CBAL: Cognitively Based Assessment of, for, and as Learning

Randy Elliot Bennett
Educational Testing Service
Princeton, New Jersey 08541
USA

Paper presented at the annual conference of the International Association for Educational Assessment, Manila, Philippines, October 2011.

Abstract

CBAL, an acronym for *Cognitively Based Assessment of, for, and as Learning*, is a research initiative intended to create a model for an innovative primary-and-secondary-school assessment system that provides summative information for policy makers, as well as formative information for classroom instructional purposes. This paper describes the major design characteristics underlying CBAL assessment prototypes and summarizes results from almost 10,000 summative pilot administrations. The major design characteristics include (1) taking a systems perspective, (2) grounding development in the results of learning-sciences research, (3) incorporating innovative, computer-delivered tasks, (4) explicitly modeling good teaching and learning practice, and (5) employing distributed summative assessment. Pilot results to date suggest that CBAL summative assessments function well on average. Even so, operating characteristics have varied widely from one form to the next, suggesting that we may not yet be able to produce innovative test forms of this type in a replicable manner.

CBAL, an acronym for *Cognitively Based Assessment of, for, and as Learning*, is a research initiative intended to create a model for an innovative primary-and-secondary-school assessment system that (1) documents what students have achieved (of learning), (2) facilitates instructional planning (for learning), and (3) is considered by students and teachers to be a worthwhile educational experience in and of itself (as learning; Bennett & Gitomer, 2009). The model is built around a system of summative and formative assessments directed at satisfying the needs of states and state consortia, as well as needs at the district, school, classroom, and individual levels. This paper briefly describes the major design characteristics underlying CBAL assessment prototypes and summarizes results from almost 10,000 summative pilot administrations.

Key Design Characteristics

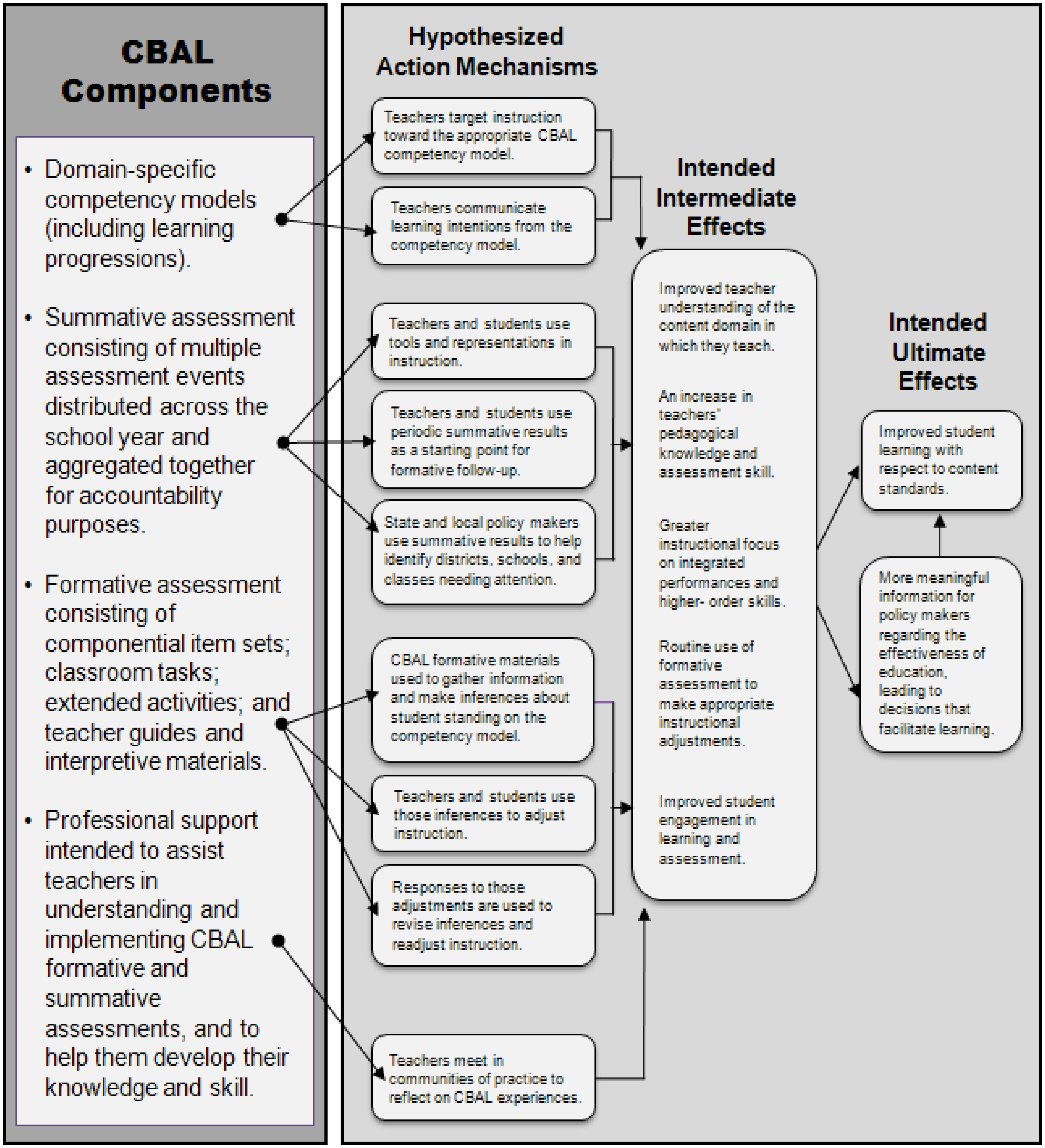
The major design characteristics underlying CBAL include: (1) taking a systems perspective, (2) grounding development in the results of learning-sciences research, (3) incorporating innovative, computer-delivered tasks, (4) explicitly modeling good teaching and learning practice, and (5) employing distributed summative assessment. Each of these design characteristics is described in turn.

Taking a Systems Perspective

Assessment exists within a larger educational context. If that context is to function effectively in educating students, its components must be coherent (Pellegrino, Chudowsky, & Glaser, 2001, p. 255). From an assessment-design perspective, coherence means devising summative assessment and formative assessment so that they mutually support one another in helping to achieve the goals of the education system (Gitomer & Duschl, 2007). For CBAL, coherence is facilitated through a detailed (but preliminary) theory of action. The CBAL theory of action directs development and research. The theory describes the system model's conceptual, summative, formative, and professional support components; the intended effects on individuals and institutions of implementing CBAL assessments; and how the system components are intended to work together to lead to those effects.

Figure 1 shows a logic diagram that summarizes the theory of action. Central to the model is the notion that, for primary and secondary education, an assessment system should have at least two central goals. One goal is to measure well because important decisions will be made on the basis of assessment results. A second goal is to have positive impact on individuals and institutions because an assessment should be a beneficial experience for all participants (Bennett & Gitomer, 2009). For a complete description of the theory of action and the research program designed to evaluate it, see Bennett (2010).

Figure 1. A logic diagram summarizing the CBAL theory of action. From Bennett (2010). Used by permission.



Grounding Development in Learning-Sciences Research

In addition to using content standards or syllabi, the CBAL system model incorporates the results of learning-sciences research as an underlying basis for assessment design. That research helps to identify the knowledge, processes, strategies, and habits of mind key to domain proficiency; the way in which those elements might be ordered for instructional purposes; and the teaching and learning practices that might help foster proficiency. CBAL assessment prototypes use this research to help exemplify and reinforce effective classroom practices for students and teachers. The domain-specific competency models, which synthesize this research and drive the design of CBAL assessment prototypes, can be found in Deane (2011), Graf (2009), and O'Reilly and Sheehan (2009).

As an example, the CBAL English language arts model crosses three modes of thought with five modes of cognitive representation (Deane, 2011). The modes of thought are interpretation, deliberation, and expression (which are closely related to reading, critical thinking, and writing, respectively). The modes of representation are social, conceptual, discourse, verbal, and print. At the intersection of each mode of thought and representation is a series of "skills foci," each of which calls into play a constellation of competencies important for a particular purpose. Associated with each skills focus is a provisional learning progression intended to mark qualitative changes in the development of that focus from late primary school to university level. Table 1 gives such a provisional progression for "argument building," a skills focus falling at the intersection of the conceptual and deliberation modes.

Table 1. A provisional learning progression for argument building. From Deane, Sabatini, & O'Reilly (2010). Used by permission.

PRELIMINARY: Can distinguish reasons from non-reasons and infer whether reasons would be used to support or oppose a position.
FOUNDATIONAL: Can self-generate multiple reasons to support an opinion.
BASIC: Can rank and select reasons by how convincing they seem; Can distinguish facts and details that strengthen a point from those that weaken it; can distinguish between reasoning that seems convincing because one agrees with it and reasoning that seems convincing because of the content of the argument.
INTERMEDIATE: Can recognize counter examples. Can distinguish valid from invalid arguments and recognize unsupported claims and obvious fallacies.
ADVANCED: Can identify and question the warrants of arguments, distinguish necessary and sufficient evidence, and synthesize a position from many sources of evidence, using that to identify key evidence and propose new lines of argument.

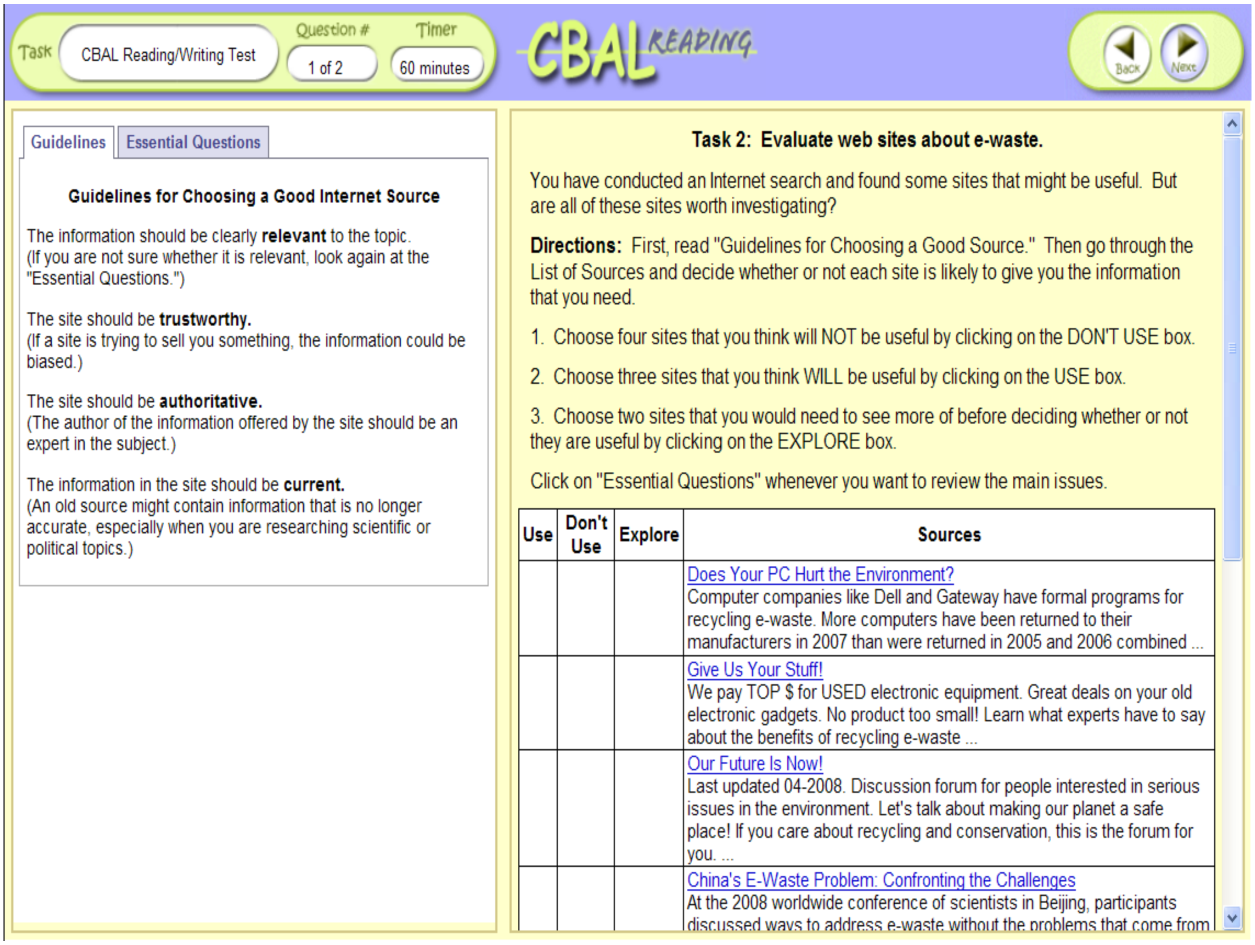
Incorporating Innovative Tasks

CBAL assessment designs incorporate innovative, computer-delivered tasks targeted at measuring both higher-order thinking and lower-level, but still essential, component skills. Each CBAL assessment prototype includes at least one “scenario-based” task set, which is composed of a series of selected-response and constructed-response questions. All questions are linked to a common context that directs the examinee toward satisfying a given goal (e.g., make a recommendation to your school principal about whether students should be required to wear uniforms). This linkage of questions to a common context, and the inclusion of an extended constructed-response task, gives CBAL scenario-based task-sets the character of structured performance exercises.

Modeling Good Teaching and Learning Practice

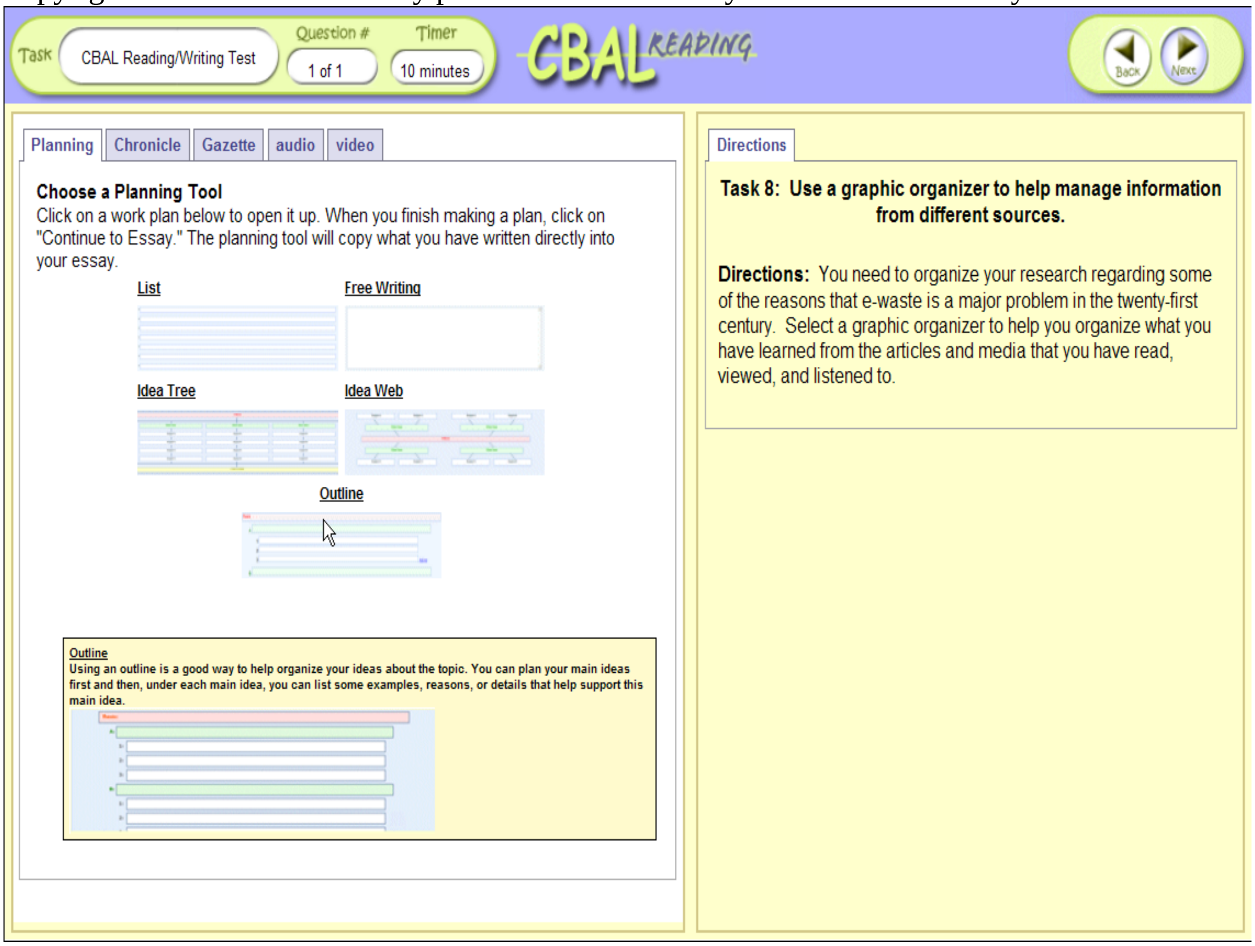
CBAL prototype assessments attempt to model good teaching and learning practice in at least four ways. First, the assessments routinely include tools and representations similar to ones proficient performers use in their domain practice. For example, proficient performers typically possess internalized standards and criteria for what constitutes good work in the domain. Skilled performers habitually judge their own work against those standards and criteria, going back and forth between an unfinished work product and the criteria more or less automatically. In line with this idea, most CBAL summative and formative assessments present conventional criteria (e.g., for evaluating the quality of a summary, of the information presented on the Internet, of a good persuasive essay) and ask students to apply those criteria *repeatedly* throughout the test. The inclusion of such criteria is aimed at encouraging their frequent use by learners (and by teachers) to the point that they become a "habit of mind." Figure 2 gives an example of how students are asked to apply criteria to the evaluation of web-site summaries. The example comes from a CBAL reading task built around the problem of electronic waste (e-waste).

Figure 2. Modeling good teaching and learning practice through the use of conventional criteria. Copyright © 2009 ETS. Used by permission. Created by Kathy Sheehan, Heather Nadelman, and Barbara Elkins.



As an additional example, proficient performers often use planning aids. Figure 3 shows a screen from a CBAL writing assessment in which students have access to several such tools for organizing their thoughts. Each tool is built around a different writing structure--including the outline, idea tree, idea web, and list--and each tool enforces the conventions of that structure (e.g., the outline tool enforces indentation). A given tool is activated by clicking on the appropriate icon. When the student is done with the tool, he or she can transfer the resulting plan into the response area for use in producing the essay.

Figure 3. A screen showing a palette of planning aids for students to use in writing assessment. Copyright © 2009 ETS. Used by permission. Created by Paul Deane and Mary Fowles.

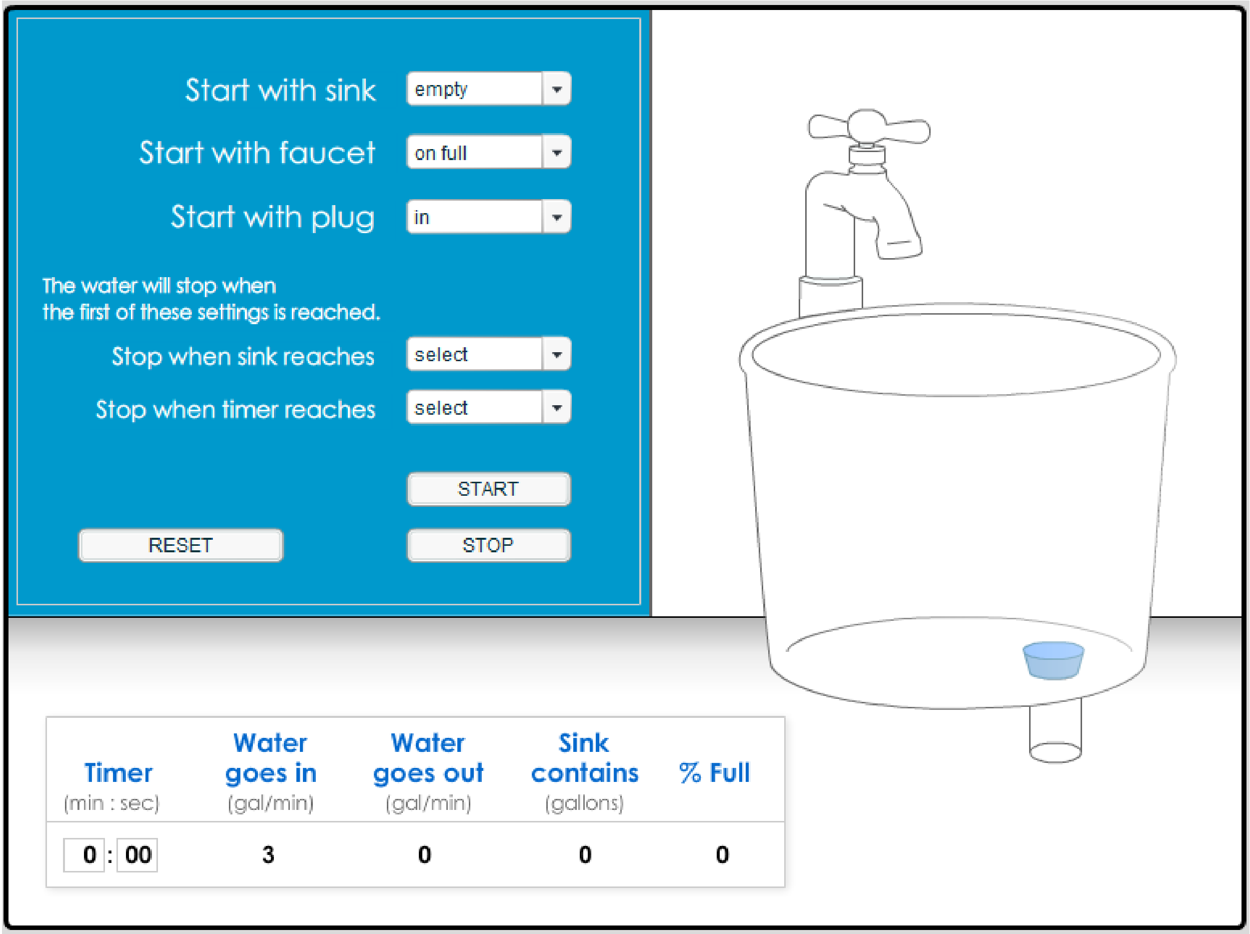


A second way in which CBAL assessment prototypes attempt to model good teaching and learning practice is by presenting reasonably realistic problem-contexts. Such contexts are intended to increase engagement and help students connect solution processes and strategies to the conditions under which they might actually be used. Figure 4 gives an example of the introduction to such a task. The task revolves around a lake that feeds water through a dam for electricity generation and irrigation of the land downstream. Because of a drought, it is uncertain as to whether the lake will recede to the point that water can no longer flow through the dam. The challenge for the students is to determine if the available data justify instituting emergency water conservation measures. To find an answer, each student must solve a series of problems involving linear functions and then formulate an argument with appropriate mathematical backing.

Figure 4. An introduction to a task presenting a reasonably realistic problem-context. Copyright © 2010 ETS. Used by permission. Created by Karen Harris.

A third approach used in CBAL to model good teaching and learning practice is to design assessment tasks to help students (and teachers) connect qualitative understanding with formalism. Such connections are especially important when using realistic problem contexts, some of which will inevitably be more or less familiar to individual students. The dam-and-lake scenario above is built around the mathematics of inflow and outflow. To help ensure that students come to the assessment task with a common qualitative understanding of inflow and outflow that can be connected with mathematical formalism, we first present students with a simulation of a sink (see Figure 5). In that simulation, the faucet represents the river feeding the lake, the basin stands for the lake, and the plug corresponds to the dam. Students can manipulate the state of the faucet, sink, and plug; understand what happens to the water level under different states; and mathematically model the relationships among the variables before applying similar modeling to the lake and dam.

Figure 5. A sink simulation used to help students (and teachers) connect qualitative understanding to mathematical formalism. Copyright © 2010 ETS. Used by permission. Created by Karen Harris.



The last approach to modeling good teaching and learning practice is realized through the structure of the CBAL scenario-based task sets. Those sets begin with a series of “lead-in” tasks and close with a performance requiring the integration of multiple competencies, called a "culminating task." The lead-in tasks have several purposes. First, they are intended to activate prior knowledge. Second, they are used to measure prerequisite competencies needed for the culminating performance. In CBAL writing, for example, the lead-in tasks target specific reading, as well as critical thinking competencies, both of which are required for the culminating performance. Finally, because they function similarly to discrete items, lead-in tasks are intended to increase generalizability, a persistent problem in performance assessment (Linn & Burton, 1994), by providing additional observations of student behavior. Because it is used in both CBAL formative and summative assessments, this scenario-based-task structure (i.e., lead-in plus culminating task) can help suggest to the teacher how the skills required for more complex performances might be decomposed. That scaffolded decomposition should be useful to the teacher for planning instruction and as a general heuristic for classroom assessment.

Employing Distributed Summative Assessment

A final design characteristic underlying the CBAL model is that it employs periodic (or distributed) summative assessment, which consists of several administrations spread across the school year (Bennett & Gitomer, 2009). This distribution should allow for the greater use of performance tasks because more time is available for assessment. Distributing assessment should also provide more timely information to teachers and, when the results are aggregated across occasions, it should create a stronger evidential basis for high-stakes decision making. That aggregation of results could be engineered such that it represents a weighted sum of student accomplishment across the year or the best estimate of the student's status at year end.

Selected Results from Piloting CBAL Summative Assessment Prototypes

From 2007 to 2010, almost 10,000 CBAL prototype summative assessments built to incorporate the design characteristics described above were administered online. Those pilot administrations were conducted in the middle-school grades in more than a dozen US states. The pilots were conducted for two main purposes: (1) to try out assessment designs and tasks so that they could be improved and (2) to gather the data needed to address scientific questions. A detailed summary of results from those pilots relating to the technical quality of the prototypes as measurement instruments can be found in Bennett (2011). That summary speaks to the theory-of-action goal that CBAL assessments "measure well." Studies relating to the second goal, that the assessments should have positive impact, will follow.

The main results reported by Bennett (2011) are shown in Table 2. In general, CBAL summative assessments appeared to work empirically as intended. They were appropriate in difficulty for the student samples examined. The assessments produced reliable scores that were associated with other measures of the same competencies to reasonable degrees. The prototypes generally measured one dimension, within and across test forms, offering the possibility for a simple approach to score aggregation. Finally, the results indicate that agreement statistics for automated scoring ranged from marginally smaller than the comparable statistics for human raters to considerably different, with the least acceptable performance occurring for short-text responses in mathematics (e.g., justifications).

Although the CBAL assessments appeared to generally function adequately well on average, Bennett (2011) also noted that considerable variability was evident from one test form to the next, particularly in the writing and math prototypes. Some random fluctuation is to be expected because most of the pilot administrations involved small samples and because new design and task ideas were tried out over time. However, fluctuation might also suggest that we do not yet possess the knowledge required to create high-quality CBAL test forms in a replicable manner.

Conclusion

This paper described the major design characteristics and pilot results associated with CBAL, a research initiative intended to create a model for an innovative primary-and-secondary-school assessment system that provides summative information for policy makers, as well as formative information for students and teachers. From four years of iterative development and data collection, we have drawn two major conclusions associated with creating the CBAL model. The first conclusion is that we believe we can *design* summative and formative assessments to: fit synergistically with one another; incorporate the results of learning-sciences research; use innovative tasks; be delivered by computer; model good teaching and learning practice; and follow a distributed structure. The second conclusion is that our pilot data offer supportive evidence for some aspects of instrument technical quality. Other aspects remain to be investigated.

Among the significant outstanding challenges is learning how to create these assessments so that they consistently function well. Being able to produce forms consistently with the desired operating characteristics is a requirement for any consequential assessment program. A second significant challenge is evaluating impact on classroom practice to verify that, in fact, the good intentions behind our assessment design efforts have been realized. Third, there are many measurement issues needing attention, including how to aggregate scores across distributed assessments, ensure fairness for special populations (e.g., students with disabilities or who are English language learners), measure growth, and provide meaningful formative information from the summative test. Finally, assuming that the above challenges can be met, significant effort must be devoted to bringing the CBAL system model from prototype to production. Such a transition will require an infrastructure that can affordably and rapidly produce, deliver, analyze, score, and report results from these innovative assessments.

Table 2

Summary of Key Results Across Three CBAL Content Areas

Content area (and # of form admini-strations)	# of tests	Median of the median *p*+ values	Median of the median % missing	Median coeff. alpha	Median cross-form correlation	Most frequent factor analytic result	Median correlation with other tests of the same skill	Median Automated-human vs. median human-human agreement
Reading (6)	3,062	.51	0%	.88	.78	1 factor within & across forms	.74[a]	.72 vs. .82 [c]
Writing (9)	5,410	.57	1%	.82	.73	1 factor within & across forms	----	.79 vs. .82 [d]
Math (12)	1,347	.45	6%	.92	----	1 factor within forms	.76[b]	.67 vs. .83 [c]

Note. Adapted from Bennett (2011). Used by permission.

[a] Median of 13 correlations, each for either a different criterion measure, scale, or grade. [b] Median of 13 correlations, each for either a different criterion measure or grade. [c] Each value is a median of a set of unweighted kappa values. [d] Each value is a median of a set of correlation coefficients.

References

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning: A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives, 8,* 70–91.

Bennett, R. E. (2011). CBAL: *Results from piloting innovative K-12 assessments* (Research Report 11-23). Princeton, NJ: Educational Testing Service.

Bennett, R. E., & Gitomer, D. H. (2009). Transforming K–12 assessment: Integrating accountability testing, formative assessment, and professional support. In C. Wyatt-Smith & J. Cumming (Eds.), *Educational assessment in the 21st century* (pp. 43–61). New York, NY: Springer.

Deane, P. (2011). *Writing assessment and cognition* (Research Report 11-14). Princeton, NJ: Educational Testing Service.

Deane, P., Sabatini, J., & O'Reilly, T. (2010). *2010 CBAL writing conceptual project report: Elaborating a joint reading/writing literacy model and linking it to the Common Core State Standards*. Unpublished manuscript, Educational Testing Service, Princeton, NJ.

Gitomer, D. H., & Duschl, R. A. (2007). Establishing multilevel coherence in assessment. In P. A. Moss (Ed.), *Evidence and decision making. The 106th yearbook of the National Society for the Study of Education, Part I* (pp. 288-320). Chicago: National Society for the Study of Education.

Graf, E. A. (2009). *Defining mathematics competency in the service of cognitively based* assessment for grades 6 through 8 (Research Report 09-42). Princeton, NJ: Educational Testing Service.

Linn, R. L., & Burton, E. (1994). Performance-based assessment: Implications of task specificity. *Educational Measurement: Issues and Practice, 13(1),* 5-8.

O'Reilly, T., & Sheehan, K. M. (2009). *Cognitively based assessment of, for and as learning: A framework for assessing reading competency* (Research Report 09-26). Princeton, NJ: Educational Testing Service.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington DC: National Academies Press.