



PADI Technical Report 18 | January 2008

# An Introduction to Design Patterns in Language Assessment



PADI | Principled Assessment Designs for Inquiry

**Hua Wei**, University of Maryland  
**Robert J. Mislevy**, University of Maryland  
**Diwakar Kanal**, University of Maryland

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.sri.com>**

**PADI 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*

Lynne Peck Theis and Ron Fried, *Documentation Designers*

Copyright © 2008 SRI International and University of Maryland. All Rights Reserved.

# An Introduction to Design Patterns in Language Assessment

Prepared by:

Hua Wei, University of Maryland  
Robert Mislevy, University of Maryland  
Diwakar Kanal, University of Maryland

---

*Acknowledgments*

This material is based on work supported by the National Science Foundation under grant REC-0129331 (PADI Implementation Grant). We are grateful to the anonymous reviewers for their constructive comments on an earlier version of this article.

*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 the National Science Foundation.

## CONTENTS

---

<b>Abstract</b>	<b>iv</b>
<b>1.0 Introduction</b>	<b>1</b>
<b>2.0 The Rationale for Design Patterns</b>	<b>4</b>
2.1 Evidence-Centered Assessment Design	4
2.2 The Role of Design Patterns	5
<b>3.0 Attributes in the Assessment Design Pattern Structure</b>	<b>7</b>
<b>4.0 Using Design Patterns in Task-Based Language Assessment</b>	<b>10</b>
<b>5.0 Examples</b>	<b>12</b>
5.1 Language for Special Purposes	12
5.2 Following Spoken Instructions	14
5.3 Story-Telling	18
<b>6.0 Conclusions</b>	<b>22</b>
<b>References</b>	<b>23</b>

---

## FIGURES

---

Figure 1.	Graphic Representation of ECD Layers	4
Figure 2.	A Story-Telling Item from the Test of Spoken English™ (TSE®)	18

---

## TABLES

---

Table 1.	Sample Design Pattern “Language for Special Purposes”	2
Table 2.	Attributes of an Assessment Design Pattern	7
Table 3.	Sample Design Pattern “Following Spoken Instructions in Japanese”	15
Table 4.	Sample Design Pattern “Story-Telling”	19

## ABSTRACT

---

Assessment *design patterns* lay out considerations for building assessment tasks that address targeted aspects or situations of language use. They differ from *test specifications* by focusing on assessment arguments at a narrative level and are organized around aspects of language use that could be detailed in different ways for different purposes or circumstances. A *design pattern* thus encapsulates experience and research about some aspect of language use, organized around the structure of assessment arguments, as a starting point for designing tasks that will in addition satisfy the constraints and purposes of the job at hand. The rationale and structure of *design patterns* are described, and their use in language assessment is illustrated with examples concerning the use of language for special purposes, contextualized listening skills, and content-specific story-tellings.

## 1.0 Introduction

---

The historically dominant forms of language assessment are testing discrete components of language knowledge (grammar, vocabulary, and phonology) or skills (listening, speaking, reading, and writing). Tests are composed of samples of items or tasks that are deemed by language experts or instruction personnel to represent significant knowledge in the domain (National Research Council, 2002). With a focus on the knowledge of language *per se*, the test items are largely decontextualized and provide at best indirect evidence about how students can use language in various real-life situations.

Recognizing that knowledge of grammar and vocabulary is not sufficient to achieve particular communicative goals, assessing “communicative language competence” (Hymes, 1972) takes into consideration the various knowledge components that are utilized in communication via language. The notion of communicative competence has begun to guide the development and use of language tests in which the various components of language competence are called upon and interact with each other in the full context of language use (Bachman, 1990; Long & Norris, 2000).

Recent research on language learning has led to improved understanding about the communicative nature of language use. Insights have been gained, for example, as to the psychological foundations of language acquisition (e.g., Robinson, 2001), the subtle dynamics of face-to-face communication (e.g., Gumperz, 1982), the cognitive processes of comprehension (e.g., Kintsch, 1998), the ways features of situations interact with individuals’ knowledge (e.g., Holland & Quinn, 1987), and other aspects of language use that have implications for assessment.

Just how to organize this mass of information in ways that can guide the assessment design process is of special interest to language testers. This report introduces and illustrates representational forms called assessment *design patterns*, which were originally developed for assessing science inquiry (Mislevy et al., 2003) but can be applied to educational measurement more generally. It presents the rationale for *design patterns*, describes the attributes in the *design pattern* structure, compares *design patterns* with other assessment design tools, discusses the use of *design patterns* in task-based language assessment, and, finally, illustrates the application of *design patterns* with three examples (each of which could be extended and further detailed).

The structure of a *design pattern* is illustrated in Table 1. Although a detailed presentation of the elements of *design patterns* does not appear until Section 3, a quick perusal of the table may help to ground the following background discussion.

**Table 1: Sample Design Pattern “Language for Special Purposes”**

<b>Attribute</b>	<b>Value(s)</b>	<b>Comments</b>
<b>Title</b>	A speaking test of English for Special Purposes (ESP)	This could be for any language, not just English. According to Douglas (2001), “Testing language for specific purpose (LSP) refers to that branch of language testing in which the test content and test methods are derived from an analysis of a specific language use situation” (p. 45). An example of the ESP test is the Japanese Language Test for Tour Guides (Brown, 1995).
<b>Summary</b>	In this design pattern, an examinee is required to speak English in response to a prompt or an interlocutor. The quality and content of his response reveal his communicative language ability and professional knowledge.	The topic and purpose of the situation are determined by the specified language use domain.
<b>Rationale</b>	In real life, people often encounter situations where they need to convey field specific information to their colleagues or laymen in spoken English. A spoken English test for special purposes is commonly administered to professionals who learn English as a second language.	
<b>Focal KSAs</b>	Communicative language ability Sufficient content knowledge	As Douglas put it, “communicative language ability and content knowledge [are] the dual components of what [is called] specific purpose language ability” (2001, p. 46).
<b>Additional KSAs</b>	1. Pragmatic and sociocultural knowledge	These aspects of knowledge may or may not be focal, depending on the purpose of a test, so choices about variable task features may be chosen to stress or avoid them accordingly.
	2. Listening ability in English	If a conversation is required, the examinee should be able to comprehend spoken English.
<b>Potential observations</b>	1. Language characteristics of the speech	More intelligible and fluent enunciation of words and sentences reveals a higher level of proficiency in speaking, and unintelligible and less fluent enunciation of words and sentences reveals a lower level of proficiency. More coherent, concise, and complete expression of ideas reveals higher proficiency, and incoherent, less concise, and less complete expression of ideas reveals lower proficiency.
	2. Amount and correctness of substantive content of the speech	The amount and correctness of the content in the speech reveal the examinee’s level of professional knowledge.
	3. Appropriateness of the speech for the situation and audience	The ability to adjust the language to a specific situation or audience reflects the examinee’s pragmatic and sociocultural knowledge.
<b>Potential work products</b>	1. A monologue (such as an oral description or explanation)	
	2. A conversation	A real or simulated interlocutor is required. If a conversation is required, the quality of the examinee’s performance depends, in part, on his (or her) listening ability and conversational competence.

<b>Potential rubrics</b>		Refer to the “development of a scoring system” for the speaking sub-test of the Occupational English Test in McNamara (1996)
<b>Characteristic features</b>	Tasks that simulate the real-life target language use situations in which English is used for a specific purpose. In other words, tasks that ask for the interaction of language ability with specialized content knowledge.	
<b>Variable features</b>	1. Type of the required response	If a conversation is required, conversational proficiency is an additional KSA assessed in the examinee.
	2. Demand for professional knowledge	If professional knowledge is not the central aspect of the construct to be assessed, it should be constrained to a level with respect to the targeted examinee population so that a lack of unnecessarily high professional knowledge does not constitute an adequate reason for poor performance on the test.
	3. Demand for sociocultural and pragmatic knowledge	
	4. Complexity of the required response	A complex response requires a good control of the language as well as sufficient professional knowledge.
	5. Time for planning and rehearsing	Providing time for rehearsal gives the examinee an opportunity to organize his ideas and language.
	6. Presence of a real or simulated interlocutor	If a conversation is the required work product. Using a conversational format induces demand for conversational competence.
<b>Exemplar tasks</b>	Refer to the “sample materials – speaking sub-test (physiotherapists)” of the Occupational English Test in McNamara (1996)	
<b>References</b>	<p>Brown, A. (1995). The effect of rater variables in the development of an occupation-specific language performance test. <i>Language Testing</i>, 12, 1-15.</p> <p>Douglas, D. (2001). Three problems in testing language for specific purposes: Authenticity, specificity, and inseparability. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, and K. O'Loughlin (Eds.), <i>Experimenting with uncertainty: Essays in honor of Alan Davies</i> (pp. 45-52). Cambridge: Cambridge University Press.</p> <p>McNamara, T. (1996). <i>Measuring second language performance</i>. London: Addison Wesley Longman Limited.</p>	

## 2.0 The Rationale for Design Patterns

---

Many factors need to be considered and coordinated throughout the process of designing a language assessment. This section begins by briefly reviewing the evidence-centered design (ECD) framework of Mislevy, Steinberg, & Almond (2002) and the perspective from which *design patterns* are introduced and contrasted with test specifications (Davidson & Lynch, 2001) and Bachman and Palmer's (1996) framework of test characteristics.

### 2.1 Evidence-Centered Assessment Design

Designing a language assessment is not an easy job. A variety of factors need to be considered throughout the design process. Two complementary ideas help to organize the effort. The first is an overarching conception of an assessment as an argument from imperfect evidence. Messick (1994, p. 16) lays out the basic narrative, saying that we "... would begin by asking what complex of knowledge, skills, or other attributes should be assessed, presumably because they are tied to explicit or implicit objectives of instruction or are otherwise valued by society. Next, what behaviors or performances should reveal those constructs, and what tasks or situations should elicit those behaviors?"

The second idea is distinguishing layers at which activities and structures appear in the assessment enterprise, all to the end of instantiating an assessment argument in operational processes (Mislevy, Steinberg, & Almond, 2002; Mislevy & Riconscente, 2006). Indicating the activities a tool supports from this perspective helps to clarify its role. The layers shown in Figure 1 focus in turn on the substantive domain, the assessment argument, the structure of assessment elements such as tasks, rubrics, and psychometric models, the implementation of these elements, and the way they function in an operational assessment.

**Figure 1: Graphic Representation of ECD Layers**

Domain Analysis
Domain Modeling <i>[Design Patterns]</i>
Conceptual Assessment Framework <i>[Templates and Task Specifications]</i>
Assessment Implementation
Assessment Delivery <i>[Four-Process Delivery System]</i>

The domain analysis layer is concerned with gathering substantive information about the domain of interest that holds meaning for assessment. This includes the content, concepts, terminology, and representational forms that people working in the domain use. It also includes conceptions about the nature of knowledge in the targeted domain as well as the way people use the knowledge to solve problems. Work that takes place in domain analysis provides the grist for an assessment argument. Bachman and Palmer's (1996) taxonomy of task characteristics is useful for both domain analysis and the assessment

framework layer described below, for describing the features of target language use and of assessment tasks and establishing a correspondence between the two.

In the domain modeling layer of assessment design, information and relationships discovered in domain analysis are organized into the form of assessment arguments. The work in this layer is a transition from specialized knowledge about the domain to specialized knowledge about the machinery of assessment, which takes place in the next layer. Toulmin diagrams are examples of tools for organizing assessment arguments at a narrative level (e.g., Kane, 1992), as are design patterns. As discussed in the following section, design patterns are meant to suggest families of tasks organized around aspects of proficiency, which could be implemented in many different ways depending on the particulars of the testing contexts.

The elements and processes that are needed to implement an assessment that embodies the argument are specified in the conceptual assessment framework (CAF). In the CAF, structures such as task templates, task specifications, and scoring mechanisms are utilized to give concrete shape to the assessments a developer needs to create. These decisions are specific and detailed, and must reflect the purposes, constraints, and resources of the intended use. Work in the CAF layer converts the assessment arguments sketched in domain modeling in operational terms. Examples of tools that have proven valuable to language testers at this layer are test specifications (Davidson & Lynch, 2001) and Bachman and Palmer's framework. Whereas a design pattern provides a systematic means for organizing issues that need to be considered whenever one assesses a targeted aspect of proficiency, test specifications provide a detailed blueprint for writing tasks with specified properties that suit the purposes, constraints, and resources of the particular testing context. Tasks created from the same specification are equivalent in terms of purpose, cost, and so on, and can be used interchangeably.

This brief discussion clarifies how design support tools, both conceptual and programmatic, can be complementary if they are tuned to different facets of the design process. A test developer with a particular assessment application in mind could browse a collection of design patterns, for example, to get ideas about how to assess aspects of proficiency germane to her project. She might use Bachman and Palmer's framework to begin to think through features of particular importance and eventually express final configurations of tasks and tests in the form of test specifications.

While the next two layers of the ECD framework are less directly related to the present discussion, they are noted for the sake of completeness as the object of assessment design and specification activities. The work in assessment implementation includes such activities as authoring tasks, calibrating items, finalizing rubrics, producing materials, producing presentation environments, and training interviewers and scorers, all in accordance with the assessment arguments and test specifications created in previous stages, all in preparation for testing examinees. In assessment delivery, tasks are presented to examinees, performance is evaluated, tests are scored, and feedback and reports are given (for a discussion of the assessment delivery architecture, see Almond, Steinberg, & Mislevy, 2002,).

## **2.2 The Role of Design Patterns**

Although each assessment application is to some degree unique in its contents, purposes, and contexts, there are certain principles and relationships that all will share simply because all are assessments. For this reason one may gain advantage by embedding these principles in processes and knowledge representations. Architect Christopher Alexander (1977) coined the term *design pattern* in the mid-1970s. A *design pattern* is the core of a solution to a problem that occurs repeatedly in our environment — but at a level of generality that the approach can be applied in many situations while adapting to the particulars of each case. The same idea was adapted by software engineers to help designers tackle programming problems that recur in different guises

(Gamma et al., 1994). For these engineers, *design patterns* provide structured insights into conceptual problems and solutions above the level of specific programming languages and implementation environments.

Analogous forms called assessment *design patterns* were developed by Mislevy et al. (2003) to support the design of tasks for assessing science inquiry in the Principled Assessment Designs for Inquiry (PADI) project. Like designing tests of communicative competence, designing science inquiry tasks is a challenge to standard assessment development practice. It calls for extended performances, cycles of hypothesis and testing, and, often, technologies such as automated scoring and computer-based simulation environments. Assessment *design patterns* organize information about the targeted proficiencies, performance, and use situations in terms of the structure of assessment arguments. They serve as an in-between layer that connects the content of an assessment argument to the structure of the argument.

In particular, each *design pattern* builds around the general form of an assessment argument, concerning the knowledge or skill one wants to address (examples in science inquiry include model-based reasoning and designing experiments), the kinds of observations that can provide evidence about acquisition of this knowledge or skill, and the features of task situations that allow students to provide this evidence. Explicating the assessment structure in a narrative form with slots to be filled, *design patterns* arrange an underlying assessment argument into elements that can subsequently be instantiated in particular operational tasks. Because the structure of a *design pattern* implicitly contains the structure of an assessment argument, filling in the slots simultaneously renders explicit the relationships among the pieces of information in terms of the roles they play in the argument.

### 3.0 Attributes in the Assessment Design Pattern Structure

---

The essential concept of *design patterns* for assessment could be fleshed out in any number of ways, and implemented in any of a variety of formats. In the Principled Assessment Designs for Inquiry (PADI) implementation, *design patterns* are created in matrix forms with the cells filled with text and links to other objects, such as other *design patterns*, Web pages, and other resources. Table 2 summarizes the key attributes of an assessment *design pattern* (see Mislevy et al., 2003 for the full list of attributes).

**Table 2: Attributes of an Assessment Design Pattern**

<b>Attributes</b>	<b>Definition</b>
Title	A short name for referring to the design pattern
Summary	Overview of the kinds of assessment situations students encounter in this design pattern and what one wants to know about their knowledge, skills, and abilities
Rationale	The underlying warrant that justifies the connection between the targeted inferences and the kinds of tasks and evidence that support them
Focal KSAs	The primary knowledge, skills, and abilities that one wants to know about students
Additional KSAs	Other knowledge, skills, and abilities that may be required
Characteristic features	Salient features of tasks that can elicit evidence about the focal KSAs
Variable features	Features of tasks that can be varied to shift the difficulty or focus of tasks
Potential observations	Features of work products that constitute evidence about the focal KSAs
Potential work products	Student responses or performances that can hold clues about the focal KSAs
Potential rubrics	Rules and instructions used to evaluate student work products
Exemplar tasks	Sample assessment tasks that are instances of this design pattern
References	Research and literature that provide backing for this design pattern

The primary organizing feature of assessment *design patterns* is the knowledge, skill, or other abilities (i.e., the KSAs) that are being targeted; these are called the Focal KSAs, and may be thought of as the (possibly multifaceted) construct that is meant to be assessed. In language assessment, as in language use, many aspects of knowledge are called into play, some being components of language knowledge and others related to knowledge about context and purpose. The Summary is an overview of the assessment situations that students encounter in this *design pattern* and what one wants to know about their knowledge, skills, and abilities. The Rationale articulates the underlying warrant that justifies the connection between the targeted inferences and the kinds of tasks and evidence that support them.

Additional KSAs are other knowledge, skills, and abilities that may be required in a task that addresses the focal KSAs. Any particular additional KSA may be called for in some tasks that can be conceived under the *design pattern* but not others, depending on design choices as to variable task features such as mode of response, kinds of stimulus materials, contextualization, and requirements for topic knowledge. Since the Additional KSAs are not what we are interested in knowing about, they represent potential threats to validity. They may be minimized or avoided in order not to interfere with the inferences about Focal KSAs and to serve the purpose of the assessment. Alternatively, if it is known that the intended examinees have sufficient levels of a given Additional KSA, it may be included in a task to add variety or realism without introducing construct-irrelevant variance.

Language knowledge interacts with other knowledge in task situations in which the demand on different aspects of knowledge varies (Robinson, 2001). Using *design patterns* can help us sort out the construct knowledge we want to make inferences about and other knowledge that is required but not what we are interested in (see Mislevy, Steinberg, & Almond, 2002, on the application of these ideas to task-based language assessment). Beyond simply identifying important aspects of knowledge that should be assessed, *design patterns* help us to focus on the targeted language knowledge by making explicit the kinds of things we want to see students doing to demonstrate their knowledge, and the characteristics of assessment tasks that would shed light on the targeted knowledge and minimize the effect of irrelevant knowledge.

*Design patterns* are thus helpful in setting up task situations that suit the assessment purposes and agree with the targeted inferences. Features of tasks are described in a *design pattern*, which may include characteristics of stimulus materials, instructions, tools, help, and so on. Some features, called Characteristic Features of tasks, must be present in some way if one is to obtain evidence about the Focal KSAs. For example, to obtain evidence about a student's selection of register (in linguistics, a register is a subset of a language used for a particular purpose or in a particular social setting, such as informal, formal, or ritualistic) in a conversation, the student must in some way consider a situation in which some registers are more appropriate than others, and a choice must be made. Variable Features of tasks are those which could vary even while the Characteristic Features are present, such as whether a task focusing on register choice involves actual conversation, choice of given options, or oral responses to fixed recorded stimuli.

A variety of types of tasks could thus be generated to provide evidence about a given set of Focal KSAs, but vary in terms of cost, difficulty, authenticity, presentation context, demand for ancillary knowledge, and so on. Indeed, the purpose of *design patterns* is to suggest a variety of ways of assessing targeted KSAs, rather than dictating a single approach. The organizing feature of *design patterns*, therefore, is the construct rather than the particulars of a test tuned to a particular purpose, context, and set of constraints.

There are a variety of possible ways of getting evidence about the targeted 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. They are things that students say, do, or make. 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. Potential Rubrics are rules and instructions that we use to evaluate student Work Products. Formulation of appropriate rubrics depends on the kinds and qualities of inferences we want to make about students. Different observations can be derived from the same Work Product as assessment purposes vary. The rubrics for arriving at these observations thus vary in accordance with the features of Work Products that are relevant to the KSAs of interest.

The remaining attributes of a *design pattern* are sample assessment tasks, and research and literature that provide background or support for the *design pattern*. (Other attributes that are not listed in Table 2 include links to other design patterns, online resources, and miscellaneous associations; see Mislevy et al., 2003, for more detailed discussion.)

## 4.0 Using Design Patterns in Task-Based Language Assessment

---

Typically, a *design pattern* is motivated by the construct to be assessed and delineates a design space that is centered on that particular construct. The purpose for which the construct is assessed is not specified in the *design pattern*. The assessment purpose is achieved by selecting and implementing a configuration of task features, work products, scoring rubrics, and other elements that are specified in the *design pattern* and match the purpose in terms of level of generality.

There are several potential sources of inspiration for creating *design patterns*. One starts from a careful consideration of target language use (TLU) in context, in order to identify the key features of situations, performances, and KSAs (Bachman & Palmer, 1996). Another is research literature on the acquisition and use of language. A third is abstracting relationships among design features of existing assessments. In the last of these, we can work backward from existing tasks, breaking them down into the essential building blocks that reveal the underlying relationships and making them explicit in a *design pattern*, and then work forward to create new tasks, possibly with rather different surface features, from the same general *design pattern*. The latter two approaches are illustrated in later sections.

At the heart of a *design pattern* is the construct to be assessed and the logic of substantiating claims about the construct with relevant and sufficient evidence. *Design patterns* draw attention to the characteristics of assessments in a way that promotes the construction of the argument of construct validity, from the perspective of construct representation (Embretson, 1983). Construct representation is concerned with identifying the underlying knowledge, skills, and abilities that affect test performance, while the more familiar perspective of nomothetic span concerns the relationships of a test to other measures of the same or similar construct. Within a *design pattern*, the strands of KSAs are listed and sorted out as focal or non-focal, and characteristics of students' Work Products that provide evidence about these KSAs are documented. The alignment between KSAs and characteristics of student performance clarifies the dependence of task performance on the KSAs that are involved in fulfilling the task and relates differences in task performance to differences in the KSAs in meaningful ways. A *design pattern* suggests a number of solutions to assessing the same construct, and the network of interrelationships among these alternative assessments constitutes evidence for exploring the utility of these assessments as measures of individual differences.

In recent years, task-based language assessment (TBLA) has gained increasing popularity among language testers and educational researchers. The notion of having students "engage in meaningful language communication" (Norris, 2002, p. 337) is the primary goal in the development and use of language tasks. Students' performance in accomplishing tasks in the testing situation is direct evidence that can be used to infer their ability to accomplish similar tasks in real-life situations or their language ability in general. No matter what types of inferences are to be made, appropriate tasks need to be developed and their characteristics clearly documented. *Design patterns* facilitate the development of language tasks in that they organize in a clear, orderly fashion important thoughts about the tasks, thoughts regarding task features, knowledge, skill, or abilities tapped by the tasks, and criteria by which the tasks are to be scored.

Authenticity is a critical quality required of the tasks in task-based assessment. It is argued explicitly that real-life activities cannot be used as test items because they may require irrelevant knowledge or experience or favor some students unfairly (Bachman, 2002). "Authentic" assessment tasks that are related to real-life tasks in a meaningful way and engage students in a goal-oriented language use setting are what language testers recommend as appropriate. By authentic we mean that the assessment tasks, by simulating the important features of the actual language use situations, resemble real-life

tasks but are practical and appropriate to administer.

A *design pattern* provides an avenue to organize and document task features that characterize a family of tasks and have them related to the other attributes of the assessment. Having task features specified in a *design pattern* minimizes the requirement of aspects of knowledge or experience that are not of interest and emphasizes the correspondence between characteristic task features and the targeted set of knowledge, skills, or abilities.

An intriguing issue in the area of task-based assessment is how to take advantage of the conceptions in cognitive psychology to investigate the varying effects of task features on students' cognitive processes and the resulting performances. Observing students in different task situations and collecting rich qualitative and quantitative data on what they say, do, or make allows us to study the relationship between task features and student behavior. *Design patterns* provide a systematic means of documenting the kinds of observations that reflect the quality of students' higher-order thinking and language production and the kinds of Work Products that adequately capture student performance. Moreover, a close correspondence can be established between Potential Observations and scoring criteria in a *design pattern* that quantifies the information and facilitates the interpretation of scores.

## 5.0 Examples

---

The following examples illustrate the basic features of *design patterns*. All could be fleshed out in considerably more detail. Many of the variables in the Bachman and Palmer framework, for example, could be considered for expanding considerations of Characteristic and Variable Features, Work Products, and Potential Rubrics and Observations. We have limited attention to a smaller number of key attributes in order to focus the discussion on central issues in each area.

### 5.1 Language for Special Purposes

Tests of “language for special purposes” concern the use of language in particular contexts of use, and encompass pragmatic, content, and sociolinguistic knowledge as well as control of language per se (Hutchinson & Waters, 1987). The *design pattern* sketched in Table 1 was motivated by the speaking subtest of the Occupational English Test (OET; Latchford, 1996), a performance-based test of English as a second language for health professionals. The test has been used in Australia to assure that examinees can function in workplaces where the ability to communicate in English is necessary. All the examinees are presumed to have had medical training overseas, and the purpose of the test is to assess the examinees’ ability to deal linguistically with the situations they may encounter in local hospitals and clinics. A sample task that motivates the *design pattern* appears in McNamara (1996, p.108):

*Setting:* Hospital clinic

*Patient:* An elderly person who is recovering from a stroke (CVA). The patient is making slow progress in learning how to walk again.

*Task:* Talk to the patient about the following pieces of equipment

A wheelchair

A walking frame

A walking stick

Explain the advantages and disadvantages of each one

You would like the patient to be as independent in his or her movements as possible. You feel the frame is not appropriate.

You want the patient to have a stick. You do not want the patient to have a wheelchair at this stage.

This task is intended for physiotherapists who are expected to be familiar with the context. Explaining to patients their physical conditions and giving medical advice is the everyday job of physiotherapists and other health professionals. The purpose of this task and tasks of a similar type is to assess whether an examinee has an adequate control of the language so that he (or she) can convey his (or her) knowledge to a patient or a colleague. Communicative language proficiency is the central aspect of the construct being assessed—in a broad sense, no different from that being assessed in many other types of English testing. But content knowledge is also assessed in this task. As Douglas (2001, p. 46) put it, “communicative language ability and content knowledge [are] the dual components of what [is called] specific purpose language ability.” A unique feature of a test of English for Special Purposes (ESP) is that the tasks must stimulate the interaction between the examinee’s language ability and specified content knowledge. As in the task described above, the examinee must have some relevant professional knowledge and be able to pass on the knowledge to his (or her) patients in comprehensible and accurate language. The two aspects of proficiency are demanded jointly in the target language use situations, and the examinee’s success in clinical settings depends critically on this synthesis of professional clinical knowledge and communicative language ability.

The level of professional knowledge required in the task is under the control of the

assessment designers—it is a key Variable Feature. In the task above, professional knowledge about the therapeutic and assistive values of wheelchairs, walking frames, and walking sticks is required, as they pertain to patients with a particular medical condition, namely recovering from a stroke. A layman may hear about or have some experience with these things, but his (or her) knowledge may not be sufficient to enable him (or her) to describe the conditions of the patient, explain the significant differences among the three types of equipment, and provide proper advice in an understandable, convincing manner.

The content knowledge tapped in this task is considered to be basic medical knowledge, and examinees, supposedly having had medical training overseas, are assumed to possess the required knowledge. In tasks where the required content knowledge is assumed or constrained to a minimum, the lack of content knowledge is not intended to be a primary reason for poor performance for the targeted examinee population, and, therefore, the evidentiary focus of the tasks is assessing language ability, but as particularized to its use in medical situations, concerning medical topics. The demand for content knowledge is determined by the purpose of the test. A basic level of substantive content knowledge is considered appropriate in the OET because subsequent profession-specific paper-and-pencil tests of professional clinical knowledge will take place to further evaluate examinees' level of professional knowledge. For a situation in which there would be no separate content knowledge tests, the demand for content knowledge in the ESP testing could be increased by including tasks that ask for a higher level of professional knowledge. The required level of medical knowledge would not be presumed among the examinee population, and by the change in this Variable Feature, the Focal KSAs would be shifted as appropriate to the revised purpose of the test.

Observations can be made on an examinee's performance in an ESP task that bear on his (or her) language ability, content knowledge, and/or combined use of both as required by the task directive. The language characteristics of the examinee's speech can be identified and used to support inferences about his (or her) language proficiency. Whether the speech is intelligible, fluent, and grammatically correct provides important clues about the examinee's level of mastery of the language. Clear, accurate, and fluent utterances with appropriate grammatical constructions and vocabulary reveal a higher level of proficiency, whereas unintelligible, less fluent utterances with frequent errors reveal a lower level of proficiency. The content of the speech can be analyzed to draw inferences about the examinee's level of professional knowledge. Whether the examinee can express the required professional knowledge and whether the knowledge he (or she) has is correct are all potential observations relevant to the ESP purpose.

Another criterion by which the examinee's speech can be evaluated is its appropriateness for the situation and the audience. The examinee's ability to handle different communicative situations and different audiences is of interest because this kind of ability is valued in real-life clinical settings. The examinee's choice of words and tone and the way he (or she) delivers the speech have important implications for the inferences that can be made about his (or her) pragmatic and sociocultural knowledge. For example, if the targeted audience is a patient or a layman, the examinee is expected to use words that are less technical. However, if the examinee is speaking to a colleague who has some experience in the field, he or she is expected to use more precise and technical vocabulary that are specific to the profession—a kind of code-switching that is required for professionals in the field.

In the sample task, the examinee is required to produce a speech in response to the written prompt. No conversation is required. In real life situations, conversations between professional physiotherapists or between a physiotherapist and a patient or a layman are commonplace. Therefore, a variant of this task, and by implication for the universe of ESP tasks, can be created by including an interlocutor who interacts with the examinee. For example, we can have in the task a simulated patient who poses questions to the examinee and expresses disagreements with the advice he (or she) has received. An

exemplar role play card that can be provided to the simulated patient is given by McNamara (1996, p. 108) as follows:

*Setting: Hospital clinic*

*Patient: You are an elderly person who is recovering from a stroke. You feel you are making painfully slow progress, and don't really expect to be able to walk again. You feel you should be allowed to have a wheelchair.*

*Task: Ask the physiotherapist when you will be given a wheelchair. Insist on your need for this equipment. Explain that you feel that the painful exercises you are doing at the moment are pointless, and that you are pessimistic about your chances of making real progress. Be difficult!*

The inclusion of a conversation requires additional proficiency of the examinee — namely, conversational competence. The task is more realistic, and it provides opportunities to observe how the examinee initiates, participates in, and maintains a conversation and uses appropriate strategies when communicative breakdowns occur. Changing this Variable Feature again changes the complex of proficiencies that constitute the Focal KSAs. Whether it is desirable depends on factors that include costs, what is known about the examinee population, and the requirements of the targeted language use.

A problem that often occurs in speaking assessment with human interlocutors is that variations in the interlocutor talk across examinees can affect the opportunity they are given to perform, the language sample they produce, and the scores they receive (Wigglesworth, 2000). One way of dealing with this problem is to provide the interlocutor a booklet where the instructional language and verbal prompts he (or she) is allowed to use throughout the test are prescribed. No matter how an examinee responds, the interlocutor must stick to what is dictated in the booklet. It controls the amount and content of the interlocutor talk, but the conversation captured in the restricted context loses the flavor of authenticity and spontaneity. Alternatively, a role play card like the one described above can be presented to the interlocutor to be used in guiding the conversation. The interlocutor is free to use his (or her) words to fulfill the tasks. The unscripted conversation adds a substantial amount of authenticity to the task, and the examinee's responses to the interlocutor will be more credible and natural. Again, both options have advantages and disadvantages. The *design pattern* brings both possibilities to the attention of the test developer, and the determination is made in accordance with the particulars of the assessment situation.

## **5.2 Following Spoken Instructions**

This example is “reverse-engineered” (Davidson & Lynch, 2001) from a sample task produced by Prof. Lyle Bachman of UCLA and his students, in the WEBLAS project (Web-delivered language assessment system; <http://www.weblas.ucla.edu/>). The sample task appears on the page for Japanese language assessment. A student is provided a map with a train station, four theaters, and several other buildings labeled in English. The student listens to a telephone recording of walking directions from the train station to the theater, and he (or she) must indicate which theater is the targeted one. Table 3 is a *design pattern* for a class of tasks that all assess the same Focal KSA in various ways, which is labeled, “Following Spoken Instructions in Japanese.”

**Table 3: Sample Design Pattern “Following Spoken Instructions in Japanese”**

<b>Attribute</b>	<b>Value(s)</b>	<b>Comments</b>
<b>Title</b>	Following spoken instructions in Japanese	Could do analogous assessment for any second language
<b>Summary</b>	In this design pattern, a student listens to a procedural description in Japanese and is then asked to carry out an action based on what he (or she) has heard.	
<b>Rationale</b>	Following oral instructions is an often-encountered real-world phenomenon. E.g., people often need to comprehend oral directives to perform an act.	Searle (1969) distinguishes four types of speech act: utterance acts, propositional acts, illocutionary acts, and perlocutionary acts. Instructions are one category of illocutionary acts. Other design patterns could be constructed for apprehending other types of speech acts.
<b>Focal KSAs</b>	Ability to comprehend spoken instructions in Japanese	
<b>Additional KSAs</b>	1. Familiarity with representational forms provided	Representational forms may be provided.
	2. Ability to read first language	There may be supporting information supplied in print in another language, presumably one that the examinee is familiar with.
	3. Sociocultural knowledge	If the examinee must carry out an action in a social setting, the task is more realistic but sociocultural knowledge is required.
<b>Characteristic features</b>	Spoken instructions in target language focused on the illocutionary use of language	
<b>Variable features</b>	<ol style="list-style-type: none"> <li>1. Speed of the oral description</li> <li>2. Length of the oral description</li> <li>3. Complexity of the oral description</li> <li>4. Cohesion of the oral description</li> <li>5. Whether replay is available</li> <li>6. Familiarity with requested actions</li> <li>7. Degree of sociocultural knowledge needed</li> <li>8. Degree of knowledge/familiarity with supporting information</li> <li>9. Whether or not a representational form is provided</li> <li>10. Change of representational forms</li> </ol>	<p>The more speeded the oral description, the harder it is for the student to comprehend it.</p> <p>The longer the oral description, the harder the task.</p> <p>The level of language and coverage of content determines the complexity of the oral description.</p> <p>Variable features 1-6 are used to control the demand on focal KSA.</p> <p>Variable features 7-10 are used to control the demand on additional KSAs.</p>
<b>Potential observations</b>	1. Correctness of action(s) required by the oral prompt	
	2. Number of attempts before submitting the answer	
	3. Number of replays before submitting the answer	Multiple replays are permitted in Bachman’s example.

<b>Potential work products</b>	1. An answer to a multiple-choice question	Alternative work products vary in accordance with cost, reliability, and realism. When answers are multiple choice, options can be closer or further from the specified action.
	2. Completion of a representational form	The representational form provided at the beginning may be incomplete and needs to be completed by the student based on the instructions he (or she) has heard.
	3. Carrying out actions requested in the oral prompt	Carrying out actions in a social setting can add requirements for sociocultural knowledge.
<b>Potential rubrics</b>		Rubrics would be based on the accuracy and completeness of the examinee's actions, as represented in the work product(s), as compared with characteristics of ideal responses.
<b>Exemplar tasks</b>	Refer to Bachman's example in the Japanese WebLAS demo ( <a href="http://www.weblas.ucla.edu/~japanese/demo/ormFrameset31.htm">http://www.weblas.ucla.edu/~japanese/demo/ormFrameset31.htm</a> )	
<b>References</b>	Searle, J. (1969). <i>Speech acts: An essay in the philosophy of language</i> . Cambridge, England: Cambridge University.	

In "Following Spoken Instructions in Japanese," the student listens to a procedural description in Japanese and then is asked to carry out a procedure based on what he (or she) has heard. Neither conversation nor speaking is required. Following oral instructions is an often encountered real-world phenomenon. People often need to comprehend oral directives to perform an act.

Ability to comprehend spoken instructions in Japanese is the Focal KSA addressed in this *design pattern*. Other kinds of abilities may also be tapped at the discretion of the developer, such as non-language proficiencies which may include familiarity with the presented representational forms and the sociocultural knowledge needed to perform the required act—in the example, familiarity with the conventions of maps as forms of knowledge representation. The example shows an interesting design choice on the part of Bachman and his students to manipulate task features to sharpen the evidentiary focus of the task. Although the language at issue is Japanese and the task would be more authentic with a Japanese map, English is used for the map captions as well as for the written directions. Reading ability in English is thus a necessary but ancillary requirement to doing well. That is appropriate for the intended population of test-takers who are native English speakers studying Japanese as a second language. They can be presumed to have enough English reading capability that the directions and map captions will not impede their work. But different choices, all consistent with the same motivating *design pattern*, might be appropriate for different populations or different purposes.

Potential Work Products in tasks generated under this *design pattern* would direct the student to carry out an action that indicates understanding of the scenario and directives. The characteristics of Work Products will then be compared with the request. An answer to a multiple-choice question, completion of a representational form, and carrying out actions requested in the oral prompt might produce evidence about the Focal KSA. In the WebLAS task, for example, the student is asked to indicate the movie theater on a map to which directions from the train station were provided orally.

In conjunction with determining the Work Product(s), the *design pattern* also prompts researchers to ponder how the Work Product(s) might be evaluated to characterize the evidence it elicits. It suggests that correctness of actions required by the oral prompt, number of attempts before submitting the answer, and number of replays before submitting the answer are potential observations that contain evidence about the examinee's level of ability to apprehend and follow oral directions. Again, considerations such as cost, purpose, and washback effects (effects of testing on learning) shape the decisions as they might be expressed in *test specifications*.

Spoken instructions in a target language focus on the illocutionary use of language. The examinee is the receiver of the message and is directed to take an action. The message is strictly one-way; there is no opportunity to ask for clarification or further information. For this reason, scenarios in which the stimulus speech sample is a recorded message are advantageous as they provide this rather uncommunicative feature while maintaining fidelity to many real-world situations.

Within the constraints noted above that define a class of listening to instructions tasks, a large number of variations can be proposed. Dimensions along which tasks might vary include the length, complexity, and structure of spoken instructions, provision to hear the speech sample multiple times, procedural or declarative information presented in the speech sample, and familiarity with the materials and situations within which the instructions are to be carried out. The two tasks described below both address the same Focal KSA — i.e. ability to comprehend instructions in Japanese. They illustrate how tasks can look different on the surface but still provide evidence about the same underlying Focal KSA. Different Variable Features, selected as design choices that suit the context of test use, cause tasks to differ in difficulty or focus.

In Task 1, a recorded speech sample gives a list, in Japanese, of movies playing at three theaters. Asked "What is the name of the movie playing in Kitayama Movie Theater?," the examinee chooses from a list of options that are presented in his (or her) first language. Note that the written materials with which the student responds are in his (or her) first language, so that reading Japanese is not required.

Settings for Variable Features for tasks within this *design pattern* include the following:

1. Factual information in oral description.
2. Relatively short speech sample.
3. No replay available (hence greater complexity, and higher demands on working memory and real-time processing than if replay were allowed)

In Task 2, a recorded message describes orally how to travel from the train station to the Kitayama Movie Theater, in terms of both map directions and route description (e.g., turn left at A street). The examinee is asked "How does one reach the Kitayama Movie Theater?" and responds with a rather different work product than that of Task 1, namely by tracing the route, with pencil, on the paper map that has been provided.

The settings for Variable Features of tasks within this *design pattern* have been determined as follows:

1. Procedural information in oral description
2. A sequencing of information is provided.
3. Replaying is allowed (so that working memory demands are decreased).
4. Both city layouts and maps as representational forms are presented in the task. Familiarity with these is therefore an Additional KSA that is required. For students who are familiar with both, this demand is not a major source of difficulty and adds to the realism of the task. However, for students who are not familiar with either, this demand constitutes a construct-irrelevant source of difficulty.

Performance on Task 1 and Task 2 can reveal proficiency in listening to instructions in Japanese, but satisfactory performance on Task 2 demands a higher level of ability than Task 1, as well as a somewhat different set of ancillary knowledge and abilities. Task 1 requires an answer to a factual question, but Task 2 demands a more complex sequence of actions. For Task 1, remembering the name is sufficient to complete the task, but for Task 2, understanding, organizing, and analyzing skills are essential to complete the task. Task 2 is harder, partly because of additional demands on listening through the greater length of the speech sample but also because a more complex action is described and must be carried out. Different choices of Characteristic Features and Work Products could be determined to produce yet other tasks that provide evidence about proficiency in listening to instructions in Japanese, each with their own cost, authenticity, and demands for ancillary knowledge.

### 5.3 Story-Telling

This *design pattern* is motivated by the story-telling tasks often used in language classrooms. These kinds of activities concentrate on the transfer of meaning rather than the language form itself. They promote second language learning by encouraging students to express their meanings in their own words, and, at the same time, control the precision of their language (Skehan & Foster, 1999). An exemplar task that motivates the *design pattern* is found in the sample Test of Spoken English™ (TSE®; TOEFL®, n.d.), shown in Figure 2. The sample design pattern for “Story-Telling” tasks is shown in Table 4.

**Figure 2: A Story-Telling Item from the Test of Spoken English™ (TSE®)**

*Please look at the six pictures below. I'd like you to tell me the story that the pictures show, starting with picture number 1 and going through picture number 6. Please take one minute to look at the pictures and think about the story. Do not begin the story until I tell you to do so.*

*Tell me the story that the pictures show. (60 seconds)*

*Note: Retrieved from*  
<http://www.ets.org/portal/site/ets/menuitem.1488512ecfd5b8849a77b13bc3921509/?vgnextoid=e3042d3631df4010VgnVCM10000022f95190RCRD&vgnnextchannel=092a7f95494f4010VgnVCM10000022f95190RCRD>

**Table 4: Sample Design Pattern “Story-Telling”**

<b>Attribute</b>	<b>Value(s)</b>	<b>Comments</b>
<b>Title</b>	Story-telling	
<b>Summary</b>	In this design pattern, a student is required to tell a story, in his (or her) own words, based on what he (or she) has heard, read, or watched.	
<b>Rationale</b>	Story-telling is an activity often used in language teaching and learning. It emphasizes the transfer of message rather than the linguistic form itself. It is relevant to students' everyday life and communication needs and is, therefore, more realistic than other tasks that assess isolated language structures.	
<b>Focal KSAs</b>	Ability to present orally the message contained in a prompt with fluent, accurate, and complex language	
<b>Additional KSAs</b>	1. Listening ability in English	If the story is presented in an oral form, listening ability is required.
	2. Reading ability in English	If the story is presented in a written form, reading ability is required.
	3. Familiarity with the context of the story	
	4. Note-taking skill	If the story is fairly long and complex and note-taking is allowed, note-taking skill is needed.
<b>Characteristic features</b>	A story based on stimuli material is presented to examinees, and the examinees must tell the story in their own words.	
<b>Variable features</b>	1. The degree to which the storyline is structured	A more structured storyline in the original narrative results in a more fluent and accurate response.
	2. The number of times the story is repeated	Repetitions of the story reduce the load on the student's working memory.
	3. The form in which the story is conveyed	The story can be presented in a written, audio, or video form.
	4. Complexity of the story	The more complex the plot, the more demanding the task.
	5. Time for planning and rehearsing	Providing time for rehearsal gives the student an opportunity to restructure the plot and organize his (or her) language.
<b>Potential observations</b>	1. Language characteristics of the response 1a) Whether the grammatical structures and vocabulary are used properly 1b) Whether the speech is coherent and fluent 1c) Whether varied and complex language structures are used to retain the subtleties in the original message	This focuses on the “language” aspect of task performance.

	2. Completeness and correctness of the response 2a) Whether the important information contained in the original message is retained 2b) Whether there is any distortion or misinterpretation with regard to the plot of the story	This focuses on the “content” aspect of task performance.
<b>Potential work products</b>	1. An oral narrative	
	2. Oral responses to questions	
<b>Potential rubrics</b>		Skehan & Foster (1999) provide definitions of fluency, accuracy, and complexity as criteria, from which rubrics could be constructed.
<b>Exemplar tasks</b>	An item in the sample Test of Spoken English ( <a href="http://www.ets.org/portal/site/ets/menuitem.1488512ecfd5b8849a77b13bc3921509/?vgnnextoid=e3042d3631df4010VgnVCM10000022f95190RCRD&amp;vgnnextchannel=092a7f95494f4010VgnVCM1000022f95190RCRD">http://www.ets.org/portal/site/ets/menuitem.1488512ecfd5b8849a77b13bc3921509/?vgnnextoid=e3042d3631df4010VgnVCM10000022f95190RCRD&amp;vgnnextchannel=092a7f95494f4010VgnVCM1000022f95190RCRD</a> )	
<b>References</b>	Skehan, P. & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. <i>Language Learning</i> , 49, 93-120.	

In a story-telling task, students are provided with prompts that suggest a situation or sequence of events (a Characteristic Feature) either in the form of a written discourse, a cartoon narrative, or perhaps a video episode (a Variable Feature). They are then asked to tell a story (the Work Product) based on the information contained in the prompts. Students’ ability to orally present a narrative in accurate and coherent language is the targeted proficiency (accuracy and coherency are Potential Observations, along with fluency, grammaticality, and other features of the language produced). Clearly, being able to comprehend the content of the prompts is a prerequisite to oral presentation. Therefore, some Additional KSAs, such as reading ability or familiarity with storyboards, will be involved in the tasks, depending on the nature of the prompts. Furthermore, students’ familiarity with the context in which the story happens also contributes to their success in the task. In the sample task, a student who has never had any experience with wet paint may have difficulty understanding what has happened to the man shown in the pictures. In designing such kinds of tasks, the designer tries to avoid tasks that will disadvantage those with less exposure to the targeted cultural background (controlling a Variable Feature), unless cultural knowledge is what the designer wants to know about (in which case it becomes focal rather than ancillary).

Students’ oral narratives can be evaluated in terms of language and content. Observations are made with regard to whether a student reproduces the content contained in the prompts without much loss of information or misinterpretation, whether he (or she) can use accurate and fluent language to express his (or her) meanings, and whether he (or she) can use varied and complex language structures to convey the subtleties of the original message. Skehan and Foster (1999) suggested that fluency, accuracy, and complexity be used as the criteria to evaluate students’ Work Products. They also pointed out that these three aspects of task performance cannot be attended to equally due to the limited attentional capacities of language learners. Therefore, tasks can be designed by varying features to increase or decrease requirements for different aspects of proficiency.

Two important features that can be varied to make the tasks more or less difficult are task structure and processing conditions, the influence of which on narrative retellings is discussed in Skehan & Foster (1999). The degree of organization of the material presented to the students has an impact on their performance. A well-structured story line helps students to allocate more attention to language form, resulting in a greater degree of fluency, accuracy, and complexity. For example, the sample task shown above furnishes an inherent time sequence among the actions shown in the pictures, with one action leading to the next in a natural and predictable manner. This allows students to focus on the speech and have more control on the language they use.

Processing conditions are another feature of the tasks that can have an influence on the quality of the oral narratives. For example, if the information given in the prompts is familiar to the students, the cognitive load of the tasks is reduced, and students' performance, in terms of the use of language and comprehension of the content, is enhanced. Moreover, if the students are given planning time before they speak, the processing burden is eased, and students are likely to use more accurate and complex language structures. A different task can be derived from the above example if planning time is not provided. The new task is harder in the sense that students need to organize their language while thinking about the meanings of the pictures. This requires a higher level of language proficiency and other non-language proficiencies.

And just as task features can be varied to place more or less stress on aspects of proficiency, a test developer can choose to highlight different aspects of quality in responses, stressing fluency, accuracy, and complexity in any weighted combination. We note in passing that it is good testing practice to make sure the examinees, as well as raters, know beforehand what aspects of their performance will be evaluated so that they can determine how to trade off their resource capabilities.

## 6.0 Conclusions

---

*Design patterns* provide a framework for explicating the assessment argument by suggesting how to elicit information about the KSAs of interest, Potential Work Products, Rubrics, and Observations, and features of the tasks. That is, information about the nature and conditions of use of various aspects of language proficiency is being organized in terms of categories that are central to assessment arguments. The representational form of a *design pattern* is at a narrative level, in the form of elements of a substantive argument, rather than a technical level, in the form of Measurement Models or details of tasks or Evaluation Procedures. *Design patterns* are meant to summarize experience and theory across some range of potential tasks, in order to provide useful starting points for new assessment projects.

The elements of the basic argument structure and their relationships are persistent across assessments of different kinds meant for different purposes. It may be noted, however, that an argument can be built from different starting points. We can start the assessment development process by asking ourselves what complex of knowledge, skills, and abilities (KSAs) should be assessed. Or we can start the argument by thinking through the Potential Observations we want to make in students' performances. Alternatively, we can start by deciding which type of Work Products we want to capture in the assessment, or which Potential Rubrics we want to use to score and evaluate student performances.

An increasingly popular approach for language assessment design is task-based language assessment (TBLA). TBLA starts when we decide which features of realistic language-use to replicate in the assessment tasks. Desired performance on real-world tasks is the construct of interest, and the real-world criterion elements are identified in the scoring rubrics (Long & Norris, 2000). The complexities of task-based language assessment motivate discussions about what we want to assess, how we create task situations, and what sorts of evidence should be collected to support our inferences. *Design patterns* can guide these discussions and help us design tasks that support inferences that serve the purpose of the assessment.

As next steps, the development of more detailed *design patterns* would prove useful to the language testing community. As research results accumulate in psycholinguistics and sociolinguistics on the one hand, and practical experience accrues across a variety of test types and contexts on the other, representational forms that harness this information for developing tests of different kinds and for different purposes can provide good starting points for developing sound assessments. We propose *design patterns* as an instance of such a tool.

## References

---

- Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A pattern language: Towns, buildings, construction*. New York: Oxford University Press.
- Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment*, 1(5). Retrieved May 20, 2007 from <http://www.bc.edu/research/intasc/jtla/journal/v1n5.shtml>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19, 453-476.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Davidson, F., & Lynch, B. K. (2001). *Testcraft: A teacher's guide to writing and using language test specifications*. New Haven: Yale University Press.
- Douglas, D. (2001). Three problems in testing language for specific purposes: Authenticity, specificity, and inseparability. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honor of Alan Davies* (pp. 45-52). Cambridge: Cambridge University Press.
- Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179-197.
- Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design patterns*. Reading, MA: Addison-Wesley.
- Gumperz, J. J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.
- Holland, D., & Quinn, N. (Eds.) (1987). *Cultural models in language and thought*. Cambridge: Cambridge University Press.
- Hutchinson, T., & Waters, A. (1987). *English for specific purposes*. Cambridge: Cambridge University Press.
- Hymes, D. H. (1972). On communicative competence. In J. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). London: Penguins Books.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527-535.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Latchford, A. (1996). Occupational English Test for health professionals. *Digest of Australian Languages and Literacy Issues*, 4, 3-4.
- Long, M. H., & Norris, J. M. (2000). Task-based language teaching and assessment. In M. Byram (Ed.), *Encyclopedia of language teaching* (pp. 597-603). London: Routledge.
- McNamara, T. F. (1996). *Measuring second language performance*. New York: Addison Wesley Longman.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.

- Mislevy, R. J., Chudowsky, N., Draney, K., Fried, R., Gaffney, T., Haertel, G., et al. (2003). *Design patterns for assessing science inquiry* (PADI Technical Report 1). Menlo Park, CA: SRI International. Retrieved May 20, 2007 from [http://padi.sri.com/downloads/TR1\\_Design\\_Patterns.pdf](http://padi.sri.com/downloads/TR1_Design_Patterns.pdf)
- Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design: Layers, structures, and terminology. In S. Downing & T. Haladyna (Eds.), *Handbook of Test Development* (pp. 61-90). Mahwah, NJ: Erlbaum.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing*, 19, 485-504.
- National Research Council (2002). Advances in the sciences of thinking and learning. In J. Pellegrino, N. Chudowsky, & R. Glaser (Eds.), *Knowing what students know* (pp. 59-110). Washington, D.C.: National Academy Press.
- Norris, J. M. (2002). Interpretations, intended uses, and designs in task-based language assessment. *Language Testing*, 19, 337-346.
- Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design. In P. Robinson, P. (Ed.), *Cognition and second language acquisition* (pp. 287-318). Cambridge: Cambridge University Press.
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49, 93-120.
- TOEFL® (n.d.) TSE® Details: Learners and Test Takers. Web page retrieved September 10, 2007, from (<http://www.ets.org/portal/site/ets/menuitem.1488512ecfd5b8849a77b13bc3921509/?vgnextoid=e3042d3631df4010VgnVCM10000022f95190RCRD&vgnnextchannel=092a7f95494f4010VgnVCM10000022f95190RCRD>)
- Wigglesworth, G. (2000). Issues in the development of oral tasks for competency based performance assessments. In G. Brindley (Ed.), *Studies in Immigrant English Language Assessment* (Volume 1). Sydney: NCELTR.





**Sponsor**

The National Science Foundation, Grant REC-0129331

**Prime Grantee**

SRI International. *Center for Technology in Learning*

**Subgrantees**

University of Maryland

University of California, Berkeley. *Berkeley Evaluation & Assessment Research (BEAR) Center and The Full Option Science System (FOSS)*

University of Michigan. *BioKIDS*

