Beyond Accessibility: Evidence Centered Design for Improving the Efficiency of Learning-Centered Assessments

Eric G. Hansen and Diego Zapata-Rivera, ETS, Princeton, NJ

Mingyu Feng, Worcester Polytechnic Institute, Worcester, MA

Paper presented at the annual meeting of the
National Council on Measurement in Education (NCME)

April 16, 2009, San Diego, CA.

Abstract

This paper illustrates how Evidence Centered Design (ECD) can be used to address the *accessibility* of learning-centered assessments, and how such efforts lay a critical foundation for improvements in other aspects of quality, particularly *learning efficiency*. Specifically, it illustrates how the same basic strategies used to ensure *accessibility* can be taken a step further to make such systems more *learning efficient* for students more generally. We draw on Cognitive Load Theory to provide a research-based rationale for this approach. If validated, this approach may lay a foundation for learning-centered systems that are not only more accessible for students with disabilities but also more learning effective and efficient, valid, and engaging for all students.

Introduction

Problem

Recent advances in computer and related technologies are helping give rise to a "new generation of technology-enabled assessments [that] offers the potential for transforming what, how, when, where, and why testing occurs, …launch[ing] a new era of integrated, *learning-centered assessment systems*" (p. 75, emphasis added) (Quellmalz & Pellegrino, 2009). Such learning-centered assessments may offer rich media, simulations, and the possibility of assessing and fostering problem-solving and critical thinking skills that are not well addressed by other means. Such systems show indications of being learning effective for students *without* disabilities. For example: (a) the Assistment system, which provides assessment and instruction supports, has been successful in helping general (nondisabled) populations learn middle school mathematics (algebra) (Razzaq & Heffernan, 2006; Feng, Heffernan, Beck, & Koedinger, 2008) and (b) students using the Diagnoser system, which provides continuous formative assessment and feedback to students and teachers, outperformed their peers on items from the state science test (Minstrell & Kraus, 2005 and Thissen-Roe, Hunt, Minstrell, & 2004, cited in Quellmalz & Pellegrino, 2009). But the outcomes do not necessarily extend to students with disabilities. In order to address the achievement gap between students with disabilities and those without disabilities, there is a great need for methods to ensure that innovative technology-based systems for assessment and learning are accessible and effective for students with disabilities. Ideally, efforts to address accessibility for students with disabilities should improve quality (e.g., learning effectiveness and efficiency, validity, and engagement) for students generally. There is a need for a framework that helps address such a broad set of concerns in a coherent and integrated manner.

Purpose

The purpose of this paper is to show how methods from Evidence Centered Design (ECD) (Mislevy, Steinberg, & Almond, 2003; Hansen, Mislevy, Steinberg, Lee, & Forer, 2005; Hansen & Mislevy, 2006; Hansen, Mislevy, & Steinberg, 2008; Hansen & Zapata-Rivera, 2008a, 2008b) can be used to support a wide range of quality criteria, including accessibility and learning efficiency. Specifically, the paper will illustrate how the procedures used to address *accessibility* can be extended to address *learning efficiency* for students more generally, thereby providing another reason to invest in systematic design for accessibility of learning-centered assessments.

Strategy

The strategy of this paper is to: (a) provide an overview of Evidence Centered Design, (b) provide an extended example illustrating how ECD is used to address a range of quality-related decisions, and (c) show how the reasoning that goes into addressing *accessibility* issues can be extended to improve *learning efficiency* for students more generally. Furthermore, the paper highlights how research on Cognitive Load Theory (CLT) helps provide an empirical basis for the ECD-based approach.

An Overview of Evidence Centered Design

Evidence Centered Design methodology is an argument-based approach, involving *claims*, such as about what students know and can do or what they have learned, along with relevant *evidence*. As shown in Figure 1, ECD originally focused on the design of assessments rather than on learning oriented systems and was primarily for *typical* students (Mislevy, Steinberg, & Almond, 2003). ECD was then extended to address *accessibility* of assessments for students with disabilities and English language learners (Hansen & Steinberg, 2004; Hansen, Mislevy, Steinberg, Lee, & Forer, 2005; Hansen & Mislevy, 2006; Hansen, Galpern, & Goodman, 2008; Hansen, Mislevy, & Steinberg, 2008). Such extensions help ensure that efforts to improve the accessibility of an assessment do not undermine the validity of the assessment results. A committee of the National Research Council that was examining assessment accommodations confirmed the need for argument-based approaches and cited extensively the ECD accessibility work in its published report (National Research Council, 2004, see Chapter 6: "Articulating Validation Arguments;" Hansen & Steinberg, 2004; Hansen Mislevy, & Steinberg, 2008).

Characteristic	ECD for assessments (Mislevy, Steinberg, & Almond, 2003)	ECD for accessibility (Hansen & Mislevy, 2005)	ECD for learning oriented products (Hansen, 2007)	ECD for learning-centered assessments (Hansen & Zapata, 2008a, 2008b)
Attends to typical students	Yes	Yes	Yes	Yes
Attends to diverse students		Yes	Yes	Yes
Attends to issues of pedagogy and learning efficiency			Yes	Yes
Provides simplified argument structure encompassing both assessment and learning				Yes

Figure 1. Some extensions to Evidence Centered Design

Building on the accessibility extensions, ECD has been further extended to address learning oriented systems (Hansen, in submission; Hansen & Zapata-Rivera, 2008a, 2008b), including those that integrate assessment and learning, such as learning-centered assessments. Essentially, these extensions for learning-oriented systems expanded the scope of the argument to include a pedagogical model, involving claims and evidence about how learners advance from one level of proficiency to the next. Hansen (in submission) also highlighted the potential value of accessibility-related methods for addressing *learning efficiency* as well as other important quality criteria (engagement, validity, learning effectiveness). Hansen, Zapata-Rivera and colleagues then provided a simplified argument structure that addresses both assessment and learning, such as for learning-centered assessments. They also piloted the methodology in various settings, including some in which accessibility was *not* the dominant consideration: (1) A project that used ECD to model a diverse set of educational products, including: (a) an online system that integrates learning and assessment (Feng, Hansen, & Zapata-Rivera, 2009), (b) an assessment of a noncognitive ability (Kyllonen, 2005), and (c) a test familiarization and practice system (Hansen & Zapata-Rivera, 2008a); (2) A project that designed a prototype learning activity based on a commercial video game (Roller Coaster 3 Platinum) and evaluated its usability (Zapata-Rivera & Hansen, 2009; Hansen & Zapata-Rivera, 2008b); and (3) A project that examined the implications of policies on the use of electronic calculators in assessment and learning settings (Hansen, Fife, Graf, & Supernavage, 2008).

In summary, ECD has been applied to educational products of a progressively more diverse set of purposes (e.g., not just pure assessment but also learning-oriented products), for which a wider range of quality criteria may be applicable. For the purpose of this paper, criteria of accessibility and learning efficiency are of special interest.

How does ECD relate to efficiency? Essentially, ECD provides a way of representing designs in a way that highlights opportunities for improving learning efficiency. Cognitive Load Theory provides a rationale for why certain design decisions should result in greater learning efficiency.

An Overview of ECD for Learning-Centered Assessments

This section provides an overview of the procedure for ECD for learning-centered assessments (ECD-LCA). The procedure consists of three major steps. Due to the importance of step 1 (argument construction), the overview provides additional detail about this step.

1. Construct an argument

    a. Define the purpose and target population for the learning-centered assessment

    b. Define appropriate quality criteria

    c. Define the purpose and structure of the argument

    d. Populate the argument structure with values. Values should be suited to a particular case (e.g., a person having a particular profile interacting with the learning-centered assessment in a particular situation).

2. Apply rules and heuristics to the argument to evaluate quality and address any quality problems. *Accessibility* and *learning efficiency* are two important aspects of quality, but are probably best understood in the context of several quality criteria which a detailed later in this paper.

3. Implement the system and validate its argument

While the steps are listed in linear sequence, the process is actually iterative and some steps can occur in different orders.

Argument Structure: An Introduction to the KSA Value Matrix

The most commonly used structure in ECD accessibility-related work is the KSA value matrix. (Other structures can be used, including Bayes nets. [Hansen, Mislevy, &

Steinberg, 2008]). This matrix is at the heart of the argument. Figure 2 shows an example of a matrix for a learning-centered assessment for middle school students learning mathematics (pre-algebra) concepts. The matrix is a way of representing key aspects of a learning-centered assessment system.

A	B	C	D	E	F	H
Row number	**KSA**	**Focal value**	**Pre-inter-vention profile value**	**Intended growth outcome**	**Require-ment value - original design**	**Post-inter-vention profile value**
1	Engage	n/a	3	n/a	3	3
2	See	n/a	0	n/a	3	0
3	Hear	n/a	3	n/a	3	3
4	Know format and delivery system	n/a	2	3	3	3
5	Know game rules	n/a	1	3	3	3
6	Know game strategy	n/a	1	2	2	2
7	Know arithmetic	n/a	2	n/a	3	2
8	Know non-math vocabulary	n/a	3	n/a	3	3
9	Know math vocabulary	n/a	2	n/a	3	2
10	Know math concept	3	2	3	3	3

Figure 2. A KSA value matrix for a game that is a learning-centered assessment

Note the header row, which indicates the content of the columns, including the column containing a list of Knowledge, Skill, and other Attributes (KSAs). Generally, the list of KSAs should include a range of KSAs that are essential parts of the targeted proficiency, such as know a particular math concept. In the context of a learning-centered assessment, the targeted proficiency refers either to: (a) the construct that the assessment is intended to measure, (b) the learning objective that the learning experience is intended to achieve, or (c) both. Also included in the list should be abilities that are necessary to perform well in the given task situation [e.g., see, hear].), even if *not* part of the targeted proficiency.

Each KSA has two or more levels, which are designated with numbers (integers), where, the higher the number, the higher is the capability. It is these levels to which the numbers populating the main body of the table refer. (Where a KSA value is not applicable, it is indicated by "n/a".)

To the right of the KSA column are five columns for different kinds of KSA values. Briefly they have the following purposes.

- Focal value column – This column defines the targeted proficiency to be measured or fostered. Populating this column involves defining each KSA as either: (a) a *focal KSA* (i.e., an essential part of the targeted proficiency), which is shown as an integer in the matrix, or (b) a *nonfocal KSA*, which is shown as "n/a" in the matrix.
- Pre-intervention profile value and post-intervention value columns – These columns provide snapshots of the values of the profile (representing the person) before and after, respectively, the experience (e.g., assessment or instruction).
- Requirement value column – This column specifies for each KSA the requirement (cognitive or other demand) level imposed on the student by the task situation.
- Intended growth outcome column – This column indicates the post-experience level for each KSA for which the profile value is intended to increase (due to the experience). Of course, for a pure assessment, there would generally be no intent to foster any focal KSA.

As indicated above, once the matrix is constructed and values populated (step 1), rules and heuristics can be applied to evaluate the quality of the argument and fixes can be made (step), and, finally, the system can be implemented and validated (step 3).

Detail Of The Procedure

This section provides a detailed description of the procedure. Learning efficiency is a major point of this example, though it is explained fairly late in the procedure that it is explained. Through this explanation, one may see that one quality criterion (e.g., accessibility) can lay an important foundation for other criteria.

1. Construct the Argument

1.a. Define the Purpose and Target Population for the Learning-Centered Assessment

Let us consider the case of a game that is a learning-centered assessment for middle school mathematics (pre-algebra). It is delivered via computer with a computer monitor. The game employs graphics, text, and sound.

Let us suppose that we want this learning-centered assessment to be usable by and effective for diverse students, including students with disabilities. For the sake of simplicity, this extended example will focus on one student who is blind who begins with a low value in the targeted proficiency, which pertains to a specific concept in middle school pre-algebra. We want to portray a situation in which that person can end with a high value in that targeted proficiency.

1.b. Define Appropriate Quality Criteria

Because the game is intended to help students learn, we are interested in learning effectiveness and efficiency and in engagement (Gee, 2003; Squire, 2002). Because it is intended to function as an assessment as well, we are also concerned about validity.

1.c. Define the Purpose and Structure of the Argument

Let us suppose that the purpose of the argument, at least initially, is to portray the intended function of the learning-centered assessment as a whole, rather than of its individual components.[1]

We do this, realizing that the argument may need to be modified as we more fully face real life constraints and that, once the game has been implemented, it still needs to be evaluated (validated) (step 3).

1.d. Populate the Argument Structure with Values

Populate the argument structure with values that are suited to a particular case (e.g., a person having a particular profile interacting with the learning-centered assessment in a particular situation)

To indicate a person who has no usable sight, we put a zero ("0") in the pre-intervention profile column for the "see" KSA. Of course, each KSA has its own set of definitions of levels. For the sake of simplicity and clarify we will elaborate very little on the meaning of specific levels of any KSAs.

---

[1] Alternatively, we could have as the purpose of portraying the actual, problematic function of a system. Regardless of our purpose, it is matter for validation to determine if that purpose has been fulfilled.

Within the argument structure, we indicate that the person has a low (as opposed to high) targeted proficiency value. We do this by ensuring that for at least one focal KSA, the pre-intervention profile value is *less* than the focal value. In this case, there is only one focal KSA—“know math concept.” Recall that focal KSAs involve matrix rows that have a number in the focal value column and nonfocal KSAs have an “n/a” in that column. For that focal KSA, the profile value of “2” is less than the focal value of “3;” therefore the matrix is consistent with our intent to portray a student who begins with a low targeted proficiency level.

Probably the most common question about the use of this matrix in accessibility-supporting versions of ECD, is: “Where do values come from?” Basically the values are based on expertise, which is derived from some combination of theory, and experience-based judgment. This principle holds whether we are concerned with profile values (pre- and post-) or requirement values, focal values, or intended growth outcome values. While much could be said in terms of heuristics for constructing and populating the matrix, we will refer the reader to other work for additional detail (see Hansen, in submission). Through later design, implementation, or empirical evaluation, the definitions of the KSAs and their various values might be revisited, refined, and corrected. But, until one has the matrix populated with their initial values (even imperfect ones), it is difficult to proceed further.

In this case, the profile shown in the pre-intervention profile column is for a person who is blind and starts with a low level in the targeted proficiency.

A	B	C	D	E	F	H
Row number	KSA	Focal value	Pre-inter-vention profile value	Intended growth outcome	Require-ment value - original design	Post-inter-vention profile value
1	Engage	n/a	3	n/a	3	3
2	See	n/a	0	n/a	3	0
3	Hear	n/a	3	n/a	3	3
4	Know format and delivery system	n/a	2	3	3	3
5	Know game rules	n/a	1	3	3	3
6	Know game strategy	n/a	1	2	2	2
7	Know arithmetic	n/a	2	n/a	3	2
8	Know non-math vocabulary	n/a	2	n/a	3	3
9	Know math vocabulary	n/a	2	n/a	3	3
10	Know math concept	3	2	3	3	3

Figure 3. Initial matrix

2. Apply rules and heuristics to the argument to evaluate quality and address any quality problems

There are many rules and heuristics that one could focus on to check various aspects of the quality of the learning-centered assessment, and we will focus on just a few of the most important ones.

2.a. Verify a Transition to a High Targeted Proficiency Value

Having earlier ensured that the person begins with a low targeted proficiency value, we need now to ensure that the person ends with a high targeted proficiency value. By examining the matrix we see that post-intervention profile value for “know math concept" is "3," which is equal to the focal value ("3"), indicating that ends with the ability that we intended. (Stated more generally, the matrix adheres to the following definition: "If, for each focal KSA, the post-intervention profile value is at least as high as the focal, then the profile is indicative of ending with a high targeted proficiency value.") Earlier we confirmed that the initial (pre-intervention) profile indicates a low targeted proficiency value and we have now seen that the final (post-intervention) profile portrays a high targeted proficiency value.

2.b. Identify Unsatisfied Nonfocal Requirements and Related Performance Issues

There are indications that the design represented by the matrix has some important quality issues. The following rule is important for indicating potential accessibility problems.

> "If, for a nonfocal KSA, the profile value is less than the requirement value, then an unsatisfied nonfocal requirement exists."

Also, a related heuristic is:

> "If an unsatisfied nonfocal requirement exists, then an accessibility problem may exist, particularly if it is due to a person's disability."

The matrix shows unsatisfied nonfocal requirements for several KSAs. First and foremost, we notice it for "see," for which the pre- or post-intervention profile value "0" is less than the requirement value of "3." Let us assume here that a value of 3 for the "see" KSA refers to a level of sight that is "good" The student's profile value of "0" does not satisfy the nonfocal requirement for "see" ("3"). Other unsatisfied nonfocal requirements (relative to the pre-intervention profile), include those for: “know format and delivery system,” “know game rules,” “know game strategy,” “know arithmetic,” “know non-math vocabulary,” and “know math vocabulary.”

One nonfocal requirement that *is* satisfied (for both pre-intervention and post-intervention profiles) is that for the “engage” KSA, which is suggestive of a consistent state of engagement on the part of the student. More formally stated, the values in the matrix mean the person’s ability to engage (indicated by profile value) satisfies the demand (requirement) for that capability, thereby resulting in a psychological state of engagement. If the task situation were less intrinsically engaging and thereby imposed a higher requirement (demand) on the student’s ability to engage, then the person may not be able to satisfy the demand, thereby resulting in a state of disengagement.

Part of the reason that an unsatisfied nonfocal requirement is *not strictly equivalent* to an accessibility problem is lack of agreement about the precise definition of “accessibility.” However, the two concepts are very closely related, and it is likely that most of what people generally think of as an accessibility problem can be modeled as involving one or more unsatisfied nonfocal requirements.

Relationship between Requirement Satisfaction and Performance. Any one of these unsatisfied nonfocal requirements is sufficient to result in poor performance, based on the pre-intervention profile, even if the student’s targeted proficiency value is high. The basic principle is that:

> “If, for each KSA, the profile value is at least as high as the requirement value, then potential performance is high; otherwise potential performance is low.”

We use the phrase potential performance (sometimes called “effective proficiency”; Hansen & Mislevy, 2006), to describe the latent (hidden) capability (similar to a “true score” in educational measurement) for performing well in a particular task situation. This use of the word “potential” acknowledges that random error can prevent performance from matching the targeted proficiency level. Another way to acknowledge random error is with the following statement:

> “If, for each KSA, the profile value is at least as high as the requirement value, then performance is likely to be high; otherwise performance is likely to be low.”

The term “likely to be” might be used to account for random variation. With sufficient reliability (such as by increasing the number of tasks), the term “likely to be” becomes essentially unnecessary. Regardless of how one treats the reality of random error, an implication is that a given experience may need to entail multiple tasks in order to acquire sufficient reliability to inform decisions within the operation of a learning-centered assessment.

2.c. Address the Unsatisfied Nonfocal Requirements

How will we address these unsatisfied nonfocal requirements? There are basically two major strategies to address an unsatisfied nonfocal requirement.

1. Increase the student’s KSA level (profile value). One can attempt to increase the student’s level in the KSA. In the case of the KSA “see” for the student who is blind, this solution is likely not workable. However, in other cases, such as where there is an unsatisfied nonfocal requirement for, say, some prerequisite math

knowledge, an appropriate pedagogical strategy (e.g., instruction) might to raise the student's level.

2. Reduce the requirement (requirement value). One can reduce or eliminate the requirement that the student cannot satisfy and instead rely on a KSA in which the student has no deficit. For example, instead of relying on sight for the math assessment, one might present the assessment via read-aloud (e.g., text-to-speech), which relies on hearing instead of sight. Of course, the solution is applicable only if the student can hear.

A possible third strategy might be considered to use both strategies 1 and 2. This example will focus on the use of one strategy or the other.

Figure 4 shows a matrix for a revised design that addressees each nonfocal requirement. This matrix has several columns added to the original matrix.

===========================

INSERT FIGURE 4 HERE

===========================

1. Column G shows the requirement values for the revised design for easy comparison to the original requirement values in column F.
2. Column I identifies the KSAs for which there is an “unsatisfied nonfocal requirement” (UNR). In this case seven such KSAs have been identified, based on the pre-intervention profile and the requirements in the original design. As

explained earlier, an unsatisfied nonfocal requirement exists where, for a nonfocal KSA, the profile value is less than the requirement value.

3. Column K describes use of the strategy of *increasing the profile value*. In the rows 4 5, and 6, this strategy is used to address an unsatisfied *nonfocal* requirement. Specifically: “know format and delivery system” was addressed through practice and familiarization materials; “know game rules” was addressed via tutorial and/or practice; and “know game strategy” was addressed in the same manner. In the last instance, which is in row 10, the strategy was also used to address the unsatisfied *focal* requirement for “know math concept” via instruction. This growth in a *focal* KSA is at the heart of the purpose of the learning-centered assessment, but generally will not occur unless the person can satisfy the nonfocal requirements.

4. Column J describes use of the other strategy, that of *reducing the requirement* to address the unsatisfied nonfocal requirement. Specifically, the requirement for “see” was reduced (actually eliminated) by providing a read-aloud feature (e.g., text-to-speech); the requirement for “know arithmetic” was reduced by providing an electronic calculator; the requirement for “know non-math vocabulary” was reduced by using simpler non-math vocabulary; and the requirement for “know math vocabulary” was reduced by providing a math glossary. The reduced requirement values shown in column G are directly attributable to use of these accessibility features.

We see now that through the use of the strategies in the revised design, the nonfocal requirements are now satisfied. In other words, using actions mentioned in column J and the features mentioned in K, the revised design ensures that, for each nonfocal KSA, the pre-intervention profile value is at least as high as the requirement value. Thus, in this revised design there is no indication of inaccessibility.

It should be emphasized that one change in a design (e.g., changing a feature of a task situation) can impact other aspects of a design. One situation of concern is when an accessibility feature that is provided to address one accessibility problem generates new accessibility problems. Of course, the resulting problems may need to be represented and addressed in a revised design. For example, provision of an assistive technology (e.g., screen reader, refreshable braille display, magnification/enlargement software) will impose requirements for knowing how to use those technologies. If the student cannot satisfy those requirements and, assuming that those are nonfocal requirements, they give rise to new accessibility problems that must be addressed.

This brings us to consider another heuristic:

> "After making a change to a design, go back to the beginning of the process (1. Construct the argument) and start again."

This is the probably the safest way to proceed. As one becomes more expert in thinking through the issues, one may be able to recapitulate the process very quickly and almost automatically.

2.d. Evaluate Validity and Learning Effectiveness

Having addressed the unsatisfied nonfocal requirements, it often makes sense to examine (or reexamine) what are arguably the most important overarching characteristics of a learning-centered assessment—validity (because it is an assessment) and learning effectiveness (because it is intended to foster growth or learning of the targeted proficiency).

Learning Effectiveness. Early on we confirmed that the pre-intervention and post-intervention profile values in the matrix are consistent with learning effectiveness. But we also noted several problems that we have since fixed. We can now give more credibility to the assertion shown in the matrix that the student's profile value increased a low targeted proficiency value (indicated by a profile value of "2") to a high targeted proficiency value (indicated by a profile value of "3").

Making the most convincing case for the learning effectiveness would involve a randomized controlled trial (What Works Clearinghouse, 2006). However, a randomized controlled trial is rarely feasible or appropriate early in design because of cost and lack of implementation.

Early in design, one promising approach would be to *describe the pedagogical strategies* that would be used to foster learning (Hansen, in submission), preferably citing the research base associated with those strategies. Column J of Figure 4 provides a very general description of pedagogical approaches, e.g., practice and familiarization (row 4), tutorial and practice (rows 5 and 6), and instruction (row 10). These strategies could be specified in greater detail and in fact may include alternative strategies for individuals with different profiles, based on an understanding how individuals with a particular profile will most readily learn. However, even at the most general level, one heuristic is:

> "For each KSA for which there is specified an 'intended growth outcome,' describe a pedagogical strategy."

Validity. Another important characteristic is validity. The Standards (1999) state:

> Validity is a unitary concept. It is the degree to which all accumulated evidence supports the intended interpretation of test scores for the proposed purpose. (p. 11).

Many kinds of evidence bear upon the determination of whether the intended inferences are supported by the evidence. Some major varieties of evidence include: (a) test content, (b) response processes, (c) internal structure, and (d) relations to other variable, and (e) consequences of testing (Standards, 1999, pp. 11-17).

ECD seems to capture this notion of supporting the intended inferences. Stated more directly, the ECD approach discussed here captures the notion that if the task situation of the assessment does *not* impose the cognitive and other demands (requirements) specified in the definition of the targeted proficiency, it is then it is doubtful that the intended inferences are supported. The rule is:

> "If, for any focal KSA, the requirement value is not equal to the focal value, then there is an indication of a lack of validity."

In this case, for the focal KSA "know math concept," the requirement value of "3" equals the focal value of "3," which is suggestive of validity (i.e., there is no indication of a lack of validity). Of course, as noted earlier, the stated requirement value would not have been credible had not all unsatisfied nonfocal requirements become satisfied.

Thus, we now have indications that the argument structure portrays a design that addresses key quality issues of accessibility, learning effectiveness, and validity.

2.e. Address Learning Efficiency by Reducing Extraneous Cognitive Load.

What about learning efficiency? Are there any additional actions that would improve learning efficiency in the learning-centered assessment? This section focuses primarily on the idea improving learning efficiency by reducing extraneous cognitive load, an approach based on Cognitive Load Theory (CLT), which has garnered considerable empirical support over the last couple of decades (Sweller, 1988; Mayer, 2008; Clark, Nguyen, & Sweller, 2006). According to Clark, Nguyen, and Sweller (2006), the key to greater efficiency is reduction in "extraneous cognitive load" (p. 12), which may be thought of as "irrelevant load" (p. 13). The lower the extraneous cognitive load, the more of the limited capacity in working memory is available for germane (relevant) load imposed by the instructional techniques that serve the learning objectives. They further note:

> Fundamentally, cognitive load theory is about efficiency. Cognitive load theory defines efficiency in terms of two variables: learner performance and learner mental effort. Instructional environments that result in higher learning outcomes with less mental effort are more efficient than environments that lead to lower outcomes with greater mental effort. (p. 19)

How do we map the concepts of CLT to ECD for learning-centered assessments? A review of some of the literature on CLT suggests that extraneous cognitive load relates to the ECD concept of nonfocal requirements while germane (relevant) load relates to the ECD concept of focal requirements.

We then make the following suggestions:

1. Nonfocal requirements are candidates. Any nonfocal requirement is a candidate for reducing extraneous cognitive load. Especially relevant might be requirements for “cognitive” KSAs (e.g., “know non-content vocabulary”) as opposed to perceptual or sensory KSAs (“see,” “hear”).

2. Learning efficiency entails greater stringency than accessibility. While *accessibility* has basically to do with ensuring that, for a nonfocal KSA, the requirement value is *no higher* than the profile value, by contrast, *learning efficiency* is *more stringent*, because it has to do with ensuring that, for a nonfocal KSA, the requirement value is *less* than the profile value.

With this in mind, consider Figure 4 and identify some promising candidates for this approach to learning efficiency.

A Design That May Have Greater Learning Efficiency. Now consider Figure 5 to see an approach for greater learning efficiency. Figure 5 is an excerpt of the argument for rows 8 and 9 for the design for learning efficiency. Column G-1, which shows an additional requirement column, has been added relative to Figure 4. Specifically, column G-1 shows the *requirement values* for a *design revision for learning efficiency*.

Column K describes the features that drive specific reductions in requirement values. In row 8 (regarding "know non-math vocabulary") we have used a "yet simpler non-math vocabulary" to further reduce the requirement value from "2" (as shown in column G) to "1," as shown in column G-1. Note that the "l" is lower than the student's initial profile value of "2." Thus, with respect to non-math vocabulary the demand (requirement) is met with very little mental effort, which may be leave greater mental capacity (e.g., working memory) to deal with the germane load (relevant load) or focal requirements imposed by the learning experiences, thereby making learning more efficient.

A	B	C	D	E	F	G	G-1	H	I	J	K
Row number	**KSA**	**Focal value**	**Pre-inter-vention profile value**	**Intended growth outcome**	**Require-ment value - original design**	**Require-ment value - revised design**	**Require-ment value - revised design for efficiency**	**Post-inter-vention profile value**	**Unsatisfied nonfocal require-ment (UNR) based on pre-inter-vention profile**	**Means of growth in profile value (pre to post)**	**Reduction in requirement value**
8	Know non-math vocabulary	n/a	2	n/a	3	2	1	2	UNR	n/a	Further reduced RV from 2 to 1 via yet simpler non-math vocabulary
9	Know math vocabulary	n/a	2	n/a	3	2	1	2	UNR	n/a	Further reduced RV from 2 to 1 via yet simpler math vocabulary as well as a math glossary
10	Know math concept	3	2	3	3	3	3	3	n/a	Instruction	n/a
	Sum	*3*	*18*	*11*	*29*			*23*			
			Targeted proficiency value (TPV) before the intervention: **Low**					Targeted proficiency value (TPV) after the intervention: **High**			

Figure 5. Excerpt of the argument for rows 8 and 9 for the design for learning efficiency, showing the modifications in column K and the additional requirement column (column G-1) [Note to reviewer: A fuller version of Figure 5 is found at the end of this document]

In row 9 (regarding "know math vocabulary") we have used a "yet simpler math vocabulary" as well as the math glossary to further reduce the requirement value from

“2” (as shown in column G) to “1,” as shown in column G-1. As with row 8, the “l” is lower than the student’s initial profile value of “2.”

Both of these design changes seek to improve learning efficiency by reducing the non-focal requirement to a level that is lower than the profile value, based on the idea that it would result in reduce extraneous cognitive load, thereby resulting in greater learning efficiency.

Another way of increasing learning efficiency—increasing the profile value—could be carried out by providing prior learning experiences that raise profile values such that they *exceed* the requirements imposed by the learning experience, thereby resulting in greater learning efficiency in the learning-centered assessment being analyzed or designed.

Regardless of how it is achieved, the key to learning efficiency (based on this application of Cognitive Load Theory) is to ensure the requirement value for the nonfocal KSA is lower—perhaps much lower—than the profile value.

Note that low profile values are not always due to a disability or some other special status (e.g., English language learner). Therefore, these efficiency-related strategies might benefit students who have a low profile value in any nonfocal KSA for other reasons, such as: (a) lack of opportunity to learn, (b) difficulty in learning that falls short of qualifying as a disability, etc. These considerations suggest that this approach to learning efficiency can benefit students generally, not just students with disabilities.

3. Implement the system and validate its argument

Once the argument has been shown to meet all quality checks, including those related to feasibility, it is then the system is implemented and the argument validated.

3.a. Identify Appropriate Implementation Strategies

Successful validation of the argument depends, among other things, on the adequacy or fidelity of the implementation (Rossi, Lipsey, Freeman, 2004)

Different implementation methodologies can be employed. For example, an incremental approach based on functional prototypes that can be evaluated may often make sense. This would allow one to explore the usability and effectiveness of one functional prototype and then use those results to develop an improved functional prototype, thereby providing more early opportunities to identify and correct design problems. With such an approach the argument may be modified at several points in response to real-world constraints and well as to design goals.

3.b. Ensure proper matching of designs to people

One of the most important challenges of implementation is that of ensuring that a given design (and its argument) will be matched to the appropriate people. This leads to the following heuristic:

> "Provide a suitable method for matching designs to people. For example, in those cases in which particular task feature is to be made available only to individuals having one of a certain set of profiles, provide a procedure for ensuring that."

In high stakes testing situations, organizations have procedures, for example, for determining who is eligible to receive those features (e.g., accommodations). Obviously, the nature of the procedures for matching people to designs will vary widely depending on the factors such as the relative seriousness of the consequences of mismatches. However, fidelity in implementation depends on making appropriate matches of designs (including their associated profiles) to actual people.

3.c. Validate the Argument for the Design

Once the system has been implemented, its argument needs to be validated. A variety of methods could be undertaken, any of which may make the argument much stronger. For example, conducting a randomized controlled trial (What Works Clearinghouse, 2006) may make the argument much stronger with respect to learning effectiveness. This takes us to the end of the basic procedure for ECD for learning-centered assessments.

The methodology is intended to be useful at virtually any stage of the process of analysis, design, implementation, or validation of a learning-centered assessment. At an early stage, the methodology might be useful to provide a set of system requirement for a learning-centered assessment for one or more person profiles. Later on, the methodology may provide a research-based description of the real system interacting with persons corresponding to those profiles.

Discussion

The extended example was intended to illustrate the kind of reasoning needed to improve the efficiency of a learning-centered assessment. It may be seen that ECD provides an argument structure that is populated with values related to a particular person (or profile) and a particular design for a learning-centered assessment. It also provides heuristics and rules for checking the quality of the argument with respect to multiple quality criteria, e.g., learning effectiveness and efficiency, validity, accessibility, and engagement. Once the argument has been refined to address all applicable quality issues, it can be implemented, then validated.

For the purposes of this paper, the key idea is that the same kind of thinking needed to address accessibility – for example, about how to address unsatisfied nonfocal requirements – is highly relevant when seeking to improve efficiency of learning associated with a learning-centered assessment.

Basically, one first addresses accessibility by ensuring that, for each nonfocal KSA, the requirement value is *no higher* than the profile value of the student, meaning that the student can satisfy all nonfocal requirements. Then one takes that line of reasoning further to make it easier for the student to satisfy nonfocal requirements by ensuring that, for one or more nonfocal KSAs, the requirement value is *less* than the profile value.

The Importance of Fine Parsing of the Targeted Proficiency

One of the issues that might occur to the reader is the extent to which obtaining the intended efficiencies may rely on a very fine parsing of the targeted proficiency. For example, the argument structure in the extended example separated out math vocabulary and non-math vocabulary, applying somewhat different strategies for addressing their respective unsatisfied nonfocal requirement. This fine parsing of the targeted proficiency would be all the more important if the definition of the targeted proficiency were revised to include as focal KSAs not only "know math concept," but also "know math vocabulary.” In that case, one would be able offer requirement-reducing features (e.g., “accommodations”) for “know non-math vocabulary,” but *not* for “know math vocabulary,” because to do so would undermine validity. (Recall that in order to ensure validity the requirement value should be *equal* to the profile value, *not* lower or higher than it. (This assumes that we are referring to a summative type of assessment function rather than a formative one, the latter of which may allow, temporarily, a lower requirement value.)

There are other important considerations that make it important to be able to parse the targeted proficiency finely—including efforts to improve accessibility for individuals with learning and cognitive disabilities. For example, while it is relatively easy to determine whether “sensory” KSAs such “see” and “hear” are part of a targeted proficiency or not, many other abilities are not so obvious (e.g., those related to executive functioning, memory, and so on). The purpose of the method is not to prescribe whether such capabilities are parts of the targeted proficiency or not, but rather to provide a framework for reasoning through the implications of whatever definitional decisions are made.

Multiple Parts of an Argument

Depending on the purpose that the designer has for the argument structure, it may become important to break the argument into several pieces or “panels.” For example, each panel might correspond to a different phases of system activity, e.g., selecting students who can benefit from an experience, addressing unsatisfied nonfocal requirements, addressing focal requirements, assessing growth (Hansen & Zapata-Rivera, 2008b, Hansen, in submission; Feng, Hansen, & Zapata-Rivera, 2009).

Conclusions

This paper has sought to illustrate how ECD and Cognitive Load Theory provide a basis for identifying specific ways to increase the efficiency of learning in a learning-centered assessment. The key methods involve ensuring that the nonfocal requirements are lower than the profile values for nonfocal KSA. This level is lower than that required for accessibility. If validated, this approach may provide a basis for a new generation of technology-based learning-centered assessments that are not only accessible for students with disabilities but also more learning efficient for students more generally. Such benefits for all students may further help justify the investment needed to clearly articulate arguments associated with learning-centered assessments and then to validate them.

References

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

Clark, R., Nguyen, F., and Sweller, J. (2006). *Efficiency in Learning: Evidence-Based Guidelines to Manage Cognitive Load.* San Francisco: Pfeiffer.

Feng, M., Beck, J., Heffernan, N., & Koedinger, K. (2008, June). Can an intelligent tutoring system predict math proficiency as well as a standardized test? In R. S. J. d. Baker, T. Barnes, & J. E. Beck (Eds.), *Proceedings of the 1st international conference on education data mining* (pp. 107-116). Montreal, Quebec, Canada.

Feng, M., & Hansen, E. G., & Zapata-Rivera, D. (2009, April). *Using Evidence Centered Design for Learning (ECDL) to examine the Assistments system.* Paper presented at the annual meeting of the American Educational Research Association (AERA), San Diego, California, April 2009.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. New York: Palgrave Macmillan.

Hansen, E. G. (in submission). Extending evidence centered design for learning oriented products. Princeton, New Jersey: ETS Research Report Series.

Hansen, E. G., Fife, J., Graf, E. A., & Supernavage, M. (2008, December). *Toward a policy for the use of electronic calculators in CBAL mathematics.* [Internal memorandum]. Princeton, NJ: ETS.

Hansen, E. G., Galpern, A. J., & Goodman, M. J. (2008, March). *Evidence centered design for improving educational games*. Paper presented as the annual meeting of the American Educational Research Association (AERA), New York City, NY, March 27, 2008.

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., & Steinberg, L. S. (2008). *Evidence centered assessment design for reasoning about testing accommodations in NAEP reading and mathematics* (ETS Research Report RR-08-28). Princeton, NJ: ETS.

Hansen, E. G., Mislevy, R. J., Steinberg, L. S., Lee, M. J., & Forer, D. C. (2005). Accessibility of tests for individuals with disabilities within a validity framework. *System: An International Journal of Educational Technology and Applied Linguistics, 33*(1), 107-133.

Hansen, E. G. & Steinberg, L. S. (2004). *Evidence Centered Assessment Design for reasoning about testing accommodations in NAEP reading and math.* Paper commissioned by the Committee on Participation of English Language Learners and Students with Disabilities in NAEP and Other Large-Scale Assessments of the Board on Testing and Assessment (BOTA) of the National Research Council.

Hansen, E. G., & Zapata-Rivera, D. (2008a). *Extending Evidence Centered Design for Learning (ECDL) for diverse educational applications.* Final report for the ECD for Learning-oriented Products Pilot Project.

Hansen, E. G., & Zapata-Rivera, D. (2008b). *Evidence Centered Design for Learning: A method for identifying educational potential in games.* Final report for the Games and Learning Project.

Kyllonen, P. (2005, September). The case for noncognitive assessments. *R & D Connections* (pp. 1-7). Retrieved August 24, 2006, from http://www.ets.org/Media/Research/pdf/RD_Connections3.pdf.

Mayer, R. E. (2008). *Learning and instruction* (2nd edition). Columbus, Ohio: Pearson Prentice-Hall.

Minstrell, J., Kraus, P. (2005). *How students learn: History, mathematics, and science in the classroom.* Chapter 11. J. Bransford and S. Donovan (Eds.). Washington, D.C.: National Academies Press.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*, 3-67.

National Research Council. (2004). *Keeping score for all: The effects of inclusion and accommodation policies on large-scale educational assessment.* Committee on Participation of English Language Learners and Students with Disabilities in NAEP and Other Large-Scale Assessments. Judith A. Koenig and Lyle F. Bachman, Editors. Washington, D.C.: National Academies Press.

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

Razzaq, L., & Heffernan, N. T. (2006). Scaffolding vs. hints in the Assistment System. In M. Ikeda, K. D. Ashley, & T. W. Chan (Eds.), *Proceedings of the 8th international conference on intelligent tutoring systems (ITS 2006),* (pp. 635-644). Berlin: Springer Verlag.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach (7th ed.).* Thousand Oaks, CA: Sage.

Squire, K. (2002, July). Cultural framing of computer/video games. Game Studies, 2(1). Retrieved from http://www.gamestudies.org/0102/squire/ on September 28, 2007

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257-285.

Thissen-Roe, A., Hunt, E. B., Minstrell, J. (2004). Behavior Research Methods & Instrumentation, 36, 234.

What Works Clearinghouse (2006). What Works Clearinghouse study design classification. Retrieved April 6, 2009 from http://ies.ed.gov/ncee/wwc/pdf/studydesignclass.pdf.

Zapata-Rivera, D., & Hansen, E. (2009, April). *Analyzing the learning potential of existing games using evidence-centered design and cognitive task analysis.* Paper presented as the annual meeting of the American Educational Research Association, San Diego, CA.

Additional Figures

A	B	C	D	E	F	G	H	I	J	K
Row number	KSA	Focal value	Pre-inter-vention profile value	Intended growth outcome	Require-ment value - original design	Require-ment value - revised design	Post-inter-vention profile value	Unsatisfied nonfocal require-ment (UNR) based on pre-inter-vention profile	Means of growth in profile value (pre to post)	Reduction in requirement value
1	Engage	n/a	3	n/a	3	3	3	n/a	n/a	n/a
2	See	n/a	0	n/a	3	0	0	UNR	n/a	Reduced RV from 3 to 0 through read-aloud
3	Hear	n/a	3	n/a	3	3	3	n/a	n/a	n/a
4	Know format and delivery system	n/a	2	3	3	3	3	UNR	Practice and familiarization	n/a
5	Know game rules	n/a	1	3	3	3	3	UNR	Tutorial and practice	n/a
6	Know game strategy	n/a	1	2	2	2	2	UNR	Tutorial and practice	n/a
7	Know arithmetic	n/a	2	n/a	3	2	2	UNR	n/a	Reduced RV from 3 to 2 through electronic calculator
8	Know non-math vocabulary	n/a	2	n/a	3	2	2	UNR	n/a	Reduced RV from 3 to 2 via simpler non-math vocabulary
9	Know math vocabulary	n/a	2	n/a	3	2	2	UNR	n/a	Reduced RV from 3 to 2 via a math glossary
10	Know math concept	3	2	3	3	3	3	n/a	Instruction	n/a
	Sum	*3*	*18*	*11*	*29*		*23*			
			Targeted proficiency value (TPV) before the intervention: **Low**				Targeted proficiency value (TPV) after the intervention: **High**			

Figure 4. Revised design, that deals with unsatisfied nonfocal requirements

A	B	C	D	E	F	G	G-1	H	I	J	K
Row number	**KSA**	**Focal value**	**Pre-inter-vention profile value**	**Intended growth outcome**	**Require-ment value - original design**	**Require-ment value - revised design**	**Require-ment value - revised design for efficiency**	**Post-inter-vention profile value**	**Unsatisfied nonfocal require-ment (UNR) based on pre-inter-vention profile**	**Means of growth in profile value (pre to post)**	**Reduction in requirement value**
1	Engage	n/a	3	n/a	3	3	3	3	n/a	n/a	n/a
2	See	n/a	0	n/a	3	0	0	0	UNR	n/a	Reduced RV from 3 to 0 through read-aloud
3	Hear	n/a	3	n/a	3	3	3	3	n/a	n/a	n/a
4	Know format and delivery system	n/a	2	3	3	3	3	3	UNR	Practice and familiarization	n/a
5	Know game rules	n/a	1	3	3	3	3	3	UNR	Tutorial and practice	n/a
6	Know game strategy	n/a	1	2	2	2	3	2	UNR	Tutorial and practice	n/a
7	Know arithmetic	n/a	2	n/a	3	2	2	2	UNR	n/a	Reduced RV from 3 to 2 via an electronic calculator
8	Know non-math vocabulary	n/a	2	n/a	3	2	1	2	UNR	n/a	Further reduced RV from 2 to 1 via yet simpler non-math vocabulary
9	Know math vocabulary	n/a	2	n/a	3	2	1	2	UNR	n/a	Further reduced RV from 2 to 1 via yet simpler math vocabulary as well as a math glossary
10	Know math concept	3	2	3	3	3	3	3	n/a	Instruction	n/a
	Sum	*3*	*18*	*11*	*29*			*23*			
			Targeted proficiency value (TPV) before the intervention: **Low**					Targeted proficiency value (TPV) after the intervention: **High**			

Figure 5. Final design, with efficiencies at rows 8 and 9. [This is a fuller version of the Figure 5 found in the main body.]

Citation:

Hansen, E. G., Zapata-Rivera, D., & Feng, M. (2009, April). *Beyond accessibility: Evidence Centered Design for improving the efficiency of learning-centered assessments.* Presentation at the annual meeting of the National Council on Measurement in Education, San Diego, California. April 16, 2009.

Version: 10 April 2009, 17:51 pm