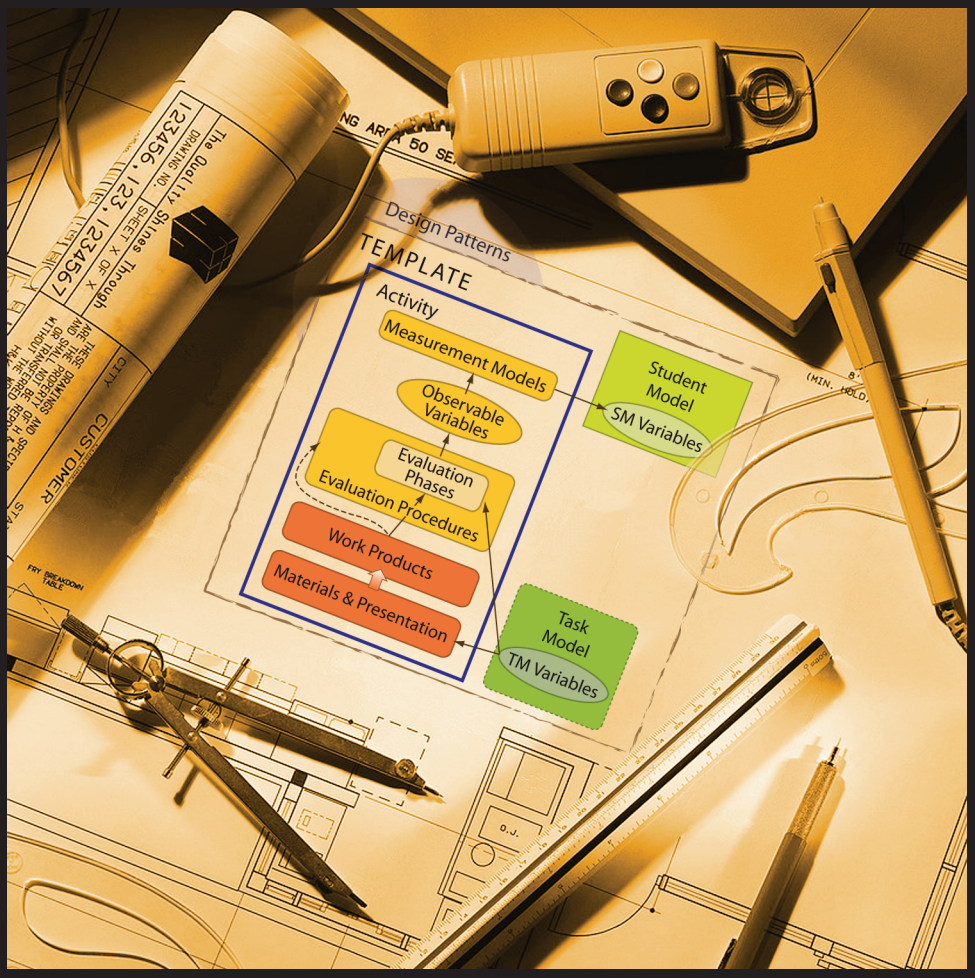


Assessment for Students with Disabilities
Technical Report 2 | January 2010

# A Design Pattern for a Spelling Assessment for Students with Disabilities

Project: Principled Science Assessment Designs for Students with Disabilities

**Ting Zhang and Robert J. Mislevy**, University of Maryland
**Geneva Haertel and Harold Javitz**, SRI International
**Elizabeth Murray and Jenna Gravel**, Center for Applied Special Technology
**Eric G. Hansen,** Educational Testing Service

Report Series Published by SRI International

**SRI International**
**Center for Technology in Learning**
**333 Ravenswood Avenue**
**Menlo Park, CA 94025-3493**
**650.859.2000**
**http://padi-se.sri.com**

**Technical Report Series Editors**
Alexis Mitman Colker, Ph.D., *Project Consultant*
Geneva D. Haertel, Ph.D., *Co-Principal Investigator*
Robert Mislevy, Ph.D., *Co-Principal Investigator*
Ron Fried, *Documentation Designer*

ASSESSMENT FOR STUDENTS WITH DISABILITIES
TECHNICAL REPORT 2

# A Design Pattern for a Spelling Assessment for Students with Disabilities

Prepared by:
Ting Zhang and Robert J. Mislevy
University of Maryland

Geneva Haertel and Harold Javitz
SRI International

Elizabeth Murray and Jenna Gravel
Center for Applied Special Technology

Eric G. Hansen
Educational Testing Service

*Acknowledgments*
This material is based on work supported by the Institute of Educational Sciences, Department of Education under Grant R324A070035 (Principled assessment designs for special education).

*Disclaimer*
Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of IES.

Contents

## *1.0 Introduction*

There has been an increasing call by educators to improve the accessibility of the educational evaluation and assessment for students with disabilities and English language learners. The Individuals with Disabilities Education Act of 1997 required states and school districts to include all students with disabilities in statewide assessment programs. These requirements were reinforced in the No Child Left Behind Act of 2001. This article uses the familiar context of a spelling bee as an example to illustrate how a principled approach (Mislevy, et al., 2003), based on the evidence-centered approach to assessment design and research on universal design for learning (Dolan, et al., 2005), can make assessment arguments more explicit and lead to the achievement of two highly-desired design goals: (1) increased accessibility of assessments to individuals with disabilities; and (2) more valid inferences about the targeted performances of these individuals.

The next section provides background on the psychology of spelling, which informs the construction of assessment arguments for assessing students' spelling capabilities. Particular attention is accorded to ways in which disabilities and cultural backgrounds can affect students' opportunities to demonstrate their spelling capabilities. The characteristics of a prototypical American spelling bee are then reviewed so as to establish a context for discussing how to assess spelling capabilities. Following that are sections that review assessment arguments, evidence-centered design (ECD), the ECD tool *design patterns*, and the principles of universal design for learning (UDL). These ideas are brought together in a section that presents and discusses a *design pattern* for developing assessments of spelling that embody the principals of universal design.

## *2.0 The Psychology of English Spelling*

Spelling is the encoding of linguistic forms into written form (Perfetti, 1997). In fact, spelling, and writing can be viewed as two types of knowledge representations: a spoken version of this information (phonetic information), and the visual representation on the printed page (graphemic information). Spelling concerns the representation of words with the necessary letters and diacritics present in an accepted standard order. Most English spellings attempt to approximate a transcription of the sounds of the language into alphabetic letters. This paper focuses on spelling in the English language, but it is important to note that some written languages employ the same modern basic Roman alphabet as English does (e.g., Spanish), some use different alphabets but handle vowels in the same way as English (e.g., Greek), and some use consonants but not vowels (e.g., Arabic). Still other languages, such as Chinese, use symbols that generally correspond to a spoken syllable with a basic meaning. A person whose first language is increasingly different from English in this sense has a greater challenge in becoming a proficient speller of English words.

Alphabets are based on correspondences between symbols and phonemes. In a perfectly phonemic system (e.g., Finnish), there is a one-to-one correspondence between symbols and sounds, and spelling is easy to learn. English, on the other hand, has been chaotic from the start (Venezky, 1976). A standardized spelling system was acquired only as late as the sixteenth century, and it was not phonemic due in part to a history of adoption of words from other languages, changing pronunciations over time, and idiosyncrasies of particular words. While rules do exist, there are more than enough exceptions to render it impossible to spell English words based solely on a small set of rules.

Barron and Strawson (1976) suggest that there are individual differences in spelling ability. People may use different strategies to process spelling. Some

people rely heavily in spelling-sound rules. Some may use word-specific associations by mapping the meaning or image of each word to its associated pronunciation. Barron (1980) points out that at least two strategies can be used with a printed word in order to obtain access to information stored in the internal lexicon. One is a phonological strategy that involves using a phonologic code generated by applying spelling–to–sound correspondence rules. The other is a visual–orthographic strategy that involves using a visual–orthographic code. During spelling, a phonological strategy can be used to produce a spelling of an item through the application of sound–to–spelling correspondence rules. A visual–orthographic strategy, on the other hand, can be used to generate the spelling of an item by retrieving information stored in the visual–orthographic entry in the lexicon. One can consider there to be two types of words: 1) regular words (e.g. GLOBE, CHURCH, SWEET) that conform to spelling rules and can be read and spelled by using either a phonological or visual–orthographic strategy; and 2) irregular words (e.g. SAID, BROAD, SWORD) that are exceptions to spelling rules and cannot be read or spelled successfully by using a phonological strategy solely. Barron (1980) indicates that application of sound–to–spelling rules to irregular words in spelling could result in the generation of a spelling word that does not correspond to a visual–orthographic entry in the lexicon. For instance, DEBT might be spelled as DET, YACHT as YOT and SWORD as SORD.

In one experiment, Barron (1979) asked children to read lists of exception words (e.g. PUT, GONE, SWORD), regular words (CUT, BONE, SWEET), and nonsense words (LUT, MONE, SWORP). Children who rely heavily on visual–orthographic associations correctly read more exception words than nonsense words. Children who rely heavily on the phonological rules were better at nonsense words. Children who rely on phonological rules were more likely to make sound–preserving errors when reading words (e.g. pronouncing the “h” in HONOR or the “w” in SWORD), while the children who use the visual–orthographic associations strategy alone tended to make meaning–preserving errors (e.g. pronouncing TWELVE as TWENTY or DONE as DID).

In summary, research shows that there is more than one way to approach spelling given the variety of English words (e.g. regular words, exceptional words, nonsense words). For a spelling assessment, the argument can be made that by selecting people who are good at using rules, the researchers also were selecting people who are better at using rules than visual–orthographic associations, and by selecting people who are poor at rules, the researchers were selecting people who may be better at specific associations than rules (Baron and Strawson, 1980).

## *3.0 Disabilities that Can Interact with Spelling*

Evidence about the capabilities that an assessment is intended to measure are confounded with other knowledge, skills, and abilities that may be required to perform the task. This can happen at the stages of apprehending the task and the information associated with it, interacting with the task during the processes of solving it, and producing a response. While the preceding section provided information about capabilities directly related to spelling English words, this section notes disabilities that, depending on the particulars of tasks and testing methods, can hinder a student's performance in assessments of spelling.

Table 1 presents a list of disabilities that are relevant to the design of assessments. Based on UDL research (CAST, 2008), they are grouped into perceptual, linguistic, motor, executive, and affect categories, The rightmost column notes how several of them can impact spelling assessment. These capabilities, which are not directly related to spelling but can affect performance in spelling assessments, will be discussed as Additional Knowledge, Skills, and Abilities (KSAs) in the section that presents the *design pattern* for creating spelling assessments.

**Table 1. UDL Categories**

| Processing Category | Common Associated Disabilities | Common Functional Limitations | Potential Impacts on Spelling |
|---|---|---|---|
| Perceptual | Blind | ▪ No functional vision (visual acuity 20/200 or poorer) | ▪ May have particular difficulty with words that are exceptions to spelling rules |
| | Low Vision | ▪ Limited functional vision (corrected visual acuity between 20/40 and 20/200) | ▪ As above |
| | 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 | ▪ May have difficulty learning to spell phonetically |
| 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) | ▪ Difficulty with spelling may be associated with difficulty decoding text and difficulty generating mental models<br>▪ May have difficulty with working memory |

| Processing Category | Common Associated Disabilities | Common Functional Limitations | Potential Impacts on Spelling |
|---|---|---|---|
| | English Language Learners | ▪ Limited English vocabulary<br>▪ Limited English syntax<br>▪ English orthography and decoding skills reduced<br>▪ Reduced comprehension in English text or oral presentation<br>▪ Background knowledge deficits (cultural and linguistic) | ▪ May have particular difficulty with words that are exceptions to spelling rules, as English has many irregularities |
| Motor | Physical Disability | ▪ Difficulty with speech<br>▪ Difficulty with movement | ▪ Communication may be unintelligible or student may be nonverbal<br>▪ May have difficulty with written spelling |
| Executive | ADHD<br>working memory | ▪ Difficulty attending<br>▪ Difficulty remembering sequence of letters | ▪ May be difficult to maintain level of concentration<br>▪ May have difficulty with working memory |
| Affect | | ▪ Uncomfortable performing in public | ▪ Anxiety may interfere with concentration, memory, etc. |

## *4.0 Spelling Bee*

We will take the familiar context of a spelling bee to ground our discussion of merged assessment design and UDL principles. We will discuss design choices that a test developer could make so that the essential assessment of spelling capabilities could be accessible to a wider range of students than the standard form alone. As a starting point, this section describes the standard form of spelling bee contests.

In the United States, capabilities in spelling are largely a product of school learning. From a very young age, students are encouraged to develop their spelling skills and sometimes take part in a form of schoolroom competition called the "spelling bee." Table 2 describes the standard form and rules of a spelling bee. Educators, teachers, and parents believe that by helping students to develop skill in spelling, they can increase students' vocabulary, learning concepts, and development of correct English usage, all of which will help students more broadly. Informal spelling bees are sometimes held in classrooms or clubs, even in early elementary grades. The E. W. Scripps Company sponsors the National Spelling Bee, an organized competition open to students under 16 years of age.

**Table 2. Rules of the National Spelling Bee in 2009 (http://www.spellingbee.com)**

| | |
|---|---|
| **Format** | The spelling bee is conducted in rounds. Each speller remaining in the spelling bee at the start of a round spells one word in each round—except in the case of a written, multiple choice, or online test. The spelling bee may be conducted orally or in writing or in a manner that is a combination of the two; however, if the spelling bee officials specify an oral format, the speller may not demand a written format. |
| **Word list** | Local spelling bee officials are responsible for selecting the word lists for use at each local spelling bee. Many local spelling bee officials use word lists generated by the Scripps National Spelling Bee. These lists include many words that appear in the current edition of *Spell It!* as well as some "end-of-bee" words. All words on Scripps National Spelling Bee word lists are entries in *Webster's Third New International Dictionary* and its addenda section, copyright 2002, Merriam-Webster, the official dictionary of the Scripps National Spelling Bee. |

| | |
|---|---|
| **Special needs** | Spelling bee officials will strive to provide accommodation for spellers who have physical challenges. All requests for spelling bee officials to accommodate special needs involving sight, hearing, speech, or movement should be directed to spelling bee officials well in advance of the spelling bee date. The judges have discretionary power to amend oral and/or written spelling requirements on a case–by–case basis for spellers with diagnosed medical conditions involving sight, hearing, speech, or movement. |
| **Pronouncer's role** | The pronouncer strives to pronounce words according to the diacritical markings in Scripps National Spelling Bee word lists and *Webster's Third New International Dictionary* and its addenda section, copyright 2002, Merriam-Webster.<br><br>Homonyms: If a word has one or more homonyms, the pronouncer indicates which word is to be spelled by defining the word.<br><br>Speller's requests: The pronouncer responds to the speller's requests for a definition, sentence, part of speech, language(s) of origin, and alternate pronunciation(s). When presented with requests for alternate pronunciations, the pronouncer or an aide to the pronouncer checks for alternate pronunciations in either *Webster's Third New International Dictionary* and its addenda section, copyright 2002, Merriam-Webster or *Merriam-Webster's Collegiate Dictionary,* eleventh edition. The pronouncer does not entertain root word questions, requests for alternate definitions, or requests for markedly slower pronunciation.<br><br>Pronouncer's sense of helpfulness: The pronouncer may offer word information—without the speller having requested the information—if the pronouncer senses that the information is helpful and the information is presented in the entry for the word in a 2008 Scripps National Spelling Bee word list or *Webster's Third New International Dictionary* and its addenda section, copyright 2002, Merriam-Webster. |
| **Judges' role** | The judges uphold the rules and determine whether or not words are spelled correctly. They also render final decisions on appeals in accordance with Rule 11. They are in complete control of the competition, and their decision is final on all questions. |
| **Speller's role** | The speller makes an effort to face the judges and pronounce the word for the judges before spelling it and after spelling it. The speller *while facing the judges* makes an effort to utter each letter distinctly and with sufficient volume to be understood by the judges. The speller may ask the pronouncer to say the word again, define it, use it in a sentence, provide the part of speech, provide the language(s) of origin, and/or provide an alternate pronunciation or pronunciations.<br><br>Misunderstandings: The speller is responsible for any misunderstanding of the word unless (1) the pronouncer never provided a correct pronunciation; (2) the pronouncer provided incorrect information regarding the definition, part of speech, or language of origin; or (3) the speller correctly spelled a homonym of the word and the pronouncer failed to either offer a definition or distinguish the homonyms. |
| **Misspelling** | Upon incorrectly spelling a word, the speller immediately drops out of the competition. |

## *5.0 ECD, Assessment Arguments, and Design Patterns*

This section provides a brief overview of the evidence-centered assessment design framework. The structure of assessment arguments is laid out with particular attention to the places where disabilities, accommodations, and UDL principles impact the validity of inferences. *Design patterns*, a tool that helps test developers make design choices in task construction, are then described in general as a prelude to the specific *design pattern* for developing spelling assessments that accord with both the principles of validity and UDL.

### *5.1 Evidence-Centered Design (ECD)*

Evidence-centered assessment design (ECD; Mislevy, Steinberg, & Almond, 2003) is a framework that makes explicit, and provides tools for, building assessment arguments (Mislevy & Riconscente, 2005). Two complementary ideas organize the effort. The first is an overarching conception of assessment as an argument from imperfect evidence. It aims to make explicit the claims (the inferences that one intends to make based on scores) and the nature of the evidence that supports those claims (Hansen & Mislevy, 2008). The second idea is distinguishing layers at which activities and structures appear in the assessment enterprise. A number of representational forms and tools have been developed to support work at various layers of work.

Both of these ideas are central to the present topic. By making the underlying evidentiary argument of an assessment explicit, the framework makes operational elements more amenable to examination, sharing, and refinement. In particular, the argument framework can be used to examine how validity is affected by accessibility features provided to students with disabilities and English language learners (Hansen & Mislevy, 2006; Hansen, Mislevy, Steinberg, Lee, & Forer, 2005). The representational form of design patterns, developed to support test developers in creating arguments for families of assessment tasks, is extended in this presentation to help test developers incorporate UDL and accommodations into tasks in concert with underlying validity arguments.

### *5.2 Assessment Arguments*

> "Validity is associated with the interpretation assigned to test scores rather than with the test scores or the test. The interpretation involves an argument leading from the scores to score–based statements or decisions, and the validity of an interpretation depends on the plausibility of this interpretive argument. The interpretive arguments associated with most test–score interpretations involve multiple inferences and assumptions. An explicit recognition of the inferences and assumptions in the interpretive argument makes it possible to identify the kinds of evidence needed to evaluate the argument. Evidence for the inferences and assumptions in the argument supports the interpretation, and evidence against any part of the argument casts doubt on the interpretation." (Kane, 1992, p. 527)

If, as Kane asserts, "most test-score interpretations involve multiple inferences and assumptions," then there are an especially large number and variety of inferences and assumptions that need to be made explicit when considering tests administered to subpopulations such as individuals with disabilities and English language learners. The accessibility extensions to ECD seek to make more visible the chains of inference and their associated assumptions. The ECD accessibility work described in this paper attempts to apply principles of evidentiary reasoning to handle the complexities of the validity argument associated with accessibility features. The key idea is to lay out evidentiary structures that capture key salient aspects of the validity argument and show the roles of assumptions about students' capabilities and task requirements (Hansen & Mislevy, 2008).

An assessment argument can be summarized as comprising: (a) a claim about a person possessing at a given level a certain targeted proficiency, (b) the data (e.g., scores) that would likely result from the specified task situation if the person possessed, at a certain level, the targeted proficiency, as well as the salient features of the task, (c) the warrant (or rationale, based on theory and experience) that tells why the person's level in the targeted proficiency

would lead to occurrence of the data, and (d) “alternative explanations” for the person’s high or low scores (i.e., explanations other than the person’s level in the targeted proficiency). A graphic illustration of an assessment argument in the context of spelling will be presented shortly.

The existence of alternative explanations that are both significant and credible might indicate that validity is threatened or being compromised (Messick, 1989). Much of the analysis that is the focus of this presentation has to do with these alternative explanations, i.e., factors that can hinder an assessment from yielding valid inferences arising from sources such as language limitations or physical or cognitive disabilities that are not the target of inference in the assessment. When such alternative explanations are recognized at the earliest stages of test design, then later reworking and retrofitting can be avoided.

An example of an alternative explanation for “poor” performance by an individual with a disability is that the individual is not able to receive the test content because there is a mismatch between the test format (e.g., visually displayed text) and the individual’s disability (e.g., blindness). An example of an alternative explanation for “good” performance would be that the accommodation eliminates or significantly reduces demand for some aspect of the targeted proficiency. The ECD accessibility effort has focused on building argument structures that might help anticipate and address key details of these alternative explanations particularly as they relate to test takers with disabilities. Once recurring kinds of threats have been identified and ways of mitigating their effects by modifying tasks in valid ways have been identified (see, for example, Cahalan-Laitusis & Cook, 2007, and CAST, 2008), these options can be built into *design patterns* centered around content or skills to provide a design space for test developers to think through how to apply them for the assessment at hand.

This paper illustrates this process in the context of a spelling assessment,

originally in the format of a traditional spelling bee. A spelling bee format can be viewed as a way to evoke evidence about students' accuracy in the production of letter spellings of words given their pronunciation. The claim of interest is that an examinee has targeted proficiency in spelling, observing data such as spelling out words given their pronunciations. The basic argument for a single item is shown by a Toulmin diagram (Toulmin, 1958) in Figure 1[1]. Figure 1 illustrates the argument associated with an incorrect spelling, with limited hearing ability as one possible alternative explanation for an incorrect response. A similar structure would depict the argument associated with a correct spelling. An alternative explanation for a correct response could be the examinee received an inappropriate accommodation (e.g., a visual hint). The course of the paper will use this framework to explore additional validity threats, how they can be addressed with options from UDL and accommodations research, and how these options impact the validity argument.

[1] Spelling bees, like most assessments, consist of multiple items. The evidence, in the form of correct and incorrect answers, may be synthesized by means of a total score or an item response theory model into a quantitative measure of the degree of proficiency the examinee has.

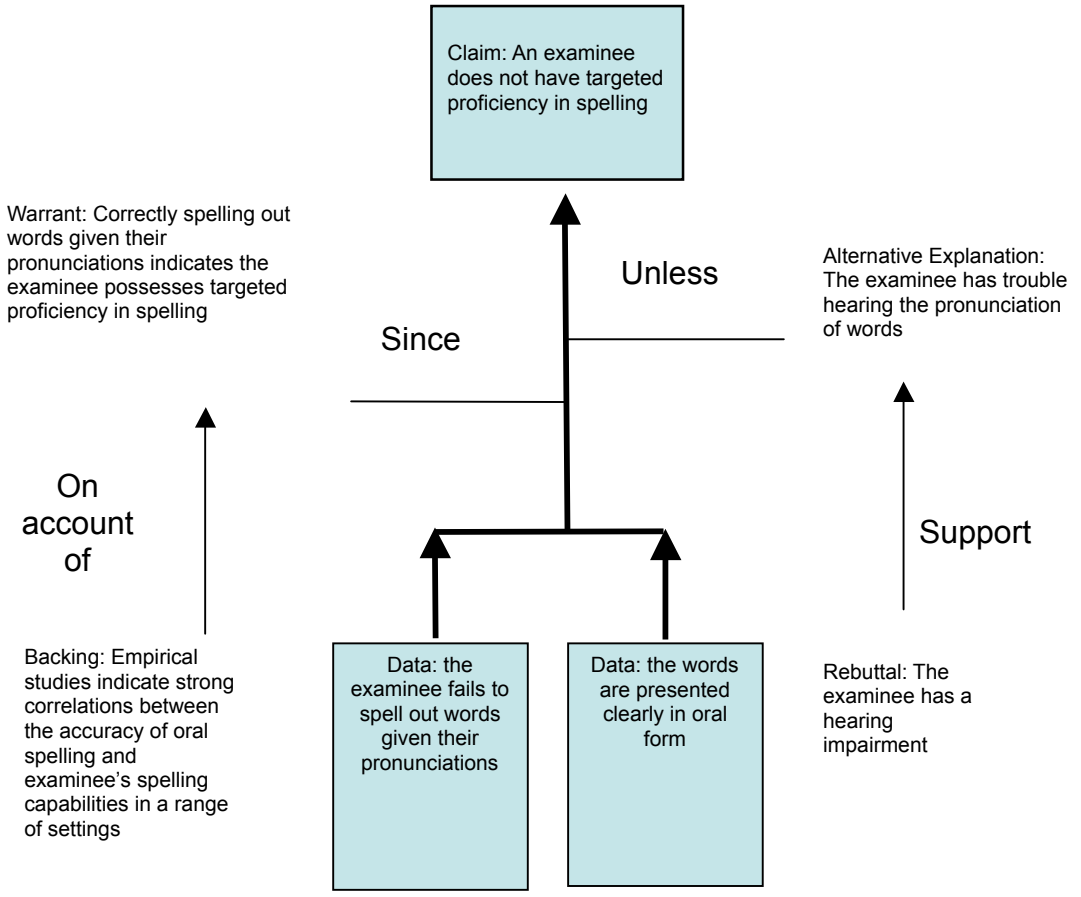


**Figure 1: Toulmin's (1958) structure for arguments for an assessment in spelling.** Inferential Reasoning flows from *data* (D) to *claim* (C) by justification of a *warrant* (W), which in turn is supported by *backing* (B). The inference may need to be qualified by *alternative explanations* (A), which may have *rebuttal evidence* (R) to support them (Mislevy, 2003).

### *5.3 Design Patterns*

The term *design pattern* was coined in the mid-1970's by Christopher Alexander, an architect, who abstracted common *design patterns* in architecture and formalized a way of describing the patterns in a "pattern language." A *design pattern* addresses both a problem that occurs repeatedly in our environment and the core of the solution to that problem—but at a level of generality that the solution can be applied many times without ever being the same in its particulars. *Design patterns* for creating assessment tasks have been developed through the National Science Foundation supported project Principled Assessment Design for Inquiry (PADI; Mislevy et al., 2003). They lay out a "design space" of options test developers can consider when

writing tasks that assess whatever knowledge or skill the *design pattern* is meant to address. With UDL insights and options incorporated, *design patterns* can help to improve the accessibility of task design goals by providing a way of representing designs that are sensitive to the issues of both validity and accessibility for test takers with disabilities (Hansen & Mislevy, 2008).

In the PADI project, *design patterns* lie in the layer in the assessment system called Domain Modeling. The previous layer, Domain Analysis, is the activity of identifying the knowledge and skills in a particular subject area to be assessed. The PADI project focused on science inquiry, but the approach can be used with any subject domain. UDL and the accommodations research noted above can be used across many domains, so the discussion of knowledge and skill requirements in light of ask features and work products applies more broadly across domains and levels. Domain Modeling specifies the relationships among the knowledge and skills in the area to be assessed. *Design patterns* are a Domain Modeling tool. A *design pattern* specifies, in non-technical terms, the evidence–centered assessment argument and bridges the content expertise and measurement expertise needed to create an operational assessment (Mislevy et al., 2003). The more technical layers of the assessment system, where the details of psychometric models, scoring rubrics or algorithms, presentation of materials, interactivity requirements, and so on, are specified, will not be addressed in this paper. This technical work can be carried out in accordance with one or more *design patterns* that lay out the substantive argument of the planned assessment (Mislevy et al., 2003; Hansen & Mislevy, 2008), and working from the *design pattern* helps ensure that the technical elements will be coordinated with each other in the service of the underlying argument.

Table 3 summarizes the key attributes of an assessment *design pattern* (see Mislevy et al., 2003 for the full list of attributes). Specifically, *design pattern* construction focuses on the identification of task requirements that indicate

proficiency on intended test constructs (Focal KSAs) and those that also contribute variance to student scores but may or may not be relevant to the construct being measured (Additional KSAs). Five key attributes, namely Focal KSAs, Additional KSAs, Characteristic Task Features, Variable Task Features, and Work Products are particularly relevant to consideration of students in connection with disabilities.

**Table 3. Attributes of an Assessment Design Pattern and Their Connection to the Assessment Argument**

| DP Attribute | Definition | Argument Element |
|---|---|---|
| Rationale | The underlying warrant that justifies the connection between the targeted inferences and the kinds of tasks and evidence that support them | Warrant |
| Focal KSAs | The primary knowledge, skills, and abilities that one wants to know about students | Claim |
| Additional KSAs | Other knowledge, skills, and abilities that may be required. | Claim if construct relevant; Alternative explanation if not |
| Characteristic task features | Salient features of tasks that can elicit evidence about the focal KSAs | Data regarding task – Needed for construct representation |
| Variable task features | Features of tasks that can be varied to shift the difficulty, focus, or demands of tasks. | Data regarding task – Design to avoid construct irrelevant demands |
| Potential work products | Student responses or performances that can hold clues about the focal KSAs | Data regarding performance – Design to avoid construct irrelevant demands |
| Potential observations | Features of work products that constitute evidence about the focal KSAs | Data regarding performance – Design to avoid construct irrelevant demands |

- Focal KSAs consist of the primary knowledge/skills/attributes of students that are addressed by assessment. Comparability of scores between individuals with and without disabilities is important, which suggests that

one should seek evidence about the same set of Focal KSAs regardless of whether the test taker has a disability or not.

- Additional KSAs consist of the other knowledge/skill/attributes that may be required in a task. A *design pattern* lists many that the task designer should consider whether to require in a task or not (by manipulating Variable Task Features and Work Product requirements, as noted below). For tests of academic subjects, the abilities to "see" and "hear" are typically Additional KSAs. On the other hand, for assessment of sight and hearing, respectively, sight and hearing will be defined as Focal KSAs. Notice that there are many disabilities that involve impairments of sight, hearing, or both (e.g., blind, low vision, color-blind, deaf, hard to hear, deaf-blind). Deficits in such Additional KSAs that are not the target of assessment can cause unduly low scores among test takers with disabilities.
- Characteristic Task Features must be present in a situation in order to evoke the desired evidence about the Focal KSAs. In the case of spelling proficiency, a Characteristic Feature of all tasks is the indication of a word which must be spelled.
- Variable Task Features are features that can be varied to shift the difficulty or focus of tasks. Variable Features have a particularly significant role with respect to test takers with disabilities and other sub-populations (e.g., speakers of minority languages). Much of our attention will be on manipulating Variable Features to reduce or eliminate demands for Additional KSAs in which there may be a deficit, while making sure (to the extent possible) that demands for Focal KSAs have not been changed. We will see that there are several ways in which spelling capabilities can be assessed by varying features of tasks, all the while maintaining the Characteristic Feature that is needed to obtain evidence about students' spelling capabilities.
- Work Products are the ways that student performance is manifest. This includes responses to multiple–choice tasks by paper and pencil, in a computer format, and pointed or spoken indications. Essays, again

written or typed, are potential Work Products; so are physical performances, drawn diagrams, completed tables, graphs, or other representational forms. Of particular importance from the perspective of UDL is that some students may not have they capabilities to produce responses in a given form even if they can interact meaningful with the substance of the task.

Regarding task features, the goal is to ensure that one assesses the targeted proficiency. A task feature that is useful across diverse students for evoking evidence as to whether the student has the Focal KSAs is classified as a Characteristic Feature. To assess capabilities in spelling, for example, it is necessary that in some way a student must be presented a word to be spelled and in some way must indicate the letters that spell it. There are many ways this can be accomplished beyond the traditional spelling bee format, and the particular combination of task features of each way brings its own requirements for apprehending the word, doing the processing of spelling, and making a response. Evidence that a task feature results in invalidity for some students and validity for others suggests that the feature (e.g., use of visual display of items in regular font) should be a Variable Feature. Such a feature could be varied for different students taking the same assessment, if the students have different profiles as to which Additional KSAs they are proficient in or have deficiencies. Given the wish to make inferences about the student's targeted proficiency, a task *should* have requirements for the Focal KSAs as evoked by Characteristic Task Features and *should not* have requirements for Additional KSAs that the student does not possess and are not meant to be assessed. That is, the task is configured in such a way as to eliminate Additional KSAs as explanations for poor performance, thereby allowing the student to demonstrate what he or she knows and can do. The section following the discussion of UDL principles will discuss the relationship of *design patterns* and realized assessment tasks from the perspective of construct validity.

Next, we consider how UDL research into task features that are available to circumvent or support deficiencies on Additional KSAs can be incorporated into design patterns.

### *5.4 Universal Design*

Originally formulated by Ron Mace at North Carolina State University (Mace, Hardie, & Place, 1996), universal design supports the creation of accessible structures by addressing the mobility and communication needs of individuals with disabilities at the design stage, a practice that has spread to areas such as civic engineering and commercial product design. Designs that from the start increase accessibility for individuals with disabilities tend to yield benefits that make everyone's experience better. The development of captioning on television provides a good example of universal design in practice. When captioning first became available, it was intended for people with hearing impairments; it now benefits not only those with hearing impairments, but also exercisers in health clubs, travelers in airports, and individuals working on their language skills. Universal design does not advocate for "one–size–fits–all" solutions, however. While one approach may work in specific instances, more common are solutions that are inherently flexible and thus provide individuals with choice in how they are used.

Well over a decade ago, researchers began to bring the concept of universal design to education, focusing not on physical objects but on curriculum (Rose & Meyer, 2000, 2002). Since then, Universal Design for Learning (UDL) has been created to inform the development of four components of education: (1) goals and standards, (2) methods, (3) materials, and (4) assessments. The principles of UDL 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; Rose, Meyer, & Hitchcock, 2005). In recent years, especially as policies have stressed the participation of populations with disabilities, varying cultural experiences, and diverse linguistic backgrounds, more flexible and universally accurate assessments have been required. To meet

these requirements, test development procedures recently have evolved to incorporate the concept of universal design (Dolan & Hall, 2001, 2007; Dolan & Rose, 2000; Ketterlin-Geller, 2005; Thompson, Johnstone, & Thurlow, 2002). UDL principles encourage a test development process that facilitates participation of the widest possible range of students and results in valid inferences about performance for all students who participate in the assessment. The "universally designed test" should consider the needs of these students from the earliest stages of test development and should involve choices in test specification, item development, test construction, and test administration that facilitate the most inclusive student participation possible, while still preserving the validity of the construct being measured.

A key challenge of applying UDL to assessment is to ensure that the designs are actually inclusive—that the needs of a full range of participating students have been addressed. One important recent advance in this regard has been the dissemination of both a framework and guidelines for UDL that articulate the range of options that must be provided in order to ensure applicability for students with the full range of abilities and disabilities. That framework (based in cognitive neuroscience) and the guidelines, (based on decades of empirical research with students who have disabilities) are key foundations that are now available (CAST, 2008).

### *5.5 Design Patterns, Construct Validity, and Specific Assessment Contexts*

Construct validity is the sine qua non of assessment properties: To what degree do the evidence and rationale for the data gathered in an assessment support the inferences or decisions that a user wants to make? In the literature on accommodated assessment, the question typically centers on whether a given alteration of a task "changes the construct" (*Standards for Educational and Psychological Testing*, AERA, APA, NCME, 1985. p. 78). Specifically, if an alteration changes the construct, then construct validity has been violated. If the alteration does not change the construct, then construct validity has not been violated.

Yet for assessment designers and developers as well as some other audiences, there is often a need to reason more deeply about the relationships between construct validity and task design. We would argue that it is important to specify more carefully what knowledge and skills, and at what levels and natures, are the essence of the intended construct to assess, and what are not. This cannot be determined simply by examining the tasks on a test, because all of the knowledge, skills, and abilities needed to do well on a test are jointly required. In a given testing application, some of these KSAs will be relevant for the inference at hand and others will not (Phillips, 1994); the target examinee population may vary on some of them and others not. It can even be the case that a given alteration on a test will introduce extraneous score variation in one application and thus reduce validity, but reduce extraneous variation in a different application of the same test, and increase validity there. It is only by knowing the purpose of a test and the intended examinee population that one can answer how a given change will impact the evidentiary value of data for the construct meant to be assessed. A series of decisions needs to be made in the course of developing a specific test for a specific purpose and testing population to reason through the question of whether a given alteration "changes the construct."

A *design pattern* helps by laying out choices to be made as appropriate to specific testing applications. It is the specific test and context to which the property of construct validity applies, and the determination of which potential sources of variance among examinees' test scores would be construct relevant or construct irrelevant.

We have discussed above how it is important for tasks intended to assess a Focal KSA to exhibit in some form the Characteristic Features denoted in the *design pattern*, and that by manipulating Variable Task Features a test developer can increase, decrease, circumvent, or support particular Additional KSAs. A key point is that exactly which Additional KSAs, at which levels, will be construct-relevant to a task in a given context is an application–specific

decision—that is, a test–for–purpose–with–population decision. The creator of the *design pattern* does not know what this decision will be because it can be validly and appropriately different for different applications.

Table 4 distinguishes what can be known at the time of creating a *design pattern* for any number of tests that in some way address the Focal KSAs, and what must be determined at the time of specifying the application to a particular test. Note that the Focal and Additional KSAs describe KSAs in the *design space* while "construct relevant" and "construct irrelevant" describe KSAs in the *application space*.

**Table 4. Focal and Additional KSAs in Design Space and Application Space**

| | **Application Space Descriptors (for thinking about KSAs for a particular test and its purpose and the intended examinee population)** | |
|---|---|---|
| **Design Space Descriptors** (for thinking about KSAs in the design pattern stage) | **Construct Relevant** | **Construct Irrelevant** |
| **Focal KSA.** A design pattern is meant to support designing tasks and assessments that assess the Focal KSAs. | (1) KSAs (which were Focal KSAs at the DP vantage point) from the design pattern, at the right level and focus for the application. | (2) KSAs which were Focal KSAs at the DP vantage point, but too hard, too easy, or off focus for the intended application. |
| **Additional KSA**. Additional KSAs may be required at the designer's discretion | (3) The designer deems certain KSAs (which were Additional KSAs at the DP vantage point) are appropriately part of the intended construct to assess. | (4) KSAs (which were Additional KSAs at the DP vantage point) that could be, and some of which will be, required to apprehend, build on, interact with, or respond to an implemented task, yet are not part of the intended construct to assess. |

Both Cell 1 and Cell 2 concern what a particular test application needs to require for KSAs that are listed in the Focal KSA attribute of a *design pattern*. A test needs to have some requirement for Focal KSAs (and perhaps some Additional KSAs as well) in order to be valid. Creating tasks that elicit these KSAs will contribute construct relevant variance in examinees' scores as long as it is done correctly.

Cell 1 addresses an implemented test application's requirement for Focal KSAs listed in the *design pattern*, in a task meant to assess the capabilities the *design pattern* is meant to support, at a level that suits the application's intended use and examinee population. This is the essence of construct relevant variance in a test: Having the intended capability makes it more likely an examinee will perform well, and lacking it makes it more likely that he or she will not perform as well.

Cell 2 concerns requirements for the Focal KSAs described in the *design pattern*, but in flawed test construction the demand for the KSAs is not the right level. For example, the word list for an in–class spelling bee for a second grade class might contain words that are much too hard for the students. The KSA of spelling English words is appropriate, but it has not been implemented appropriately for the intended use.

Cells 3 and 4 concern a particular test application's requirement for KSAs that are listed in the Additional KSA attribute of a *design pattern*. These demands may or may not contribute to construct relevant variance in that application, depending on the purpose and examinee population.

In Cell 3, the designer deems certain Additional KSAs are appropriately part of the intended construct to assess. For example, it may be decided that working memory capability needed to spell words without writing them along the way is appropriate for an in–class spelling bee because it is intended to give the students feedback on how well they would do in the upcoming spelling bee competition that does not allow writing while spelling. More generally, prerequisite knowledge is often considered "fair game" in assessing school skills. For example, on a test of standards at a given grade, including requirements for knowledge from standards for earlier grades are often considered appropriate and construct–relevant reasons for poor performance, and are therefore not scaffolded (i.e., the Variable Task Feature "scaffolding" has been set to none for these Additional KSAs).

Cell 4 concerns Additional KSAs that are required to apprehend, interact with, or respond to an implemented task but are not part of the intended construct to assess. For example, the standard spelling bee requires a spoken response, and the KSA of speaking is almost certainly not of the essence of the capability at issue. It is a potential explanation of poor performance. Allowing for typed, written, or pointed–to spelling of words as a task feature is a UDL approach to mitigating this problem. In general, requirements in a task for physical and cognitive KSAs that are not construct relevant can lead to poor performance and mask the KSAs that are the intent of assessment (Focal KSAs, plus Additional KSAs that are construct relevant in the application at hand). They are thus potentially sources of construct irrelevant variation.

Note that the Additional KSA of being able to say letters aloud—i.e., to produce a Work Product in the form of a spoken sequence of letters—is not universally construct relevant or construct irrelevant in and of itself, but only in light of the purpose of a given test application. The *design pattern* cannot provide the answer, but it can alert the test developer to the question and offer suggestions for UDL and accommodation strategies when the Additional KSA is deemed construct irrelevant for the application and there are examinees in the test population who may not have the Additional KSA at the required levels.

The phrase “potentially construct irrelevant sources of variation” highlights the role of the intended examinee population in determining whether a requirement for a construct–irrelevant Additional KSA contributes to invalid inferences in a given application. Being able to speak letters in a spelling bee is a construct irrelevant requirement, but if it is known a priori that everyone in the class is able to spell words aloud, this will not be a source of poor performance for this population. But it might be for a different class that has a student who has difficulty responding in this manner. An alternative way of responding in that class, perhaps used only by that student, would be

necessary in order to remove a construct irrelevant source of variance in the second class.

## *6.0 A Design Pattern for Creating Spelling Assessments with UDL Infused*

This section presents a PADI *design pattern* built up from the familiar context of a spelling bee, expanded to reveal the considerations of assessment design and UDL. Specifically, we address issues associated with how to identify Focal KSAs, Additional KSAs, and task features, and how UDL research into task features that are available to circumvent or support deficiencies on Additional KSAs when they are construct irrelevant can be incorporated into *design patterns*.

The Appendix illustrates a *design pattern* for an expanded design space to create spelling bee contests that are more accessible to a wider range of students. The assessment design space reflects an awareness of spellers who have special needs involving sight, hearing, speech, or movement as well as individuals who are nondisabled.

### *6.1 Focal KSAs and Characteristic Task Features*

To develop an assessment from *design patterns*, we begin by defining targeted proficiency as consisting of one or more Focal KSAs, which are central to the "claim" that one wishes to make about what a person know or can do and may be thought of as the construct that is meant to be assessed. Once a Focal KSA has been defined as a measurement target, it should be held constant across diverse test takers across various forms of a task. In our spelling assessment, the Focal KSA that we are seeking is the speller's ability to encode a non–alphabetic representation of a word to its oral or written alphabetic form. This then serves as the target proficiency/construct knowledge that we want to make inferences about regardless of whether the speller has a disability or not.

After defining the construct of assessment through Focal KSAs (and any Additional KSAs that may need to be incorporated into the construct for the particular assessment application), we need to develop tasks that incorporate Characteristic Task Features across diverse students in order to evoke evidence as to whether the student has the Focal KSAs. Before or during a spelling bee contest, there are several Characteristic Task Features that we can present to spellers with or without disabilities. For example, a list of the to–be–tested words is provided to all students in a form they can apprehend. It contains the same amount of information to every speller, so we can ensure that all spellers can have equal opportunity of learning before the contest. Another important Characteristic Task Feature is English words are presented to spellers in the contest in a way that their spelling is not communicated. We also want to make sure that this presentation of the word to the student is clear and unambiguous.

### *6.2 Additional KSAs and Variable Task Features*

Additional KSAs are knowledge, skills and abilities that may or may not be required in assessment tasks that elicit evidence about the Focal KSAs that a *design pattern* addresses. As discussed above, sometimes Additional KSAs will be deemed part of the mix of KSAs that are the measurement construct in a particular application. Many times they are not construct relevant in this way but are skills which, depending on how tasks are constructed, may or may not be required to apprehend, interact with, or respond to the task. The *design pattern* lists Additional KSAs to prompt a task designer to think through which ones, and at which levels, may be construct relevant in his or her application, which ones are not construct relevant but might be required, and how to support or avoid ones for particular students for whom they would pose unintended difficulties.

An Additional KSA corresponds to the “alternative explanation” in the assessment argument if the Additional KSA is irrelevant to the intended construct of a given assessment application. In most assessment applications, the mental and physical Additional KSAs needed to apprehend,

interact with, and respond to tasks will turn out to be construct irrelevant. A UDL–infused *design pattern* highlights these kinds of Additional KSAs, as well as Additional KSAs such as content knowledge, prerequisite knowledge, and familiarity with representational forms that will be deemed construct relevant in some applications but not in others. Whether construct–irrelevant Additional KSAs in a given application are to be supported or not (e.g., glossary, background facts, equation list) is a decision to be made in that application, depending in part on resources, testing purposes, and test population (Phillips, 1994) either by the assessment design team, either at the level of the testing program, or at the level of the individual task if that is appropriate in the testing program.

The following Additional KSAs in a spelling bee *design pattern* are likely to be construct relevant in most spelling contest applications:

- Knowledge of root words and etymologies
- Knowledge of correct English usages
- Knowledge of foreign languages from which English draws
- Knowledge of English phonics, conventions, and rules
- Knowledge of English vocabularies

All of these Additional KSAs are knowledge that research and experience has shown to be important to develop a high degree of proficiency in spelling. Even though these Additional KSAs are not the primary targeted proficiency that we aim to measure, they are directly associated with the Focal KSAs. Deficits in such Additional KSAs can cause unduly low scores among participants regardless of whether they have disabilities. There is no Variable Task Feature provided in the *design pattern* to reduce or eliminate the demands for these almost certainly construct–relevant Additional KSAs.

Another set of Additional KSAs, on the other hand, will generally be deemed construct-irrelevant but may be involved in tasks generated under this *design pattern*. The task author can consider offering support, presenting material, or

specifying work products that reduce or avoid requirements for these Additional KSAs for some or all students, either through accommodation forms of a tasks or UDL principles. Many of these Additional KSAs are linked to Variable Task Features or Potential Work Products for suggestions on how to do this.

The PADI project team reviewed relevant background information on ECD and UDL to determine the intersection between UDL principles and PADI *design patterns*. Based on this analysis, six UDL categories are now used to categorize types of likely construct–irrelevant Additional KSAs that may be required in spelling tasks and can influence student performance.  Definitions of UDL categories are provided in Table 5.

**Table 5. Definitions of Six UDL Categories by Principle**

| Principle 1 – Multiple Means of Representation | |
|---|---|
| **UDL Category Name** | **Definition of UDL Category** |
| **Perceptual** | To reduce barriers to accurate assessment it is important to ensure that key information is equally perceptible to all students by: 1) providing the same information through different sensory modalities (e.g. through vision, or hearing, or touch); 2) providing information in a format that will allow for adjustability by the user (e.g., text that can be enlarged, sounds that can be amplified). Such multiple representations not only ensure that information is accessible to students with particular sensory and perceptual disabilities, but also easier to access for many others. |
| **Language and Symbols** | Students vary in their facility with different forms of representation – both linguistic and non-linguistic. As a result, inequalities arise when information is presented to all students through a single form of representation. An important assessment strategy is to ensure that alternative representations are provided not only for accessibility but also for clarity and comprehensibility for all students. |
| **Cognitive** | Individuals differ greatly in their information processing skills and in their access to prior knowledge by which they can assimilate new information. Proper design and presentation of information can help to ensure that assessments accurately measure student knowledge. |
| **Principle II –Multiple Means of Action and Expression** | |
| **Skill and Fluency** | It is important to provide materials with which all students can interact, yet there is no medium of expression that is equally suited for all students or for all kinds of communication. Assessments should be designed so that the medium of response to questions does not interfere with an accurate demonstration of knowledge. |
| **Executive** | Executive functions include setting and maintaining goals, developing plans of action, managing information, and monitoring progress toward a goal, using working memory to assist in the process. Although many assessments may want to include measures of these skills, it is important to decide if any are essential to the focal KSA. |
| **Principle III – Multiple Means of Engagement** | |
| **Affect** | Students differ markedly in the ways in which they can be engaged or motivated. In assessment, it is important that students be willing to engage in the task. The level of challenge, students' interest in a topic, the provision of choice, and students' level of frustration can all impact student motivation. To prevent a lack of engagement from influencing the construct being measured, it is important to provide multiple ways in which to engage all students. |

In our *design pattern* for constructing spelling bee assessments, spellers with disabilities involving sight, hearing, speech, or movement are of concerns. Following are likely construct-irrelevant Additional KSAs that are selected from the six UDL categories infused in a *design pattern*:

- Sight
- Hearing
- Speech
- Movement

Specifically, “Sight” and “Hearing” were chosen under the UDL category of “Perceptual.” “Speech” and “Movement” were selected from the UDL category called “Skill and Fluency.”

Much of our attention in this report is on manipulating Variable Task Features to reduce or eliminate demands for Additional KSAs for which students may have a deficit, while making sure (to the extent possible) that demands for construct relevant KSAs have not been changed. In a *design pattern*, if assessment designers identify the possibility that some students may lack abilities that are likely to be construct irrelevant, the designers can link appropriate Variable Features to these Additional KSAs to give the task designer information to reduce or eliminate the requirement for these Additional KSAs. This helps ensure that the deficits will not be the cause of poor performance on the assessment.

Motivated by the Additional KSAs within each of the six UDL categories, the PADI project team added additional UDL–based Variable Task Features into *design patterns*. Once test developers decide which of the aforementioned six categories may add challenging construct irrelevant requirements for certain students, developers can choose to support these different categories through a variety of Variable Task Features (See Appendix).

Let us consider how Additional KSAs and Variable Task Features interact with respect to sight. If a speller is blind and cannot satisfy the requirement of sight

for the spelling bee contest, the designer can link Variable Task Features to reducing or eliminating the demands for sight. For instance, by providing the speller who is blind with a list of words in Braille before the contest, the speller is able to access and study the same word list as nondisabled participants. For spellers having limited vision, a large font size on a word list can be provided to remove this accessibility barrier. An example of linking between Additional KSAs and Variable Task Features with respect to sight is illustrated in the *design pattern* in the Appendix by bold and large font size letters.

If a speller has hearing disability, the assessment designer could link the Additional KSA of hearing to UDL–infused Variable Task Features so as to call a task developer's attention to other strategies (e.g., visual graphics, video animation, or tactile graphics) to invoke a speller's response of spelling corresponding English words. For spellers with speech and communication disorders, the designer could access the list of Variable Task Features for Skills and Fluency to find draw tools or keyboards so that participants can write or type their response out.

### *6.3 Potential Work Products and Potential Observations*

There are a variety of possible ways of acquiring evidence about the Focal KSAs from what students say, do, or make in the task situations. In a *design pattern*, Potential Work Products are student responses or performances that can hold clues about the Focal KSAs. Potential Observations are features of Work Products that constitute evidence about the Focal KSAs. They describe qualities, strengths, or degrees of characteristics of realized Work Products. In a spelling bee contest with only nondisabled spellers involved, the common format of Work Product is the oral spelling of each word. However, for students with disabilities, especially speech disabilities, other forms of Potential Work Products should be considered. In our spelling bee *design pattern*, we anticipate following Potential Work Products:

- Oral spelling of word
- Written spelling of word
- Typed or otherwise manual spelling of word with no visual feedback

- Selection of spelling word from choices

The Work Products can be analyzed to draw inferences about the speller's level of proficiency in spelling based on following Potential Observations:

- Correctness of spelled word (right/wrong is usual)
- How close response is to target
- If word spelled incorrectly, how closely it follows phonics rules

Especially, we want to point out that the second Potential Observation can be chosen by examiners to evaluate closeness of incorrect spelling.

### *7.0 Conclusion*

In sum, the key attributes of a *design pattern* lay out a design space for building assessment tasks for candidates both with and without disabilities. The Focal KSAs and the Characteristic Task Features are relevant to all potential examinees, and the *design pattern* provides guidance for how tasks might build in requirements for Focal KSAs by controlling Characteristic Task Features. The assessment designer must anticipate the ways in which Variable Task Features drive requirements for Additional KSAs so that the these features can be manipulated and students will not be confronted by requirements for Additional KSAs that have been deemed construct irrelevant in the testing application at hand.

For students without any disability, we typically establish a set of default (or standardized) Characteristic Features and settings of Variable Features. We typically do this based on the assumption or knowledge of these features being appropriate given the state of Additional KSAs that characterize nondisabled students (Hansen & Mislevy, 2008). For student with disabilities, the purpose of providing UDL through ECD in our design is to remove unfair disadvantage while at the same time addressing the possibility of unfair advantages for the person who receives that accommodation. The six UDL categories within Additional KSAs along with the accompanying UDL Variable Task Features guide designers to consider the diverse needs of all students. A similar extension of Potential Work Products that would support a range of ways of responding to tasks is being developed and linked with appropriate UDL–motivated KSAs. By infusing UDL into the PADI design system, assessment designers are able to create flexible *design patterns* that provide a more accurate measure of student learning.

`**References:**

American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC, American Psychological Association.

Barron, J. (1979). Orthographic and word-specific mechanisms in children's reading of words. *Child Development, 50,* 60-72.

Barron, J. & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 386-393.

Barron, R. W. (1980). Visual and phonological strategies in reading and spelling. In U. Frith (Ed.), *Cognitive process in spelling* (pp. 195-214). New York, NY: Academic Press INC.

Barron, J., Treiman, R. Wilf, J. & Kellman, P. (1980). Spelling and reading by rules. In U. Frith (Ed.), *Cognitive process in spelling* (pp. 159-194). New York, NY: Academic Press INC.

Cahalan-Laitusis, C., & Cook, L. (Eds.). (2007). *Accommodating students with disabilities on state assessments: What works?* Arlington, VA: Council for Exceptional Children.

CAST (2008). *Universal design for learning guidelines version 1.0.* Wakefield, MA: Author.

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 Large-scale Assessment and Accommodations: What Works.* L. L. Cook and C. C. Cahalan. Arlington, VA, Council for Exception Children**:** 95-111.

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, and Assessment 3* (7). Available online at http://escholarship.bc.edu/jtla/vol3/7

Dolan, R. P., & Rose, D. H. (2000). Accurate assessment through Universal Design for Learning. *Journal of Special Education Technology, 15*(4).

Hansen, E. G., & Mislevy, R. J. (2006). Accessibility of computer-based testing for individuals with disabilities and English language learners within a validity framework. In M. Hricko & S. Howell (Eds.), *Online assessment and measurement: Foundation, challenges, and issues*. Hershey, PA: Idea Group Publishing, Inc.

Hansen, E. G. & Mislevy, R. J. (2008) Design Patterns for Improving Accessibility for Test Takers With Disabilities. ETS Research Report No. RR-08-49.

Hansen, E. G., Mislevy, R. J., Steinberg, L. S., Lee, M. J., & Forer, D. C. (2005). Accessibility of tests within a validity framework. *System: An International Journal of Educational Technology and Applied Linguistics, 33,* 107-133.

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

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

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education/Macmillan.

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

Mislevy, R., & Riconscente, M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology (PADI Technical Report 9).* Menlo Park, CA: SRI International.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of

educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*.

Perfetti, C. A. (1997). The psycholinguistics of spelling and reading. In C. A. Perfetti, L. Reiben & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 21-38). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Phillips, S. E. (1994). High-stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education, 7,* 93–120.

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.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large-scale assessments* (NCEO Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Education Outcomes.

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

Venezky, R. L. (1976). *Notes on the history of English spelling*. Visible Language, 10, 351-365.

***Appendix:*** **Spelling Bee Assessment Design Pattern on PADI Online Design System**

Spelling Bee Assessment | Design Pattern 2106 [ View Tree | Duplicate | Permit | Export | Delete ]

| | | |
|---|---|---|
| Title: | [ Edit ] | Spelling Bee Assessment |
| Summary | [ Edit ] | The DP uses spelling Bee as an example to illustrate how advances in measurement with the combination of cognition science can improve the validity and fairness of the tests, with a particular eye toward accessibility for individuals with disabilities. |
| Focal Knowledge, Skills, and Abilities | [ Edit ] | FK1. Ability to encode linguistic form of a word to its oral or written alphabetic form details |
| Rationale | [ Edit ] | R1. Correctly spelling out words given their pronunciations indicates the examinee possesses targeted proficiency in spelling |
| Additional Knowledge, Skills, and Abilities | [ Edit ] | AK1. ********** The following Additional KSAs are associated knowledge that are often involved with tasks that address the Focal KSAs. Whether they are to be supported or not (e.g., glossary, background facts, equation list) is a decision to be made -- either by the assessment design team, either at the level of the testing program or at the level of the individual task if that is appropriate in the testing program ********** |
| | | AK2. Knowledge of root words and etymologies |
| | | AK3. Knowledge of correct English usages |
| | | AK4. Knowledge of foreign languages from which English draws |
| | | AK5. Knowledge of English phonics, conventions, and rules |
| | | AK6. Knowledge of English vocabularies |

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

AK8. **Perceptual**
**. vision** **details**

AK9. . hearing details

AK10. Skill and fluency
. dexterity, strength, and mobility details

Potential observations [ Edit ]

Po1. Correctness of spelled word (right/wrong is usual)

Po2. How close is response to target? (could choose to evaluate closeness of incorrect spellings)

Po3. Is the word spelled incorrectly, but in a way that follows phonic rules?

Potential work products [ Edit ]

Pw1. Oral spelling of word

Pw2. Written spelling of word

Pw3. Typed or otherwise manually spelling of word with no visual feedback

Pw4. Selection of spelling word from choices

Potential rubrics [ Edit ]

| | | |
|---|---|---|
| Characteristic features | [ Edit ] | Cf1. English word presented to examinee, in a way that its spelling is not communicated.<br>Cf2. Clear and unambiguous expression of the word.<br>Cf3. Word lists |
| | | Vf4. Language and Symbols (1): Supports for Vocabulary and Symbols<br>- Pre-taught vocabulary and symbols<br>- Embedded support for key terms (e.g. technical glossary, hyperlinks/ footnotes to definitions, illustrations, background knowledge)<br>- Embedded support for non-technical terms (e.g. non-technical glossary, hyperlinks/ footnotes to definitions, illustrations, background knowledge)<br>- Embedded alternatives for unfamiliar references (e.g. domain specific notation, jargon, figurative language, etc.)<br>Vf5. Language and Symbols (2): Supports for Syntactic Skills and Underlying Structure<br>- Alternate syntactic levels (simplified text)<br>- Grammar aids<br>- Highlighted syntactical elements (e.g. subjects, predicates, noun-verb agreement, adjectives, phrase structure, etc.)<br>- Highlight structural relations or make them more explicit |
| | | Vf6. Language and Symbols (3): Supports for English Language<br>- All key information in the dominant language (e.g. English) is also available in prevalent first languages (e.g. Spanish) for second language learners and in ASL for students who are deaf<br>- Key vocabulary words have links to both dominant and non-dominant definitions and pronunciations<br>- Domain-specific vocabulary (e.g. "matter" in science) is translated for both special and common meanings<br>- Electronic translation tools, multi-lingual glossaries |

Vf7. Language and Symbols (4): Supports for Decoding and Fluency
- Digital text with automatic text to speech
- Digital Braille with automatic Braille to speech

Vf8. Cognitive Features (1): Supports for Background knowledge
- Advanced organizers, pre-teaching, relevant analogies and examples
- Links to prior knowledge (e.g. hyperlinks to multimedia, concrete objects in students' environments)
- Provision of an example

Vf9. Cognitive Features (2): Supports for Critical features, Big Ideas, and Relationships
- Concept maps, graphic organizers, outlines
- Highlight features in text, diagrams, graphics, and illustrations
- Reducing the field of competing information or distractions, masking
- Using multiple examples and non-examples to emphasize critical concepts

Vf10. Cognitive Features (3): Options that Guide Information Processing
- Explicit prompts for each step in a sequential process
- Interactive models that guide exploration and inspection
- Graduated scaffolds that support information processing strategies
- Multiple entry points and optional pathways through content
- Chunking information into smaller elements, progressive release of information, sequential highlighting
- Discrete question (s) or scenario-based text presentation
- Complexity of the scientific investigation presented in the scenario
- Cognitive complexity (Webb's Depth of Knowledge Levels)
- If selected response, distractors based on misconceptions/typical errors vs. non-misconceptions

Vf11. Cognitive Features (4): Supports for Memory and Transfer
- Checklists, organizers, sticky notes, electronic reminders
- Prompts for using mnemonic strategies and devices
- Templates, graphic organizers, concept maps to support note-taking
- Scaffolding that connects new information to prior knowledge
- Embedding new ideas in familiar ideas and contexts, use of analogy, metaphor, example

Vf12. Skill and Fluency (1): Supports for Manipulations
- Virtual manipulatives, Snap-to constraints
- Nonstick mats, Larger objects

Vf13. Skill and Fluency (2): Supports for Navigation
- Alternatives for physically interacting with materials: by hand, by voice, by single switch, by keyboard, by joystick, by adapted keyboard

Vf14. Skill and Fluency (3): Alternatives to Writing
- Voice recognition, Audio taping, Dictation, Video, Illustration

Vf15. Skill and Fluency (4): Supports for Composition
- Keyboarding and alternative keyboards, Onscreen keyboard,
- Wider lines, Larger paper, Pencil grips
- Drawing tools - with shapes, lines, etc.
- Blank tables, charts, graph paper
- Spellcheckers, calculators, sentence starters, word prediction, dictation (voice recognition or scribe), symbol-to-text, sentence strips

Vf16. Executive Features (1): Support for Goal and Expectation Setting
- Prompts and scaffolds to estimate effort, resources, and difficulty
- Animated agents that model the process and product of goal-setting
- Guides and checklists for scaffolding goal-setting

Vf17. Executive Features (2): Supports for Goal Maintenance and Adjustment
- Maintain salience of objectives and goals (e.g. reminders, progress charts)
- Adjust levels of challenge and support (e.g. adjustable leveling and embedded support, alternative levels of difficulty, alternative points of entry)

Vf18. Executive Features (3): Supports for Planning and Sequencing
- Embedded prompts to "stop and think" before acting
- Checklists and project planning templates for setting up prioritization, schedules, and steps
- Guides for breaking long-term objectives into reachable short-term objectives

Vf19. Executive Features (4): Supports for Managing Information
- Graphic organizers and templates for organizing information
- Embedded prompts for categorizing and systematizing
- Checklists and guides for note-taking

Vf20. Executive Features (5): Supports for Working Memory
- Note-taking, Mnemonic aids
- Locate items near relevant text

Vf21. Executive Features (6): Supports for Monitoring Progress
- Guided questions for self-monitoring
- Representations of progress (e.g. before and after photos, graphs and charts)
- Templates that guide self-reflection on quality and completeness
- Differentiated models of self-assessment strategies

Vf22. Affect Features (1): Supports for Intrinsic Motivation (Challenge and/or Threat)
- Offer individual choice
- Enhance relevance, value, authenticity (e.g. contextualize to students' lives, provision of an example)
- Options to vary level of novelty and risk (e.g. options in peer and adult support, alternatives to competition, alternatives to public display or performance, alternative consequences)
- Options to vary sensory stimulation (e.g. shortened work periods, frequent breaks, noise buffers, optional headphones, alternative settings, presentation of fewer items at a time)

Vf23. Affect Features (2): Supports for Sustaining Effort and Persistence
- Maintain salience of goals (e.g. explicit display of goals, periodic reminders, replacement of long-term goals with short-term objectives, prompts for visualization)
- Adjustable levels of challenge and support
- Encourage collaboration and support
- Communicate on-going, mastery-oriented feedback

Vf24. Affect Features (3): Support for Self-regulation
- Guide motivational goal-setting
- Scaffold self-regulatory skills and strategies
- Develop emotional self-assessment and reflection

| | |
|---|---|
| I am a kind of | [ Edit ] |
| These are kinds of me | [ Edit ] |
| These are parts of me | [ Edit ] |
| Educational standards | [ Edit ] |
| Templates | [ Edit ] |
| Exemplar tasks | [ Edit ] |
| Online resources | [ Edit ] |
| References | [ Edit ] |

**Sponsor**
The U.S. Department of Education, Grant No. R324A070035

**Prime Grantee**
SRI International. *Center for Technology in Learning*

**Subgrantees**
University of Maryland
CAST