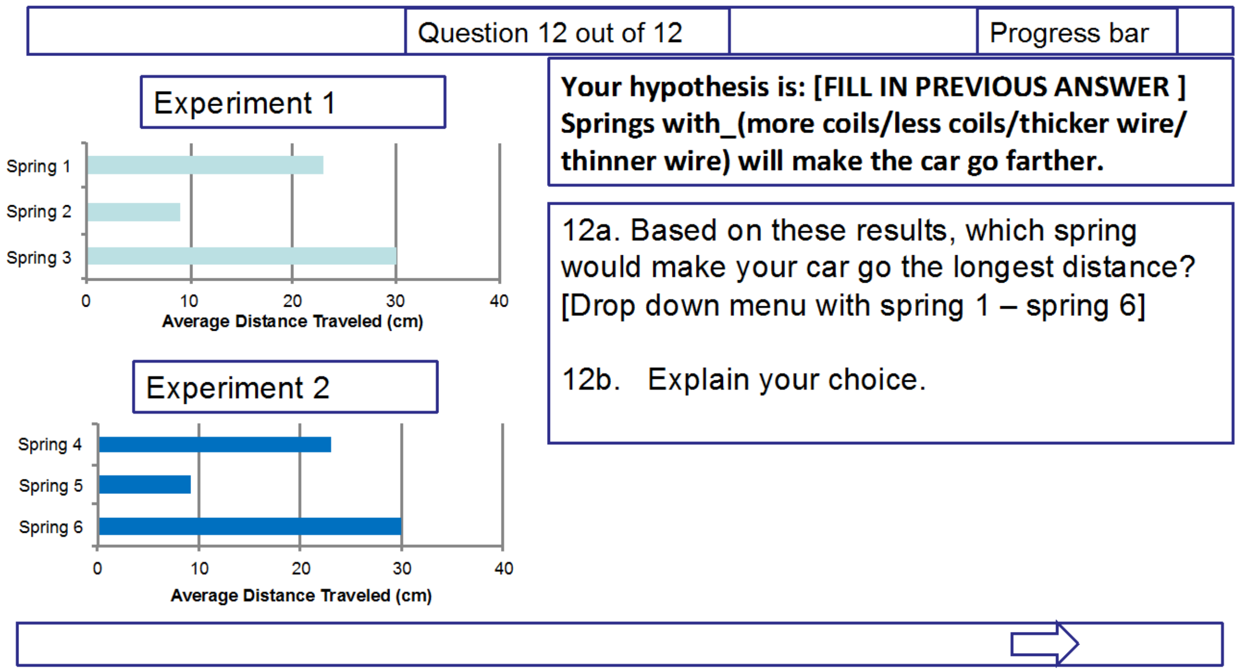


Figure A11. Scene 11 - Question 12.

At the end, the examinee/user is informed about which settings would have made the car go the longest distance.

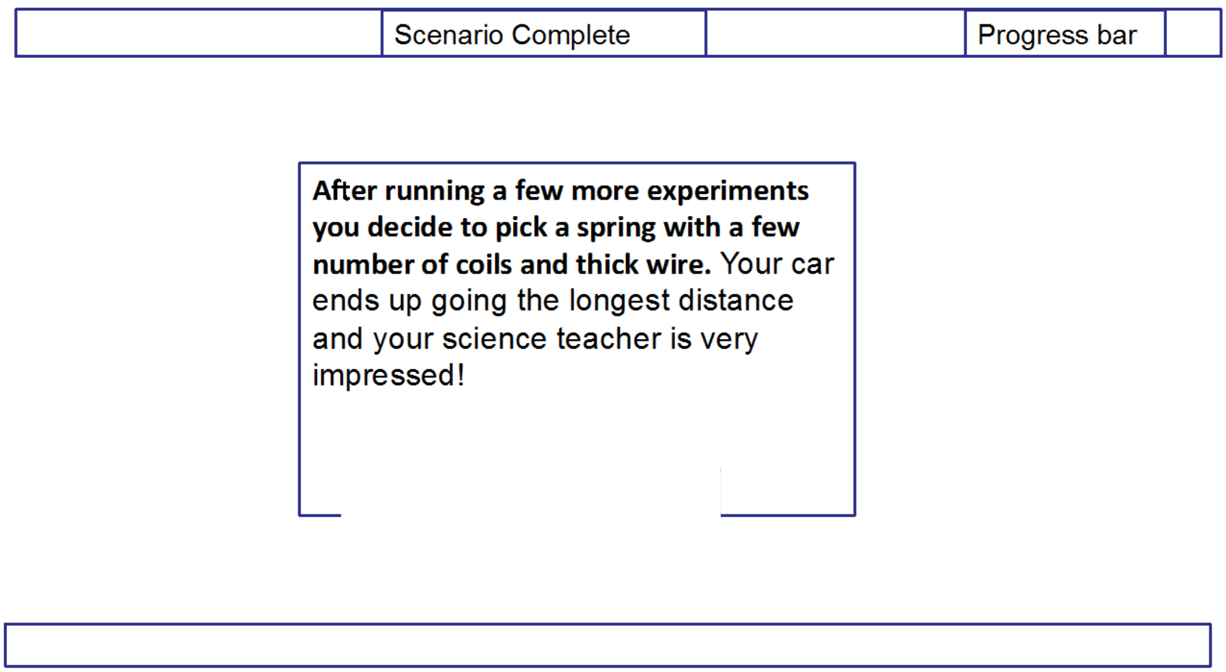


Figure A12. Scene 12 – Scenario complete.

Appendix B: Exemplar II. Science Tasks: Discrete, Multiple Choice Items, Online, UDL-Infused—Bicycle Rider

Design Pattern Used for Bicycle Rider

[NV] Interpreting Data in Tables, Charts, and Graphs - AERA 2011 | Design Pattern 2130

[| Permit | Delete | View: View (vertical)]

Title [Edit]

[NV] Interpreting Data in Tables, Charts, and Graphs - AERA 2011

Overview [Edit]

This Design Pattern describes key components of tasks that might be designed to measure students' ability to understand relationships among data as represented in canonical science and mathematical forms (i.e., tables, charts and graphs). Webb's Depth of Knowledge (DOK) framework is used throughout to scaffold design of items that tap this ability at each level of Webb's framework.

Focal Knowledge, Skills, and Abilities [Edit]

FK1. Ability to identify data points in one or more representational forms.

FK2. Ability to compare and /or contrast multiple representations and the data represented therein.

FK3. Ability to extrapolate or interpolate data points from given data.

FK4. Ability to describe simple mathematical relationships or trends among data.

FK5. Ability to draw conclusions or make predictions based on data.

Rationale [Edit]

R1. A key activity of science inquiry is working with data in many forms or representations. Students' ability to analyze relationships among data is integral to their participation in science inquiry and to represent and think critically about relationships among experimental variables, observed phenomena, etc.

Additional Knowledge, Skills, and Abilities [Edit]

AK1. ==
The following Additional KSAs are prerequisite knowledge that can be required for tasks that address the Focal KSA. Whether they are to be supported or not (e.g., glossary, background facts, equation list) is a decision to be made -- either by the assessment design team, at the level of the testing program, or at the level of the individual task if that is appropriate in the testing program.
==

AK2. Awareness of different representational forms

AK3. Knowledge of what data are

AK4. Ability to identify dependent and independent variables

AK5. Knowledge of mathematics

AK6. Scientific content knowledge

AK7. ==
The following Additional KSAs are generally construct-irrelevant knowledge, skills, or other attributes that may be involved in tasks generated under this design pattern. The task author can consider offering supports, presenting material, or getting work products that reduce or avoid requirements for these Additional KSAs, either through accommodated forms of a task or UDL principles. Many of these Additional KSAs are linked to Variable Task Features or Potential Work Products for suggestions on how to do this.
==

AK8. Perceptual
. vision
. hearing
. touch

AK9. Language and symbols
. vocabulary and symbols
. syntax and underlying structure
. English-language proficiency
. decoding text or math notation
. decoding charts, graphs, or images

AK10. Cognitive
. background knowledge
. concepts and categories
. information processing strategies
. memory and transfer

AK11. Skill and fluency

		. dexterity, strength, and mobility . navigation and object manipulation . automaticity (e.g., calculations, writing) . familiarity with media . facility with tools AK12. Executive (problem solving) . goal and expectation setting . goal maintenance and adjustment . planning and sequencing steps in a process . managing information and resources . working memory . monitoring progress AK13. Affective . intrinsic, task-specific motivation (challenge and/or threat, interest) . sustaining effort and persistence . coping skills and frustration management
Potential observations	[Edit]	Po1. Correct identification of the location of a data point in chart or graph OR the accurate identification of a value to complete a data table. Po2. Identification of representational forms of data that communicate the same mathematical relationships among data (or trends in data). Po3. Accuracy of conclusions drawn from data that are intended to inform predictions. Po4. Appropriateness of inferences drawn from data tables, charts, and graphs.
Potential work products	[Edit]	Pw1. Selection of inference or prediction (selected response) Pw2. Written interpretation of data from one or more representational forms Pw3. Written prediction based on interpretation of data
Potential rubrics	[Edit]	Pr1. Key for selected response items Pr2. Partial credit rubric for scoring of written responses
Characteristic features	[Edit]	Cf1. The presentation contains numeric data Cf2. The presentation includes at least one representational form Cf3. The presented data are in a scientific context
Variable features	[Edit]	Vf1. Number of representations Vf2. Type(s) of representations Vf3. Amount of data Vf4. Complexity of representational form(s) Vf5. Number of variables represented in the table, graph, or chart Vf6. Provision of an example Vf7. Amount of content knowledge required Vf8. Presence of color(s) in table, graph, or chart Vf9. Data source (student collected vs. provided) Vf10. Perceptual Features (1): Representational Format - Flexible size of text and images - Flexible amplitude of speech or sound - Adjustable contrast - Flexible colors - Flexible layout Vf11. Perceptual Features (2): Auditory Information - Text equivalents (e.g. captions, automated speech to text) - Visual graphics or outlines - Virtual manipulatives, video animation - Verbal descriptions - Tactile graphics, objects Vf12. Perceptual Features (3): Visual Information

- Spoken equivalents for text and images
- Automatic text to speech
- Tactile graphics
- Braille

Vf13. Language and Symbols (1): Supports for Vocabulary and Symbols
- Pre-taught vocabulary and symbols
- Embedded support for key terms (e.g. technical glossary, hyperlinks/ footnotes to definitions, illustrations, background knowledge)
- Embedded support for non-technical terms (e.g. non-technical glossary, hyperlinks/ footnotes to definitions, illustrations, background knowledge)
- Embedded alternatives for unfamiliar references (e.g. domain specific notation, jargon, figurative language, etc.)

Vf14. Language and Symbols (2): Supports for Syntactic Skills and Underlying Structure
- Alternate syntactic levels (simplified text)
- Grammar aids
- Highlighted syntactical elements (e.g. subjects, predicates, noun-verb agreement, adjectives, phrase structure, etc.)
- Highlight structural relations or make them more explicit

Vf15. Language and Symbols (3): Supports for English Language
- All key information in the dominant language (e.g. English) is also available in prevalent first languages (e.g. Spanish) for second language learners and in ASL for students who are deaf
- Key vocabulary words have links to both dominant and non-dominant definitions and pronunciations
- Domain-specific vocabulary (e.g. "matter" in science) is translated for both special and common meanings
- Electronic translation tools, multi-lingual glossaries

Vf16. Language and Symbols (4): Supports for Decoding and Fluency
- Digital text with automatic text to speech
- Digital Braille with automatic Braille to speech

Vf17. Cognitive Features (1): Supports for Background knowledge
- Advanced organizers, pre-teaching, relevant analogies and examples
- Links to prior knowledge (e.g. hyperlinks to multimedia, concrete objects in students' environments)
- Provision of an example

Vf18. Cognitive Features (2): Supports for Critical features, Big Ideas, and Relationships
- Concept maps, graphic organizers, outlines
- Highlight features in text, diagrams, graphics, and illustrations
- Reducing the field of competing information or distractions, masking
- Using multiple examples and non-examples to emphasize critical concepts

Vf19. Cognitive Features (3): Options that Guide Information Processing
- Explicit prompts for each step in a sequential process
- Interactive models that guide exploration and inspection
- Graduated scaffolds that support information processing strategies
- Multiple entry points and optional pathways through content
- Chunking information into smaller elements, progressive release of information, sequential highlighting
- Discrete question(s) or scenario-based text presentation
- Complexity of the scientific investigation presented in the scenario
- Cognitive complexity (Webb's Depth of Knowledge Levels)
- If selected response, distractors based on misconceptions/typical errors vs. non-misconceptions

Vf20. Cognitive Features (4): Supports for Memory and Transfer
- Checklists, organizers, sticky notes, electronic reminders
- Prompts for using mnemonic strategies and devices
- Templates, graphic organizers, concept maps to support note-taking
- Scaffolding that connects new information to prior knowledge
- Embedding new ideas in familiar ideas and contexts, use of analogy, metaphor, example

Vf21. Skill and Fluency (1): Supports for Manipulations
- Virtual manipulatives, Snap-to constraints
- Nonstick mats, Larger objects

Vf22. Skill and Fluency (2): Supports for Navigation

- Alternatives for physically interacting with materials: by hand, by voice, by single switch, by keyboard, by joystick, by adapted keyboard

Vf23. Skill and Fluency (3): Alternatives to Writing
- Voice recognition, Audio taping, Dictation, Video, Illustration

Vf24. Skill and Fluency (4): Supports for Composition
- Keyboarding and alternative keyboards, Onscreen keyboard,
- Wider lines, Larger paper, Pencil grips
- Drawing tools - with shapes, lines, etc.
- Blank tables, charts, graph paper
- Spellcheckers, calculators, sentence starters, word prediction, dictation (voice recognition or scribe), symbol-to-text, sentence strips

Vf25. Executive Features (1): Support for Goal and Expectation Setting
- Prompts and scaffolds to estimate effort, resources, and difficulty
- Animated agents that model the process and product of goal-setting
- Guides and checklists for scaffolding goal-setting

Vf26. Executive Features (2): Supports for Goal Maintenance and Adjustment
- Maintain salience of objectives and goals (e.g. reminders, progress charts)
- Adjust levels of challenge and support (e.g. adjustable leveling and embedded support, alternative levels of difficulty, alternative points of entry)

Vf27. Executive Features (3): Supports for Planning and Sequencing
- Embedded prompts to "stop and think" before acting
- Checklists and project planning templates for setting up prioritization, schedules, and steps
- Guides for breaking long-term objectives into reachable short-term objectives

Vf28. Executive Features (4): Supports for Managing Information
- Graphic organizers and templates for organizing information
- Embedded prompts for categorizing and systematizing
- Checklists and guides for note-taking

Vf29. Executive Features (5): Supports for Working Memory
- Note-taking, Mnemonic aids
- Locate items near relevant text

Vf30. Executive Features (6): Supports for Monitoring Progress
- Guided questions for self-monitoring
- Representations of progress (e.g. before and after photos, graphs and charts)
- Templates that guide self-reflection on quality and completeness
- Differentiated models of self-assessment strategies

Vf31. Affect Features (1): Supports for Intrinsic Motivation (Challenge and/or Threat)
- Offer individual choice
- Enhance relevance, value, authenticity (e.g. contextualize to students' lives, provision of an example)
- Options to vary level of novelty and risk (e.g. options in peer and adult support, alternatives to competition, alternatives to public display or performance, alternative consequences)
- Options to vary sensory stimulation (e.g. shortened work periods, frequent breaks, noise buffers, optional headphones, alternative settings, presentation of fewer items at a time)

Vf32. Affect Features (2): Supports for Sustaining Effort and Persistence
- Maintain salience of goals (e.g. explicit display of goals, periodic reminders, replacement of long-term goals with short-term objectives, prompts for visualization)
- Adjustable levels of challenge and support
- Encourage collaboration and support
- Communicate on-going, mastery-oriented feedback

Vf33. Affect Features (3): Support for Self-regulation
- Guide motivational goal-setting
- Scaffold self-regulatory skills and strategies
- Develop emotional self-assessment and reflection

I am a kind of [Edit]

These are kinds of me [Edit]

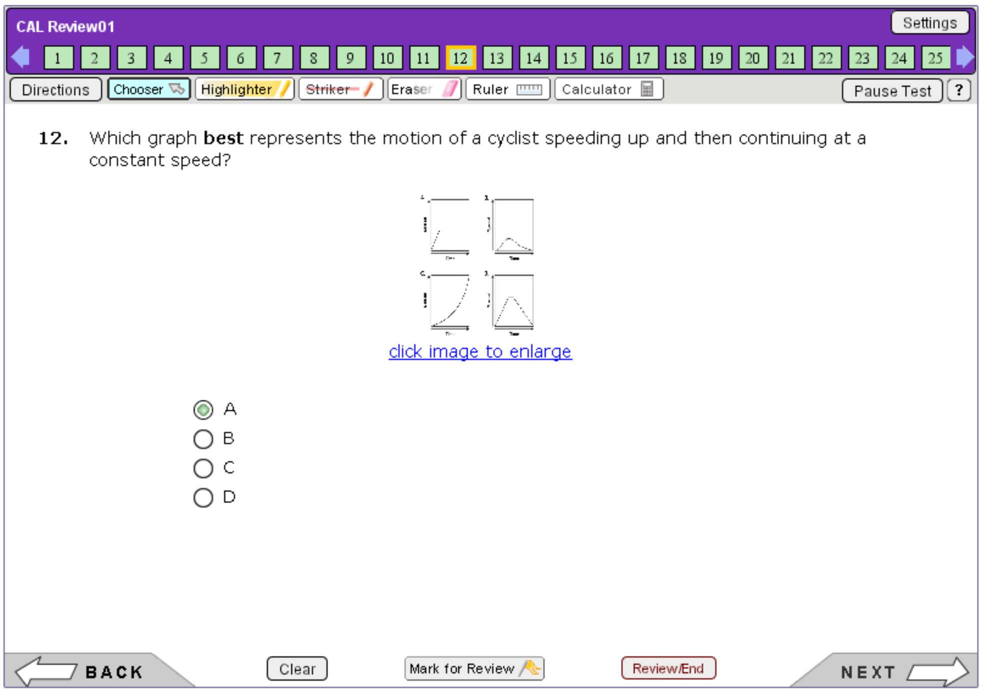
These are [Edit]

parts of me		
Educational standards	[Edit]	NV (3) Inquiry Standard N.8.A.1. Students know how to identify and critically evaluate information in data, tables, and graphs
Templates	[Edit]	
Exemplar tasks	[Edit]	
Online resources	[Edit]	
References	[Edit]	

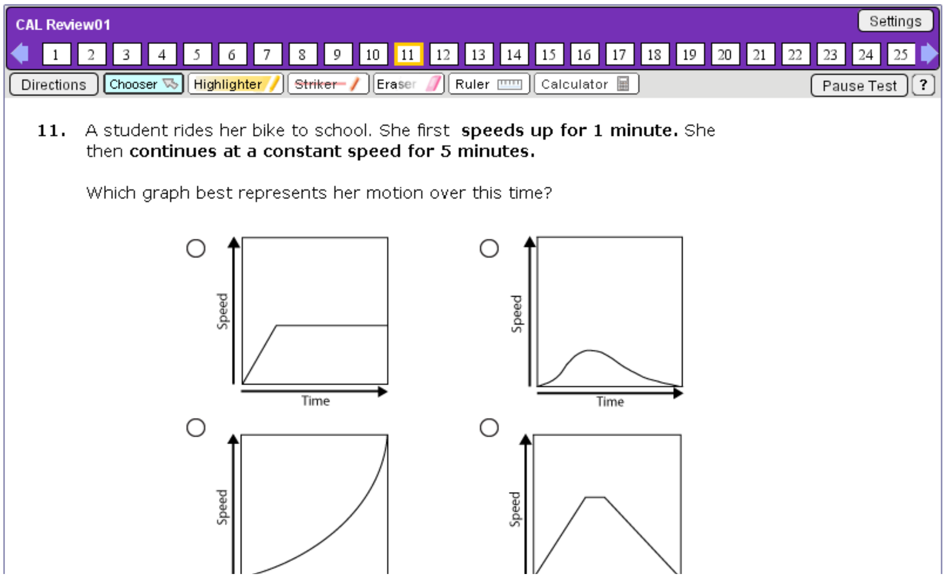
Tags [Add Tag]

(No tags entered.)

Original Bicycle Rider Item



Revised Bicycle Rider Item



Appendix C. Exemplar III. Science Tasks: Suite of Items, Graduated Degree of Complexity, UDL-Infused for Students with Significant Cognitive Disabilities—Recycling Suite of Items

Design Pattern Used for Recycling Suite of Tasks

PADI | Design Patterns | Education Standards | Exemplars | Templates | Task Specifications | Student Models | Student Model Variables | Activities | Meas. Models | Observable Variables | Eval. Procedures | Evaluation Phases | Work Products | Materials & Presentation | Task Model Variables

Hello csanford
Account Settings
Logout
Edit Model
Export to PDF

TEAS Symposium: Data Analysis and Probability B1 (grades 9-12) | Nu Design Pattern 3116 [Duplicate | Permit | Delete | View: View (vertical)]

Title	[Edit]	TEAS Symposium: Data Analysis and Probability B1 (grades 9-12)
Overview	[Edit]	For univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics
Rationale	[Edit]	R1. Students in grades 9-12 must learn how to use appropriate methods to analyze data. Being able to display, describe, and calculate summary statistics for univariate data is an important precursor to doing the same for bivariate data.
Focal KSAs	[Edit]	FK1. Ability to identify and/or calculate summary statistic to answer a question when given a data set and a question about the data FK2. Ability to answer a question about data by identifying, creating, and using a graphical display, and calculating the summary statistic
Add'l KSAs: Cognitive Background Knowledge	[Edit]	AK1. Knowledge of what histograms, dot plots, stem-and-leaf and box plots are AK2. Ability to create graphical representations (e.g., histograms, dot plots, stem-and-leaf and box plots) of a set of data AK3. Ability to use (read and interpret) graphical representations (e.g., histograms, dot plots, stem-and-leaf and box plots) to answer questions about the data AK4. Ability to select representations (e.g., histograms, dot plots, stem-and-leaf and box plots) of different types of data (e.g., frequency data, univariate data) to answer specific questions about the data AK5. Knowledge of prerequisite vocabulary and symbols, and basic understand of concept (e.g., median, outliers, spread, minimum and maximum, range, box plot, stem and leaf, mode) AK6. Knowledge of what data are (e.g. a number that represents a property of some item) AK7. Ability to add, subtract, multiply and divide
Add'l KSAs: Perceptual (Receptive)	[Edit]	Ability to perceive images in the stimulus material and question. (e.g., through print, objects, holistic description, Braille, audio description, tactile images) (Image in this case means a picture, drawing, table, map, graph, or photograph and not a mental image) Ability to perceive physical objects required for the task. (e.g., see physical objects used to relate a story) Ability to perceive the linguistic components of the stimulus material and question. (e.g., through print, objects, audio, Braille, tactile images)

Add'l KSAs: Skill and Fluency (Expressive)	[Edit]	Ability to communicate response. (e.g., respond verbally, by using pictures, by making a selection from a group) Ability to express a response in text. (e.g., by writing, drawing, using Braille, using a scribe, using Dragon Dictate) Ability to manipulate digital/electronic equipment. (e.g., assistive technology) Ability to manipulate physical materials. (e.g., dexterity, strength, and mobility) Knowledge of how to use physical materials or digital/electronic equipment. (e.g., familiarity, assistive technology)
Add'l KSAs: Language and Symbols	[Edit]	Ability to comprehend text, symbols, images, or objects. (Image in this case means a picture, drawing, table, map, graph, or photograph, and not a mental image) Ability to decode text, symbols, tactile images, images, or objects. (Image in this case means a picture, drawing, table, map, graph, or photograph, and not a mental image) Ability to recognize text, symbols, tactile images, images, or objects. (Image in this case means a picture, drawing, table, map, graph, or photograph, and not a mental image) Ability to understand English vocabulary and syntax. (If the student doesn't have the linguistic competency then it would be hard to support. If a student speaks another language then a bilingual translator can be used)

Add'l KSAs: Executive	[Edit]	Ability to monitor goals and progress. (e.g., breaking long-term objectives [i.e., objectives with multiple parts that can be administered over several sessions] into short-term goals) Ability to plan and sequence. (e.g., for items with a sequence of steps that must be completed in a particular order (could be a single step problem) that are likely to be administered in one session)
Add'l KSAs: Affective	[Edit]	Ability to engage. (e.g., task-specific motivation) Ability to persist and sustain effort.

Potential Observations	[Edit]	PO1. Given data and a question, student correctly identifies the statistic appropriate to answer the question, and correctly calculates the summary statistic PO2. Given data and a question specifying the statistic to be calculated, student correctly calculates the summary statistic PO3. Given data and a question, student correctly selects a summary statistic from a list of different types of summary statistics to answer the question PO4. Given data and a question, student correctly selects the value that represents the summary statistic PO5. Given data and a questions, student correctly selects the calculation method (formula) to calculate the summary statistic PO6. Given data and a question, student (1) correctly determines the type of display to be created; (2) correctly creates the display; (3) correctly interprets the display; and (4) correctly calculates or selects the summary statistic that answers the question NOTE: Order of steps can vary depending on the question and type of representation to be created, e.g., in order to create a box plot, one needs to first calculate the median)

Potential Work Products	[Edit]	PW1. Identification of a summary statistic; calculation of a summary statistic based on data presented in a question about the data PW2. Calculation of a summary statistic based on data presented in a question about the data PW3. Selection of a type summary statistic based on data presented in a question about the data PW4. Selection of a value for a summary statistic based on data presented in a question about the data PW5. Selection of a method of a summary statistic based on data presented in a question about the data PW6. Selection/expression of the type of display to be created; and creation of data display; and expression of relationships among variables represented in the data display; and selection/calculation of the summary statistic NOTE: Multi-step problems can be scored by assigning a score to each sub-part of the item and then aggregating those scores to reflect the Focal KSA
Characteristic Features	[Edit]	CF1. Items will present a data set about which questions can be answered through graphic displays and/or summary statistics CF2. The presentation of a framing question that incorporates familiar context CF3. Only univariate measurement data will be used
Variable Features: Cognitive Background Knowledge	[Edit]	VF1. Provide student with definitions and examples of various types of data and their corresponding types of graphical representations (e.g., histograms, dot plots, stem-and-leaf and box plots) VF2. Provide familiar materials that support student's understanding of histograms, dot plots, stem-and-leaf and box plots; pre-taught vocabulary and symbols VF3. Provide student with graph paper VF4. Provide student with a glossary (including illustrations) of pre-taught vocabulary and symbols VF5. Provide mechanism to highlight key features of a graph to support a consideration or an understanding of the appropriateness of different representations VF6. Provide mechanism to highlight or mask parts of the data set VF7. Provide a calculator
Variable Features: Perceptual (Receptive)		• Delivery mechanisms by which the question is perceived. (e.g., read aloud verbatim/read aloud paraphrase, pictures, large print, printed text, Braille, text, symbols, rebuses, concrete objects, description of objects or images, text to speech, signing, auditory amplification, closed captioning, CCTV - close circuit TV to increase size of font, vary contrast, etc.) • Delivery parameters for oral presentation of material. (e.g., speed of reading, volume, amount of expression used, student ability to pause, stop and/or repeat information read aloud) • Supports for the use of equipment required for the task. (e.g., communication board, CD player; Possible to reprogram communication board to include punctuation, capitalization, etc.)
Variable Features: Skill and Fluency (Expressive)		• Supports for manipulating physical materials. (e.g., use of velcro, size of materials, teacher manipulation of materials; In writing, student can manipulate cards with punctuation symbols on them and velcro on back to apply correct punctuation to a sentence) • Supports for manipulating digital/electronic equipment. (e.g., pointers, teacher manipulation of equipment, spoken commands, stylus for input, larger keyboard/buttons, adaptive mouse) • Supports for composing a response in text. (e.g., speech to text, written by teacher, keyboarding, word prediction software) • Practice with familiar equipment. • Response mode options. (e.g., pointing, speech and verbalization, writing, signing, switch or other assistive device/augmentative communication device, eye gaze; for lowest functioning students: predictable behavioral response, tolerate assistance such as hand over hand) • Practice tutorials with unfamiliar physical materials or digital/electronic equipment. (Practice tutorials can be used to introduce students to new item formats or modeled examples using materials that are not construct relevant or new tools to support test taking)

Variable Features: Language and Symbols	• Embedded support for vocabulary and symbols. (e.g., technical and non-technical glossary, hyperlinks/footnotes to definitions, illustrations, background knowledge) • Digital text with or without automatic text to speech. • Highlight essential elements, words, or phrases. • All key information in the dominant language (e.g., English) is also available in prevalent first languages (e.g., Spanish) for second language learners. • All key information available in sign language for students who are deaf. • Digital Braille with or without automatic Braille to speech. • Alternate syntactic levels (simplified text). • Level of abstraction required of student. (e.g., concrete objects, images, text) • New vs. pre-taught vocabulary and symbols. • Use of multiple representations. (e.g., physical models, demonstrations, acting out scenarios) • Read language and symbols aloud.
Variable Features: Cognitive	• Options for guiding exploration and information processing: use consistent signals/cues. (Signals/cues may include designations in assessments such as line numbers in passages, symbols for directions [e.g., stop signs to stop, arrows to continue], or behavioral gestures indicating where a student should mark a response) • Options for supporting critical features, big ideas, and relations: provide graphic organizers. • Options for guiding exploration and information processing: provide sequential highlighting. (Definition: to emphasize or make information prominent as it appears in a sequence by differentiated use of color, lighting, sound, or tactile surface [e.g., highlight the paragraph in yellow and highlight each word as it is read in blue]) • Options for supporting critical features, big ideas, and relations: provide a response template. • Options for supporting critical features, big ideas, and relations: outline information. • Options for guiding exploration and information processing: allow viewing of stimuli from previous stages and parts. • Options for guiding exploration and information processing: provide a guide or checklist for prioritization of steps in multi-step problems. • Options for guiding exploration and information processing: provide multiple entry points. • Options for guiding exploration and information processing: mask part of the information. (Masking incorrect response in a selected response item [aka strike out]. Student selects the incorrect response to be masked. Teacher presents all response options at first trial and then if response is incorrect the teacher masks the student's incorrect response item [see Florida approach; note: state test level decision on how to deal with incorrect responses when there is multiple response options]) • Options for supporting memory and transfer: note-taking. • Options for supporting memory and transfer: locate items near relevant text. • Options for supporting memory and transfer: reread question/stimulus.
	• Options for supporting critical features, big ideas, and relations: highlight information.
Variable Features: Executive	• Representations of progress. (e.g., before and after photos, graphs and charts) • Prompts, scaffolds, and questions to monitor progress, to "stop and think", and for categorizing and systematizing. • Prompts and scaffolds to estimate effort, resources, and difficulty. • Guides, checklists, graphic organizers, and/or templates for goal setting, prioritizing, breaking long-term objectives into reachable short-term goals, self-reflection, and self-assessment. • Adjust levels of challenge and support. (e.g., adjustable leveling and embedded support, alternative levels of difficulty, alternative points of entry)

Variable Features: Affective		• Task options for engagement: variety of stimuli. • Task options for engagement: item/task format. (e.g., selected response vs. constructed response, performance) Task options for writing: Student writes 2-3 sentences, Present a written sentence and student corrects it, Compose sentences using words and punctuation from words/punctuation represented on cards, Technology-enhanced writing tasks • Task options for engagement: heighten salience. • Task options for engagement: enhance relevance, value, and authenticity of tasks. (task refers to the assessment items, stimulus "story", and materials) In writing: create a letter to a friend, use stories with their own names or names of classmates, content out of students' personal life • Teacher options for providing supports for attention and engagement: provide varied levels of challenge and support. • Teacher options for providing supports for attention and engagement: provide supports to reduce student frustration. (e.g., noise reduction, extended test taking time, contingencies, number of items administered at one time) • Teacher options for providing supports for attention and engagement: provide optimal student positioning (positions which encourage alertness, not recumbent). • Teacher options for providing supports for attention and engagement: provide feedback to support engagement. • Teacher options for providing supports for attention and engagement: prompt student to engage/re-engage. • Teacher options for providing supports for attention and engagement: cover up part of text so student isn't overwhelmed. • Task options for engagement: vary amount of context supporting tasks. (e.g., discrete tasks vs. scenarios) • Teacher options for providing supports for attention and engagement: administer assessment at optimal time of day for student engagement. • Teacher options for providing supports for attention and engagement: provide verbal/gestural prompts.
Educational Standards	[Edit]	State 1 Standard 1. Identify when data from part of a group (sample) should not be used to make predictions regarding the whole group State 1 Standard 2. Identify problems with inaccurate counting when collection data and use strategies to correct mistakes State 1 Standard 3. Identify a missing part of objects, pictures, or symbols in real-world situations State 1 Standard 4. Describe information in bar graphs, circle graphs, and single-line graphs representing data from real-world situations State 1 Standard 5. Collect data and display in single-line graphs, circle graphs, and bar graphs State 1 Standard 6. Determine the mode by identifying the number that occurs most often and the mean by finding the average State 1 Standard 7. Calculate the range and median for data from real-world situations State 1 Standard 8. Identify information in simple pictographs and bar graphs that represent data from real-world situations State 1 Standard 9. Organize data in pictographs and bar graphs and identify the labels for categories State 1 Standard 10. Identify the number that occurs most frequently (mode) in a set of data with up to nine numbers

		State 1 Standard 11. Find the difference between the largest and smallest numbers n a set of data (range) and the median in a real-world situation State 1 Standard 12. Identify quantity in data sets of 10, by counting objects, pictures, or symbols, and identify which category has more, less, or none State 2 Standard 1. Read and interpret tables, charts, and graphs, including line graphs, bar graphs, frequency tables, or circle graphs State 2 Standard 2. Use basic statistical concepts, including mean, median, mode or range State 2 Standard 3. Make predictions and draw conclusions based on a simple set of data and its statistical measures State 3 Standard 1. Collect, record, and display data using appropriate methods (pictograph, table, bar graph, line plot, scatterplot) to answer a question (e.g. How much rain did we get each day this week?) State 3 Standard 2. Use data to make a prediction, see a pattern, or draw a conclusion (e.g. Are there more boys than girls? Why? What is the favorite ice cream flavor of the class?) State 3 Standard 3. Use a graph to compare sets of data and draw conclusions (e.g. Analyze a scatter plot of rain in St. George and in Salt Lake City over the course of a year. Which city has the most rain in July?)
Variable Features: Cognitive		• Options for guiding exploration and information processing: use consistent signals/cues. (Signals/cues may include designations in assessments such as line numbers in passages, symbols for directions [e.g., stop signs to stop, arrows to continue], or behavioral gestures indicating where a student should mark a response) • Options for supporting critical features, big ideas, and relations: provide graphic organizers. • Options for guiding exploration and information processing: provide sequential highlighting. (Definition: to emphasize or make information prominent as it appears in a sequence by differentiated use of color, lighting, sound, or tactile surface [e.g., highlight the paragraph in yellow and highlight each word as it is read in blue]) • Options for supporting critical features, big ideas, and relations: provide a response template. • Options for supporting critical features, big ideas, and relations: outline information. • Options for guiding exploration and information processing: allow viewing of stimuli from previous stages and parts. • Options for guiding exploration and information processing: provide a guide or checklist for prioritization of steps in multi-step problems. • Options for guiding exploration and information processing: provide multiple entry points. • Options for guiding exploration and information processing: mask part of the information. (Masking incorrect response in a selected response item [aka strike out]. Student selects the incorrect response to be masked. Teacher presents all response options at first trial and then if response is incorrect the teacher masks the student's incorrect response item [see Florida approach; note: state test level decision on how to deal with incorrect responses when there is multiple response options]) • Options for supporting memory and transfer: note-taking. • Options for supporting memory and transfer: locate items near relevant text. • Options for supporting memory and transfer: reread question/stimulus. • Options for supporting critical features, big ideas, and relations: highlight information.
Variable Features: Executive		• Representations of progress. (e.g., before and after photos, graphs and charts) • Prompts, scaffolds, and questions to monitor progress, to "stop and think", and for categorizing and systematizing. • Prompts and scaffolds to estimate effort, resources, and difficulty. • Guides, checklists, graphic organizers, and/or templates for goal setting, prioritizing, breaking long-term objectives into reachable short-term goals, self-reflection, and self-assessment. • Adjust levels of challenge and support. (e.g., adjustable leveling and embedded support, alternative levels of difficulty, alternative points of entry)

Variable Features: Affective	• Task options for engagement: variety of stimuli. • Task options for engagement: item/task format. (e.g., selected response vs. constructed response, performance) Task options for writing: Student writes 2-3 sentences, Present a written sentence and student corrects it, Compose sentences using words and punctuation from words/punctuation represented on cards, Technology-enhanced writing tasks • Task options for engagement: heighten salience. • Task options for engagement: enhance relevance, value, and authenticity of tasks. (task refers to the assessment items, stimulus "story", and materials) In writing: create a letter to a friend, use stories with their own names or names of classmates, content out of students' personal life • Teacher options for providing supports for attention and engagement: provide varied levels of challenge and support. • Teacher options for providing supports for attention and engagement: provide supports to reduce student frustration. (e.g., noise reduction, extended test taking time, contingencies, number of items administered at one time) • Teacher options for providing supports for attention and engagement: provide optimal student positioning (positions which encourage alertness, not recumbent). • Teacher options for providing supports for attention and engagement: provide feedback to support engagement. • Teacher options for providing supports for attention and engagement: prompt student to engage/re-engage. • Teacher options for providing supports for attention and engagement: cover up part of text so student isn't overwhelmed. • Task options for engagement: vary amount of context supporting tasks. (e.g., discrete tasks vs. scenarios) • Teacher options for providing supports for attention and engagement: administer assessment at optimal time of day for student engagement. • Teacher options for providing supports for attention and engagement: provide verbal/gestural prompts.

Educational Standards [Edit]	State 1 Standard 1. Identify when data from part of a group (sample) should not be used to make predictions regarding the whole group State 1 Standard 2. Identify problems with inaccurate counting when collection data and use strategies to correct mistakes State 1 Standard 3. Identify a missing part of objects, pictures, or symbols in real-world situations State 1 Standard 4. Describe information in bar graphs, circle graphs, and single-line graphs representing data from real-world situations State 1 Standard 5. Collect data and display in single-line graphs, circle graphs, and bar graphs State 1 Standard 6. Determine the mode by identifying the number that occurs most often and the mean by finding the average State 1 Standard 7. Calculate the range and median for data from real-world situations State 1 Standard 8. Identify information in simple pictographs and bar graphs that represent data from real-world situations State 1 Standard 9. Organize data in pictographs and bar graphs and identify the labels for categories State 1 Standard 10. Identify the number that occurs most frequently (mode) in a set of data with up to nine numbers State 1 Standard 11. Find the difference between the largest and smallest numbers n a set of data (range) and the median in a real-world situation State 1 Standard 12. Identify quantity in data sets of 10, by counting objects, pictures, or symbols, and identify which category has more, less, or none State 2 Standard 1. Read and interpret tables, charts, and graphs, including line graphs, bar graphs, frequency tables, or circle graphs State 2 Standard 2. Use basic statistical concepts, including mean, median, mode or range State 2 Standard 3. Make predictions and draw conclusions based on a simple set of data and its statistical measures State 3 Standard 1. Collect, record, and display data using appropriate methods (pictograph, table, bar graph, line plot, scatterplot) to answer a question (e.g. How much rain did we get each day this week?) State 3 Standard 2. Use data to make a prediction, see a pattern, or draw a conclusion (e.g. Are there more boys than girls? Why? What is the favorite ice cream flavor of the class?) State 3 Standard 3. Use a graph to compare sets of data and draw conclusions (e.g. Analyze a scatter plot of rain in St. George and in Salt Lake City over the course of a year. Which city has the most rain in July?)

Appendix C2. Task Template for Recycling Suite of Items
Data Analysis and Probability B1 (Grades 9–12)

PADI | Design Patterns | Education Standards | Exemplars | Templates | Task Specifications | Student Models | Student Model Variables | Activities | Meas. Models | Observable Variables | Eval. Procedures | Evaluation Phases | Work Products | Materials & Presentation | Task Model Variables

Hello csanford
Account Settings
Logout
Edit Model
Export to PDF

TEAS Symposium: Data Analysis and Probability B1 (grades 9-12) Task | Task Family 3135

[Duplicate | Permit | Delete]

Title [Edit]	TEAS Symposium: Data Analysis and Probability B1 (grades 9-12) Task
Nu Design Pattern [Edit]	TEAS Symposium: Data Analysis and Probability B1 (grades 9-12) For univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics
Grade Level Activities [Edit]	GL1. Describe center, spread, and shape of univariate and bivariate data GL2. Use a variety of summary statistics and graphical displays to analyze the characteristics of the data GL3. Construct histograms, dotplots, stem-and-leaf plots, box plots, or scatter plots and select from among them to assist in understanding data GL4. Ability to comment on overall shape of the plots and on points that do not fit the general shape GL5. Explain differences in measures of center (e.g., mean or median) and spread (e.g., standard deviation or interquartile range) GL6. Recognize the sample mean and median can different greatly for a skewed distribution GL7. Understand that for data that are identified by categories (e.g., gender, favorite color, ethnic origin), bar graphs, pie charts, and summary table often display information about the relative frequency or percent in each category

	Item 4	Item 3	Item 2	Item 1
Depth of Knowledge (DOK)	Do1. Application (Organize, Compute, Order, Construct)	Do1. Performance (Perform, Demonstrate, Locate, Follow procedures)	Do1. Recall (Identify)	Do1. Recall (Identify)
Selected Focal KSAs	• Ability to answer a question about data by identifying, creating, and using a graphical display, and calculating the summary statistic	• Ability to answer a question about data by identifying, creating, and using a graphical display, and calculating the summary statistic		

Focal KSA Notes				
Selected KSA for Items 1 and 2			• Ability to use (read and interpret) graphical representations (e.g., histograms, dot plots, stem-and-leaf and box plots) to answer questions about the data	• Knowledge of what histograms, dot plots, stem-and-leaf and box plots are
KSA for Items 1 and 2 Notes				
Associated AKSAs, Cognitive Background Knowledge	• Knowledge of what histograms, dot plots, stem-and-leaf and box plots are • Ability to create graphical representations (e.g., histograms, dot plots, stem-and-leaf and box plots) of a set of data • Ability to use (read and interpret) graphical representations (e.g., histograms, dot plots, stem-and-leaf and box plots) to answer questions about the data • Ability to select representations (e.g., histograms, dot plots, stem-and-leaf and box plots) of different types of data (e.g., frequency data, univariate data) to answer specific questions about the data • Knowledge of prerequisite vocabulary and symbols, and basic understand of concept (e.g., median, outliers, spread, minimum and maximum, range, box plot, stem and leaf, mode) • Knowledge of what data are (e.g. a number that represents a property of some item) • Ability to add, subtract, multiply and divide	• Knowledge of what histograms, dot plots, stem-and-leaf and box plots are • Ability to create graphical representations (e.g., histograms, dot plots, stem-and-leaf and box plots) of a set of data • Ability to use (read and interpret) graphical representations (e.g., histograms, dot plots, stem-and-leaf and box plots) to answer questions about the data • Ability to select representations (e.g., histograms, dot plots, stem-and-leaf and box plots) of different types of data (e.g., frequency data, univariate data) to answer specific questions about the data • Knowledge of prerequisite vocabulary and symbols, and basic understand of concept (e.g., median, outliers, spread, minimum and maximum, range, box plot, stem and leaf, mode) • Knowledge of what data are (e.g. a number that represents a property of some item) • Ability to add, subtract, multiply and divide	• Knowledge of what histograms, dot plots, stem-and-leaf and box plots are • Ability to select representations (e.g., histograms, dot plots, stem-and-leaf and box plots) of different types of data (e.g., frequency data, univariate data) to answer specific questions about the data • Knowledge of prerequisite vocabulary and symbols, and basic understand of concept (e.g., median, outliers, spread, minimum and maximum, range, box plot, stem and leaf, mode) • Knowledge of what data are (e.g. a number that represents a property of some item) • Ability to add, subtract, multiply and divide	

Potential Observations	• Given data and a question, student (1) correctly determines the type of display to be created; (2) correctly creates the display; (3) correctly interprets the display; and (4) correctly calculates or selects the summary statistic that answers the question NOTE: Order of steps can vary depending on the question and type of representation to be created, e.g., in order to create a box plot, one needs to first calculate the median)	• Given data and a question, student (1) correctly determines the type of display to be created; (2) correctly creates the display; (3) correctly interprets the display; and (4) correctly calculates or selects the summary statistic that answers the question NOTE: Order of steps can vary depending on the question and type of representation to be created, e.g., in order to create a box plot, one needs to first calculate the median)		
Potential Observation Notes (based on selected KSA)				
Potential Work Products	• Selection/expression of the type of display to be created; and creation of data display; and expression of relationships among variables represented in the data display; and selection/calculation of the summary statistic NOTE: Multi-step problems can be scored by assigning a score to each sub-part of the item and then aggregating those scores to reflect the Focal KSA	• Selection/expression of the type of display to be created; and creation of data display; and expression of relationships among variables represented in the data display; and selection/calculation of the summary statistic NOTE: Multi-step problems can be scored by assigning a score to each sub-part of the item and then aggregating those scores to reflect the Focal KSA		

Potential Work Product Notes (based on selected KSA)				
Characteristic Features	• Items will present a data set about which questions can be answered through graphic displays and/or summary statistics • The presentation of a framing question that incorporates familiar context • Only univariate measurement data will be used	• Items will present a data set about which questions can be answered through graphic displays and/or summary statistics • The presentation of a framing question that incorporates familiar context • Only univariate measurement data will be used	• Items will present a data set about which questions can be answered through graphic displays and/or summary statistics • The presentation of a framing question that incorporates familiar context • Only univariate measurement data will be used	• Items will present a data set about which questions can be answered through graphic displays and/or summary statistics • The presentation of a framing question that incorporates familiar context • Only univariate measurement data will be used
Associated Variable Features, Cognitive Background Knowledge	• Provide student with definitions and examples of various types of data and their corresponding types of graphical representations (e.g., histograms, dot plots, stem-and-leaf and box plots): **Yes** • Provide familiar materials that support student's understanding of histograms, dot plots, stem-and-leaf and box plots; pre-taught vocabulary and symbols: **Yes** • Provide student with graph paper: **Yes** • Provide student with a glossary (including illustrations) of pre-taught vocabulary and symbols: **Yes** • Provide mechanism to highlight key features of a graph to support a consideration or an understanding of the appropriateness of different representations: **Yes** • Provide mechanism to highlight or mask parts of the data set: **Yes** • Provide a calculator: **Yes**	• Provide student with definitions and examples of various types of data and their corresponding types of graphical representations (e.g., histograms, dot plots, stem-and-leaf and box plots): **Yes** • Provide familiar materials that support student's understanding of histograms, dot plots, stem-and-leaf and box plots; pre-taught vocabulary and symbols: **Yes** • Provide student with graph paper: **Yes** • Provide student with a glossary (including illustrations) of pre-taught vocabulary and symbols: **Yes** • Provide mechanism to highlight key features of a graph to support a consideration or an understanding of the appropriateness of different representations: **Yes** • Provide mechanism to highlight or mask parts of the data set: **Yes** • Provide a calculator: **Yes**	• Provide student with definitions and examples of various types of data and their corresponding types of graphical representations (e.g., histograms, dot plots, stem-and-leaf and box plots): **Yes** • Provide familiar materials that support student's understanding of histograms, dot plots, stem-and-leaf and box plots; pre-taught vocabulary and symbols: **Yes** • Provide student with graph paper: **No** • Provide student with a glossary (including illustrations) of pre-taught vocabulary and symbols: **Yes** • Provide mechanism to highlight key features of a graph to support a consideration or an understanding of the appropriateness of different representations: **Yes** • Provide mechanism to highlight or mask parts of the data set: **Yes** • Provide a calculator: **No**	• Provide student with definitions and examples of various types of data and their corresponding types of graphical representations (e.g., histograms, dot plots, stem-and-leaf and box plots): **No** • Provide familiar materials that support student's understanding of histograms, dot plots, stem-and-leaf and box plots; pre-taught vocabulary and symbols: **No** • Provide student with graph paper: **No** • Provide student with a glossary (including illustrations) of pre-taught vocabulary and symbols: **No** • Provide mechanism to highlight key features of a graph to support a consideration or an understanding of the appropriateness of different representations: **No** • Provide mechanism to highlight or mask parts of the data set: **No** • Provide a calculator: **No**

Selected Variable Features: Perceptual	• Delivery mechanisms by which the question is perceived eg: **Yes** • Delivery parameters for oral presentation of material eg: **Yes** • Supports for the use of equipment required for the task eg: **Yes**	• Delivery mechanisms by which the question is perceived eg: **Yes** • Delivery parameters for oral presentation of material eg: **Yes** • Supports for the use of equipment required for the task eg: **Yes**	• Delivery mechanisms by which the question is perceived eg: **Yes** • Delivery parameters for oral presentation of material eg: **Yes** • Supports for the use of equipment required for the task eg: **Yes**	• Delivery mechanisms by which the question is perceived eg: **Yes** • Delivery parameters for oral presentation of material eg: **Yes** • Supports for the use of equipment required for the task eg: **Yes**
Selected Variable Features: Skill and Fluency	• Supports for manipulating physical materials eg: **Yes** • Supports for manipulating digital/electronic equipment eg: **Yes** • Supports for composing a response in text eg: **Yes** • Practice with familiar equipment: **No** • Response mode options eg: **Yes** • Practice tutorials with unfamiliar physical materials or digital/electronic equipment eg: **No**	• Supports for manipulating physical materials eg: **Yes** • Supports for manipulating digital/electronic equipment eg: **Yes** • Supports for composing a response in text eg: **Yes** • Practice with familiar equipment: **No** • Response mode options eg: **Yes** • Practice tutorials with unfamiliar physical materials or digital/electronic equipment eg: **No**	• Supports for manipulating physical materials eg: **Yes** • Supports for manipulating digital/electronic equipment eg: **Yes** • Supports for composing a response in text eg: **No** • Practice with familiar equipment: **No** • Response mode options eg: **Yes** • Practice tutorials with unfamiliar physical materials or digital/electronic equipment eg: **No**	• Supports for manipulating physical materials eg: **No** • Supports for manipulating digital/electronic equipment eg: **Yes** • Supports for composing a response in text eg: **No** • Practice with familiar equipment: **No** • Response mode options eg: **Yes** • Practice tutorials with unfamiliar physical materials or digital/electronic equipment eg: **No**

Selected Variable Features: Affective	• Task options for engagement: variety of stimuli: **Implemented** • Task options for engagement: item/task format eg: **No** • Task options for engagement: heighten salience: **No** • Task options for engagement: enhance relevance, value, and authenticity of tasks eg: **Implemented** • Teacher options for providing supports for attention and engagement: provide varied levels of challenge and support: **No** • Teacher options for providing supports for attention and engagement: provide supports to reduce student frustration eg: **Yes**	• Task options for engagement: variety of stimuli: **Implemented** • Task options for engagement: item/task format eg: **No** • Task options for engagement: heighten salience: **No** • Task options for engagement: enhance relevance, value, and authenticity of tasks eg: **Implemented** • Teacher options for providing supports for attention and engagement: provide varied levels of challenge and support: **No** • Teacher options for providing supports for attention and engagement: provide supports to reduce student frustration eg: **Yes**	• Task options for engagement: variety of stimuli: **Implemented** • Task options for engagement: item/task format eg: **No** • Task options for engagement: heighten salience: **No** • Task options for engagement: enhance relevance, value, and authenticity of tasks eg: **Implemented** • Teacher options for providing supports for attention and engagement: provide varied levels of challenge and support: **No** • Teacher options for providing supports for attention and engagement: provide	• Task options for engagement: variety of stimuli: **Implemented** • Task options for engagement: item/task format eg: **No** • Task options for engagement: heighten salience: **No** • Task options for engagement: enhance relevance, value, and authenticity of tasks eg: **Implemented** • Teacher options for providing supports for attention and engagement: provide varied levels of challenge and support: **No** • Teacher options for providing supports for attention and
	• Teacher options for providing supports for attention and engagement: provide optimal student positioning (positions which encourage alertness, not recumbent): **Yes** • Teacher options for providing supports for attention and engagement: provide feedback to support engagement: **Yes** • Teacher options for providing supports for attention and engagement: prompt student to engage/re-engage: **Yes** • Teacher options for providing supports for attention and engagement: cover up part of text so student isn't overwhelmed: **No** • Task options for engagement: vary amount of context supporting tasks eg: **No** • Teacher options for	• Teacher options for providing supports for attention and engagement: provide optimal student positioning (positions which encourage alertness, not recumbent): **Yes** • Teacher options for providing supports for attention and engagement: provide feedback to support engagement: **Yes** • Teacher options for providing supports for attention and engagement: prompt student to engage/re-engage: **Yes** • Teacher options for providing supports for attention and engagement: cover up part of text so student isn't overwhelmed: **No** • Task options for engagement: vary amount of context supporting tasks eg: **No** • Teacher options for	supports to reduce student frustration eg: **Yes** • Teacher options for providing supports for attention and engagement: provide optimal student positioning (positions which encourage alertness, not recumbent): **Yes** • Teacher options for providing supports for attention and engagement: provide feedback to support engagement: **Yes** • Teacher options for providing supports for attention and engagement: prompt student to engage/re-engage: **Yes** • Teacher options for providing supports for attention and engagement: cover up	engagement: provide supports to reduce student frustration eg: **Yes** • Teacher options for providing supports for attention and engagement: provide optimal student positioning (positions which encourage alertness, not recumbent): **Yes** • Teacher options for providing supports for attention and engagement: provide feedback to support engagement: **Yes** • Teacher options for providing supports for attention and engagement: cover up part of text so student isn't overwhelmed: **No** • Task options for engagement: vary amount of context
	providing supports for attention and engagement: administer assessment at optimal time of day for student engagement: **Yes** • Teacher options for providing supports for attention and engagement: provide verbal/gestural prompts: **No**	providing supports for attention and engagement: administer assessment at optimal time of day for student engagement: **Yes** • Teacher options for providing supports for attention and engagement: provide verbal/gestural prompts: **No**	part of text so student isn't overwhelmed: **No** • Task options for engagement: vary amount of context supporting tasks eg: **No** • Teacher options for providing supports for attention and engagement: administer assessment at optimal time of day for student engagement: **Yes** • Teacher options for providing supports for attention and engagement: provide verbal/gestural prompts: **No**	supporting tasks eg: **No** • Teacher options for providing supports for attention and engagement: administer assessment at optimal time of day for student engagement: **Yes** • Teacher options for providing supports for attention and engagement: provide verbal/gestural prompts: **No**

Item Complexity Notes	IC1. DOK Level: Application (Organize, Compute, Order, Construct) Number of component steps of Focal KSA included: 4 Numeric data versus categorical data: Numeric Complexity of distribution (number of data points): 15 Complexity of data: Triple and double digit numbers Type of data presented in table: Raw Number and complexity of categories in the display: 3 grade categories Degree of contextualized information: Medium Presence of a framing question that incorporates familiar context [Framing question does not require student response]: No Type of sub-questions: Use of a procedural (a, c) and conceptual (b, d) question Type of graphical display to be created: Histogram Summary statistic used for calculation: Mean Amount and Type of Scaffolding: I. Presentation of problem context and prompts is hierarchally arranged [A asks calculation of the mean and, B asks to use calculation to answer another question] structure the student through the problem II. More information presented as student is finished with each part of the item III. Simple language	IC1. DOK Level: Performance (Perform, Demonstrate, Locate, Follow procedures) Number of component steps of Focal KSA included: 2 Numeric data versus categorical data: Numeric Complexity of distribution (number of data points): 3 Complexity of data: Single digit numbers Type of data presented in table: Categorized Number and complexity of categories in the display: 3 grade categories Presence or absence of a framing question that incorporates familiar context Type of sub-questions: Use of a procedural (a) and conceptual (b) question Type of graphical display to be created: Histogram Summary statistic used to answer question: Mode Amount and Type of Scaffolding: I. Presentation of problem context and prompts is hierarchally arranged [A asks calculation of the mean and, B asks to use calculation to answer another question] structure the student through the problem II. More information presented as student is finished with each part of the item III. Simple language IV. Graph paper to support creation of histogram V. Pre-categorized data	IC1. DOK Level: Recall (Identify) Complexity of distribution (number of data points): N/A Numeric data versus categorical data: Categorical Complexity of data: N/A Type of data presented in table: Categorized Number and complexity of categories in the display: 3 Type of framing question: None Type of graphical display to be created: None Summary statistic used to answer question: Mode Amount and Type of Scaffolding: I. Simplified language II. Pre-categorized data III. Most frequent prize is obvious (5 is higher than 4 and 3)	IC1. DOK Level: Recall (Identify) Complexity of distribution (number of data points): N/A Numeric data versus categorical data: Categorical Complexity of data: N/A Type of data presented in table: Categorized Number and complexity of categories in the display: 3 Type of framing question: None Type of graphical display to be created: None Summary statistic used to answer question: Mode Amount and Type of Scaffolding: I. Simplified language II. Pre-categorized data III. Most frequent prize is obvious (5 is higher than 4 and 3)

Item Directive	ID1. Item C	ID1. Item B	ID1. Item A1	ID1. Item A2
	Examiner presents the student with a table displaying data for 10 students and the number of aluminum cans they collected. Examiner says, "In a recycling contest, students collected aluminum cans. This data table shows how many aluminum cans each student collected." a) "What is the mean number of cans collected?" Examiner presents a second table and uses gestural prompts as he/she says, "If a student collects 90-100 cans the prize is an mp3 player [DVD]. If a student collects 80-89 cans, the prize is a video game [gift card]. If a student collects 70-79 cans, the prize is a pair of movie tickets." b) "Which student has a score that is the same as the mean? What is the prize that student would receive?" c) "Examiner presents a card with the names of three kinds of graphical displays: histogram, dot plot, and pie chart. Examiner says, "Here is a list of three kinds of graphical displays - a histogram, dot plot, and pie chart. Choose one of these and create a graph to show the number of students who received each prize." d) "Which prize was received by the most students?"	Examiner presents student with a frequency table of the number of students who received each type of prize. Examiner says, "Look at the data table. Three students received an mp3 player [DVD]; four students received a video game [board game] and five students each received a pair of movie tickets." a) "Create a histogram that shows how many students received each prize." Present student with a piece of graph paper they can use to create their histogram. b) "Which prize is the mode?"	Examiner presents the student with a histogram and says, "Which prize was given to the most students?"	If student cannot respond, the examiner says, "Look at/touch the histogram."

Correct Answer	❶	CA1. a) 84 cans b) Jennifer/Video game [gift card] c) Histogram [created using the data in table b] d) Video game [gift card]	CA1. a) Student creates a histogram b) Pair of movie tickets	CA1. Pair of movie tickets	CA1. Student looks at/touches the histogram
Materials for Examiner	❶	Mf1. - Task worksheet that describes item and delivery instructions - Task data sheet or other method to record student's response	Mf1. - Task worksheet that describes item and delivery instructions - Task data sheet or other method to record student's response	Mf1. - Task worksheet that describes item and delivery instructions - Task data sheet or other method to record student's response	Mf1. - Task worksheet that describes item and delivery instructions - Task data sheet or other method to record student's response

Description of Stimulus Materials	❶	Do1. Stimulus Material 1: Table with two columns: Student Name and Number of Cans Collected and ten rows of data: Angela 98 Chris 96 Fred 74 Geneva 82 Jennifer 84 Jose 72 Kathryn 100 Kavita 82 Ken 72 Kim 80 Stimulus Material 2: Table with two columns: Number of cans and Prize and three rows of data: 90-100 Mp3 player 80-89 Video game 70-79 Pair of movie tickets	Do1. Stimulus Material 1: Table with two columns: Prize and Number of Students and three rows of data: Mp3 3 Video game 4 Pair of movie tickets 5	Do1. Stimulus Material 1: Histogram displaying the following: y-axis - Number of students (0-5); x-axis - Prize (Mp3, 3 students; Video game, 4 students; Pair of movie tickets, 5 students)	Do1. Stimulus Material 1: Histogram displaying the following: y-axis - Number of students (0-5) x-axis - Prize (Mp3, 3 students; Video game, 4 students; Pair of movie tickets, 5 students)

Notes	N1. Question presentation individualized (e.g., verbal/gestural prompts, related in sign language, use concrete objects). Response format individualized based on student communication system (e.g., method used to create display: paper based vs. computer based vs. concrete manipulatives). Remind student of prior instructional experiences to orient student to what's being asked, this is not reteaching and does not include detailed examples and analogies ["Remember when we did the lesson on calculating the mean?" Verbal prompts [Examiner says, "and then what?"]. Gestural prompts [Examiner points to particular drawings regarding the question being asked]. Sign language [entire content of the question is signed].	N1. Question presentation individualized (e.g., verbal/gestural prompts, related in sign language, use concrete objects). Response format individualized based on student communication system (e.g., method used to create display: paper based vs. computer based vs. concrete manipulatives). Remind student of prior instructional experiences to orient student to what's being asked, this is not reteaching and does not include detailed examples and analogies ["Remember when we did the lesson on calculating the mean?"]	N1. Question presentation individualized (e.g., verbal/gestural prompts, related in sign language, use concrete objects). Response format individualized based on student communication system (e.g., method used to create display: paper based vs. computer based vs. concrete manipulatives). Remind student of prior instructional experiences to orient student to what's being asked, this is not reteaching and does not include detailed examples and analogies ["Remember when we did the lesson on calculating the mean?"]. Discuss with states about the DOK level if item is Recall (Identify).	

Depth of Knowledge, Browder, Flowers, andWakeman[6]

Attention: touch, look, listen, repeat what the teacher said, vocalize, respond, attend, recognize

Memorize/recall: list, describe (facts), state math facts, identify, state, define, match, recognize, label, follow a pattern

Performance: answer, follow 1 step directions, find answer, present, read, separate, spell, tell time, map, model demonstration, perform, demonstrate, follow, choose, count, locate, group by given attributes, solve simple (one computation skill) problems, measure

Comprehension: understand, extend a pattern, sketch, ask and answer questions, categorize/group by unknown attributes, explain, conclude, group, restate, review, translate, classify/sort with understanding, simplify (equivalent forms)

Application: compute, organize, collect (such as data), apply, revise, construct, solve complex (multiple computation skills) problems, use given formulas in novel situations (formula may or may not be identified), explain a process, conduct research

Analysis, Synthesis, Evaluation: create a complex pattern, analyze, compare, contrast, compose, predict, plan, judge, evaluate, interpret data, generalize findings, create hypotheses

---

[6] Browder, D. M., Flowers, C., & Wakeman, S. Y. (2008). Facilitating participation in assessments and the general curriculum: Level of symbolic communication classification for students with significant cognitive disabilities. *Assessment in Education: Principles, Policy, and Practice, 15*(2): 137–151.

Materials for Administering Recycling Suite of Items

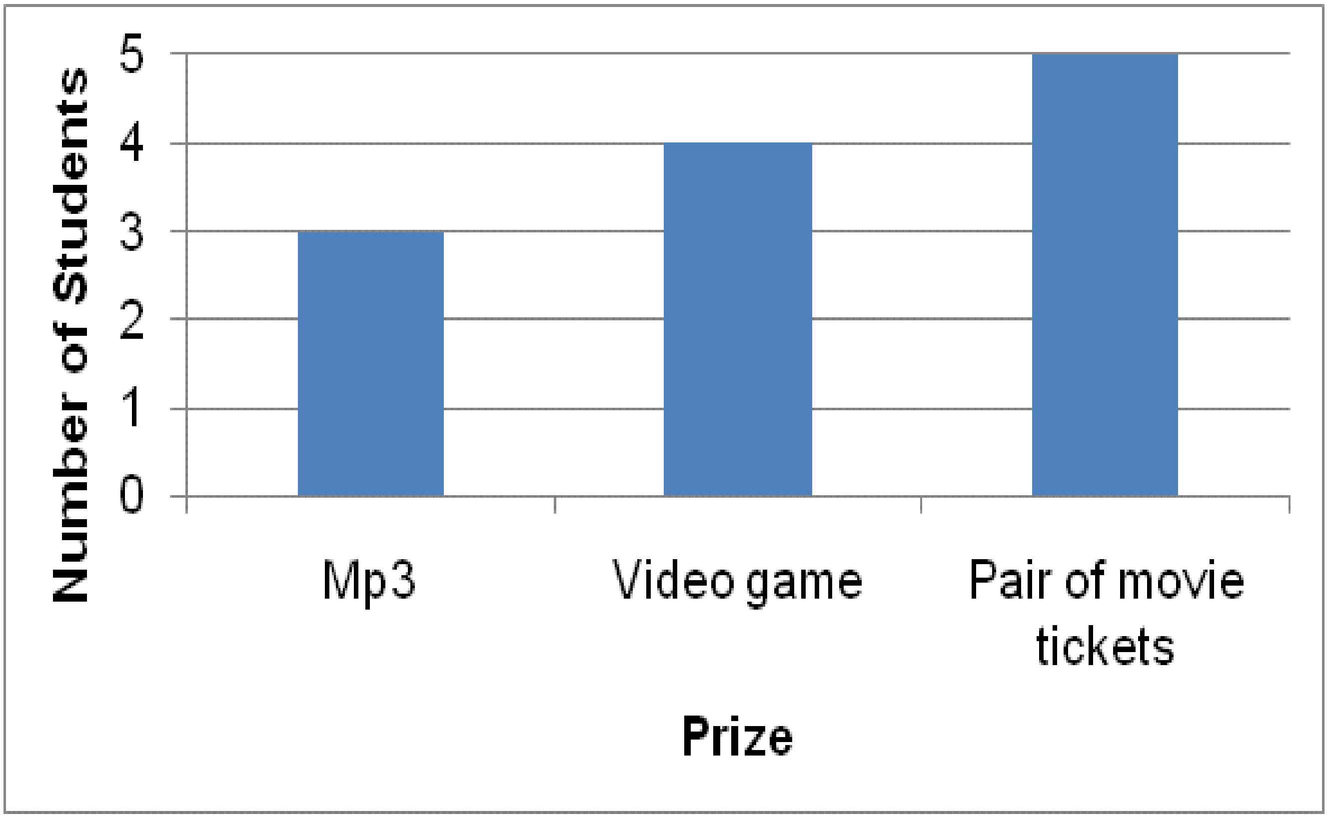


Figure C1. Data & Probability B1, Grades 9-12 Item A1 & A2.

Table C1. Data and Probability B1, Grades 9-12, Item B

Frequency table	
Prize	Number of students
Mp3	3
Video game	4
Pair of movie tickets	5

Table C2.Data and Probability B1, Grades 9-12 Item C

Number of cans collected by students	
Student name	Number of cans collected
Angela	98
Chris	96
Fred	74
Geneva	82
Jennifer	84
Jose	72
Kathryn	100
Kavita	82
Ken	72
Kim	80

Table C3. Data and Probability B1, Grades 9-12 Item C

Cans for prizes	
Number of cans	Prize
90 - 100	MP3 player
80 - 89	Video game
70 - 79	Pair of movie tickets

Histogram
Dot plot
Pie chart

Detailed Instructions for Administering Recycling Suite of Items

Table C4. Item A1

Steps	Scripts, materials, and student responses	Directions for scoring, reporting, and moving to the next item
Directions: What the teacher says (bold script) and does (regular text)	The student is presented with a histogram, and the examiner says, "Which prize was given to the most students?"	If you change the directions to meet individual needs of the student, indicate the changes in Section 1 and answer the questions about accommodations in Section 2 of the Data Collection Booklet.
Materials: What the student perceives	Histogram representing prizes given: Mp3, Video Game, or Pair of movie tickets.	If you change materials, indicate the changes in Section 1 and answer the questions about accommodations in Section 2 of the Data Collection Booklet.
Student Correct response	Student says, "Pair of movie tickets." Student points to the "Pair of movie tickets" column.	Mark "Correct" in Section 1 of the Data Collection Booklet if the response was independent and consistent with the student's typical response mode. MOVE TO ITEM B.
Student Incorrect response†	Student says, "MP3" or "Video game." Student points to the first or second column.	Mark "Incorrect" and record how the student responded in the appropriate field in Section 1 of the Data Collection Booklet. MOVE TO ITEM A2.
Student No Response—doesn't respond to question†	Student claps hands. Student stares at wall. Student hums with eyes closed.	Obtain the student's attention and repeat the directions. If the student still doesn't respond, record the student's lack of response in Section 1 of the Data Collection Booklet. MOVE TO ITEM A2.
Student Refused to Respond†	Student says (signs or gestures) "No" or "I don't want to." Student pushes task materials away.	Mark "Refused" and record how the student responded in the appropriate field in Section 1 of the Data Collection Booklet. STOP TASK

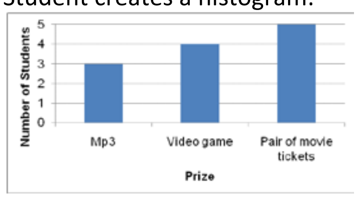
Note. These are just some examples of potential student responses and are not exhaustive lists.

Table C5. Item A2

Steps	Scripts, materials, and student responses	Directions for scoring, reporting, and moving to the next item
Directions: What the teacher says (bold script) and does (regular text)	If student cannot respond, the examiner says, "Look at/touch the histogram."	If you change the directions to meet individual needs of the student, indicate the changes in Section 1 and answer the questions about accommodations in Section 2 of the Data Collection Booklet.
Materials: What the student perceives	Histogram representing prizes given: Mp3, Video Game, or Pair of movie tickets.	If you change materials, indicate the changes in Section 1 and answer the questions about accommodations in Section 2 of the Data Collection Booklet.
Student correct response	Student indicates the histogram.	Mark "Correct" in Section 1 of the Data Collection Booklet if the response was independent and consistent with the student's typical response mode. STOP TASK
Student incorrect response†	Student says, "I don't know." Student echoes "histogram."	Mark "Incorrect" and record how the student responded in the appropriate field in Section 1 of the Data Collection Booklet. STOP TASK
Student no response—doesn't respond to question†	Student claps hands. Student stares at wall. Student hums with eyes closed.	Obtain the student's attention and repeat the directions. If the student still doesn't respond, record the student's lack of response in Section 1 of the Data Collection Booklet. STOP TASK
Student refused to respond	Student says (signs or gestures) "No" or "I don't want to." Student pushes task materials away.	Mark "Refused" and record how the student responded in the appropriate field in Section 1 of the Data Collection Booklet. STOP TASK

Note. These are just some examples of potential student responses and are not exhaustive lists.

Table C6. Item B

Steps	Scripts, materials, and student responses	Directions for scoring, reporting, and moving to the next item
Directions: What the teacher says (bold script) and does (regular text)	The student is presented with a frequency table of the number of students who received each type of prize, and the examiner says, "Look at the data table. Three students received an mp3 player [DVD]; four students received a video game [board game] and five students each received a pair of movie tickets. Create a histogram that shows how many students received each prize. Present student with a piece of graph paper they can use to create their histogram. Which prize is the mode?	If you change the directions to meet individual needs of the student, indicate the changes in Section 1 and answer the questions about accommodations in Section 2 of the Data Collection Booklet.
Materials: What the student perceives	Frequency table with number of students receiving each type of prize. Graph paper for student created histogram.	If you change materials, indicate the changes in Section 1 and answer the questions about accommodations in Section 2 of the Data Collection Booklet.
Student correct response	Student creates a histogram:  b) Pair of movie tickets	Mark "Correct" in Section 1 of the Data Collection Booklet if the response was independent and consistent with the student's typical response mode. MOVE TO ITEM C.
Student incorrect response†	Student reads or recreates the frequency table. Student says, "I don't know."	Mark "Incorrect" and record how the student responded in the appropriate field in Section 1 of the Data Collection Booklet. MOVE TO ITEM C.
Student no response—doesn't respond to question†	Student claps hands. Student stares at wall. Student hums with eyes closed.	Obtain the student's attention and repeat the directions. If the student still doesn't respond, record the student's lack of response in Section 1 of the Data Collection Booklet. MOVE TO ITEM C.
Student refused to respond	Student says (signs or gestures) "No" or "I don't want to." Student pushes task materials away.	Mark "Refused" and record how the student responded in the appropriate field in Section 1 of the Data Collection Booklet. STOP TASK

Note. These are just some examples of potential student responses and are not exhaustive lists.

Table C7. Item C

Steps	Scripts, materials, and student responses	Directions for scoring, reporting, and moving to the next item
Directions: What the teacher says (bold script) and does (regular text)	The examiner presents the student with a table displaying data for 10 students and the number of aluminum cans they collected. Examiner says, "In a recycling contest, students collected aluminum cans. This data table shows how many aluminum cans each student collected. What is the mean number of cans collected?" Examiner presents a second table (table be below) and uses gestural prompts as he/she says, "If a student collects 90-100 cans the prize is an mp3 player [DVD]. If a student collects 80-89 cans, the prize is a video game [gift card]. If a student collects 70-79 cans, the prize is a pair of movie tickets. Which student has a score that is the same as the mean? What is the prize that student would receive?" Examiner presents a card with the names of three kinds of graphical displays: histogram, dot plot, and pie chart. Examiner says, "Here is a list of three kinds of graphical displays – a histogram, dot plot, and pie chart. Choose one of these and create a graph to show the number of students who received each prize. d) Which prize was received by the most students?"	If you change the directions to meet individual needs of the student, indicate the changes in Section 1 and answer the questions about accommodations in Section 2 of the Data Collection Booklet.

<table>
<tr><th>Steps</th><th>Scripts, materials, and student responses</th><th>Directions for scoring, reporting, and moving to the next item</th></tr>
<tr><td>Materials: What the student perceives</td><td>a)
<table>
<tr><td>Student Name</td><td>Number of Cans Collected</td></tr>
<tr><td>Angela</td><td>98</td></tr>
<tr><td>Chris</td><td>96</td></tr>
<tr><td>Fred</td><td>74</td></tr>
<tr><td>Geneva</td><td>82</td></tr>
<tr><td>Jennifer</td><td>84</td></tr>
<tr><td>Jose</td><td>72</td></tr>
<tr><td>Kathryn</td><td>100</td></tr>
<tr><td>Kavita</td><td>82</td></tr>
<tr><td>Ken</td><td>72</td></tr>
<tr><td>Kim</td><td>80</td></tr>
</table>
b)
<table>
<tr><td>Number of cans</td><td>Prize</td></tr>
<tr><td>90-100</td><td>Mp3 player</td></tr>
<tr><td>80-89</td><td>Video game</td></tr>
<tr><td>70-79</td><td>Pair of movie tickets</td></tr>
</table>
c) Card with names of three types of graphical displays: histogram, dot plot, and pie chart.</td><td>If you change materials, indicate the changes in Section 1 and answer the questions about accommodations in Section 2 of the Data Collection Booklet.</td></tr>
<tr><td>Student correct response</td><td>84 cans
Jennifer/Video game [gift card]
Histogram (created using the data in table b)
d) Video game [gift card]</td><td>Mark “Correct” in Section 1 of the Data Collection Booklet if the response was independent and consistent with the student’s typical response mode.
STOP TASK ⬣</td></tr>
<tr><td>Student incorrect response†</td><td>Any response besides 84 or “I don’t know.”
Any response besides Jennifer/Video game or “I don’t know.”
A frequency table, a representation of a display not selected, or “I don’t know.”
A response of “Mp3 player,” “Pair of movie tickets,” or “I don’t know.”</td><td>Mark “Incorrect” and record how the student responded in the appropriate field in Section 1 of the Data Collection Booklet. STOP TASK ⬣</td></tr>
<tr><td>Student no response—doesn’t respond to question†</td><td>Student claps hands.
Student stares at wall.
Student hums with eyes closed.</td><td>Obtain the student’s attention and repeat the directions.
If the student still doesn’t respond, record the student’s lack of response in Section 1 of the Data Collection Booklet. STOP TASK ⬣</td></tr>
</table>

Steps	Scripts, materials, and student responses	Directions for scoring, reporting, and moving to the next item
Student refused to respond	Student says (signs or gestures) “No” or “I don’t want to.” Student pushes task materials away.	Mark “Refused” and record how the student responded in the appropriate field in Section 1 of the Data Collection Booklet. STOP TASK ⬣

Note. These are just some examples of potential student responses and are not exhaustive lists.

Table C8. General Alignment Between AKSAs and Variable Features

AKSAs	Variable features
Perceptual (Receptive) AP1. Ability to perceive the linguistic components of the stimulus material and question (e.g., through print, objects, audio, Braille) (P1, P2, P3) AP2. Ability to perceive images in the stimulus material and question (e.g., through print, objects, holistic description, Braille) (P1, P2, P3) AP3. Ability to perceive physical objects required for the task (e.g., see physical objects used to relate a story) (P1, P2)	Perceptual (Receptive) P1. Delivery mechanisms by which the question is perceived (e.g., read aloud verbatim/read aloud paraphrase, pictures, large print, printed text, Braille, text, symbols, concrete objects, description of objects or images, text to speech, signing, auditory amplification, CCTV – close circuit TV, to increase size of font, vary contrast, etc.) P2. Supports for the use of equipment required for the task (e.g., communication board, CD player) P3. Delivery parameters for oral presentation of material (e.g., speed of reading, volume, amount of expression used, student ability to pause, stop, and/or repeat information read aloud)
Skill and Fluency (Expressive) AS1. Ability to communicate response (e.g., respond verbally, by using pictures, by making a selection from a group) (S1, S2, S3, S4, S5, S6) AS2. Ability to compose or express a response in text (e.g., by writing, using Braille) (S1, S2, S3, S4, S5, S6) AS3. Ability to manipulate physical materials (e.g., dexterity, strength and mobility) (S1, S2, S3, S4, S5, S6) AS4. Ability to manipulate digital/electronic equipment (S1, S4, S5, S6) AS5. Knowledge of how to use physical materials or digital/electronic equipment (e.g., familiarity) (S5, S6)	Skill and Fluency (Expressive) S1. Response mode options (e.g., pointing, speech and verbalization, writing, signing, switch or other assistive device/augmentative communication device, eye gaze, for lowest functioning students – predictable behavioral response, tolerate assistance – e.g., hand over hand) S2. Supports for composing a response in text (e.g., speech to text, written by teacher, keyboarding) S3. Supports for manipulating physical materials (e.g., use of velcro, size of materials, teacher manipulation of materials) S4. Supports for manipulating digital/electronic equipment (e.g., pointers, teacher manipulation of equipment, spoken commands, stylus for input, larger keyboard/buttons, adaptive mouse) S5. Practice tutorials with unfamiliar physical materials or digital/electronic equipment S6. Practice with familiar equipment
Language and Symbols AL1. Ability to recognize text, symbols, or images (L2, L4, L5, L8, L9, L10, L11) AL2. Ability to decode text, symbols, or images (L1, L2, L3, L4, L5, L8, L9, L10, L11) AL3. Ability to comprehend text, symbols, or images (L1, L2, L3, L4, L5, L6, L7, L8) AL4. Ability to understand English vocabulary and syntax (L2, L3, L4, L5, L7, L8)	Language and Symbols L1. Level of abstraction required of student (e.g., concrete objects, images, text) L2. New vs. pre-taught vocabulary and symbols L3. Embedded support for vocabulary and symbols (e.g., technical and non-technical glossary, hyperlinks/footnotes to definitions, illustrations, background knowledge) L4. All key information in the dominant language (e.g., English) is also available in prevalent first languages (e.g., Spanish) L5. All key information in sign language for students who utilize this mode of

AKSAs	Variable features
Cognitive AC1. Ability to attend to stimuli (DOK level 1) (C37, C38, C39, C40, C41, C42, C43, C44, C45, C46) AC2. Ability to recall related background knowledge (DOK level 2) (C5, C6, C7, C8, C9, C10, C11, C12) AC3. Ability to perform (e.g., answer questions, solve simple problems) (DOK level 3) (C11, C12, C13, C19, C20, C29, C30, C33) AC4. Ability to comprehend (e.g., provide an explanation) (DOK level 4) (C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19) AC5. Ability to apply information (e.g., organize information) (DOK level 5) (C13, C14, C15, C16, C17, C18, C19, C20, C21) AC6. Ability to analyze, synthesize, or evaluate information (DOK level 6) (C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, C25, C26, C27, C28, C29, C30, C31) AC7. Ability to understand the meaning of an example (C16, C24) AC8. Ability to process multi-step problems (C13, C14, C15, C20, C22, C23, C24, C25, C26, C27, C28, C31, C32, C34, C35) AC9. Ability to recall and use information presented in a task/item (working memory) (C32, C33, C34, C35, C36) AC10. Ability to understand the structure of "organizers" used to present information or to scaffold responses (e.g., understand meaning of headers, subtitles, etc. in text) (C11, C21, C24, C29, C30) AC11. Ability to understand the purpose of highlighted features in text or illustrations (C21, C25)	communication L6. Use of multiple representations (e.g., physical models, demonstrations, acting out scenarios) L7. Alternate syntactic levels (simplified text) L8. Highlight essential elements, words, or phrases L9. Digital text with automatic text to speech L10. Digital Braille with automatic Braille to speech L11. Read language and symbols aloud Cognitive C1. Depth of knowledge of the content – SELECTED IN EVERY DESIGN PATTERN AND TASK C2. Complexity of the content (e.g., length of story, number of supporting details included, richness of context) – SELECTED IN EVERY DESIGN PATTERN AND TASK C3. Item/task format (selected response vs. constructed response, performance, etc.) C4. Adjustable levels of challenge (teacher able to adjust) Options for supporting background knowledge: C5. Pre-teach background content (pre-teach definitions of unfamiliar words or concepts unrelated to the standard; pre-teach means teaching a student for the first time the definition of a word or concept that is included in the narrative of a test item but not part of the construct being measured) C6. Provide analogies and examples C7. Provide hyperlinks to multi-media C8. Provide links to related information C9. Provide links to familiar materials C10. Provide concept maps C11. Remind student of prior experiences C12. Remind student of materials or activities used to teach foundational reading/English language arts skills Options for supporting critical features, big ideas, and relations: C13. Provide graphic organizers C14. Outline information C15. Highlight information C16. Provide alternative forms of key concepts C17. Provide multi-media glossaries C18. Provide translation tools C19. Provide modeled prompts (on non-construct relevant content) C20. Provide a response template

AKSAs	Variable features
Executive AE1. Ability to set goals and expectations (E1, E4, E5) AE2. Ability to monitor goals and progress (E1, E2, E3, E4, E5) AE3. Ability to plan and sequence (E1, E4, E5) AE4. Ability to self-regulate and reflect during problem solving (E1, E2, E3, E4, E5) Affective AA1. Ability to engage (e.g., task-specific motivation) (A1, A2, A3, A4, A5, A6, A7, A8, A11, A12, A13, A14, A15) AA2. Ability to persist and sustain effort (A1, A2, A3, A4, A5, A6, A7, A8, A11, A12, A13, A14, A15)	C21. Remind student of the function of tools/features designed to aide comprehension and processing of information (e.g., highlighting, graphic organizers, captions, and headings) Options for guiding exploration and information processing: C22. Provide multiple entry points C23. Allow viewing of stimuli from previous stages and parts C24. Use familiar materials C25. Use consistent signals/cues C26. Provide sequential highlighting C27. Chunk information into smaller elements C28. Mask part of the information C29. Provide modeled prompts (on non-construct relevant content) C30. Provide a practice item or task C31. Provide a guide or checklist for prioritization of steps in multi-step problems Options for supporting memory and transfer: C32. Note-taking C33. Mnemonic aids C34. Locate items near relevant text C35. Reread question/stimulus C36. Present items as a discrete unit or embed in a scenario Executive E1. Prompts and scaffolds to estimate effort, resources, and difficulty E2. Prompts, scaffolds, and questions to monitor progress, to “stop and think”, and for categorizing and systematizing E3. Representations of progress (e.g., before and after photos, graphs and charts) E4. Guides, checklists, graphic organizers, and/or templates for goal setting, prioritizing, breaking long-term objectives into reachable short-term goals, self-reflection, and self-assessment E5. Adjust levels of challenge and support (e.g., adjustable leveling and embedded support, alternative levels of difficulty, alternative points of entry) Affective Teacher options for providing supports for attention and engagement: A1. Cover up part of text so student isn’t overwhelmed A2. Prompt student to engage/re-engage A3. Provide verbal/gestural prompts A4. Provide feedback to support engagement

AKSAs	Variable features
	A5. Provide supports to reduce student frustration (e.g., noise reduction, extended test taking time, contingencies, number of items administered at one time) A6. Provide varied levels of challenge and support A7. Provide optimal student positioning (positions which encourage alertness, not recumbent) A8. Administer assessment at optimal time of day for student engagement Task options for engagement (task refers to the assessment items, stimulus “story”, and materials): A9. Provide students with choices for personal control of age-appropriate content when construct is not impacted (e.g., choice of topic or theme) MAY NOT BE APPLICABLE FOR STATEWIDE ASSESSMENTS A10. Provide students with choices for personal control of task context when construct is not impacted NOT MAY NOT BE APPLICABLE FOR STATEWIDE ASSESSMENTS A11. Enhance relevance, value, and authenticity of tasks A12. Heighten salience A13. Variety of stimuli A14. Vary amount of context supporting tasks (e.g., discrete tasks vs. scenarios) A15. Item/task format (selected response vs. constructed response, performance, etc.)

Table C9. Technology Requirements of Task Template Displaying for Items A2, A1, and B

Task Template Fields	Task Template Subfields	Overall Specifications for Task
Title		**Recycling Task**
Purpose		The purpose of this task template is to specifiy the details for the recycling task. This task will take place on-line and may be administered along with other tasks.
Item authoring system		This task will take place in an item authoring system. This system allows for separate versions of a test, an examiner version and a student version. Tasks will be developed to work within this system.
Programming Environment (the environment in which the actual programming of the task will take place)		HTML 5 and Java web app
Task Logic and Presentation (requirements for the overall pieces of the task)	**Format requirements/constraints (screen layout (header bars..), placement of where stimuli occur, responses are allowed to occur, tool bars…)**	• UDL toolbar • item name • Footer bar with a next button • Stimulus will appear on the left • Response options on the right • There is a student version, which shows the student response capture and there is a separate version for the examiner which would display the examiner response capture • Header bar with progress bar
	Visual display requirements (such as number of items per screen, width and height of the screen, resolution, background)	• One item on a screen at a time • 16x9 resolution screen • White background with black text
	Graphical requirements/constraints (settings of the graphs and the images)	• Make sure that spaces occur between each bar • Greyscale is used • Graphical displays should be maximized to fit the screen
	Movement between screens (how to get to the next screen, when it is appropriate to move between screens)	• Students (or examiner) will click on the Next button on the bottom of the screen If the student (or the examiner) has not answered the question an error message will appear and they will not be allowed to move to the next item.
Adaptivity (movement between pieces)	**Number of items (and names)**	• 4 items (C, B, A1 and A2)
	Starting item	• A1

Adaptivity (movement between pieces)	**Branching**	• If A1 is incorrect, then move to A2 • If A1 is correct, then move to B • After the student sees B they move to C • After the student sees C the task ends • After the student sees A2 the task ends			
	Number of scenes per item	• A1: One scene • A2: One scene • B: Two scenes • C: Four scenes			
		Item A2: Scene 1	**Item A1: Scene 1**	**Item B: Scene 1**	**Item B: Scene 2**
Task Stimuli	**Font characteristics**	Ariel 18pt to start, with options to increase	Ariel 18pt to start, with options to increase	Ariel 18pt to start, with options to increase	Ariel 18pt to start, with options to increase
	Image specifics (details on the specific images to be included)	Student is presented with a histogram with 3 bars, one for MP3 (3), one for Video game (4) and one for pair of movie tickests (5)	Student is presented with a histogram with 3 bars, one for MP3 (3), one for Video game (4) and one for pair of movie tickests (5)	The student is presented with a frequency table of the number of students who received each type of prize There is text that states, “Look at the data table. Three students received an mp3 player; four students received a video game, and five students each received a pair of movie tickets."	Student is presented with a histogram with 3 bars, These bars would be the same as the student response in scene 1.
	Interitem dependency	NA	NA	NA	Histogram that was created in scene 1 is presented. If no histogram was created, then the histogram might not exist. (students may have created the histogram outside of the online format)
Response Capture	**Student Response capture (type such as text box, radio button, drop down)**	Text showing: " Look at/touch the histogram."	Text showing: " 1) Which prize was given to the most students? Select the correct response." Selection for the following options "A. MP3 player B. Video game C. Pair of movie tickets"	Text showing: " Create a histogram that shows how many students received each prize. " A graphical interface is presented that would allow the student to create a histograph with 3 bars.	Text showing: " 1) Which prize is the mode? Select the correct response." Selection for the following options "A. MP3 player B. Video game C. Pair of movie tickets"
	Examiner Response capture	Examiner Response box has a note that says "Have the student respond to the question. Click on the box that most appropriately shows the students response." Response options for the teacher are "A. Student correct response, B. Student incorrect response, C. Student has no response, D. Student refused to respond"	Examiner Response box has a note that says "Have the student respond to the question. Click on the box that most appropriately shows the students response." Response options for the teacher are "A. Student correct response, B. Student incorrect response, C. Student has no response, D. Student refused to respond"	Examiner Response box has a note that says "Have the student respond to the question. Click on the box that most appropriately shows the students response." Response options for the teacher are "A. Student correct response, B. Student incorrect response, C. Student has no response, D. Student refused to respond"	Examiner Response box has a note that says "Have the student respond to the question. Click on the box that most appropriately shows the students response." Response options for the teacher are "A. Student correct response, B. Student incorrect response, C. Student has no response, D. Student refused to respond"

Appendix D. Identifying Accessibility Barriers

Identifying Accessibility Barriers: ***Pinball Car Race***

	Description of Steps	Selected Focal KSAs			Selected Non-focal KSAs[1]		
Step 1	Present background information including definitions, image of students to support engagement				Perceive and comprehend text and image (E) (V) H M (A)	Use rollover to display definitions E (V) H (M) (A)	 E V H M A
Step 2	Play animation, determine in what time segment spring has greatest potential energy and greatest kinetic energy	Knowledge of potential energy	Knowledge of kinetic energy	Ability to determine point at which spring has greatest potential and kinetic energy	Perceive and comprehend text and animation (E) (V) H M (A)	Use mouse clicks to operate animation and make selection from menu E (V) H (M) A	Maintain persistence throughout task E V H M (A)
Step 3	Present background information on experiment and springs, introduce two variables				Perceive and comprehend text and images (E) (V) H M (A)	Understand the structure of information presented in tables E (V) H (M) A	Perceive and comprehend levels of variables (E) (V) H (M) A
Step 4	Select hypothesis to test, explain reasoning	Ability to identify a hypothesis	Ability to evaluate hypothesis		Perceive and comprehend text (E) (V) H M (A)	Use mouse clicks to make selections E (V) H (M) A	Compose a response in text (E) (V) H (M) (A)

[1] Note: The chosen categories are only examples and do not represent all categories of students with special needs. Categories not presented here include Autism, emotional disturbance, intellectual disabilities, multiple disabilities, other health impairments, specific learning disabilities, speech language impairment, and traumatic brain injury.

Identifying Accessibility Barriers: ***Pinball Car Race***

	Description of Steps	Selected Focal KSAs			Selected Non-focal KSAs		
Step 5	Select number of coils and thickness of wire to test in experiment, explain why the characteristics are appropriate to the hypothesis	Ability to identify variables of the scientific situation that could be controlled	Ability to recognize an experiment as manipulating 1 variable and measuring effect on another	Ability to evaluate hypothesis	Perceive and comprehend text (E) (V) H M (A)	Understand the structure of information presented in tables (E) (V) H M (A)	Perceive and comprehend levels of variables (E) (V) H (M) (A)
Step 6	Present results of Experiment 1, describe features of table and graph				Perceive and comprehend text (E) (V) H M (A)	Read data tables and bar graphs (E) (V) H M (A)	Knowledge of metric system of measurement (E) V H M A
Step 7	Read data table and bar graph for Experiment 1, select whether hypothesis is supported or contradicted, explain answer	Ability to recognize an experiment as manipulating 1 variable and measuring effect on another	Ability to identify variables of the scientific situation that could be controlled	Ability to evaluate hypothesis	Perceive and comprehend text, data tables, and bar graphs (E) (V) H M (A)	Select radio button indicating support or contradiction of hypothesis E (V) H (M) (A)	Compose a response in text (E) (V) H (M) (A)
Step 8	Review table and bar graph and determine if experimental settings are appropriate, generate changes in settings to better reflect hypothesis	Ability to evaluate a hypothesis	Ability to interpret results of experiment	Ability to identify variables to be controlled to prevent misleading information	Perceive and comprehend text, data table, and bar graph (E) (V) H M (A)	Select radio button indicating whether or not selected settings were appropriate E (V) H (M) (A)	Compose a response in text (E) (V) H (M) (A)

Identifying Accessibility Barriers: ***Pinball Car Race***

	Description of Steps	Selected Focal KSAs			Selected Non-focal KSAs		
Step 9	Run experiment again with same settings or new settings, select settings and explain why these settings are appropriate	Ability to evaluate a hypothesis	Ability to identify variables to be controlled to prevent misleading information		Perceive and comprehend levels of variables (E) (V) H M (A)	Select radio buttons that indicate levels of variables (E) (V) H (M) (A)	Compose a response in text (E) (V) H (M) (A)
Step 10	Read data table and bar graph for Experiment 2, select whether hypothesis is supported or contradicted, explain answer	Ability to recognize an experiment as manipulating 1 variable and measuring effect on another	Ability to identify variables of the scientific situation that could be controlled	Ability to evaluate hypothesis	Perceive and comprehend text, data tables, and bar graphs (E) (V) H M (A)	Select radio button indicating support or contradiction of hypothesis (E) (V) H (M) (A)	Compose a response in text (E) (V) H (M) (A)
Step 11	Select spring to use in race, explain choice	Ability to interpret results of two experiments	Ability to formulate conclusions from results		Perceive and read a bar graph (E) (V) H M (A)	Compose a response in text (E) (V) H (M) (A)	 E V H M A

Identifying Ways to Enhance Accessibility: ***Pinball Car Race***

For each step: Top row: identify relevant populations (1 per column). Middle: specify the difficulty. Bottom: suggest solution.							Comments
Step 1	Population →	ELL	Visually Impaired	Hearing Impaired	Motor Impaired	Attention Deficits	
	Difficulty →	Difficulty decoding and comprehending text in English, lack of familiarity with dominant culture	Difficulty perceiving text, and images		Difficulty using mouse for roll-over	Engagement and persistence difficulties, may not read material carefully, may lose track of place in text, cognitive overload	
	Solution →	Produce test in dominant language, provide English/non-English language glossary of terms, use culturally sensitive scenarios	Use large font, maximum contrast, provide magnifier, read aloud, provide text to speech, describe images (verbally or in Braille)		Provide hard copy of definitions, integrate augmentative communication devices (e.g., adaptive mouse)	Use of scenario, illustration, and animation to increase engagement and understanding, provide progress bar, provide sequential highlighting	

Identifying Ways to Enhance Accessibility: *Pinball Car Race (continued)*

For each step: Top row: identify relevant populations (1 per column). Middle: specify the difficulty. Bottom: suggest solution.							Comments
Step 2	Population →	ELL	Visually Impaired	Hearing Impaired	Motor Impaired	Attention Deficits	
	Difficulty →	Difficulty decoding and comprehend-ing text in English, lack of familiarity with dominant culture	Difficulty perceiving text, images, and animation		Difficulty using mouse to initiate animation or select time frame	Engagement and persistence difficulties, may not read material carefully, may lose track of place in text, cognitive overload	
	Solution →	Produce test in dominant language, provide English/ non-English language glossary of terms, use culturally sensitive scenarios	Use large font, maximum contrast, provide magnifier, read aloud, provide text to speech, describe images and animations (verbally or in Braille)		Integrate augmentative communication devices (e.g., adaptive mouse), provide physical assistance to make response, allow student to respond verbally, speech recognition software	Use of scenario, illustration, and animation to increase engagement and understanding, provide opportunity to rerun animation, provide progress bar, provide sequential highlighting	

Identifying Ways to Enhance Accessibility: ***Pinball Car Race*** *(continued)*

For each step: Top row: identify relevant populations (1 per column). Middle: specify the difficulty. Bottom: suggest solution.							Comments
Step 3	Population →	ELL	Visually Impaired	Hearing Impaired	Motor Impaired	Attention Deficits	
	Difficulty →	Difficulty decoding and comprehending text in English, lack of familiarity with dominant culture	Difficulty perceiving text and images		Difficulty using mouse to select characteristics of springs	Engagement and persistence difficulties, may not read material carefully, may lose track of place in text, cognitive overload	
	Solution →	Produce test in dominant language, provide English/ non-English language glossary of terms, use culturally sensitive scenarios, provide multiple representations of springs	Use large font, maximum contrast, provide magnifier, read aloud, provide text to speech, describe images (verbally or in Braille)		Integrate augmentative communication devices (e.g., adaptive mouse), provide physical assistance to make response, allow student to respond verbally	Provide interactive system and illustrations of springs to increase engagement and understanding, provide progress bar, provide sequential highlighting	

Identifying Ways to Enhance Accessibility: *Pinball Car Race* (continued)

For each step: Top row: identify relevant populations (1 per column). Middle: specify the difficulty. Bottom: suggest solution.							Comments
Step 4	Population →	ELL	Visually Impaired	Hearing Impaired	Motor Impaired	Attention Deficits	
	Difficulty →	Difficulty decoding and comprehending text in English, lack of familiarity with dominant culture	Difficulty perceiving text		Difficulty using mouse to select radio buttons, difficulty composing text	Engagement and persistence difficulties, may not read material carefully, may lose track of place in text, cognitive overload	
	Solution →	Produce test in dominant language, provide English/ non-English language glossary of terms, use culturally sensitive scenarios	Use large font, maximum contrast, provide magnifier, read aloud, provide text to speech, describe images (verbally or in Braille)		Integrate augmentative communication devices (e.g., adaptive mouse), provide physical assistance to make response, allow student to respond verbally, speech recognition software	Provide progress bar, provide sequential highlighting, provide drop down menus	

Identifying Ways to Enhance Accessibility:
Pinball Car Race *(continued)*

For each step: Top row: identify relevant populations (1 per column). Middle: specify the difficulty. Bottom: suggest solution.							Comments
Step 5	Population →	ELL	Visually Impaired	Hearing Impaired	Motor Impaired	Attention Deficits	
	Difficulty →	Difficulty decoding and comprehending text in English, lack of familiarity with dominant culture	Difficulty perceiving text and images		Difficulty using mouse to select radio buttons, difficulty composing text	Engagement and persistence difficulties, may not read material carefully, may lose track of place in text, cognitive overload	
	Solution →	Produce test in dominant language, provide English/ non-English language glossary of terms, use culturally sensitive scenarios	Use large font, maximum contrast, provide magnifier, read aloud, provide text to speech, describe images (verbally or in Braille)		Integrate augmentative communication devices (e.g., adaptive mouse), provide physical assistance to make response, allow student to respond verbally, speech recognition software	Provide interactive system and illustrations of springs to increase engagement and understanding, provide progress bar, provide sequential highlighting, provide hypothesis from prior item supporting short term memory	

Identifying Ways to Enhance Accessibility: ***Pinball Car Race*** *(continued)*

For each step: Top row: identify relevant populations (1 per column). Middle: specify the difficulty. Bottom: suggest solution.							Comments
Step 6	Population →	ELL	Visually Impaired	Hearing Impaired	Motor Impaired	Attention Deficits	
	Difficulty →	Difficulty decoding and comprehending text in English, unfamiliar with abbreviations	Difficulty perceiving text and images			Engagement and persistence difficulties, may not read material carefully, may lose track of place in text, cognitive overload	
	Solution →	Produce test in dominant language, provide English/ non-English language glossary of terms, use of arrows to point out salient features of tables and graphs	Use large font, maximum contrast, provide magnifier, read aloud, provide text to speech, describe images (verbally or in Braille), use of arrows to point out salient features of tables and graphs			Use of arrows to point out salient features of tables and graphs, use of progress bar	

Identifying Ways to Enhance Accessibility:
Pinball Car Race *(continued)*

For each step: Top row: identify relevant populations (1 per column). Middle: specify the difficulty. Bottom: suggest solution.								Comments
Step 7	Population →	ELL	Visually Impaired	Hearing Impaired	Motor Impaired	Attention Deficits		
	Difficulty →	Difficulty decoding and comprehending text in English, unfamiliar with abbreviations, difficulty composing text in English	Difficulty perceiving text, table, bar graph		Difficulty using mouse to select radio buttons, difficulty composing text	Engagement and persistence difficulties, may not read material carefully, may lose track of place in text, cognitive overload		
	Solution →	Produce test in dominant language, provide English/ non-English language glossary of terms, use of arrows to point out salient features of tables and graphs	Use large font, maximum contrast, provide magnifier, read aloud, provide text to speech, describe images (verbally or in Braille)		Integrate augmentative communication devices (e.g., adaptive mouse), provide physical assistance to make response, allow student to respond verbally, speech recognition software	Provide hypothesis supporting short term memory, use of progress bar, multiple representations		

Identifying Ways to Enhance Accessibility:
Pinball Car Race *(continued)*

For each step: Top row: identify relevant populations (1 per column). Middle: specify the difficulty. Bottom: suggest solution.							Comments
Step 8	Population →	ELL	Visually Impaired	Hearing Impaired	Motor Impaired	Attention Deficits	
	Difficulty →	Difficulty decoding and comprehending text in English, unfamiliar with abbreviations, difficulty composing text in English	Difficulty perceiving text, table, bar graph		Difficulty using mouse to select radio buttons, difficulty composing text	Engagement and persistence difficulties, may not read material carefully, may lose track of place in text, cognitive overload	
	Solution →	Produce test in dominant language, provide English/ non-English language glossary of terms, use of arrows to point out salient features of tables and graphs	Use large font, maximum contrast, provide magnifier, read aloud, provide text to speech, describe images (verbally or in Braille)		Integrate augmentative communication devices (e.g., adaptive mouse), provide physical assistance to make response, allow verbal responses, speech recognition	Use of progress bar, provide hypothesis from prior item supporting short term memory, use of scenarios to increase engagement, multiple representations	

Identifying Ways to Enhance Accessibility: ***Pinball Car Race*** *(continued)*

For each step: Top row: identify relevant populations (1 per column). Middle: specify the difficulty. Bottom: suggest solution.							Comments
Step 9	Population →	ELL	Visually Impaired	Hearing Impaired	Motor Impaired	Attention Deficits	
	Difficulty →	Difficulty decoding and comprehending text in English, , difficulty composing text in English	Difficulty perceiving text, table, bar graph		Difficulty using mouse to select radio buttons, difficulty composing text	Engagement and persistence difficulties, may not read material carefully, may lose track of place in text, cognitive overload	
	Solution →	Produce test in dominant language, provide English/ non-English language glossary of terms, use of arrows to point out salient features of tables and graphs	Use large font, maximum contrast, provide magnifier, read aloud, provide text to speech, describe images (verbally or in Braille)		Integrate augmentative communication devices (e.g., adaptive mouse), provide physical assistance to make response, allow student to respond verbally, speech recognition software	Use of progress bar, provide hypothesis from prior item supporting short term memory, use of scenarios to increase engagement, underline and bold salient terms	

Identifying Ways to Enhance Accessibility:
Pinball Car Race *(continued)*

For each step: Top row: identify relevant populations (1 per column). Middle: specify the difficulty. Bottom: suggest solution.							Comments
Step 10	Population →	ELL	Visually Impaired	Hearing Impaired	Motor Impaired	Attention Deficits	
	Difficulty →	Difficulty decoding and comprehending text in English, , difficulty composing text in English	Difficulty perceiving text, table, bar graph		Difficulty using mouse to select radio buttons, difficulty composing text	Engagement and persistence difficulties, may not read material carefully, may lose track of place in text, cognitive overload	
	Solution →	Produce test in dominant language, provide English/ non-English language glossary of terms,	Use large font, maximum contrast, provide magnifier, read aloud, provide text to speech, describe images (verbally or in Braille)		Integrate augmentative communication devices (e.g., adaptive mouse), provide physical assistance to make response, allow verbal responses, speech recognition	Use of progress bar, provide hypothesis from prior item supporting short term memory, multiple representations	

Identifying Ways to Enhance Accessibility: ***Pinball Car Race*** *(continued)*

For each step: Top row: identify relevant populations (1 per column). Middle: specify the difficulty. Bottom: suggest solution.							Comments
Step 11	Population →	ELL	Visually Impaired	Hearing Impaired	Motor Impaired	Attention Deficits	
	Difficulty →	Difficulty perceiving text, table, bar graph	Difficulty perceiving text, table, bar graph		Difficulty using mouse to select drop down menus, difficulty composing text	Engagement and persistence difficulties, may not read material carefully, may lose track of place in text, cognitive overload	
	Solution →	Use large font, maximum contrast, provide magnifier, read aloud, provide text to speech, describe images (verbally or in Braille)	Use large font, maximum contrast, provide magnifier, read aloud, provide text to speech, describe images (verbally or in Braille)		Integrate augmentative communication devices (e.g., adaptive mouse), provide physical assistance to make response, allow verbal responses, speech recognition	Use of progress bar, provide hypothesis from prior item supporting short term memory.	

K-12 Center at ETS

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.