

From Usable to Useful Assessment Knowledge and Evaluation

In this article contrasts are made between usable and useful knowledge. Usable knowledge represents knowledge that someone finds pertinent. Useful knowledge has a function and can lead to specific action. The applications to educational assessment and evaluation are clear in that much assessment knowledge is developed at a level that is nominally usable, but provides no real guidance in the improvement of learning. Knowledge management through the measurement of social and organizational capital is suggested as a strategy to augment accountability policies based exclusively on test scores. This approach would work as a way to set targets and monitor the development of schools as an institution, responsible for performance by adults and students that extend beyond measured performance.

Keywords: knowledge management, assessment, social capital, accountability.

Eb016

Eva L. Baker¹

UCLA Graduate School of
Education & Information
Studies.

Center for the Study of
Evaluation National Center
for Research on Evaluation,
Standards, and Student
Testing (CRESST)
eva@ucla.edu

Del conocimiento utilizable al conocimiento útil en evaluación

En este trabajo se contraponen los conceptos de conocimiento utilizable y conocimiento útil en evaluación. Un conocimiento utilizable es aquel que alguien encuentra pertinente, mientras

¹ The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305A050004, as administered by the National Center for Education Research, Institute of Education Sciences, U.S. Department of Education, and under Office of Naval Research Award Number #N00014-04-1-0209, as administered by the Office of Naval Research.

The findings and opinions expressed herein do not reflect the positions and policies of the National Center for Education Research, the Institute of Education Sciences, the U.S. Department of Education, or the Office of Naval Research.

que un conocimiento útil es aquel que tiene una función y puede conducir a una acción específica. Las implicaciones de esto para la evaluación educativa son claras dado que la evaluación del conocimiento se desarrolla a un nivel que es nominalmente utilizable, pero que no ofrece una orientación real para la mejora del aprendizaje. Se sugiere que la gestión del conocimiento a través de la medida del capital social y organizacional, puede ser una estrategia que aumente las políticas de rendición de cuentas basadas exclusivamente en las puntuaciones de los tests. Este enfoque podría ser útil para establecer objetivos y monitorizar el desarrollo de las escuelas como instituciones, responsables del rendimiento de adultos y estudiantes, más allá del rendimiento simplemente medido.

Palabras clave: gestión del conocimiento, evaluación, capital social, rendición de cuentas.

1. THE RATIONALE

Policymakers are fond of saying that we have enough research knowledge especially about assessment. They see the problem, rather as a lack of effective application of the knowledge we already have. Let's focus on the range of knowledge needed to improve students' learning and, in particular, what to do about it in the assessment and testing arena. Start with a consideration of how knowledge is supposed to be used, and where assessment falls short. Then review potential solutions, some unsurprisingly drawn from my own research.

There are types of knowledge that are characterized as basic or fundamental, that may or may not be applied in the foreseeable future. I am interested in education, where application of new knowledge is vitally needed and cannot be put off for the next –or the following– generation of 10-year-olds. If research knowledge is to be applied in a reasonable period of time, say in five years, it should possess certain properties. The knowledge must be both *usable* and *useful*. Some time will be given to the distinctions between these two ideas, in order to make a case for how we should design our systems and to help us act with greater intelligence for the good of all students.

Let's start with the negative. Each of us knows what *unusable* knowledge is. It includes, but is not limited to, unrecalled names of colleagues unexpectedly encountered at conferences; comments that, while accurate, are far too impolite to utter; random information we know but can't remember why; and fractions of ideas that we can't seem to mold into a functional whole. The most frustrating variant of unusable knowledge is the one in which we know part or all that should be done, with inherent goals valuable to attain, but we remain nonetheless unable to make significant progress applying the partial forms of knowledge that we have at our disposal. This type of unusable knowledge seems to pertain to many aspects of educational improvement.

"Usable knowledge" was a term used by Lindblom and Cohen (1979) in a book that focused on the relationship of the social sciences and problem solving in the real world. If you see the actual book, you will notice that in the title, the words "usable knowledge" are

not capitalized, either a sign of the authors' sense of its rarity, its salience, or an e. e. cummings allusion. Remember, Lindblom and Cohen were specifically considering research and evaluation knowledge, and they found it wanting. Rarely could they find instances that were usable for the specific social problems at hand. Thirty years later, what's new?

Is there a kind of knowledge that is valuable for our current situations? Let's exclude obviously wrong information, or an answer to a question not asked. It is helpful to make the distinction between *usable* and *useful* knowledge. For knowledge to be *usable*, it needs to be understood and then translated into practical terms so that it has the potential to be applied to a particular situation or problem. To be *useful*, knowledge permits us to *act* by changing the problem into a solvable form, and leading to the development of a solution or greater insight. The difference between usable and useful knowledge is analogous to the contrast between potency and action, a revered St. Thomas Aquinas formulation. Useful knowledge is effective.

In a reference of the same vintage as Lindblom and Cohen, Carol Weiss (1977) discussed how research and evaluation findings could be usable, even though they might not directly lead to a decision about a given situation. Their utility is indirect and functions to illuminate a murky or tangled problem so that new approaches are possible. Illumination literally means that we see the problem in a new light and are enabled to reconceive it. For a homely example, we know that it takes many measures and complex equations in order to forecast the weather. If they were not transformed, combined, and summarized verbally, most of us would have no idea what they meant. Yet, this complex knowledge can be translated into an understandable, usable form, such as "Get ready for rain". If there were no rain, the information would have been usable, but not useful. So the best kind of knowledge would be first usable (able to be translated into relevant and understandable language) and then useful, to help us to act, to improve demonstrably our situation (get a working umbrella).

2. USABLE AND USEFUL KNOWLEDGE IN SCHOOL REFORM

Why are some schools or institutions successful in finding knowledge and making it both usable and useful? I believe, besides luck, that the people in these organizations exhibit certain recurring predispositions that add up to effectiveness. First, they focus on their primary business: In schools, that is *learning* of many types, both by students and by educators. Second, they embrace information of both formal and informal types of knowledge, finding ways to integrate sources of information into a kind of coherence. Third, their staffs make information public and exchangeable. And last, they take pride in the outcomes they achieve.

The remainder of this article will link the concepts of usable and useful knowledge to a specific feature of school reform: assessment -that is, the testing of children or other students, for the purposes of certification, instructional improvement, system monitoring, the evaluation

of educational services, and in some cases, accountability and the consequences that follow. How does and how could assessment knowledge play out in a school or other training environment?

3. ASSESSMENT, THE EVER-NEW SOLUTION

Without question, policymakers in education and training at all levels, worldwide, share the belief that assessment knowledge will help them and people in the educational systems to solve problems of teaching, learning, and management. Interest in assessment, of course, preceded and now continues to extend far beyond its use in education. Assessment suffuses the fields of health environment, transportation, criminal justice, and social services practices and policies. Because assessment results (how well individuals or institutions score on tests or other indicators) are often decisive in determining access to an improved economic status, it is easy to see why regions, states, and countries have ascribed the great power to test scores -a way to project winners and losers. Nor is it lost on us that both the private and government sectors have picked up the assessment talisman, for situations ranging from matters of convenience to those of survival. For instance, we rate restaurant cleanliness, nursing homes, quality of intensive care in major hospitals, and environmental factors such as sunshine and amount of pollution. No one in Los Angeles would dream of going to a restaurant with less than an "A" rating, even if the air, inside or out, were unbreathable. We pay attention to one rating because it seems to be related to actions people take. It is useful. In the case of smog, we have a greater challenge. We have usable (the pollution rating) but not really useful knowledge, except under extreme conditions (it is too polluted to drive, play, or run outside).

For those directly interested in education and training, among policymakers and planners, the level of enthusiasm for assessment as a management tool is palpable (e.g., No Child Left Behind Act of 2001, 2002). Policymakers' interests include precollegiate schools, early education, university and college education, formal preparation for the workplace, and training and development in business, the military, and other technical sectors. Why? Managers have well learned the lessons of numbers. Assessment information provides a quantitative measure that allows them to make distinctions among people and organizations. These differences can be easily summarized. They can, it is believed, distinguish between the better and the best. They use the classifications that they create (novice, proficient, expert) as a convenient way to allocate services and sanctions. The assessment procedures they have adopted are but an inexpensive fraction of the cost of total services, and usually selected the least costly alternative. Are the numbers valid? Not an issue.

We should note that a massive change in the role of assessment has occurred in the last two decades. The measurements (tests and so on) have had their usability nominally transformed. Instead of just showing (albeit approximately) how well students and organizations perform, with the understanding that the performance is merely an estimate or a sample of what people might do, given longer times, greater flexibility, or a slightly different

problem set, now the scores on the test have become ends in themselves. There is a widespread belief that there is no better way to measure learning than to obtain a test score, any test score, and that good things should always be inferred from test scores going up, and all bad things accrue from test scores going down. As a result, in many sectors, a second-stage transformation has occurred that legitimates only those activities congruent with the test content and form, with a result that topics or areas that are untested also remain untaught. And so, policymakers have made what was the *measurement* into the educational or training *intervention* (Baker, 1982; Resnick & Resnick, 1992). The public has embraced this notion, thinking that if we had good tests, then this natural focus would serve both children and the educational systems in which they participate.

Additional justification for the use of test performance as the key measure of quality is the evidence that economic success is somewhat connected to prior test performance. For regions, states, or countries in competition with one another, attracting families and businesses with high test scores has become part of the regular sales pitch for real estate people, with the result that children will be with children like themselves, at least on an academic measure or on a larger scale, investment in states or nations with good test scores (Organisation for Economic Co-Operation and Development, 2005).

So we have, as some might say, a neat, closed system. A relatively cheap set of instruments allows us to distinguish among people and among institutions, on the basis of proficiency of some sort, and the rewards and sanctions that follow (known as accountability) have made performance on these instruments paramount. What could possibly be wrong with this approach? One way to pose the question is this: Are assessment results generating unusable, usable, or useful knowledge? Or in other words, are assessments solving problems directly related to learning?

Let's take a moment to consider what the assessment of learning is all about, how it may work, and how it often does work. Then we will apply to assessment the standards of unusable, usable, or useful knowledge.

Assessment is a feature of all learning, even if it is only the informal self-questioning that occurs when one has patiently turned book leaves and realizes that not one idea has been remembered (Dembo & Junge, 2005). Assessment is also the process in which tender inquiries are made by teachers asking a child to explain why a particular strategy was proposed (Black & Wiliam, 1998). More commonly, assessments are formal examinations, generated internally or externally, and are usually intended to hold consequences for the examinee.

These consequences may involve receiving a teacher's mark or grade, a score allowing access to a particular course, a diploma with or without endorsement, or on the downside, a signal to reconfigure an individual's goals, plans, and hopes. Assessment in its purest form gives feedback, and the more adapted the assessment is to what the learner is experiencing

and the capacities and learnings brought into the assessment situation, the more likely the assessment will promote growth and accomplishment. It will be useful. At least that is the story (Baker, 2007a; Nyquist, 2003; Pellegrino, Chudowsky & Glaser, 2001; Resnick & Resnick, 1992).

Reality, unfortunately, stands apart. Assessment data, when used on a large scale for system monitoring purposes, may be neither adaptive nor appropriate to the learner, and may not provide information that can be used by the student or the teacher. Some assessments may likely conflict with other data and create a problem in understanding what it all means. Furthermore, assessment data may not be in a form that is easily digested. Moreover, because formal assessments have a quantitative flavor -statistical transformations and the like- they exude science, and that alone may swamp the credibility of other sources of assessment knowledge. We must get inside the makings of tests to see how they relate to learning and what sense can really be made from data rather than adopting a passive position and assuming that experts have done the thing right.

4. WHAT TYPES OF KNOWLEDGE ABOUT ASSESSMENT ITSELF DO WE REALLY NEED AS EDUCATORS?

I claim that we first must know the purpose, or purposes, of one or more assessments (tests American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999). Who is really the primary audience? Who is to make use of the information, and what decisions will hinge on it? These questions are rapidly followed by questions of what to assess, who gets examined and how frequently, how to design assessments, how to interpret results, and how to determine whether the results are to be trusted, in the light of the purposes they serve (or validity). Are examinations evaluating a particular reading program appropriate to gauge school effectiveness? Should university admissions test scores (as in the United States) be used to judge the quality of secondary schools? These are validity questions.

To address this set of problems, I believe that we must make sure that the assessments given for *whatever purpose*, whether large scale or in the classroom, serve first the learning and the learners. At the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), we have been working for 20 years or so on a strategy to design and implement assessments so that they meet three criteria: a) They lead to coherent, sustained learning; b) they support a spiral or progression of learning, each enhancing and linking what has come before; and c) they direct students to knowledge and skills are adaptive, that that can be transferred (in psychological terms) or applied to new or unforeseen situations. The formats in or authorities under which the assessments are given are less important than the learning they actually stimulate (Baker, 1997, 2007a, 2008; Baker, Abedi, Linn & Niemi, 1996).

In brief, CRESST models are based on research knowledge. To design a test or assessment, we first focus on desired student cognition and learning (Wittrock & Baker, 1991). We then

ESTUDIOS

FROM USABLE TO USEFUL
ASSESSMENT KNOWLEDGE
AND EVALUATION

reverse the usual way tests are built: Instead of starting with subject matter –world history, for instance– we begin with the cognitive expectations –are we focusing on communication, content understanding, problem solving, or some combination? We employ a structure that transforms content goals, cognitive demands, and what some call 21st century skills (principally being adaptable and able to transfer knowledge and skills to unknown settings). We call this structure ontology; it is software based and allows for transparency in the key elements (such as principles) and their relationships to other kinds of knowledge. After deciding on the family of cognitive demands, we use a template (representing a translation of research into a usable form, Baker, 2003a, 2007b). Then, we return to the subject matter domain and apply a template or structure of the assessment model to it, substituting content or examples as needed. In history, we could address Asia in the 19th century, the Spanish expansion, or the complexities of the Cold War using the same general model and specific template (see Baker, Freeman & Clayton, 1991). This approach forces a level of coherence among sets of assessment tasks, among subject matters, and among authorities (bureaucratic levels) administering the test. It allows teachers to "line up" their instruction and formative assessments with external mandates without corrupting the tests or their view of teaching. It also supports vertical integration from grade to grade. It saves money because it allows important task architecture to be reused (Baker et al., 1996). While this isn't the forum for detailed procedures, let me provide a brief sketch of how this works.

Here is a representation of the approach, with learning in the middle of five key families of intellectual skills or cognitive demands: content understanding, problem solving, communication, metacognition (or actively controlling your own learning) and teamwork and collaboration (Baker & Mayer, 1999; see Figure 1 in Appendix). I'll illustrate the content understanding model in part.

5. THE TEMPLATES

First we figured out what the key elements are, from research, that describe significant content understanding. Accept for the moment that they are a) understanding the big ideas in a domain, b) seeing their relationships, c) avoiding misconceptions, and d) using prior knowledge and resources to convey meaning (Baker, Herman & Linn, 2004). In content understanding, I'll illustrate two examples of templates drawn from the model that produce different looking tasks but share the same deep infrastructure. We at CRESST (with teacher advice) have agreed that we want children to read or encounter real text, or representations of artifacts, whether historical or current, or literary, scientific, or artistic. Thus the specification for the task requires the presentation of primary source materials. We also need –as the research supports– students to demonstrate that they can integrate specific prior knowledge with higher principles or themes. This process of translating research into models, models into templates, and templates into coherent assessments represents our strategy for making research knowledge "usable." We have decided, again based on substan-

tiated cognitive research (Chi, Glaser & Farr, 1988) and a modicum of logic, that students' work should be scored on models based on experts' performance rather than extracting from the discussion of the abstract idea of what "good" work is (Baker, 2003a).

Each cognitive model can generate multiple templates (Baker, 2003b; Vendlinski, Baker & Niemi, 2008). In the first example of a template for content understanding (see Figure 2 in Appendix), students are given a writing task in history, and after reading primary source materials of considerable length, they construct an answer evaluated by a scoring rubric based on expert performance. In an example in Hawaiian history, students first read instructions for the task (see Figure 3 in Appendix). Figure 4 (see Appendix) shows an excerpt of one of the longer documents the students (12-year-olds) would read. The scoring rubric is shown in Figure 5 (see Appendix). We have used this framework in Grades 2 through university (with appropriate modifications), and in subject matters ranging from chemistry to humanities to mathematics. Now remember, the model is about deep understanding. The second template to help generate multiple assessment tasks asks students to use (usually on a computer) a graphical task to show relationships within a domain (see Figure 6 in Appendix). There are a number of ways, sometimes more than one in any student's work, to organize a field. In this second history example, students were given primary source materials to read (writings of a Depression era United States president and his opponent) and asked to map their understanding (see Figure 7 in Appendix). The same approach has been used in secondary school genetics and in an adult literacy measure. These representations, just as in the template for content understanding (see Figure 2 in Appendix), are scored by using experts' responses to the questions. The cognitive demands of the tasks are similar, even though the formats of the tasks differ. The relationship between the written and graphic tasks is about 0.6 -approximately the same relationship as between parents' education level and students' achievement. Making research knowledge usable required our abstracting from the fields of learning, psychometrics, and psychology, conducting some of our own studies, and trying out the approach on a small scale, statewide, and as a regular part of the annual assessments of 360,000 children.

To take the next step, and transform knowledge into a useful form, assessment alone is not sufficient. Assessments need to be timed so that the data can be used when needed. Those teaching and those expected to make interpretations should possess high levels of content knowledge (in fact, teachers' self-report of their own topical knowledge has recurrently shown up as a big predictor of student performance on complex academic tasks, at elementary, middle, and secondary school levels, (Baker et al., 1996). But we can help make assessment knowledge useful by providing help that ensures teachers and other educators know as much as they can about their students. In systems where good records are kept, teachers and other instructional leaders have to learn how to combine test results and other sources of information, and how to weigh or value different information. When performance falls short of expectations or requirements, teachers need to know where to

find help, and they need assistance in knowing what to do (too often they may fall back on a failed method). These last two points extend beyond the assessment remit, but are related to the careful documentation of the models guiding assessment design (Chung, Niemi & Bewley, 2003; Niemi, Chung & Bewley, 2003). Teachers can see whether children performed poorly because they didn't have sufficient prior knowledge, had difficulty integrating new and old information, or perhaps could not organize their thoughts into more global principles.

6. SOCIAL CAPITAL AND SCHOOL IMPROVEMENT

CRESST's efforts have been focused on the key components of skill acquisition and academic learning, the bread and butter of schools. Our cognitive models have been created to optimize the attention and the coverage of the types of learning schools (and businesses and the military) are most likely to develop. But in the context of assessments deliberately used for school improvement, there is another important component to be considered. Hargreaves (2003) describes this as *social and organizational capital*. In education, it means identifying the components of management, climate, and personal behavior that support the acquisition of learning and sustain the development of effectively functioning people and institutions. It is a bigger order than raising test scores, but essential to even that goal. It is the bases upon which schools should be evaluated as effective organizations.

Social capital is a term that summarizes a set of affective and behavioral constructs that serve to support the human-to-human element in schools, the parts that make school fun instead of only challenging, personal rather than distant, and organic in development rather than lurching by bits. Elements of social capital vary, whether considered from a broader community or from a school perspective. On the interpersonal level, social capital would include safety and security; in both the physical and psychological sense (see Figure 8). Operationally, participants would trust one another, and trust would allow them to share productive options (rather than keeping useful tricks to themselves), as well as to disclose areas of difficulty and to seek help with impunity. Social capital involves the development of individual propensities like motivation and self-confidence, or goal setting and monitoring progress. On a social/organizational level, social capital involves the development of a shared sense of efficacy, a valuing of collective processes, and the active search for connections and networks to pursue development.

In other words, social capital may be like glue of coherence, or the agar in which intellectual capital is nurtured. Notice, no discussion of the merits of cooperative versus competitive approaches is provided as cultural and traditional contexts play important framing roles in the social capital accumulation.

7. SHOULD SOCIAL CAPITAL BE MEASURED? IF SO, HOW?

The logic of present reform is that if it is important, it should be monitored, and monitoring these days implies measurement of one sort or another. Social capital can be

measured directly through the use of instruments that provide self-reports of states of mind of individuals, derived either through their response to direct questions or their solution to particular scenarios. Such measures may be best used in a bottom-up, developmental way, where collective responses are shared among local respondents and serve to identify problems or needs (Baker & Goldschmidt, 2005).

The use of social capital self-report measures to respond to external requirements, from states or other authorities, are likely to yield socially desirable responses. Social capital development can be inferred from behavioral data to meet top-down requirements (Huang et al., 2007). If course offerings were intended to be equally distributed among different segments of the school population, boys and girls, rich and poor, majority or minority group member, such data could be acquired through archival means. If one were interested in the impact of programs related to coherence and stability of a program, one could monitor changes in absences of both teachers and students over time. If one were interested in the sense of collective efficacy and networking, one could monitor the formation and composition of study groups of students. Teacher behavior could also reveal the development of social capital, including the induction of new teachers to the school, the sharing of resources, the voluntary exchange of approach and information. In short, social capital is best measured by deeds rather than standardized words.

8. INTERPRETING RESULTS

One area that both teachers and school managers have little experience with (at least in the United States) is serious interpretation of assessment results. Questions need to be answered about how good the results are (in comparison to what), how to integrate classroom and other sources of information, and how to think reflectively in order to infer a reasonable next step.

One approach is to use what business calls decision-support systems. These use software that allows easy query and manipulation of data. They work in part like browsers and in part like spreadsheets. The problem with many of these systems is that they are not sensitive to educators' needs. CRESST has tried again to translate research knowledge into tools that help make this type of analysis easy, productive, and even fun. These systems can meet one of the biggest challenges, that of incoherent information, allowing the identification of conflicting or similar data among different sets of students, tests by kind of task, subject matter, or instructional history. These systems are clearly adjuncts for the teacher, rather than machines that spit out right answers. They depend on the insightful questions that a good teacher may think to ask in order to explain information. CRESST has created the Quality School Portfolio (QSP), originally just to show a prototype of what could be done. QSP was used in more than 1000 schools and in every state in the U.S. It has been transformed into a Web-based system and tried in states with varying kinds of accountability systems, including New Jersey, Indiana, and Nebraska. Following these pilots, emphasizing the "webbiness" of the system and the classroom and parent interface, the system went

nationwide, for free. The system components have local, school, classroom, and parent functions. There is also a place to access student work (Heritage, Lee, Chen & La Torre, 2005). Our studies suggest that teachers and principals find great value in the system, especially those with little external support for data analysis (Heritage et al., 2005). The creation of an individual record for a student is a boon for teacher, student, and parent. Again, we have tried to transform data (some of it usable) into a more useful form. However, commercial versions with new platforms and better technical assistance will swamp our efforts. We also have a product "iStat" (Kim & Baker, 2007) that will allow more intensive analyses and similar QSP graphics to interface with any database system.

9. UTILITY IS CONTEXT DEPENDENT

We believe that researchers can go a long way to help make their findings more usable - at least capable of being understood and tried in a variety of settings. Providing tools such as assessment templates (assessment authoring systems, see Baker & Niemi, 2001) will help teachers by raising the quality of some of what they do without raising the time expenditure commensurately.

But the hardest parts of knowledge-based reform are both general and specific. A fundamental change is required of many teachers - a shift from a chronological perspective of "what I will do Monday, or in March", to "what should each learner be doing". Such a cultural shift needs leadership, tools, time, and collaboration to succeed. Moreover, it requires that the administration of schools, at all levels, be willing to take a chance on change, and be ready to revise if well-thought-out plans fail to yield results. The context for success of knowledge-based reform is key. Knowledge must be locally owned and valued, and the infrastructure must allow enough stability for trials. Staffs need the capacity to investigate, including time and tools. Learning must be the major outcome, and where differences exist between local and external policies, a way to reach congruence or a temporal peace must be pursued from all sides. Evidence is useless, however, if fear of sanctions stun individuals into safe, incremental, and destructive practices such as drill and practice of test content and formats.

Whether assessment and evaluation knowledge is ultimately useful, of course, depends on the pudding. In other words, research and assessment and data interpretation form part of the foundation for change. For assessment knowledge and results to be useful, context, capacity, and communication of the teaching and learning system are key. Unless assessment and evaluation knowledge is demonstrably useful to students who do the learning, it is no more than a comforting management exercise. Useful knowledge must go to the heart of why, what, and how students learn.■

Manuscript received: September 20th, 2008
Revised manuscript received: December 4th, 2008

REFERENCIAS

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Baker, E. L. (1982). Tests & instruction: An historical overview. In J. L. Schwartz & M. S. Garet (Eds.), *Assessment in the service of instruction*. Cambridge, MA: Massachusetts Institute of Technology.
- Baker, E. L. (1997). Model-based performance assessment. *Theory Into Practice*, 36, 247-254.
- Baker, E. L. (2003a). Multiple measures: Toward tiered systems. *Educational Measurement: Issues & Practice*, 22(2), 13-17.
- Baker, E. L. (2003b, April). *Templates, objects, and assessment models: We're not in Kansas anymore*. Presentation at the annual meeting of the American Educational Research Association, Chicago, IL.
- Baker, E. L. (2007a). Model-based assessments to support learning and accountability: The evolution of CRESST's research on multiple-purpose measures. *Educational Assessment*, 12(3 & 4), 179-194.
- Baker, E. L. (2007b). The end(s) of testing (2007 AERA Presidential Address) [Electronic version], *Educational Researcher*, 36(6), 309-317.
- Baker, E. L. (2008, September). *Measuring and training 21st century skills*. Presentation to the Office of Naval Research and the Naval Service Training Center, Los Angeles.
- Baker, E. L., Abedi, J., Linn, R. L. & Niemi, D. (1996). Dimensionality and generalizability of domain-independent performance assessments. *Journal of Educational Research*, 89(4), 197-205.
- Baker, E. L., Freeman, M. & Clayton, S. (1991). Cognitive assessment of history for large-scale testing. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition* (pp. 131-153). Englewood Cliffs, NJ: Prentice-Hall.
- Baker, E. L. & Goldschmidt, P. (2005, April). *In search of school quality and accountability: Moving beyond the California academic performance index (API)*. Paper presented at the annual meeting of the American Educational Research Association, Montreal.
- Baker, E. L., Herman, J. L. & Linn, R. L. (2004). *Center for research on evaluation, standards, and student testing. RFA goal: One-assessment, standards, and accountability* (Proposal submitted to the U.S. Department of Education, Institute of Education Sciences). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Baker, E. L. & Mayer, R. E. (1999). Computer-based assessment of problem solving. *Computers in Human Behavior*, 15, 269-282.
- Baker, E. L. & Niemi, D. (2001). *Assessments to support the transition to complex learning in science* (Proposal to the National Science Foundation). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Baker, E. L., Niemi, D., Herl, H., Aguirre-Muñoz, Z., Staley, L., Linn, R. L. et al. (1996). *Report on the content area performance assessments (CAPA): A collaboration among the Hawaii Department of Education, the Center for Research on Evaluation, Standards, and Student Testing (CRESST) and the teachers and children of Hawaii* (Final Deliverable). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

ESTUDIOS

FROM USABLE TO USEFUL
ASSESSMENT KNOWLEDGE
AND EVALUATION

- Black, P. J. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice*, 5(1), 7-74.
- Chi, M. T. H., Glaser, R. & Farr, M. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- Chung, G. W. K., Niemi, D. & Bewley, W. L. (2003, April). *Assessment applications of ontologies*. Presentation at the annual meeting of the American Educational Research Association, Chicago, IL.
- Dembo, M. H. & Junge, L. G. (2005). Learning strategies. In H. F. O'Neil (Ed.), *What works in distance learning* (pp. 41-63). Greenwich, CT: Information Age Publishing.
- Hargreaves, D. (2003, January). *From improvement to transformation*. Keynote address to the International Congress for School Effectiveness and Improvement 2003 conference "Schooling the Knowledge Society", Sydney, Australia.
- Heritage, H. M., Lee, J. J., Chen, E. & La Torre, D. (2005). *Upgrading America's use of information to improve student achievement*. *CSE Report 661*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Huang, D., Miyoshi, J., La Torre, D., Marshall, A., Perez, P. & Peterson, C. (2007). *Exploring the intellectual, social and organizational capitals at LA's BEST*. *CSE Report 714*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Kim, K-S. & Baker, E. L. (2007). iStat [Computer software]. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Lindblom, C. E. & Cohen, D. K. (1979). *Usable knowledge: Social science and social problem solving*. New Haven, CT: Yale University Press.
- Niemi, D., Chung, G. W. K., Bewley, W. L., Vallone, J. & Zhang, L. (2003, April). *Assessment design using ontologies: Linking assessments, content, and cognitive demands*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).
- Nyquist, J. B. (2003). *The benefits of reconstructing feedback as a larger system of formative assessment: A meta-analysis*. Unpublished master's thesis, Vanderbilt University, Nashville, TN.
- Organisation for Economic Co-Operation and Development. (2005). *PISA 2003 data analysis manual: SAS® users*. Paris: OECD Publications. Available online at <http://www.pisa.oecd.org/dataoecd/53/22/35014883.pdf>
- Pellegrino, J. P., Chudowsky, N. & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Resnick, L. B. & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. G. Gifford & M. C. O'Conner (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.
- Vendlinski, T. P., Baker, E. L. & Niemi, D. (2008). Templates and objects in authoring problem-solving assessments. In E. L. Baker, J. Dickieson, W. Wulfbeck & H. F. O'Neil (Eds.), *Assessment of problem solving using simulations* (pp. 309-333). New York: Erlbaum.
- Weiss, C. H. (1977). Research for policy's sake: The enlightenment function of social research. *Policy Analysis*, 3, 531-45.
- Wittrock, M. C. & Baker, E. L. (Eds.). (1991). *Testing and cognition*. Englewood Cliffs, NJ: Prentice-Hall.

ESTUDIOS
EVA L. BAKER

Figure 1.
Model-based assessment example

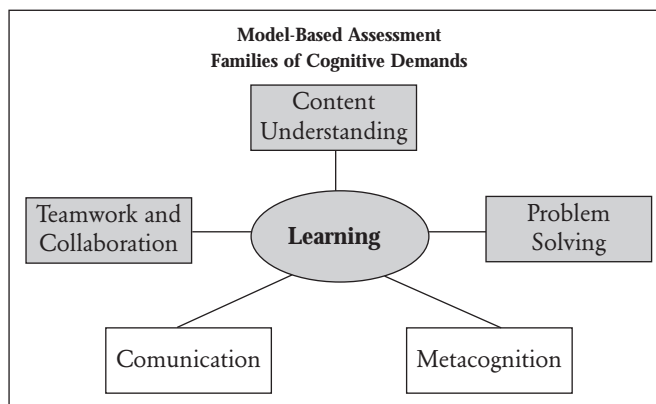


Figure 2.
Template 1 - explanation example

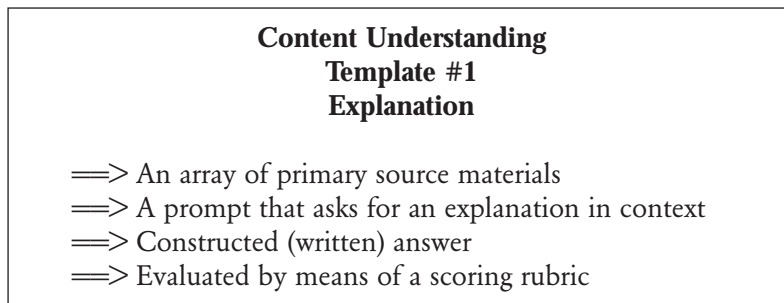


Figure 3.

Hawaiian history writing assignment example

**Hawaiian History Writing Assignment:
Bayonet Constitution**

Imagine you are in a class that has been studying Hawaiian history. One of your friends, who is a new student in the class, has missed all the classes. Recently, your class began studying the Bayonet Constitution. Your friend is very interested in this topic and asks you to explain everything that you have learned about it.

Write an essay explaining the most important ideas you want your friend to understand. Include what you have already learned in class about Hawaiian history, and what you have learned from the texts you have just read. While you write, think about what Thurston and Liliuokalani said about the Bayonet Constitution, and what is shown in the other materials.

Your essay should be based on two major sources:

1. The general concepts and specific facts you know about Hawaiian history, and especially what you know about the period of the Bayonet Constitution.

2. What you have learned from the readings yesterday.

Be sure to show the relationships among your ideas and facts.

ESTUDIOS

FROM USABLE TO USEFUL
ASSESSMENT KNOWLEDGE
AND EVALUATION

Figure 4.

Hawaiian history primary source document example

**Excerpts from Hawaiian History
Primary Source Documents
Liliuokalani**

For many years our sovereigns had welcomed the advice of American residents who had established industries on the islands. As they became wealthy, their greed and their love of power increased. Although settled among us, and drawing their wealth from resources, they were alien to us in their customs and ideas, and desired above all things to secure their own personal benefit.

Kalakaua valued the commercial and industrial prosperity of his kingdom highly. He sought honestly to secure it for every class of people, alien or native. Kalakaua's highest desire was to be a true sovereign, the chief servant of a happy, prosperous, and progressive people.

And now, without any provocation on the part of the king, having matured their plans in secret, the men of foreign birth rose one day en masse, called a public meeting, and forced the king to sign a constitution of their own preparation, a document which deprived [him] of all power and practically took away the franchise from the Hawaiian race.

Figure 5.

History explanation scoring rubric example

**History Explanation
Scoring Rubric**

- General impression of content quality
- Principles or concepts
- Prior knowledge
- Use of available resources
- Misconceptions (negative)
- Argumentation (domain appropriate)
- English mechanics

Template 2 - knowledge representation example

- Key aspects of ideas, supporting facts and views and their relationships
- Relationship is explicit
- Organizational options
 - Core and peripheral
 - Hierarchical
 - Cause and effect
 - Chronological
- Expert scoring

History mapper example

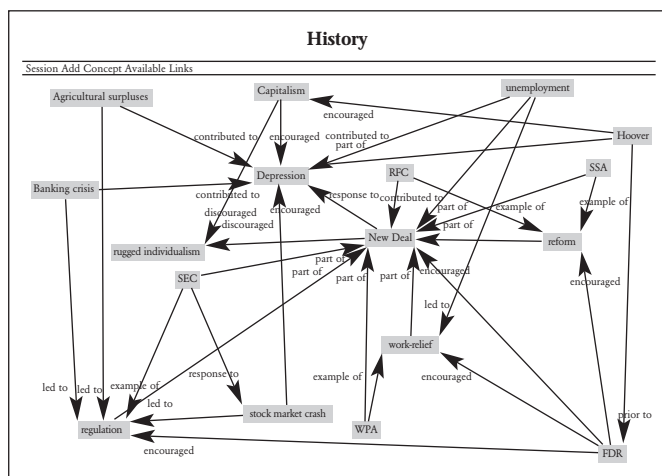
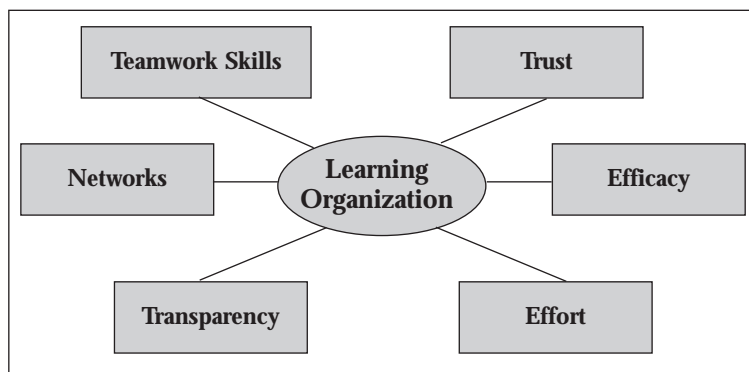


Figure 8.
Social capital in knowledge management



Copyright of Estudios Sobre Educacion is the property of Servicio de Publicaciones de la Universidad de Navarra, S.A. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.