# A Universal Design for Learning-based Framework for Designing Accessible Technology-Enhanced Assessments

Research Report

Robert P. Dolan
Kelly Burling
Michael Harms
Ellen Strain-Seymour
Walter (Denny) Way
*Pearson*

David H. Rose
*Center for Applied Special Technology*

April 2013

**Abstract**

The increased capabilities offered by digital technologies offer new opportunities to evaluate students' deeper knowledge and skills and on constructs that are difficult to measure using traditional methods. Such assessments can also incorporate tools and interfaces that improve accessibility for diverse students, as well as inadvertently introduce new accessibility barriers. Designing these technology-enhanced tasks according to universal design principles is one way to address these accessibility concerns, but requires a grounded understanding of students' diverse abilities and the ways they interact with the tasks. A thorough consideration of the factors that impact construct validity, with an emphasis on identifying and eliminating sources of construct-irrelevant variance, is essential to this process. This report proposes a framework based on the principles of Universal Design for Learning (UDL) for defining task design guidelines consistent with the goals of universal design and thus accessible to a wide range of students, including those with disabilities and who are English learners.

*Keywords:* computer-based testing; technology-enhanced items; accessibility; students with disabilities; English learners; universal design; UDL

**Acknowledgements**

The authors thank Emily Lai and Paul Nichols for their valuable feedback on earlier versions of this manuscript.

# A Universal Design for Learning-based Framework for Defining Guidelines for Accessible Technology-Enhanced Assessments

## Introduction

The increasingly high-stakes associated with large-scale testing programs, especially those in the K-12 arena, have also elevated the stakes for test designers. As greater consequences have been associated with success or failure on such tests, the demand for accurate and valid measures of student progress has intensified. Raising the bar is only part of the challenge: both the No Child Left Behind Act of 2001 (NCLB) and the Individuals with Disabilities Education Act Reauthorization of 2004 (IDEA) have also broadened the range of students that must be included consistently and reliably in high-stakes assessments. The inclusion of—and increased attention to—students with disabilities, students who are English learners (EL), and other "disaggregated groups" for whom traditional means of testing may be inappropriate or invalid, has brought new challenges to the center of the testing industry.

These policy-driven changes are not the only challenge. Advances in the cognitive sciences, learning sciences, and neurosciences underscore the need for assessments that are more discriminating, differentiated, and dynamic than existing instruments, both to meet the needs of accountability and to meaningfully inform instruction.

To meet these challenges this report argues that large-scale assessments must take much greater advantage of advances in two related fields: computer-based testing and universal design. From this argument a framework is proposed that directs the definition of guidelines for accessible computer-based task design.

**The Potential of Digital Technologies**

Computer use for educational assessment and instruction continues to accelerate as technology becomes more powerful and readily available. For the most part, however, the prevailing use of computers in assessment is reflective of an early stage of technology adoption in which new technologies are used to do old things. Indeed, in most current computer-based assessments, computers are used to deliver, manage, or score the same kinds of assessments as would formerly have been delivered through traditional paper-and-pencil formats. The core components of an assessment—the constructs being measured, the methods of measurement, the content, media, and modalities used—have remained essentially the same, still resting largely on a print-based set of assumptions (Bennett, 2002; Pailliotet, Semali, Rodenberg, Giles, & Macaul, 2000; Smagorinsky, 1995).

Digital technologies, on the other hand, have the potential to expand assessments to include the wider range of concepts and skills now demanded by modern cognitive science and initiatives such as the Common Core State Standards (CCSS; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) and the Partnership for 21st Century Skills (Bennett, 2001; Dede, 2009; Dolan, Goodman, Strain-Seymour, Adams, & Sethuraman, 2011; Jodoin, 2003; Kane, 1992; Parshall, Harmes, Davey, & Pashley, 2010; Quellmalz & Pellegrino, 2009; Russell, 2002; K. Scalise & B. Gifford, 2006; Strain-Seymour, Way, & Dolan, 2009; Wendt & Harmes, 2009; Zenisky & Sireci, 2002). For example, digital technology extends the assessment of concepts that cannot be represented well using static text or images, such as those that require sound, motion, or transformation. Similarly, these innovative or technology-enhanced tasks[1] provide the basis for assessing a wider range of skills and means of expression, as well as broaden tremendously the situations in which we can collect this evidence (Behrens & DiCerbo, 2013; Shute, 2011; Wainess, Koenig, & Kerr, 2011). Furthermore, the dynamic and transformative qualities of digital media allow the evaluation of not only "product" but "process", such as timing and sequence, the order and automaticity of skills and strategies, and the use of supports. These qualities will allow students the opportunity, for example, to "conduct virtual lab experiments, in which their actual manipulations of data, technologies, and substances would demonstrate their understanding of processes, methods, and outcomes more clearly than any written or verbal responses could" (Rose & Meyer, 2002, p.148). Thus digital technology provides a more flexible platform for constructing assessments. That flexibility can be used to accurately measure both a broader and deeper range of constructs and a broader range of students. The key is to harness these potentials appropriately.

## Universal Design and Universal Design for Learning

The limits of traditional assessment are dramatically more apparent for some students than for others. For those students "at the margins"—those students doing poorly in traditional classrooms and for whom assessment is arguably most important—traditional assessments may well be least accurate. For too many students academic achievement as measured by assessments is confounded with their ability to use the medium of assessment. According to Rose & Meyer (2002), "traditional assessments tend to measure things that teachers aren't trying to measure (visual acuity, decoding ability, typing ability, motivation), thus confounding the results and leading us to make inaccurate inferences about students' learning" (p.143).

[1] We depart from the more commonly used term "technology-enhanced items" to emphasize the broader designs and roles for these evaluative activities and the situations in which they can be embedded, ranging from traditional computer-based tests to digital games.

Because the expressive medium used for an assessment can influence performance independent of students' knowledge of the content or a skill (e.g. Russell & Haney, 1997, 2000), evaluation must be sensitive to its true purpose, and to the strengths and weaknesses of the learner that may not be germane to the learning being assessed. For example, the creative expression or knowledge gained by students with motor difficulties will not be accurately evaluated via handwritten assessments. For another, the acquisition of content knowledge in social studies or mathematics will not be measured accurately on a print-based multiple-choice test for an EL student. A more flexible approach is needed not only to improve the accuracy of assessments for students "at the margins" but also to enhance the meaningfulness of assessments for all students.

In recent years, especially as policies have stressed the participation of populations with disabilities, varying cultural experiences, and diverse linguistic backgrounds, more flexible and broadly accurate assessments have been required. Universal design, which was originally formulated to create accessible structures and devices by addressing the mobility and communication needs of individuals with disabilities at the design stage (Mace, Hardie, & Place, 1996), has recently been applied to large-scale educational assessment (Almond et al., 2010; Dolan & Hall, 2001, 2007; Ketterlin-Geller, 2005; Thompson, Johnstone, & Thurlow, 2002). These applications have arisen either directly from universal design, as in the case of universal design of assessment (Thompson, et al., 2002), or indirectly through Universal Design for Learning (UDL; Orkwis & McLane, 1998; Rose & Meyer, 2002).

Universal design of assessment (Thompson, et al., 2002) describes seven basic principles derived directly from Mace's original principles, and as such addresses directly neither the role of digital technologies in implementing assessments nor the importance of considering cognitive and pedagogical factors during design. By contrast, UDL preserves the tenets of universal design while building upon the principles and foundations of learning. As such, UDL extends universal design from a physical space to a pedagogical space.

There are two facets to UDL. The first is a conceptual model from which a set of principles and practices are derived. That model, based in the cognitive sciences and neurosciences, establishes the domain to be addressed when applying universal design to learning. The second facet of UDL is the set of specific practices and guidelines by which universal design is actually accomplished. These two facets will now be described in turn.

The UDL model is based in the cognitive neuroscience of learning, and its principles emphasize three key aspects of pedagogy: the means of representing information, the means for the expression of knowledge, and the means of engagement in learning (Rose & Meyer, 2002; Vygotsky, 1934, 1962). Within the learning brain, it is common to identify three broad regions that are evident in both the anatomy and physiology of the brain (Cytowic, 1996). These regions are referred to as recognition, strategic, and affective networks to reflect their individual specializations. Briefly, the recognition networks (located in posterior lobes of the neocortex) receive and analyze information (the "what" of learning), the strategic networks (located in the frontal lobes) plan and execute actions (the "how" of learning), and the affective networks (located in central neocortex and limbic system) evaluate and set priorities (the "why" of learning). Collectively, these networks coordinate how people work and learn (Rose & Lapinski, 2011; Rose & Meyer, 2000, 2002). In addition to these major domains, executive processes are commonly identified in the prefrontal cortex that perform the functions of setting goals, conceiving plans to reach those goals, and monitoring progress. Prefrontal cortex is generally considered to function by biasing or controlling the processing in the other networks so that there is unity of action and identity.

These networks and the pathways that connect them differ among individuals. Variability in functional capabilities across and within these brain networks, and differences in executive processes, are assumed to be a primary contributor to observable learning and performance differences when demonstrating learning. Altogether these basic divisions contribute to the various kinds of learning that are central to both education and assessment.

From a design standpoint, the major principles of UDL reflect the major divisions of learning and the necessary responses to individual differences. By providing *multiple means of representation,* a UDL curriculum addresses the severe limitations in any single representation of the information in a problem. To provide basic access for students with sensory disabilities or other challenges and multiple routes to meaning for all students (e.g., representing a math concept both textually and graphically), it is necessary to provide multiple, redundant, and varied representations of concepts and information.

To support diversity in *strategic networks*, a UDL curriculum provides *multiple means of expression*, giving students flexible models of skilled performance to learn from, opportunities to practice skills and strategies in a supported environment, relevant and ongoing feedback, and flexible opportunities for demonstrating skill using a variety of media and styles. While many students may write (or type or dictate) essays, other alternatives may include rich mixes of writing, illustrating, speaking, animating, and video-making. With UDL, the method of evaluation suits the task and the means. Students are required to meet a higher standard of expressive literacy—knowing in what contexts (for which purposes and for which audiences) to use text, images, sound, video, or combinations of media[2]. At the same time, these options enable students for whom one medium may be a barrier to find a more effective and engaging medium for their purpose.

Regarding the *affective* facets of learning, a UDL curriculum provides *multiple means of engagement* to address the centrality of motivation in learning and the individual differences that underlie motivation and engagement. Offering a choice of content and tools, providing adjustable levels of challenge and support, offering a variety of rewards or incentives, and offering a choice of learning context are effective strategies to support affective learning. There is no single solution to the problem of engaging students because of individual differences; there are many different reasons for students' lack of engagement. Students with disabilities can provide insights into this problem. For example, the same design that would likely engage a student with attention deficit hyperactivity disorder (a high degree of novelty and surprise, for example) might be terrifying (and thus disengaging) to a student with autism spectrum disorder, for whom predictability is paramount.

In terms of supporting the design of accurate and accessible large-scale assessments, UDL offers three general types of guidance. First, it can help inform the definition of constructs in ways that represent the different kinds of learning that assessments are likely to address. Second, UDL can promote assessment designs that include a fully representative consideration of the kinds of students who are likely to be assessed. Third, it can provide options for administering assessments that are flexible enough to accurately assess every student, with "fallback" to accommodations only when necessary.

---

[2] This is consistent with CCSS English Language Arts standards that emphasize the skilled use of multiple media for communication purposes.

## Current Approach

The goal of this report is to devise a research-informed, validity-based, and student-centric UDL framework that can support the definition of guidelines that adequately leverage the potential of digital technologies to better evaluate an increased range of students on an increased range of constructs. This framework is defined in three sections. The first section introduces a theoretical basis for the framework through a discussion of construct validity. The framework is built in the second section. The final section provides suggestions on defining task design guidelines based on the UDL framework.

## Framing Guidelines in Terms of Construct Validity

Construct validity—evidence of a task's ability to measure the appropriate domains of interest that it purports to measure—is necessary to support appropriate and grounded inferences about student knowledge and skills (AERA, APA, & NCME, 1999). Furthermore, such validity must be present across the entire range of students being assessed, regardless of their overall abilities and the challenges they face during assessment. In K-12 educational testing, construct validity is best represented by the degree of alignment between a test, item, or task and one or more individual learning standards or goals, such as those contained within CCSS. Construct validity can also be achieved through the absence of two major threats: construct underrepresentation and construct-irrelevant variance (Messick, 1994):

> *In the threat to validity known as 'construct under-representation,' the assessment is too narrow and fails to include important dimensions or facets of the construct. In the threat to validity known as 'construct-irrelevant variance,' the construct is too broad, containing excess reliable variance that is irrelevant to the interpreted construct" (p.8).*

Of these two validity threats, construct-irrelevant variance best represents the challenges created for diverse learners by the current techniques used in traditional large-scale assessment, since it can consistently inflate or reduce scores for certain individuals or groups in ways that are unrelated to what the task is intended to measure.

The benefits of digital technology-based tasks described earlier can be addressed in terms of construct validity, and specifically in terms of how they can reduce construct-irrelevant variance that arises from overreliance on fixed, inflexible forms of media present in traditional testing. But how can the mechanisms by which these digital technologies can potentially enhance assessment be understood? Large-scale assessment design processes currently address threats to validity using practices such as test accommodations, bias reviews, and differential item functioning (DIF) analyses, and have begun to address some threats to validity at the item or task level through techniques such as the use of accessible design principles and simplified language. However, none of these can sufficiently ensure that digital technologies are used to the fullest extent, namely to provide the flexibility to measure a broader range of students against a broader range of constructs. An alternate perspective for addressing the potential for digital technologies to improve construct validity is provided by Bejar et al. (2003) in their discussion of Embretson (1983): “Construct representation [is] a key aspect of test validity concerned with understanding the cognitive mechanisms related to the item solution and item features that call on these mechanisms.”

As a conceptual model built upon principles of cognitive mechanisms, UDL contributes understanding necessary to fully harness the features of digital technology and thus reduce construct-irrelevance present in traditional testing. Furthermore, UDL can also help mitigate the potential for digital technologies to inadvertently *introduce* sources of construct-irrelevant variance. The power of using UDL lies in its application of a pedagogical approach toward understanding instruction and learning to the processes involved in students' understanding, strategizing, and responding to evaluative tasks. During instruction and assessment, any factors that impede student achievement and the demonstration of student achievement, respectively—due to inappropriate assumptions about students' background knowledge and skills—are, by definition, construct-irrelevant.

The central challenge in creating more accessible tasks is identifying design issues representing sources of construct-irrelevant variability for student populations with different needs. The UDL framework is proposed as a means for identifying and addressing these design issues. At a practical level, the primary goal of the framework is to inform the definition of guidelines such that they provide an easy-to-use reference for designers of tasks and evaluative situations—which may include traditional assessment administration—to relate higher-level theoretical considerations to current and proposed designs. As such the framework must provide a classification system that organizes current understanding in a way that facilitates linking design guidelines to specific design contexts. In addition the framework must encourage systematic evaluation of task and task situation design proposals by identifying strengths and weakness for specific student populations.

A second goal of the framework is to extend UDL foundations to include current human performance and interface design research relevant to the novel task components possible with online evaluation. A necessary requirement of the task creation processes is the continual investigation of how students think about and respond to specific task features (Leighton & Gokiert, 2005). Although extensive research has been performed to examine the statistical properties of tasks, less effort has been directed at understanding the potential for task features to introduce and/or reduce construct-irrelevant variance (Barton, 2007; Ferrara et al., 2003; Haladyna & Downing, 2004; Leighton & Gokiert, 2005). Hence this process of applying and extending UDL principles to evaluate task design from a pedagogical perspective is grounded in cognitive theory.

## Defining the UDL Framework

The UDL framework proposed in this report consists of three modules: 1) phases of student-task interaction, 2) factors that impact construct validity, and 3) categories of student processing abilities, as illustrated in Figure 1. Each of these will now be described.

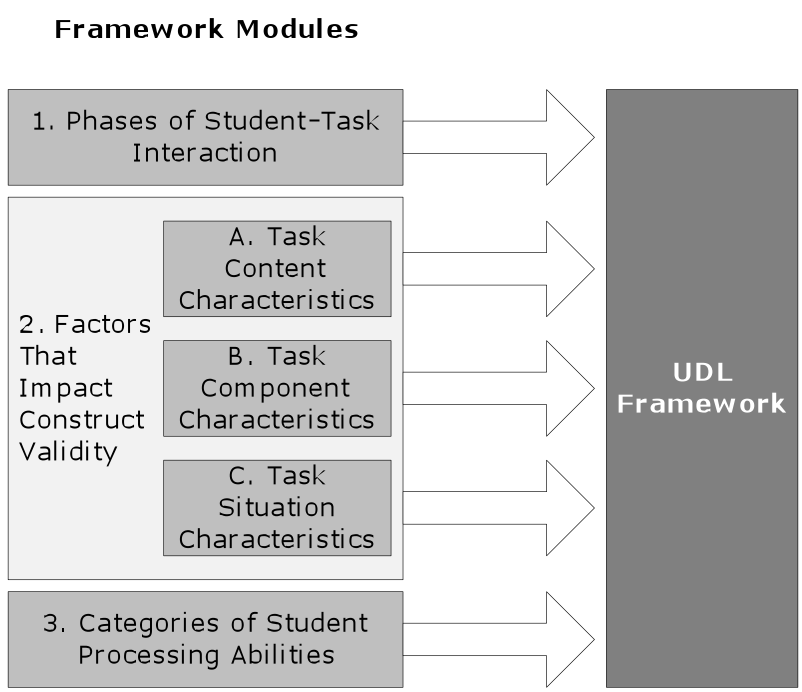


Figure 1. The modules that comprise the UDL framework.

### Module 1: Phases of Student-Task Interaction

The first UDL framework module provides an understanding of the ways that students interact with evaluative tasks. The interaction between student and task is dynamic and often iterative. A task is rarely a mere stimulus that solicits a response, but rather can require students to construct goals and objectives, priorities, situated meanings, ways of responding, and even ways of evaluating performance. Even for relatively “simple” tasks, such as a traditional multiple choice test item designed to evaluate recall, the interaction of the student with the task can be quite complex and dynamic, especially from cognitive, executive, and motivational perspectives.

To help explain this interaction, the major processing steps or phases applicable when a student responds to any task have been articulated. The phases are based on a simple model of a student progressing through a task from its initial presentation until completion of a selected or constructed response. These phases of student-task interaction are shown in Table 1.

Table 1
*Phases of student-task interaction*

| Phase of interaction | Student Task |
|---|---|
| **Task presentation** | Recognize and understand the information presented in the task directions and stimulus/stimuli |
| **Strategic interaction** | Manipulate, reorganize, modify, and/or combine the information in the prompt and strategically apply prior knowledge and skills |
| **Response action** | Plan, organize, and produce a response based upon an understanding of what the response entails |

Within these three phases occur steps in which information is filtered, transformed, constructed, and integrated with existing knowledge and skills. While these phases are clearly distinguishable logically, they are less clearly distinguishable in time or place; the processes of gathering information, checking, re-checking, and constructing new information are all highly iterative and interactive. More importantly, while each phase may emphasize some processes over others (e.g., perceptual processing during task presentation, motoric processing during response action), students’ various processing abilities are impacted in each of the three phases.

As an example, consider how a student's affective processing abilities impact their interactions during the three phases. During the task presentation phase, the various components of any task are likely to be differentially effective in recruiting interest and attention. Whether a student attends to one or more sources of information will, in turn, significantly affect the remainder of his or her performance. However, initial engagement is only one aspect; during the strategic processing phase, affective processes are at work as well. Whether a student sustains engagement and effort through the process of problem solving is an affective issue—motivation and emotional state play key roles in whether continued problem solving is prioritized or not. Finally, in response action stage, affective processing plays a similarly important role in the "quality" of response, including the degree to which the student may review and/or edit his or her response.

## Module 2: Factors that Impact Construct Validity

The second UDL framework module consists of three categories of factors impacting construct validity: a) task content characteristics, b) task component characteristics, and c) task situation characteristics. Each of these three categories will now be described.

### A. Task Content Characteristics

The goal of the UDL framework is to support identification of design considerations related to the construct-relevant and construct-irrelevant factors that impact task validity[3]. In many ways the construct definition is the most challenging aspect of task design and the most fundamental to task validity. The relevance of task content to the task's construct, and the clarity with which the content is presented, directly impact task validity. If the task content is misaligned to the construct or is not clearly contextualized and presented, the functionality of the

[3] The term "task validity" will serve as shorthand for the validity of inferences that can be drawn from student responses to that task.

task itself hardly matters because task validity has already been compromised. The relationship between task content and construct validity presented in the UDL framework is based on an accepted definition of validity (Messick, 1993), traditional "item writing" guidelines (Haladyna, 1999; Haladyna, Downing, & Rodriguez, 2002), research into sources of construct-relevant variance in assessment tasks (Leighton & Gokiert, 2005), and characteristics and manipulable task features described in evidence-centered design (ECD; Mislevy, Almond, & Lukas, 2003).

Multiple factors must be considered to promote the use of appropriate, clearly conveyed task content. The seven task content considerations presented below support designers of tasks and task situations in considering those aspects most appropriate for the current measurement purposes and the implications of using multiple aspects simultaneously. Each consideration has associated questions to remind designers of factors that will influence the task validity. These task content considerations are not intended to replace item or task writing guidelines for specific design processes; instead, they represent common concerns that can be used in conjunction with or incorporated into guidelines defined for specific task situations.

1. **Relevant**—The task measures the content it intends to measure without extraneous content.
    - Does the task clearly address knowledge and skills identified in the test or task situation specifications?
    - Is the content of the task clearly related to the objectives the task is supposed to measure?
2. **Representative**—Task content is aligned with test or task situation specifications; task elements correspond to materials and/or environments used in the classroom.

- Do the content and structure of the task align with quality instructional methods as defined by teachers and experts?
- Does the task look like something students will have seen or used in the classroom?
- Is task content too narrowly construed or presented, and thus in risk of undersampling the breadth of intended constructs?

3. **Realistic**—Unambiguous relationship between media or virtual environment and its real-world counterpart.
    - Are media used to represent actual processes or events sufficiently realistic to be easily identifiable?
    - Could the media representation be mistaken for something else—or be too removed from an actual representation—such that matching the media to what they represent introduces construct-irrelevant cognitive load?
4. **Synergistic**—Task elements complement one another in conveying meaning.
    - Do multiple task stimulus elements stimulate the same processing category simultaneously, or do they compete?
    - Are there multiple simultaneous visual or auditory stimuli that might overwhelm students?
5. **Clear and Unambiguous**—Task intent and the process for responding to task are clearly conveyed and contextualized.
    - Do the instructions clearly convey the scope and intent of the task?
    - Are all the steps necessary to respond fully to the item, and is it clear how to proceed through the steps?
    - Is the context sufficiently defined?

6. **Free of Bias**—Task is sensitive to the full population of students.
   - Is the task sensitive to cultural, socio-economic, gender, age, language, disability, and regional issues?
   - Will construct-irrelevant prior knowledge and/or skills unfairly advantage one group over another?
   - If post field-testing, is DIF analysis likely to detect any existing bias?
7. **Appropriate Time and Task Load**—The time required to view and interact with task elements has been considered and is appropriate to the intended difficulty and level of inquiry of the task. The impact of the task on student's time or energy to complete the rest of the test or task situation has been considered.
   - Is the duration of any multi-media elements appropriate for the difficulty of the task and the level of inquiry (e.g. recall versus problem solving)?
   - Is the time required to interact with an task appropriate for the difficulty and level of inquiry of the task?
   - Do design features, such as multiple screens, increase task load inappropriately (e.g. multiple screens taxing working memory when the measurement focus is drawing inferences between texts)?
   - Will the time or task load of an task negatively impact the amount of time or energy a student will have available for other tasks?

### B. Task Component Characteristics

While an assessment task—especially a traditional "test item"— is often considered an atomic entity, it can instead be conceptualized as a set of standardized components. Separating tasks into constituent parts as such facilitates the identification of sources of variance. Instead of analyzing a combinatorically large list of all possible task types for sources of variance, it is simpler to identify the sources of variance arising from each component individually, based on the assumption that to a first approximation the presence of these elements results in independent effects.

Based on a series of structured interviews conducted with elementary and secondary students with disabilities and their teachers (Burling et al., 2006), interface components were identified with which students interact while responding to technology-enhanced tasks. Part of the rationale for studying students with disabilities was to elucidate interface challenges potentially impact all students (Meyer & Rose, 2005). These findings were combined with a review of educational assessment technology practice and research (Bennett et al., 1999; Rose & Meyer, 2002; Kathleen Scalise & Bernard Gifford, 2006; Zenisky, 2005), with emphasis on those approaches that expand assessments to include the wider range of concepts and skills that modern cognitive science now demands (e.g. Bennett, 2001; Dede, 2009; Quellmalz & Pellegrino, 2009; Russell, 2002; Zenisky & Sireci, 2002). As such, this review included selected response, constructed response, technology-enhanced, performance tasks, and simulation tasks.

As a result of these efforts, 11 task components were identified that serve as unique sources of variance in student interaction, and thus represent the range of construct validity considerations. These components are listed in Table 2. While it is tempting to categorize each task component as belonging to one of the three phases of student-task interaction (e.g., image components as belonging to task presentation; constructed response math components as belonging to response action) in reality interactions with task components occur across two or three phases. For example, while text components within an task stimulus (e.g., a reading passage) are primarily used by students during task presentation phase (initial reading), a student is likely to secondarily interact with the text during strategic interaction phase (referring back), and may even quote a passage from the source text in their response and thus use text components during their response action phase. Table 2 lists both the primary and secondary phase with which each task component is generally associated, to help clarify what each task component category encompasses.

Note that the response options task component covers a broad range of activities and interactions and vary across the types of responses solicited by the tasks. Selected response tasks include multiple choice, multiple response multiple choice, hot-spot or figural response, ordering or sequencing a list of tasks in accordance with some rule, and sorting or categorizing problems or ranking items by correctness. As such, student actions can range from selecting a checkbox or radio button to clicking on a graphic or text to dragging icons or text. Constructed responses include typing a numerical answer to a quantitative question and figural responses where the student marks on, assembles, or interacts with a figure (e.g., build a circuit, plot points on a grid, correct errors in a passage).

Note, too, that to the extent constructed responses involve textual and mathematical expressions, two additional task components have been defined to further specify sources of variance entailed during student interaction. Additional specific constructed responses, such as chemical equations, certainly could also be defined but were not included in the current investigation.

Table 2
*Task components and the phases of interaction with which they are primarily and secondarily associated.*

| Task Component | Definition | Associated Phases of Interaction | | |
|---|---|---|---|---|
| | | Task Presen-tation | Strategic Inter-action | Response Action |
| Text | Language-based terms or concepts in task instructions, stimuli, and response choices | 1° | 2° | |
| Images | Static images (e.g., photos, artwork, maps, cartoons), icons (images on interface elements representing functionality), and symbols (images commonly understood to represent a particular concept) | 1° | 2° | |
| Audio | Independent audio recordings or an audio track accompanying a video or animation | 1° | 2° | |
| Tables and Graphs | Tables used to organize information, convey structure and relationships. Graphs used to represent data visually | 1° | 2° | |
| Mathematical Numbers and Symbols | Mathematical expressions, scientific expressions, scientific notation, scientific elements, numbers and symbols | 1° | 2° | |
| Video and Animation | Visual representations that contain action | 1° | 2° | |
| Response Options | Activities and interactions associated with generating selected and constructed responses | | 2° | 1° |
| Active Objects / Links | Words or icons that result in an action or take student to a different location; pictures with multiple active regions that take student to different locations | 2° | 1° | |
| Multi-stage / Multi-part Tasks | Multiple actions or responses required within one task. Screen elements or environment changes at each stage of multi-stage tasks. Multi-part tasks have a different page for each part. | | 1° | 2° |
| Constructed Response: Text | Language-based composition ranging from fill-in-the-blank to essays | | 2° | 1° |
| Constructed Response: Math | Input a response ranging from a single number to complex proofs or displays of work | | 2° | 1° |

### C. Task Situation Characteristics

In addition to the task-level factors that influence construct validity, factors must be considered at the level of the entire evaluative situation, whether a traditional large-scale assessment administration, an in-class quiz, or completion of an instructional project. In traditional paper-based testing, guidelines are established for how objectives will be measured, including determination of eligible content and format (Millman & Greene, 1993; Smisko, Twing, & Denny, 2000). Additional guidelines detail the layout of traditional test items in a test booklet, the tools that will be available to students, and how students will progress through the test booklet. However, technology-based task writers are faced with numerous tools, media, and functionalities beyond those that have been considered under traditional assessments.

Without guidance and boundaries to establish consistency, every task in an evaluative situation could require students to learn a new interface and new functions, a proposition that would introduce significant sources of construct-irrelevant variance. The following set of task situation considerations, based largely upon existing computer-based instructional and testing design guidelines and research (e.g. Allan, Bulla, & Goodman, 2003; Allman, 2004; Association of Test Publishers, 2002; Consortium, 2002; Dolan, Hall, Banerjee, Chun, & Strangman, 2005), are proposed to guide the design of task situations that uphold task-level construct validity by reducing construct-irrelevant factors at the task situation level:

1. Develop a standardized user experience interface design.
2. Provide an interface template to all task writers and designers so they have a sense of screen design, layout, and real estate.
3. Develop specifications for the presentation and combination of components in tasks with multiple components in the stimulus materials.

4. Develop specifications for the interface and functionality of interactive components.
5. To the extent that students will be able to take notes, draft responses, have access to "scratch paper", use graphic organizers, etc., develop specifications for how these additional materials will be accessed, and when and how they will be visible (e.g., toggle, minimize, available only as a separate page, available superimposed over material on the screen).
6. Based upon curriculum standards and common materials and instructional approaches, determine what activities, simulated environments, and reference materials are relevant for assessment.
7. Based on the activities, simulated environments, and reference materials, design a set of tools, a list of appropriate simulation environments, and a list of activities to be used by task writers.
8. Use the specifications developed from the previous guidelines to create a style sheet for task writers and designers/programmers.
9. Adhere to current best-practices for accessible user interface design.
10. Make available operating system-level accessibility features or functionally replicate them.

**Module 3: Categories of Student Processing Abilities**

The third UDL framework module provides a thorough understanding of the diverse set of abilities and challenges that impact students' effective interaction with tasks in demonstrating their construct-relevant knowledge and skills. Given the wide range of relevant student characteristics, it is important to derive a usable structure for considering these characteristics during assessment design.

This structure, in the form of a set of processing abilities students apply when interacting with the assessment components, was informed by the neurosciences, cognitive sciences, learning sciences, and human information processing field, as well as by the principles of UDL. Such a "cognitive" approach is by no means novel to large-scale assessment and can be traced back to the pioneering work by Spearman (1904) and Cattell (1971), More recently, Haladyna & Downing (2004) proposed that knowledge and information correspond to curricular content or domain knowledge, while skills and abilities refer to cognitive or fluid abilities. Similarly the categories of student processing abilities derived for the UDL framework are based on differences in the ways students interact with their knowledge, skills, and novel information.

From the neurosciences and cognitive/learning sciences underlying UDL, two global assumptions can be derived. First, modern theories of the brain and cognitive science emphasize how distributed the process we call "learning" is across the recognition, strategic, and affective networks in the brain. Where learning was formerly considered a singular aspect of brain function, it is now understood that learning is pervasive, evident, and widely distributed throughout the whole brain (e.g. Duncan, Burgess, & Emslie, 1995; Fuster, 2003; Goldberg, 2001; Lane & Nadel, 2000; Mountcastle, 1998; Shaywitz et al., 1998). Indeed, there are as many facets or types of learning as there are divisions in the brain. Second, modern theories of the brain and cognitive science emphasize process over content. For example, whereas memory was formerly conceived in terms of physical objects or content (a bucket or long-term storage device where there are memory traces), memory is now understood much more dynamically as a process—or more accurately processes—by which the brain changes itself. The same areas of the brain responsible for perception are the ones responsible for memory; these two functions share the same processes and the same anatomy.

Earlier the three networks underlying UDL—recognition, strategic, and affective—were described. Corresponding to these networks are three broad types of processing (and thus learning) as originally described by (Vygotsky, 1934, 1962), namely:

1. **Representational Processes**—those processes located primarily in the posterior neocortex by which the brain takes information in and makes meaning out of it.
2. **Strategic Processes**—those processes located primarily in anterior neocortex by which the brain organizes and executes actions (both mental and physical).
3. **Affective Processes**—those processes located primarily in central neocortex by which the brain prioritizes and places value on information and action.

This parallel, differential processing model concurs with others' theories of cognitive processing and human information processing, most notably those arising from the field of artificial intelligence (Anderson, 1983; Simon, 1979). For example the model of Rumelhart & McClelland (1986) states that information is processed simultaneously by several different parts of the overall system including perceptual, strategic, affective, and executive, and the results of distributed processing are combined into the higher-order processes extending across phases in a task and across tasks within an overall assessment situation. This model also incorporates four general principles shared among most modern theories of human information processing (Huitt, 2003):

1. There is an *assumption of a limited processing capacity* within sensory and cognitive systems. This implies that the amount of information that can be processed by the system is constrained in observable and predictable ways. Bottlenecks, or restrictions in the transmission and processing of information, frequently occur at very specific points and under specific conditions.

2. A *control mechanism is required* to oversee the encoding, transformation, processing, storage, retrieval and utilization of information. That is, not all of the processing capacity of the system is available; an executive function that oversees this process will use up some of this capability. When one is learning a new task or is confronted with a new environment, the executive function requires more processing power than when one is doing a routine task or is in a familiar environment.
3. There is a *two-way flow of information* as we interpret information about the world around us. We use a combination of information captured through the senses (often referred to as bottom-up processing) and information we have stored in memory (often called top-down processing) in a dynamic process as we construct meaning about our environment and our relations to it. As a result, our knowledge and experience shape our perceptions of our current environment.
4. Humans are *physiologically predisposed to process and organize information in specific ways*. For example, it is widely accepted that processing associated with language is functionally distinct from processing that occurs on pictures and imagery. This is supported by the fact that deficits in one skill do not necessarily cause corresponding reductions in all others. Therefore, categories can be created that represent relatively independent forms of processing that occur during the performance of complex tasks.

So as to embody the central tenet of UDL gleaned from recent discoveries in the neurosciences, we propose a fifth principle be added, namely that there is significant *individual variability* in the level of performance across different categories of information processing. As such, the categories of processing abilities defined within the UDL framework characterize variations in task performance as a function of individual student differences. This variability is assumed to result from both the biological and developmental predispositions inherent to the individual and the effects of situation-specific influences.

Using as a catalyst the three somewhat abstract main categories in the UDL conceptual model—representational processing, strategic processing, and affective processing—and considering the additional perspectives described above, the following set of six categories of student processing abilities are proposed, as shown in Table 3.

Table 3
*Student processing ability categories*

| Category | Definition |
|---|---|
| **Perceptual processing** | Those activities involved in converting and categorizing the many sensations that reach our brain into stable and valid representations—percepts—of the external world. Through such processes we are able to recognize and remember objects and events in the environment in spite of variation in their sensory features (e.g., their size, color, location, in different contexts). Individuals differ in their ability to sense and categorize information between different modalities (e.g., vision, touch) and within different aspects of any modality (e.g., pitch, loudness, duration). |
| **Linguistic processing** | Those specific perceptual activities involved in recognizing the patterns of auditory, visual, and tactile stimuli (e.g., Braille) that constitute language. Through specialized processes devoted to language, we are able to recognize and remember the elements from which meaning can be derived, such as vocabulary, syntax, visual word recognition, and text structure (e.g., letter, play, and poem). Individuals vary in their ability to process language and linguistic elements that is separable from the variance associated with their overall perceptual and cognitive capacities. |
| **Cognitive processing** | Those skills and strategies by which an individual constructs meaning from the elements of perception and language to interact with his or her environment. Such "meaning making" typically involves the connection and comparison of one element (e.g., an object, a word) with other elements in memory (prior knowledge), in the environment (context), etc. Comprehending text, as opposed to merely recognizing its elements, involves cognitive processing. Individuals differ both in the prior knowledge they can bring to bear in making meaning, and in the strategies and skills they have available to construct that meaning. |

| Category | Definition |
|---|---|
| **Motoric processing** | Those processes through which meaningful patterns of action can be constructed. Such meaningful patterns of action, a complement to the processes of perception, involve many different forms or modalities of expression, such as pointing, speaking, and writing. Each of these forms of expression involves complex patterns of motoric activity, from motor planning to actual execution. Individuals differ in their ability to express themselves both within modalities (e.g., spelling versus composition) and across modalities (e.g., speaking versus writing). |
| **Executive processing** | Those processes by which an individual sets and maintains goals, devises plans and strategies for reaching those goals, allocates and organizes the effort and mental resources that would be necessary for implementing those strategies, and monitors progress in reaching goals so that plans can be revised or extended as results warrant. Typically executive processes also involve the processes by which goals and tactics are "held in mind" (e.g., working memory) and the processes by which potential tactics are "tried out" in mind with the intent of predicting their outcomes before actual concrete action is taken. Executive functions are emphasized during novel or unstructured tasks, when more than fact retrieval or routine operations are required. Consider an adult walking in a novel, challenging, or dangerous environment, pausing even their gum chewing to focus attention and cautiously plan the next step. |
| **Affective processing** | Those processes by which an individual evaluates the importance or significance of events, objects, or plans. Beyond mere recognition of an object, affective processing evaluates its "value." An object's value is determined not by properties of the object but by the interaction between the individual's status—his or her goals, fears, needs, biological states (e.g., hunger)—and the object's properties. While some affective reactions to the environment are fast, hardwired, and "instinctual," research has also shown that there is a hierarchy of structures which, like the other systems, provide higher levels of integration and opportunities to learn affectively. This provides a means for emotional "background knowledge" and new self regulation strategies to be brought to bear on our affective responses to the world and our ability to control and manipulate our emotions productively. |

As an example of how these categories of student processing abilities can be used in addressing construct-relevance, consider a task that intentionally measures a student's executive processing skills, such as whether she can plan and organize an effective experiment to answer a question in science. In this case, her executive processes are—by definition—construct-relevant. Scaffolding those executive processes—by including a template, model, or guide for planning an experiment, for example—would interfere with construct measurement and compromise the validity of interpretation drawn from her response. This would be true whether or not she has a diagnosed executive function disability.

However, other students may bring functional weaknesses in other categories that are construct-*irrelevant*. An EL student, for example, may do poorly on such a task only because the stimulus and/or response are exclusively in English; he or she would be subject to the additional cognitive load associated with linguistic translation, a requirement not present for other students who happen to have mastery of the English language.

**Consolidating the Framework**

Given the complexity of the UDL framework as described, defining guidelines from it that are comprehensible and usable remains a challenge. To simplify this process, a model is proposed to support understanding of task variance from the perspective of the components contained within the task. This component design model uses two framework elements as two of its dimensions: 1) **task component characteristics** and 2) **categories of student processing abilities**, and is illustrated in Figure 2. At the intersection of a task component and a category of processing abilities there is a set of sources of variance. For each set of sources of variance, a set of design recommendations can be made.

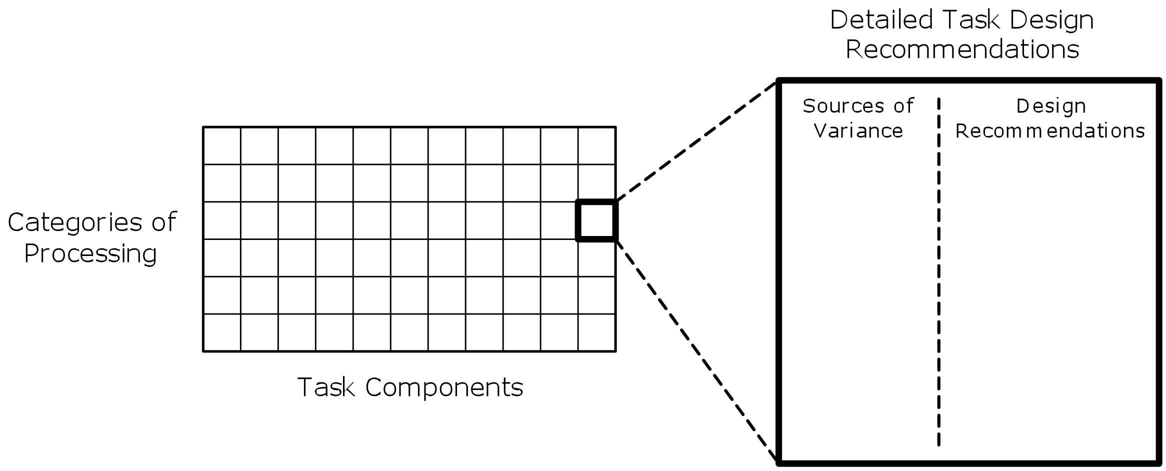


Figure 2. The UDL framework component design model. The component design model can be envisioned as a two dimensional matrix, with task component and category of processing abilities representing the two axes. At the intersection of a particular pair of task component and category of processing abilities, a set of sources of variance can be identified together with design recommendations for reducing this variance to the extent that it is construct-irrelevant.

As an example, consider an audio component used as part of a task stimulus from the perspective of perceptual processing abilities. At the intersection of these two axes, sources of variance, together with design recommendations for reducing this variance should it be construct-irrelevant, can be hypothesized. *Whether a source of variance is construct-irrelevant and thus should be reduced is fully a function of the task's intended construct(s)*. Hearing ability is clearly a source of variance in this situation, but is it construct-relevant? If so—as it would be in the case of a hearing test—design recommendations to reduce variance due to hearing ability are inappropriate, since reduction of variance would invalidate inferences drawn from student responses. On the other hand, in the case of a history task which contains an audio clip of a famous speech, the variance caused by hearing ability is construct-irrelevant. As such, the design recommendations support a task designer in creating a task in which construct-irrelevance due to hearing ability is reduced.

## Applying the UDL Framework to Task Design Guidelines Definition

Task design guidelines defined by the UDL framework would contain three tiers (see Figure 1 for reference). The first tier is a set of task situation considerations and would be defined directly from the **task content characteristics**. The second tier would be a set of task content and considerations defined from the **task situation characteristics**. The third tier would be a set of component level considerations defined directly from the **component design model** presented above in Figure 2. Across all three tiers, the **phases of student-task interaction** support an understanding of interactions between student and task components necessary to identify sources of variance and associated design recommendations.

Together these tiers would provide a flexible, systematic approach toward guiding task design from multiple perspectives and at different levels of detail. Designers often use combinations of checklists, rating scales, guidelines, and structured review processes to direct and support design efforts. However, the effectiveness of the process is only as good as the quality and appropriateness of the approach that is used. The categorical structure of the UDL framework and the resulting task design guidelines are appropriate to the problem context and framed at the appropriate level of detail.

It is important to note that design guidelines must consider all of the various processes and players in the online task development process. While this report has focused on task design, other processes are also key, such as online testing platform design and development. It is up to the individual guidelines to call out various processes and ensure the guidelines address them appropriately.

The categories in the component level considerations suggest the questions to be asked about each component of a task. It is important to define, as specifically as possible, the individual factors within each category that affect student interaction with each component by considering questions such as "How does a student interact with an image on a perceptual level?" and "What factors distinguish between students who perceive an image well and those who do not?" The individual factors constitute sources of variance. Sources of variance define the ways in which students differ in interacting with and responding to task content, the media through which it is conveyed, and the physical interface.

Task designers must avoid creating overly complex interfaces by attempting to address too many design considerations simultaneously in a given task. It is thus important to apply the considerations described under **task situation characteristics** when determining the total number of supports to include in a task and the potential for interaction between them, especially in terms of cognitive load and limited screen real estate.

From a process perspective, the task presentation and response action phases are where the interface issues are most distinct. However, the higher-level information processing activities that occur throughout the task interaction process all have potential to introduce variance. Factors such as cognitive load, anxiety, and differences in media-specific information processing requirements interact in ways that are not directly observable. Time requirements and the efficiency with which a student can interact with a task can also have subtle impacts on validity at both a task and overall evaluative situation level.

It should be noted that the UDL framework can provide a structure for the design of studies intended to identify remaining sources of construct-irrelevant variance both quantitatively and qualitatively, such as through DIF and expert review.

Preliminary guidelines have been created according to the UDL framework and are publicly available at http://pearsonassessments.com/udcbt/. Appendix B below provides one-page summary checklists contained within these preliminary guidelines. These checklists can be used to identify sources of construct-irrelevant variance in task designs as a function of each task component, and point toward design solutions that can remedy them. As such, they provide an high-level list of “what” needs to be done, for which the guidelines themselves provide the “how.”

**Additional UDL Guideline Considerations**

In addition to the considerations addressed under the UDL framework’s task content characteristics, it is critically important that tasks—especially those involving complex scenarios, interfaces, and/or instructions—be designed according to a sound evidentiary-based validity framework or principled assessment design approach, such as ECD (Mislevy, et al., 2003; Rupp, Gushta, Mislevy, & Shaffer, 2010). In fact, preliminary efforts have explored ways to directly combine ECD and UDL during assessment design (Haertel et al., 2010).

Beyond supporting design decisions and review processes, guidelines developed from the UDL framework can provide a blueprint for tagging of tasks according to appropriate accessibility supports from a validity perspective. By forcing explication of intended constructs, it becomes readily apparent when administration supports might compromise validity—whether or not that support is an “accessibility” support (e.g., use of calculators or formula sheets). This helps link design decisions with actual administration conditions.

Guidelines based on the UDL framework must also be grounded on solid understanding of research and instructional and assessment best practices, whenever possible (Almond, et al., 2010). Furthermore, it is important to make extensive use of existing, relevant accessibility standards and specifications, including the following:

- World Wide Web Consortium (W3C) Web Content Accessibility Guidelines and User Agent Accessibility Guidelines
  http://www.w3.org/WAI/guid-tech.html
- Section 508 of the U.S. Rehabilitation Act
  http://www.section508.gov/index.cfm?FuseAction=Content&ID=12
- Accessible Digital Media Guidelines
  http://ncam.wgbh.org/invent_build/web_multimedia/accessible-digital-media-guide/
- Trace Research and Development Center at the University of Wisconsin Application Software Design Guidelines
  http://trace.wisc.edu/docs/software_guidelines/toc.htm
- IMS Global Learning Consortium Guidelines for Developing Accessible Learning Applications
  http://www.imsproject.org/accessibility/accv1p0/imsacc_guidev1p0.html
- IMS Global Learning Consortium Accessible Portable Item Protocol (APIP)
  http://www.imsglobal.org/apip.html

In addition to directly supporting design of tasks, guidelines based on the UDL framework could be used in developing sets of reusable design templates that support the generation of families of tasks with similar functionality and/or supports. This could help address the challenges in the efficiency of applying the guidelines, assist in institutionalizing the principles, and facilitate consistent application across similar types of tasks.

**Evaluating Task Designs Using UDL Framework-based Guidelines**

Identifying and defining the sources of variance present in a task allows an assessment designer to distinguish between those that are construct-relevant from those that are construct-irrelevant. While it is impossible to remove all sources of construct-irrelevant variance from an assessment, awareness of their impact increases the ability to make valid interpretations of student performance.

The following is an example process for analyzing a task design using guidelines defined from the UDL framework. This process could serve as a model for organizing test or task situation design guidelines, and could be incorporated into or even replace components of existing test review processes.

1. **Evaluate the task design for construct validity.**

   This analysis determines whether the appropriate content was selected given the task specification. The task-level construct should clearly define the knowledge and skills the task intends to measure. The content should be evaluated for construct validity through the following two steps:

   a. Assess the construct using the task content and situational considerations within the test or task situation design guidelines.

   b. Assess the construct using assessment situation-specific task writing guidelines and task review processes. Assessment-specific task writing guidelines should suggest the scope of the construct and the methods and materials that would be used to best teach the constructs, provide exemplars for how the constructs should be measured, define task formats, and identify potential sources of variance (Smisko, et al., 2000).

2. **Evaluate the task design for sources of construct-irrelevant variance.**

   This analysis determines whether the task design and chosen interfaces interfere with measurement of the intended construct by adding additional knowledge and/or skill requirements due to the interaction between the student and the medium.

   a. Identify the components within task design (refer to Table 2). For each component, examine the knowledge and skills likely required by that component as a function of each processing abilities category, as listed in the test or task situation design guidelines. Identify which if any of these knowledge or skills factors are construct-irrelevant.

3. **Revise the task design to incorporate appropriate design recommendations and re-evaluate using the guidelines**.

   a. For all sources of construct-irrelevant variance identified above, consider the proposed design solutions suggested in the test or task situation design guidelines to minimize their influence on measurement. The solutions should be considered according to feasibility, cost-effectiveness, and student populations[4] affected.

4. **Upon completion of design changes identified through application of the guidelines, the same process should again be applied to any changed elements to identify additional sources of construct-irrelevant variance that may have been introduced inadvertently.**

The outcome of applying the guidelines analysis process would be the identification of elements of the task design likely to introduce construct-irrelevant variance together with proposed methods for minimizing that variance. While no design evaluation can guarantee that all potential sources of construct-irrelevant variance are eliminated, it is expected that a

[4] Ideally tasks would be not designed to cater to or exclude particular groups of students. However, this may not be feasible at the task level in all cases.

systematic, comprehensive evaluation based on research and best-practices will identify a greater number of issues early. This will facilitate the creation of tasks that are accessible to the widest range of students and more completely achieve the goals of universal design.

## Conclusions

Effective design of technology-based evaluative tasks requires consideration of issues from multiple perspectives. The increased opportunity to create tasks effective for a wider range of student populations also increases potential threats to validity that must be considered to minimize inadvertent introduction of that construct-irrelevant variance. The addition of new capabilities and the availability of new media types require application of guidelines and task review methodologies that identify sources of construct-irrelevant variance not present in traditional assessment.

The conceptual distinction between factors that impact construct validity, categories of student processing abilities, and phases of student-task interaction was proposed as a basis for identifying and mitigating sources of construct-irrelevant variance, as illustrated in Figure 1. As a result, the framework represents a comprehensive approach for analyzing and understanding a wide range of factors that influence construct validity based upon principles of UDL.

The UDL framework addresses *all* students, even those who are native English speakers and without disabilities. A central premise of UDL and thus the framework is that design solutions should be effective for the entire range of student needs; a design effective for a student with a specific need or functional challenge should also be effective for students without any special requirements. The category structure is intended to identify more detailed sources of variability associated with commonly defined populations of students. This increases the likelihood that the needs of a particular population are being considered, and to suggest specific accommodations in cases where this is not possible. Students presently “at the margins” highlight weaknesses in our present testing environments. Considering the issues that these students bring to the testing environment inspires solutions that are applicable to all students (Rose, Meyer, & Hitchcock, 2005). Students with specific challenges help pinpoint sources of construct-irrelevant variance and guide task designers toward better designs for all students.

**References**

AERA, APA, & NCME. (1999). *Standards for Educational and Psychological Testing*. New York: AERA.

Allan, J. M., Bulla, N., & Goodman, S. A. (2003). *TEST ACCESS: Guidelines for Computer Administered Testing*. Louisville, Kentucky: American Printing House for the Blind.

Allman, C. B. (2004). *Making tests accessible for students with visual impairments: A guide for test publishers, test developers, and state assessment personnel*. Louisville, KY: American Printing House for the Blind (APH).

Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke, J., et al. (2010). Technology-enabled and universally designed assessment: Considering access in measuring the achievement of students with disabilities—A foundation for research. *Journal of Technology, Learning, and Assessment, 10*(5), 1-52.

Anderson, J. R. (1983). *The Architecture of Cognition.* Cambridge, MA: Harvard University Press.

Association of Test Publishers. (2002). *Guidelines for Computer-Based Testing*. Washington, D.C.

Barton, K. E. (2007). Validity and Accommodations: The Journey Toward Accessible Assessments. In L. Cook & C. C. Laitusis (Eds.), *Large-Scale Assessment and Accommodations: What Works?* (pp. 81-93). Arlington, VA: Council for Exception Children.

Behrens, J. T., & DiCerbo, K. E. (2013). *Technological Implications for Assessment Ecosystems: Opportunities for Digital Technology to Advance Assessment*. Princeton, NJ: The Gordon Commission on the Future of Assessment.

Bejar, I. I., Lawless, R. R., Morley, M. E., Wagner, M. E., Bennett, R. E., & Revuelta, J. (2003). A feasibility study of on-the-fly item generation in adaptive testing. *Journal of Technology, Learning and Assessment, 2*(3).

Bennett, R. E. (2001). How the Internet will help large-scale assessment reinvent itself. *Education Policy Analysis Archives, 9*(5).

Bennett, R. E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Journal of Technology, Learning, and Assessment, 1*(1).

Bennett, R. E., Goodman, M., Hessinger, J., Kahn, H., Ligget, J., Marshall, G., et al. (1999). Using Multimedia in Large-Scale Computer-Based Testing Programs. *Computers in Human Behavior, 15*(3-4), 283-294.

Burling, K. S., Beck, R., Jude, J., Murray, E. A., Dolan, R. P., & Harms, M. (2006, April 10). *Constructing innovative computer-administered tasks and items according to universal design: Illustrative examples with pilot data.* Paper presented at the National Council on Measurement in Education Annual Meeting, San Francisco, CA.

Cattell, R. B. (1971). *Abilities: Their Structure, Growth, and Action*. Boston: Houghton Mifflin.

Consortium, I. G. L. (2002). *IMS Guidelines for Developing Accessible Learning Applications* (White paper No. Version 1.0): IMS Global Learning Consortium.

Cytowic, R. E. (1996). *The Neurological Side of Neuropsychology*. Cambridge, MA: MIT Press.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science, 323*(5910), 66-69.

Dolan, R. P., Goodman, J., Strain-Seymour, E., Adams, J., & Sethuraman, S. (2011). *Cognitive Lab Evaluation of Innovative Items in Mathematics and English Language Arts Assessment of Elementary, Middle, and High School Students* (Research Report). Iowa City, IA: Pearson.

Dolan, R. P., & Hall, T. E. (2001). Universal Design for Learning: Implications for large-scale assessment. *IDA Perspectives, 27*(4), 22-25.

Dolan, R. P., & Hall, T. E. (2007). Developing accessible tests with universal design and digital technologies: Ensuring we standardize the right things. In L. L. Cook & C. C. Cahalan (Eds.), *Large-scale Assessment and Accommodations: What Works* (pp. 95-111). Arlington, VA: Council for Exception Children.

Dolan, R. P., Hall, T. E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *Journal of Technology, Learning & Assessment, 3*(7), 1-32.

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia, 33*(3), 261-268.

Ferrara, S., Duncan, T. G., Perie, M., Freed, R., McGivern, J., & Chilukuri, R. (2003, April). *Item construct validity: Early results from a study of the relationship between intended and actual cognitive demands in a middle school science assessment.* Paper presented at the American Educational Research Association, Chicago, IL.

Fuster, J. M. (2003). *Cortex and Mind: Unifying Cognition*. New York: Oxford University Press.

Goldberg, E. (2001). *The Executive Brain: Frontal Lobes and the Civilized Mind*. New York: Oxford University Press.

Haertel, G., DeBarger, A. H., Cheng, B., Blackorby, J., Javitz, H., Ructtinger, L., et al. (2010). *Using Evidence-Centered Design and Universal Design for Learning to Design Science Assessment Tasks for Students with Disabilities*: SRI International.

Haladyna, T. M. (1999). *Developing and Validating Multiple Choice Items*. Hillsdale, NJ: Erlbaum.

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 17-27.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 13*(3), 309-344.

Huitt, W. (2003). The Information processing approach to cognition. *Educational Psychology Interactive* Retrieved May 15, 2006, from http://chiron.valdosta.edu/whuitt/col/cogsys/infoproc.html

Jodoin, M. G. (2003). Measurement efficiency of innovative item formats in computer-based testing. *Journal of Educational Measurement, 40*(1), 1-15.

Kane, M. T. (1992). The assessment of professional competence. *Evaluation and the Health Professions, 15*, 163-182.

Ketterlin-Geller, L. R. (2005). Knowing What All Students Know: Procedures for Developing Universal Design for Assessment. *Journal of Technology, Learning, and Assessment, 4*(2), 1-23.

Lane, R. D., & Nadel, L. (2000). *Cognitive Neuroscience of Emotion*. New York: Oxford University Press.

Leighton, J. P., & Gokiert, R. J. (2005, April). *The cognitive effects of test item features: Informing item generation by identifying construct irrelevant variance.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Montreal, Quebec, Canada.

Mace, R. L., Hardie, G. J., & Place, J. P. (1996). *Accessible Environments: Toward Universal Design*. Raleigh, NC: Center for Universal Design.

Messick, S. (1993). Validity. In R. Linn (Ed.), *Educational Measurement* (pp. 13-104). Phoenix, AZ: American Council on Education & Oryx Press.

Messick, S. (1994). *Alternative Modes of Assessment, Uniform Standards of Validity* (No. ETS RR-94-60). Princeton, NJ: Educational Testing Service.

Meyer, A., & Rose, D. H. (2005). The future is in the margins: The role of technology and disability in educational reform. In D. H. Rose, A. Meyer & H. Charles (Eds.), *The universally designed classroom: Accessible curriculum and digital technologies* (pp. 13-35). Cambridge, MA: Harvard Education Press.

Millman, J., & Greene, J. (1993). The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 335-366). New York: Macmillan.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A Brief Introduction to Evidence-Centered Design*. Silver Springs, MD: University of Maryland.

Mountcastle, V. B. (1998). *Perceptual Neuroscience*. Cambridge, MA: Harvard University Press.

National Center on Educational Outcomes. (2009). Online Accommodations Bibliography. Retrieved December 10, 2009, from http://apps.cehd.umn.edu/nceo/accommodations/

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). Common Core State Standards Washington, D.C.: NGA & CCSSO.

Orkwis, R., & McLane, K. (1998). *A curriculum every student can use: Design principles for student access. ERIC/OSEP Topical Brief.* (No. ED423654). Reston, VA: U.S. Department of Education, Office of Special Education Projects.

Pailliotet, A. W., Semali, L., Rodenberg, R. K., Giles, J. K., & Macaul, S. L. (2000). Intermediality: Bridge to critical media literacy. *Reading Teacher*, *54*(2), 208.

Parshall, C. G., Harmes, J. C., Davey, T., & Pashley, P. J. (2010). Innovative item types for computerized testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of Adaptive Testing* (pp. 215-230). New York: Springer.

Quellmalz, E. S., & Pellegrino, J. W. (2009). Technology and Testing. *Science, 323*(5910), 75-79.

Rose, D. H., & Lapinski, S. (2011). What can technology learn from the brain? In T. Gray & H. Silver-Pacuilla (Eds.), *Breakthrough Teaching and Learning: How Educational and Assistive Technologies are Driving Innovations* (pp. 53-70). New York, NY: Springer.

Rose, D. H., & Meyer, A. (2000). Universal Design for Learning. *Journal of Special Education Technology, 15*(1), 67-70.

Rose, D. H., & Meyer, A. (2002). *Teaching Every Student in the Digital Age: Universal Design for Learning*. Alexandria, VA: ASCD Press.

Rumelhart, D., & McClelland, J. (Eds.). (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press.

Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. . *Journal of Technology, Learning, and Assessment, 8*(4).

Russell, M. (2002). How computer-based technology can disrupt the technology of testing and assessment *Technology and assessment: Thinking ahead -- Proceedings of a workshop*. Washington, D.C.: National Academy Press.

Russell, M., & Haney, W. (1997). Testing Writing on Computers: An Experiment Comparing Student Performance on Tests Conducted via Computer and via Paper-and-Pencil. *Education Policy Analysis Archives, 5*(3).

Russell, M., & Haney, W. (2000). Bridging the Gap between Testing and Technology in Schools. *Education Policy Analysis Archives, 8*(19).

Scalise, K., & Gifford, B. (2006). Computer-based assessment in e-learning: A framework for constructing “Intermediate Constraint” questions and tasks for technology platforms. . *Journal of Technology, Learning, and Assessment, 4*(6).

Scalise, K., & Gifford, B. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing “Intermediate Constraint” Questions and Tasks for Technology Platforms. *Journal of Technology, Learning, and Assessment, 4*(6).

Shaywitz, S. E., Shaywitz, B. A., Pugh, K. R., Fulbright, R. K., Constable, R. T., Mencl, W. E., et al. (1998). Functional disruption in the organization of the brain for reading in dyslexia. *Proc Natl Acad Sci U S A, 95*(5), 2636-2641.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer Games and Instruction* (pp. 503-524). Charlotte, NC: Information Age Publishing.

Simon, H. A. (1979). *Models of Thought*. New Haven, London: Yale University Press.

Sireci, S. G., Li, S., & Scarpati, S. (2003). *The effects of test accommodation on test performance: A review of the literature* (No. Center for Educational Assessment Research Report no. 485). Amherst, MA: University of Massachusetts Amherst, School of Education.

Smagorinsky, P. (1995). Constructing meaning in the disciplines: Reconceptualizing writing across the curriculum as composing across the curriculum. *American Journal of Education, 103*(2), 160-184.

Smisko, A., Twing, J., & Denny, P. (2000). The Texas model for content and curricular validity. *Applied Measurement in Education, 12*(4), 333-342.

Spearman, C. E. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology, 15*, 201-293.

Strain-Seymour, E., Way, W. D., & Dolan, R. P. (2009). *Strategies and Processes for Developing Innovative Items in Large-Scale Assessments* (Research Report). Iowa City, IA: Pearson Education.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large-scale assessments* (No. NCEO Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Education Outcomes.

Tindal, G., & Fuchs, L. S. (1999). *A summary of research on test changes: An empirical basis for defining accommodations*. Lexington, KY: Mid-South Regional Resource Center / OSEP.

Vygotsky, L. S. (1934, 1962). *Thought and Language* (E. Hanfmann & G. Vikar, Trans.). Cambridge, MA: MIT Press.

Wainess, R., Koenig, A., & Kerr, D. (2011). *Aligning instruction and assessment with game and simulation design*. Los Angeles, CA: The National Center for Research on Evaluation, Standards, Student Testing (CRESST).

Wendt, A., & Harmes, J. C. (2009). Evaluating innovative items for the NCLEX, Part 1: Usability and pilot testing. *Nurse Educator, 34*(2), 56-59.

Zenisky, A. L. (2005). And Now For Something Completely Different Completely Different… Selected Novel Test Item Formats for Selected Novel Test Item Formats for Computerized Assessment. Amherst, MA: UMass Amherst.

Zenisky, A. L., & Sireci, S. G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education*, *15*(4), 337-362.

## Appendix A. Understanding Students with Disabilities and English Learners

A central tenet of universal design is to consider—and design for—the broadest range of individuals. Limiting accessibility considerations to particular disaggregated groups is antithetical to this approach, as it discourages the application of broad, flexible solutions suitable for the general population. As such, the intention of the UDL framework is to move beyond the mindset of designing—or retrofitting through accommodations—assessments for particular groups of students. However, the field of education is still evolving in its ability to support diverse learners, and both instructional and assessment techniques are often defined according to students with particular disabilities or their native-versus-instructional linguistic abilities. As such, this appendix is provided to foster an understanding of common functional challenges faced by English learners (EL) and students with disabilities during testing, as a function of the six processing abilities categories, to support a transition to a truly universally designed approach in testing.

Table 4 illustrates common situational challenges faced by EL students during testing, together with some common test presentation and response accommodations that are currently used to reduce construct-irrelevant variance potentially introduced by these challenges. It is important to note that these challenges are created by testing students in a language they are still acquiring; as such these challenges are as much—or more—about the situation in which the students is placed than they are about the students themselves.

Table 4
*Common functional challenges faced by EL students during testing*

| **Processing Abilities Category** | Common Situational Challenges | Example Assessment Presentation & Response Accommodations |
|---|---|---|
| **Linguistic** | ▪ Decoding, fluency, comprehension challenges during reading<br>▪ Comprehension of syntactic and semantic meaning<br>▪ Integrating information, making inferences<br>▪ Connecting text<br>▪ Poor meta-cognitive skills<br>▪ Difficulty generating mental models needed for comprehension (reading, listening)<br>▪ Difficulty with written expression (planning, revising, self-regulating, writing mechanics) | ▪ Grammatical support tools<br>▪ Text-to-speech systems<br>▪ Access to dictionaries and glossaries, including bilingual<br>▪ Dual language testing materials |

Table 5 illustrates common functional and situational challenges faced by students with disabilities during testing. The organizational structure is based upon the disability categories defined under IDEA and relevant to students taking general assessments, the common functional and situational challenges that students with these disabilities face, and common test presentation and response accommodations that are currently used to reduce construct-irrelevant variance potentially introduced by these challenges (for reviews see: National Center on Educational Outcomes, 2009; Sireci, Li, & Scarpati, 2003; Tindal & Fuchs, 1999).

Table 5

*Common functional challenges faced by students with disabilities during testing*

| Processing Abilities Category | Common Associated Disabilities | Common Functional / Situational Challenges | Example Assessment Presentation & Response Accommodations |
|---|---|---|---|
| **Perceptual** | Blind | ▪ No functional vision (visual acuity 20/200 or poorer) | ▪ Braille Embosser<br>▪ Refreshable Braille Display<br>▪ Nemeth Code<br>▪ Screen Readers/Talking Browsers<br>▪ Text-to-speech systems<br>▪ Optical Character Recognition<br>▪ Haptic Devices |
| | Low Vision | ▪ Limited functional vision (corrected visual acuity between 20/40 and 20/200) | ▪ Screen Magnification<br>▪ Screen Readers/Talking Browsers<br>▪ Text-to-speech systems<br>▪ Optical Character Recognition |
| | Deaf / Hard of Hearing | ▪ No functional hearing, limited functional hearing<br>▪ Often corresponding delays in linguistic, social, emotional and cognitive development<br>▪ Literacy problems, especially delays in reading and writing, and difficulty with decoding and comprehension<br>▪ Differences between ASL and English syntax | ▪ Volume Controls<br>▪ Signing Avatars<br>▪ Grammatical support tools |

| Processing Abilities Category | Common Associated Disabilities | Common Functional / Situational Challenges | Example Assessment Presentation & Response Accommodations |
|---|---|---|---|
| **Linguistic** | Learning Disability: Reading / Language | ▪ Decoding, fluency, comprehension challenges during reading<br>▪ Comprehension of syntactic and semantic meaning<br>▪ Integrating information, making inferences<br>▪ Connecting text<br>▪ Poor meta-cognitive skills<br>▪ Difficulty generating mental models needed for comprehension (reading, listening)<br>▪ Difficulty with written expression (planning, revising, self-regulating, writing mechanics) | ▪ Grammatical support tools<br>▪ Text-to-speech systems |
| | Primary Language Disorders | ▪ Limited vocabulary<br>▪ Limited pragmatic use of language<br>▪ Limited comprehension of syntactical structures<br>▪ Limited background knowledge and social development<br>▪ Reading disorders and delays | ▪ Grammatical support tools<br>▪ Text-to-speech systems<br>▪ Embedded vocabulary support<br>▪ Supplemental background knowledge |
| **Cognitive** | Mild Mental Retardation | ▪ Impaired functioning across subject areas<br>▪ Longer time to accomplish tasks, achieve mastery<br>▪ Generalizing skills<br>▪ Comprehension<br>▪ Expression<br>▪ Task Switching<br>▪ Time perception | ▪ Grammatical support tools<br>▪ Text-to-speech systems |
| **Motoric** | Physical Disability | ▪ Physical mobility: fine motor skills, difficulty manipulating materials, limited functional visual field due to constrained head and eye movements<br>▪ Difficulty maintaining body positions, fatigue<br>▪ Potential for corresponding developmental brain disturbance | ▪ Alternative Keyboards<br>▪ Alternative Mouse Systems<br>▪ Voice Recognition Systems<br>▪ Screen Readers/Talking Browsers<br>▪ Text-to-speech systems<br>▪ On-screen keyboards<br>▪ Word-prediction |

| Processing Abilities Category | Common Associated Disabilities | Common Functional / Situational Challenges | Example Assessment Presentation & Response Accommodations |
|---|---|---|---|
| | Dysgraphia / Dyspraxia (fine motor issues) | ▪ Handwriting and drawing<br>▪ Writing fluency<br>▪ Manipulating materials<br>▪ Fine motor skills | ▪ Alternative Keyboards<br>▪ Alternative Mouse Systems<br>▪ Snap-to constraints |
| **Executive** | Attention Deficit / Hyperactivity Disorder | ▪ Setting and maintaining goals<br>▪ Sustaining attention and effort<br>▪ Organizing strategies<br>▪ Task switching (both perseveration and distraction)<br>▪ Monitoring progress and reacting to feedback | ▪ Grammatical support tools<br>▪ Text-to-speech systems<br>▪ Graphic organizers, checklists<br>▪ Alternative contexts for testing (reduce distractions, etc.)<br>▪ Alternative timing (smaller sections) |
| | Learning Disability: Math | ▪ Automaticity, fact retrieval<br>▪ Problem solving is interrupted due to concentration on fact retrieval<br>▪ Representations of word problems | ▪ Grammatical support tools<br>▪ Text-to-speech (MathML)<br>▪ Calculator |
| **Affective** | Autism Spectrum Disorder: Asperger's Syndrome | ▪ Anxiety<br>▪ Hypersensitivity and compulsions (routines, stimulation)<br>▪ Communication, listening comprehension<br>▪ Comprehending abstract concepts and language<br>▪ Ascertaining relevance<br>▪ Motor involvement<br>▪ Distractibility<br>▪ Time perception | ▪ Text-to-speech systems<br>▪ Grammatical support tools<br>▪ Alternative contexts for testing (reduce distractions, etc.)<br>▪ Alternative timing (smaller sections) |
| | Emotional Disturbance | ▪ Difficulty setting goals and objectives<br>▪ Difficulty in sustaining concentration and effort<br>▪ Difficulty in adopting norms and values of testing<br>▪ Difficulty in monitoring progress<br>▪ Difficulty in recognizing affective signals and markers | ▪ Alternative contexts for testing (reduce distractions, etc.)<br>▪ Alternative timing (smaller sections) |

## Appendix B. Guideline Checklists

The following one-page summary checklists have been pulled out of preliminary guidelines defined according to the UDL framework and available publicly at http://www.pearsonassessments.com/udcbt/. These checklists can be used to identify sources of construct-irrelevant variance in task designs as a function of task component, and point toward design solutions that can remedy them. As such they provide a high-level list of "what" needs to be done, for which the guidelines themselves provide the "how." Note that only relevant categories of processing are considered for each task component (e.g., no motoric factors are considered for text or image components).

## Task Component: Text

| Task Component: Text | | |
|---|---|---|
| **Category of Processing** | **If this task *does not* intend to measure:** | **Then consider the following design options to minimize measurement of unintended constructs (construct irrelevant variance):** |
| **Perceptual** | Visual Ability | Refreshable Braille, Screen Reader, TTS |
| | Visual Acuity | Flexible size text |
| | Visual Discrimination | Flexible fonts, flexible contrast |
| **Linguistic** | English Language Proficiency | Alternate languages (natural, ASL, non-English) |
| | Vocabulary | Vocabulary links (dictionary & thesaurus, talking, multiple languages) |
| | Syntactic Skills | Grammar aids, simplified syntax |
| | Word Decoding Skills | TTS for individual words (talking dictionary) |
| | Reading Fluency | TTS with synchronous highlighting |
| | Knowledge of Text Structure | Graphic organizers, explicit indicators of text structure |
| **Cognitive** | Background Knowledge | Links to background knowledge |
| | Comprehension Strategies | Prompts and supports for strategies |
| | Categorical and Conceptual Skills | Advance organizers, concept maps |
| | Attention and Concentration skills | Prompting, breaking text into smaller sections, locate prompts near relevant text |
| **Executive** | Goal Setting Ability | Explicit Instructions, goal-setting supports |
| | Goal Maintenance and Adjustment | Reminders, prompts |
| | Monitoring Progress | Extrinsic scaffolds for monitoring |
| | Working Memory | Note-taking, mnemonic aids, text complement, locate prompts near relevant stimuli |
| **Affective** | Self-regulation | Scaffolds for self-regulation |
| | Intrinsic Task-Specific Motivation | Alternative content for interest |
| | Extrinsic Incentives | Individualized rewards, repercussions |
| | Test Conditions | Alternative settings and conditions (time, sessions, location); review for racial, cultural, ethnic, & gender bias; differential item functioning; age appropriate content |

## Task Component: Images

| Task Component: Images | | |
|---|---|---|
| **Category of Processing** | **If this task *does not* intend to measure:** | **Then consider the following design options to minimize measurement of unintended constructs (construct irrelevant variance):** |
| **Perceptual** | Visual Ability | Tactile display, 3-D manipulatives, text equivalents, longdesc |
| | Visual Acuity | Flexible image size, zoom |
| | Visual Discrimination | Flexible contrast |
| | Color Perception | User specified color options, avoid common color-blindness combinations, redundant presentation of information conveyed in color |
| | Shape Recognition | Alternative visual options, description, tactile option |
| **Cognitive** | Visual Processing Skills | Highlight critical features |
| | Knowledge of Graphic Conventions | Alternative descriptions or depictions |
| | Knowledge of Iconic Conventions | Alternatives for icons (rollover descriptions, legend, customizable icons for tools or commands) |
| | Visual Syntax Fluency | Highlight critical relationships |
| | Background Knowledge | Links to background knowledge |
| | Cognitive Strategies | Prompts and supports for viewing and interpretation strategies |
| | Planning and Organizing Skills | Graphic organizer, planning templates |
| | Attention and Concentration | Simplified images, prompts |
| **Executive** | Goal Setting Ability | Explicit Instructions, goal-setting supports |
| | Goal Maintenance and Adjustment | Reminders, prompts |
| | Monitoring Progress | Extrinsic Scaffolds for monitoring |
| | Working Memory | Note-taking, mnemonic aids, text complement, locate prompts near relevant stimuli |
| **Affective** | Self-regulation | Scaffolds for self-regulation |
| | Intrinsic Task-specific Motivation | Alternative content for interest |
| | Extrinsic Incentives | Individualized rewards, repercussions |
| | Test Conditions | Alternative settings and conditions (time, sessions, location); review for racial, cultural, ethnic, & gender bias; differential item functioning; age appropriate content |

## Task Component: Audio

| Task Component: Audio | | |
|---|---|---|
| **Category of Processing** | **If this task *does not* intend to measure:** | **Then consider the following design options to minimize measurement of unintended constructs (construct irrelevant variance):** |
| **Perceptual** | Hearing Ability | Visual alerts, captions (SMIL, etc.) |
| | Auditory Threshold | Adjustable volume |
| | Auditory Processing Speed | Adjustable rate |
| | Auditory Discrimination | Highlight critical features, emphasize discriminants |
| **Linguistic** | English Language Proficiency | Alternate languages (natural, ASL, non-English) |
| | Receptive Vocabulary | Vocabulary links to predetermined lists or dictionary with word prediction for spelling |
| | Syntactic skills | Alternate syntactic levels (simplified text) |
| | Prosody Recognition | Alternative prosodic emphasis |
| | Idiomatic Expression Familiarity | Alternatives for idiomatic expressions |
| **Cognitive** | Background Knowledge | Links to background knowledge |
| | Cognitive Listening Skills | Prompts and supports for listening and interpretation strategies |
| | Planning and Organizing Skills | Graphic organizer, planning templates |
| | Attention and Concentration | Increased segmentation, navigation control (pause, forward, reverse, replay, and search features |
| **Motoric** | Navigation of Audio File | Keyboard alternatives for all on screen navigation commands, assistive device compatibility, do not disable OS functions |
| **Executive** | Goal Setting Ability | Explicit Instructions, goal-setting supports |
| | Goal Maintenance and Adjustment | Reminders, prompts |
| | Monitoring Progress | Extrinsic Scaffolds for monitoring |
| | Working Memory | Note-taking, mnemonic aids, text complement, locate prompts near relevant stimuli |
| **Affective** | Self-regulation | Scaffolds for self-regulation |
| | Intrinsic Task-specific Motivation | Alternative content for interest |
| | Extrinsic Incentives | Individualized rewards, repercussions |
| | Test Conditions | Alternative settings and conditions; review for racial, cultural, ethnic, & gender bias; differential item functioning; age appropriate content |

## Task Component: Tables and Graphs

| Task Component: Tables and Graphs | | |
|---|---|---|
| **Category of Processing** | **If this task *does not* intend to measure:** | **Then consider the following design options to minimize measurement of unintended constructs (construct irrelevant variance):** |
| **Perceptual** | Visual Ability | Identify row and column headers using appropriate mark up language; identify subheads and data cells using appropriate mark up language; provide a linearized version for Screen Readers that cannot read tables or side by side text; do not use structural markup for visual formatting, it interferes with screen readers; provide summaries of tables for VI; static tables can be rendered with pre-produced audio (MathSpeak, National Braille Association Tape Recording Manual) |
| | Visual Acuity | Flexible size, zoom, SVG or similar for static tables/graphs |
| | Visual Discrimination | Flexible Contrast |
| | Color Perception | User specified color options, avoid common color-blindness combinations, redundant presentation of information conveyed in color |
| | Display Complexity | Present only necessary information; explicit labeling and formatting |
| **Cognitive** | Knowledge of Conventions | Highlight critical (but construct-irrelevant) features; Alternative descriptions or depictions |
| | Visual syntax Fluency | Highlight critical relationships |
| | Background Knowledge | Links to background knowledge |
| | Cognitive Strategies | Prompts and supports for viewing and interpretation strategies |
| | Planning and Organizing Skills | Graphic organizer, planning templates |
| | Attention and Concentration | Tools for orientation within table/graph |
| **Motoric** | Navigating Tables and Graphs (Static and Dynamic | Keyboard alternatives for all on screen navigation commands, long desc of commands for active tables or graphs including result of action; assistive device compatibility, do not disable OS functions |
| **Executive** | Goal Setting Ability | Explicit Instructions, goal-setting supports |
| | Goal Maintenance and Adjustment | Reminders, prompts |
| | Monitoring Progress | Extrinsic Scaffolds for monitoring |
| | Working Memory | Note-taking, mnemonic aids, text complement, locate items near relevant screen elements |
| **Affective** | Self-regulation | Scaffolds for self-regulation |
| | Intrinsic Task-specific Motivation | Alternative content for interest |
| | Extrinsic Incentives | Individualized rewards, repercussions |
| | Test Conditions | Alternative settings and conditions (time, sessions, location); review for racial, cultural, ethnic, & gender bias; differential item functioning; age appropriate content |

**Task Component: Mathematical and Scientific Notation**

| Task Component: Mathematical and Scientific Notation | | |
|---|---|---|
| **Category of Processing** | **If this task *does not* intend to measure:** | **Then consider the following design options to minimize measurement of unintended constructs (construct irrelevant variance):** |
| **Perceptual** | Visual Ability | Nemeth Code, MathML (www.w3c.org/Math) LaTeX, ChemML, CML (http://www.xml-cml.org/information/position.html), AsTeR (http://www.cs.cornell.edu/Info/People/raman/aster/aster-toplevel.html); flexible size, pictorial representations |
| | Visual Acuity | Flexible Fonts, Zoom, SVG or similar technology |
| | Visual Discrimination | Flexible Contrast |
| **Linguistic** | Mathematical Syntax | Highlight order of operations |
| | Mathematical Fluency | Simplified numbers, retain concept |
| **Cognitive** | Background Knowledge | Links to background knowledge |
| | Calculations Complex | Calculator, scrap paper, simplified numbers |
| | Expressions | Make expressions accessible to screen readers with text descriptions using MathML or LaTeX, following guides for spoken mathematics (NCAM). Embed text and audio files, gets tricky with complex expressions because each part needs to be accessible separately as well wholly. Concatenated speech is awkward for complex expressions. AsTeR reads LaTeX, creates audio, and allows navigation (http://www.cs.cornell.edu/Info/People/raman/aster/aster-toplevel.html) |
| **Executive** | Goal Setting Ability | Explicit Instructions, goal-setting supports |
| | Goal Maintenance and Adjustment | Reminders, prompts |
| | Monitoring Progress | Extrinsic Scaffolds for monitoring |
| | Working Memory | Note-taking, mnemonic aids, text complement, locate prompts near relevant stimuli |
| **Affective** | Self-regulation | Scaffolds for self-regulation |
| | Intrinsic Task-specific Motivation | Alternative content for interest |
| | Extrinsic Incentives | Individualized rewards, repercussions |
| | Test Conditions | Alternative settings and conditions (time, sessions, location); review for racial, cultural, ethnic, & gender bias; differential item functioning; age appropriate content |

## Task Component: Video/Animation

| Task Component: Video/Animation | | |
|---|---|---|
| **Category of Processing** | **If this task *does not* intend to measure:** | **Then consider the following design options to minimize measurement of unintended constructs (construct irrelevant variance):** |
| **Perceptual** | Visual Ability | Rich description |
| | Visual Acuity | Flexible image size, zoom |
| | Visual Discrimination | Black and White/Greyscale options, flexible contrast |
| **Cognitive** | Visual Processing Skills | Highlight critical features |
| | Knowledge of Graphic Conventions | Alternative descriptions or depictions |
| | Knowledge of Iconic Conventions | Alternatives for icons |
| | Visual Syntax Fluency | Highlight critical relationships |
| | Background Knowledge | Links to background knowledge |
| | Cognitive Strategies | Prompts and supports for viewing and interpretation strategies |
| | Planning and Organizing Skills | Graphic organizer, planning templates |
| | Attention and Concentration | Summary of action, prompts, navigation control (pause, forward, reverse, replay and search features), highlighting in graphic organizer synchronized to stages of action |
| **Motoric** | Navigation of Animation or Video | Keyboard alternatives for all on screen navigation commands; assistive device compatibility, do not disable OS functions |
| **Executive** | Goal Setting Ability | Explicit Instructions, goal-setting supports |
| | Goal Maintenance and Adjustment | Reminders, prompts |
| | Monitoring Progress | Extrinsic Scaffolds for monitoring |
| | Working Memory | Note-taking, mnemonic aids, text complement, locate prompts near relevant stimuli |
| **Affective** | Self-regulation | Scaffolds for self-regulation |
| | Intrinsic Task-specific Motivation | Alternative content for interest |
| | Extrinsic Incentives | Individualized rewards, repercussions |
| | Test Conditions | Alternative settings and conditions (time, sessions, location); review for racial, cultural, ethnic, & gender bias; differential item functioning; age appropriate content |

## Task Component: Response Options

| Task Component: Response Options | | |
|---|---|---|
| **Category of Processing** | **If this task *does not* intend to measure:** | **Then consider the following design options to minimize measurement of unintended constructs (construct irrelevant variance):** |
| **Perceptual** | Ability to Distinguish Stimulus and Response Components | Explicit labeling; physical and functional separation of stimulus and response areas, with supports for navigation between them by screen readers and single switch devices |
| | Ability to Ascertain Actions Required for Response | Simple, clear instructions; highlight all enabled elements; consistent methodology for highlighting enabled elements; simultaneous highlighting of functionally related enabled elements; clear relationship between mouse active behaviors and mouse keys, tab navigation, single switch navigation; animations to model required actions |
| **Linguistic** | English Language Proficiency | Alternate languages (natural, ASL, non-English), |
| | Vocabulary | Vocabulary links (dictionary & thesaurus, talking, multiple languages) |
| | Syntactic Skills | Grammar aids, simplified syntax |
| | Word Decoding Skills | TTS for individual words (talking dictionary) |
| | Reading Fluency | TTS with synchronous highlighting |
| | Knowledge of Text Structure | Graphic organizers, explicit indicators of text structure |
| **Cognitive** | Understanding Response Requirements | Animation of tools required for response, practice test training, consistent tool use across tasks |
| | Planning and Organizing Skills | Graphic organizer, planning templates |
| | Attention and Concentration | Divide task into discrete steps |
| **Motoric** | Navigation Abilities | Do not disable OS functions to provide that assistive device and software compatibility; keyboard alternatives for all on screen and mouse active commands, tab navigation, Voice activation |
| | Selection | Dwell time selection Assigned key (tab, space) |
| | Keyboarding | Alternate Keyboard, Screen Keyboard (fatiguing), Dictation (Scribe or Voice Recognition) |
| | Drag and Drop | Assigned keys for select, hold, drop, Keyboard equivalents, Structured navigation with tabs, , Snap to constraints |
| **Executive** | Goal Setting Ability | Explicit Instructions, goal-setting supports |
| | Goal Maintenance and Adjustment | Reminders, prompts |
| | Monitoring Progress | Extrinsic Scaffolds for monitoring |
| | Working Memory | Note-taking, mnemonic aids, text complement, locate prompts near relevant stimuli |
| | Self-regulation | Scaffolds for self-regulation |

| **Task Component: Response Options** | | |
|---|---|---|
| **Category of Processing** | **If this task *does not* intend to measure:** | **Then consider the following design options to minimize measurement of unintended constructs (construct irrelevant variance):** |
| **Affective** | Self-regulation | Scaffolds for self-regulation |
| | Intrinsic Task-specific Motivation | Alternative content for interest |
| | Extrinsic Incentives | Individualized rewards, repercussions |
| | Test Conditions | Alternative settings and conditions (time, sessions, location); review for racial, cultural, ethnic, & gender bias; differential item functioning; age appropriate content |

## Task Component: Active Objects and Links

| Task Component: Active Objects and Links | | |
|---|---|---|
| **Category of Processing** | **If this task *does not* intend to measure:** | **Then consider the following design options to minimize measurement of unintended constructs (construct irrelevant variance):** |
| **Perceptual** | Visual Ability | Text equivalents for all non-text elements |
| | Visual Acuity | Flexible image size, zoom |
| | Visual Discrimination | User specified color options, avoid common color-blindness combinations, redundant presentation of information conveyed in color, flexible contrast |
| | Shape Recognition | Alternative visual options, description |
| **Cognitive** | Visual Processing Skills | Redundant text links, client side image maps, hot spots should be rendered as list of hypertext links |
| | Knowledge of Active Object Conventions/ Hypertext Syntax Fluency | Explicit instructions, semantic info of objects conveyed through text, objects that represent controls/tools or other programmatic elements must be used consistently throughout, clearly indicate the target of all links, highlight critical relationships (color or highlighting to indicate related enabled elements, redundant text or auditory indicators for VI) |
| | Attention and Concentration | Prompts, explicit descriptions, focus indicator |
| **Motoric** | Navigation Abilities | Keyboard alternatives for all on screen navigation or action; assistive device compatibility, do not disable OS functions |
| **Executive** | Goal Setting Ability | Explicit Instructions, goal-setting supports |
| | Goal Maintenance and Adjustment | Reminders, prompts |
| | Monitoring Progress | Extrinsic Scaffolds for monitoring |
| | Working Memory | Note-taking, mnemonic aids, text complement, locate prompts near relevant stimuli |
| **Affective** | Self-regulation | Scaffolds for self-regulation |
| | Intrinsic Task-specific Motivation | Alternative content for interest |
| | Extrinsic Incentives | Individualized rewards, repercussions |
| | Test Conditions | Alternative settings and conditions (time, sessions, location); review for racial, cultural, ethnic, & gender bias; differential item functioning; age appropriate content |

## Task Component: Constructed Response: Text Composition

| Task Component: Constructed Response: Text Composition | | |
|---|---|---|
| **Category of Processing** | **If this task *does not* intend to measure:** | **Then consider the following design options to minimize measurement of unintended constructs (construct irrelevant variance):** |
| **Perceptual** | Visual Ability | Alternate Input Devices, Braille displays, TTS or self-voicing read-back of student composition |
| | Visual Acuity | Flexible size text |
| | Visual Discrimination | Flexible fonts, flexible contrast |
| **Linguistic** | English Language Proficiency | Alternate languages for composition (natural, ASL, non-English), TTS or self-voicing read-back |
| | Vocabulary | Vocabulary links (dictionary & thesaurus, talking, multiple languages, ASL translator) |
| | Syntactic skills | Grammar check, TTS or self-voicing read-back |
| **Cognitive** | Medium Familiarity/Dexterity | Alternate response options |
| | Planning and Organizing Writing | Graphic organizers, access to rubric |
| | Writing Fluency | Models, virtual mentors |
| **Motoric** | Production Dexterity | Alternate input devices, dictation (voice recognition, scribe), do not override OS functions |
| | Navigation Abilities | Keyboard alternatives, assistive device compatibility (do not override OS functions) |
| | Strength and Mobility | Assistive Device Compatibility (do not override OS functions), physical setting flexibility |
| | Automaticity | Variable or no time constraints |
| **Executive** | Goal Setting Ability | Explicit Instructions, goal-setting supports |
| | Goal Maintenance and Adjustment | Reminders, prompts |
| | Monitoring Progress | Extrinsic Scaffolds for monitoring |
| | Working Memory | Note-taking, mnemonic aids, text complement, locate prompts near relevant stimuli |
| **Affective** | Self-regulation | Scaffolds for self-regulation |
| | Intrinsic Task-specific Motivation | Alternative content for interest |
| | Extrinsic Incentives | Individualized rewards, repercussions |
| | Test Conditions | Alternative settings and conditions; review for racial, cultural, ethnic, & gender bias; differential item functioning; age appropriate content |

**Task Component: Constructed Response: Math**

| Task Component: Constructed Response: Math | | |
|---|---|---|
| **Category of Processing** | **If this task *does not* intend to measure:** | **Then consider the following design options to minimize measurement of unintended constructs (construct irrelevant variance):** |
| **Perceptual** | Visual Ability | Alternate Input Devices, Braille displays, particularly complex for math to be read-back, complex for students entering Nemeth Code |
| | Visual Acuity | Flexible size font |
| | Visual Discrimination | Flexible fonts, flexible contrast |
| **Linguistic** | Creating Graphs or Tables | Templates, edit or reorganize vs. create, spreadsheets, graph + scrap paper |
| | Creating Diagrams or Drawings | Drawing palettes |
| | Creating Equations | Equation palettes, on-screen calculators |
| **Cognitive** | Calculations | On screen calculators, simplified calculations |
| | Problem Solving | Models, virtual mentors, calculation focus |
| **Motoric** | Production Dexterity | Alternate input devices, dictation (voice recognition, scribe), do not override OS functions |
| | Navigation Abilities | Keyboard alternatives, assistive device compatibility (do not override OS functions) |
| | Strength and Mobility | Assistive Device Compatibility (do not override OS functions), physical setting flexibility |
| | Automaticity | Variable or no time constraints |
| **Executive** | Goal Setting Ability | Explicit Instructions, goal-setting supports |
| | Goal Maintenance and Adjustment | Reminders, prompts |
| | Monitoring Progress | Extrinsic Scaffolds for monitoring |
| | Working Memory | Note-taking, mnemonic aids, text complement, locate prompts near relevant stimuli |
| **Affective** | Self-regulation | Scaffolds for self-regulation |
| | Intrinsic Task-specific Motivation | Alternative content for interest |
| | Extrinsic Incentives | Individualized rewards, repercussions |
| | Test Conditions | Alternative settings and conditions (time, sessions, location); review for racial, cultural, ethnic, & gender bias; differential item functioning; age appropriate content |

## Task Component: Multi-stage/Multi-part Tasks

| Task Component: Multi-Stage/Multi-Part Tasks | | |
|---|---|---|
| **Category of Processing** | **If this task *does not* intend to measure:** | **Then consider the following design options to minimize measurement of unintended constructs (construct irrelevant variance):** |
| **Perceptual** | Distinguishing Intra-task Navigation Actions | Consistent highlighting of all enabled intra-element navigation indication and control elements; clear relationship between mouse active behaviors and mouse keys, tab navigation, single switch navigation |
| | Identifying Stimulus and Response Components | Consistent, distinguishable stimulus and response designs; explicit labeling; physical and functional separation of stimulus and response areas, with supports for navigation between them by screen readers and single switch devices |
| **Linguistic** | Constructing Meaning From Text | Provide clear concise description of stages and parts; provide alternate representations of the stages and parts |
| **Cognitive** | Understanding Response Requirements | Provide explicit instruction to indicate required steps; use elements consistently; provide practice; be consistent with instructional practices |
| | Planning and Organizing Skills | Allow actions to be reversed; allow reverse navigation; allow modification of previous stages and parts; allow selective clearing of work |
| | Hypertext Syntax Fluency | Use alt-text; indicate link targets |
| **Motoric** | Navigation Abilities | Provide keyboard alternatives for intra-task navigation and mouse actions; provide sensible tabbed navigation; do not disable OS functions to allow assistive device and software compatibility |
| **Executive** | Goal Setting Ability | Explicit Instructions, goal-setting supports |
| | Goal Maintenance and Adjustment | Reminders, prompts |
| | Monitoring Progress | Extrinsic Scaffolds for monitoring |
| | Working Memory | Note-taking, mnemonic aids, text complement, locate prompts near relevant stimuli |
| **Affective** | Self-regulation | Scaffolds for self-regulation |
| | Intrinsic Task-Specific Motivation | Alternative content for interest |
| | Extrinsic Incentives | Individualized rewards, repercussions |
| | Test Conditions | Alternative settings and conditions (time, sessions, location); review for racial, cultural, ethnic, & gender bias; differential item functioning; age appropriate content |