# *Assessment: A Vision*

*Valerie Shute, FSU; Diego Zapata, ETS; Dave Kuntz, Pragmatics, Inc.; Roy Levy, ASU; Ryan Baker, CMU; Joe Beck, WPI; Robert Christopher, ASU*

## *1. Preamble*

> *Standardized tests are monstrously unfair to many kids. We're creating a one-size-fits-all system that needlessly brands many young people as failures, when they might thrive if offered a different education whose progress was measured differently*. ~Robert Reich

Educational systems (in the U.S. and around the world) face huge challenges that require bold and creative solutions to prepare students for success in the 21st century. Part of the solution will require a new focus on students developing the ability to solve complex problems in innovative ways, as well as the ability to think clearly about systems. We need to identify ways to fully engage students through learning environments that meet their needs and interests. When coupled with online collaboration with other students (locally and distally), such environments have the potential to develop students' communication skills and creative abilities as they become exposed to diverse cultures and viewpoints. Assessment results can and should have important implications for instruction, positively influencing both the teaching and learning sides of the equation. Currently, assessments are too often used for purposes of grading, promotion, and placement, but not for learning. That needs to be fixed. The stance we take on assessment is that it should: (a) support, not undermine, the learning process for learners and teachers (as well as online agents); (b) provide more formative than summative information (i.e., give useful feedback during the learning process instead of a single judgment at the end); and (c) be responsive to what is known about how people learn, generally and developmentally.

This attractive vision of assessment has its primary goal to improve learning (e.g., Black & Wiliam, 1998; Shute, 2007). Stiggins (2002; 2006) has suggested that assessment should be used for two reasons: to gather evidence to inform instructional decisions, and to encourage learners to try to learn. It is this vision of educational assessment that we find to be exciting, powerful, and absolutely critical to support the kinds of learning outcomes and processes necessary for students to succeed in the 21st century. This type of assessment is referred to as "formative assessment," which may be thought of as assessment *for* learning, in contrast to "summative assessment" (or assessment *of* learning).

Current approaches to assessment are usually divorced from learning. That is, the typical educational cycle is: Teach. Stop. Administer test. Go loop (with new content). Now consider the following metaphor representing an important shift that occurred in the world of retail outlets (from small businesses to supermarkets to department stores), suggested by Pellegrino, Chudhowsky, & Glaser (2001, p. 284). No longer do these businesses have to close down once or twice a year to take inventory of their stock. Rather, with the advent of automated checkout and barcodes for all items, these businesses have access to a continuous stream of information that can be used to monitor inventory and the flow of items. Not only can business continue without interruption, but the information obtained is far richer, enabling stores to monitor trends and aggregate the data into various kinds of summaries, as well as to support real-time, just-in-time inventory management. Similarly, with new assessment technologies, schools should no longer have to interrupt the normal instructional process at various times during the year to administer external tests to students. Instead, assessment should be continual and invisible to students, supporting real-time, just-in-time instruction. This comprises a main feature of our vision of assessment in support of learning, 20 years hence.

There are two other key aspects of our vision. First, assessment will cover not only the particular domain skills, but also key competencies and attributes that are important for success in the 21st century. Specifically, we propose assessing general cognitive competencies such as problem-solving ability as well as non-cognitive competencies and attributes such as motivation, frustration, and open-mindedness. Since all of these learner characteristics affect learning, it is essential to start thinking about how we can assess them. Second, rather than being used just for the current lesson, assessment data will be made widely available, for interpretation by a wide variety of researchers, and for use by a broad community of stakeholders.

In summary, our three themes for assessing learning, described below, include: (a) comprehensive profiles/models (*what* and *how* to assess), (b) seamless and ubiquitous assessment (*when* and *where* to assess), and (c) assessment information for decision making (*who* is using the assessment data). Each of these will be briefly explained below.

## 2. Three Themes for Assessing Learning

### 2.1. Comprehensive Models

For the first theme (Comprehensive Models), 20 years from now, we envision a well-mapped landscape of traditional and emerging competencies as well as other personal attributes. To achieve this goal, research needs to be conducted in the areas of identifying, modeling, and assessing these attributes. Identifying refers to determining which sets of characteristics of learners help to direct or develop a learner's education. In addition to domain-specific knowledge and skills (e.g., reading and math skills), some examples of other relevant attributes include: creative problem solving, systems thinking, self-regulation, information-seeking skills, compassion, and ability to transfer skills to new contexts. Examples of state-based attributes include boredom, frustration, and excitement, which can relate to a student's success. Research is needed to derive a taxonomy of relevant competencies and attributes that are optimally suited for our rapidly changing world as instantiated in a variety of contexts. Along the lines of research already underway in the "lifelong learning user modeling" community (e.g., Kapoor & Horvitz, 2007; Kay, 2008), we can imagine combined and evolving profiles representing a comprehensive synopsis of what's known, what can be done, what is believed, what is preferred, and so on.

Modeling refers to establishing conceptual and computational representations of each key competency and attribute, which will require considerable research. Existing modeling tools like Bayesian networks, artificial neural networks, genetic algorithms, and Markov decision processes are promising; new tools will likely be invented and new research uncovered that will no doubt produce new techniques that are even more effective and efficient than those available today. Because models should (a) work across a range of students, (b) be validated by experts, and (c) be able to be applied within new environments (e.g., games, simulations, computer tutors) additional research will be needed on the transfer, portability and integration of models across contexts. Toward that end, we see the need to explicitly include context information in the models (e.g., the what-who-when-where-how-why data surrounding the learning).

Assessing generally refers to the process of gathering important information about competencies and attributes, both in formal and informal learning contexts. This should lead to valid and reliable inferences, both diagnostic and predictive. Assessment uses may be summative (e.g., for purposes of accountability and promotion) or formative (to support learning). However, we note that data from a given assessment can, in fact, be used for both formative and summative purposes through appropriate design and other analytical methods. Such a perspective aligns with viewing assessment as a dynamic agent in student learning over time. Research is needed on the possible automation and streamlining of these approaches. Related research on longitudinal assessment is necessary to support this dynamic perspective of the role of assessment in student learning and progress. Recent innovations in measurement include the development of principled frameworks that explicitly integrate specific theories regarding the domain, cognition, and learning into task and assessment design (Mislevy, Steinberg, & Almond, 2003) and analytic techniques that support inferences about students along multiple dimensions and at multiple grain sizes (Leighton & Gierl, 2007). Such developments have focused on traditional, cross-sectional assessments, but have the potential for longitudinal measurement (e.g., learning trajectories over time) as occurs in student modeling. In short, assessment should be driven by the definition of terms (*identification*) and the rules of interaction (*model*). In this way, assessments can be developed as the definitions and models are created.

## 2.2. *Seamless and Ubiquitous Assessment*

For the second theme (Seamless and Ubiquitous Assessment), 20 years from now, we envision a continuous process that fuses assessment and learning, similar to the metaphor about the stores-inventory mentioned earlier. Seamless refers to the removal of the false boundaries between learning and assessment that characterize the current Teach/Stop/Test model. Ubiquitous refers to the constant nature and need to feed back the results and implications of assessment into learning, anywhere anytime. The current state of affairs is characterized by a few, illustrative examples of the rich potential for assessment to be fully integrated into the educational enterprise, and considerably more instances of a stark divide between assessment and other aspects of education. At present that divide is crossed by teachers and other first-responders. Our goal is to remove the load from the teacher, creating tools that are easy to incorporate in the daily lesson plan, and which include actionable information.

Research necessary to accomplish this integration includes development of, and evaluation on the design, implementation, and interpretation of the ensuing data from seamless and ubiquitous assessment and learning systems. Some examples of existing stealth-like approaches include the Reading Tutor (Mostow & Aist, 1997), which listens to students read stories aloud, and uses automated speech recognition to assess their reading proficiency as accurately as one of the best (Woodcock Reading Mastery Test) instruments (Beck & Sisson, 2006). Another example is assessing creative problem solving while a player is immersed in the game Oblivion (Shute et al., 2009). Immediate next steps would include use of existing tools to collect and report on educationally-valuable competencies (e.g., using MS Word to track spelling errors to make inferences about mechanical or comprehension reading challenges). Seamless assessment will involve models and procedures for supporting inferences across contexts (e.g., time, domains, and developmental levels). Constructing principled methods of data management to enable the integration of these diverse sources of information and sharing them among stakeholders is an important new challenge, described next.

## 2.3. *Assessment Information for Decision Making*

For the third theme (Assessment Information for Decision Making), 20 years from now, we envision stakeholders (e.g., students, parents, teachers, administrators, policy makers, researchers, funding agencies) will receive the assessment information they need to make informed evidence-based decisions. Recognizing that each stakeholder has different information needs, assessment models should provide each stakeholder with access to assessment information in meaningful forms. Current research shows that assessment information can enhance decision-making processes at different levels (e.g., student, class, school, and district levels). Policy makers, for example, require aggregate information about the strengths and weaknesses of students to make informed evidence-based decisions (Coburn & Talbert, 2005; Honig & Coburn, 2008). As more complex environments evolve, research would be needed regarding the kind of information required by each stakeholder and the kinds of external representations that would communicate assessment information in effective ways. Transparency regarding the characteristics of the assessment models used in these new learning environments facilitates acceptance and wider adoption by the community. Table 1 shows information of interest to different stakeholders.

**Table 1**. Assessment information and usage across stakeholders

| Stakeholder | Assessment Information | Assessment Usage[1] |
|---|---|---|
| *Policy makers, administrators* | • Validity and reliability of inferences based on assessments<br>• Kind of information being collected on the students<br>• Types of assessment claims supported | Policy makers need supporting evidence to decide whether the current policies with respect to education are effective and appropriate. Assessment data can provide some of this evidence, although the information would have to be rolled up and summarized accordingly to be useful at this high level of abstraction. |
| *Teachers, mentors, tutors* | • What is being learned (content, competencies, other attributes)<br>o Relative to other students (normative)<br>o Relative to self (learning trajectories)<br>o Relative to standards (criteria) | Teachers have a diverse set of needs, ranging from needing to know individual student progress, sub-groups, and the whole class. In addition, progress can be measured in relation to the learner, a larger group, or a criterion. Finally, teachers can use assessment results to determine what works and what does not to inform future teaching. |
| *Students* | • Strengths and weaknesses regarding valued competencies.<br>• Levels/types of other personal attributes. | Students can use assessment results to learn content, hone skills, and learn about learning. Research shows benefits in terms of enhanced metacognitive skills and student learning[2]. |
| *Parents* | • Same as teachers/mentors, but at simpler level of interpretation | Parents can use assessment results to answer questions such as: Does my child need help? Should I talk to or complain about the teacher/curriculum? Should we switch schools? |

[1] Jean-Marc Wise (personal communication, 2009) suggested this third column on *usage* for Table 1.

[2] (e.g., Brna et al., 1999; Bull & Pain, 1995; Hartley & Mitrovic, 2002, Kay, 1998; Dimitrova, 2003; Zapata-Rivera & Greer, 2004).

| | | |
|---|---|---|
| *Funding agencies* | • Validity and reliability of inferences based on assessments<br>• Types of assessment claims supported<br>• Current challenges and future research agenda | Funding agencies would like to base their decisions about what to fund on the needs and the relative effectiveness of the programs that apply for funding. The proposed assessments can provide relevant information. Again, there would have to be an appropriate level of aggregation. Note that real-time automatic aggregation may not be appropriate in all cases, as some methods require a period of maturation before yielding results. |

Having briefly defined our three themes comprising our vision of assessment to support learning in the near (5-20 year) future, we now focus on the benefits and barriers to this envisionment.

# *3. Benefits and Barriers of Implementing This Vision*

How and what we learn is rapidly changing. No longer are students spending years in classrooms that have changed little in the past two hundred years, learning skills that have also changed little in the past two hundred years. Now, students are graduating with the understanding that their educational process has not yet been completed, rather, it is just beginning as lifelong learning is a reality for most citizens. Similarly, there is a shift in how instructional content is delivered. No longer must we learn our content from an instructor at the front of a room; more and more content is being delivered electronically via software. This shift in content type and delivery mechanism presents many challenges; how can we assess learners in these new skills? Are we able to perform better assessments using these new resources? Will assessment become more challenging with a variety of technological educational resources, or can we streamline the process? All of these challenges also represent exciting opportunities for improving the educational process. We begin with the benefits of this proposed approach.

## *3.1. Benefits of the Assessment Vision*

Constructing seamless, ubiquitous assessments across multiple learner dimensions, with data accessible by diverse stakeholders is expected to yield several direct educational benefits as well as other indirect ones. First, the time spent administering the test, handling make-up exams, and going over test responses is not particularly conducive to learning. Approximately 10% of class time is spent on assessment activities. Given the primacy of time on task as a predictor of learning, reallocating that 10% into activities that are more educationally productive is a potentially large benefit that would apply to almost all students in all classes, and would be equivalent to giving students graduating high school an extra year of instruction.

Second, by having assessments that are continuous and ubiquitous, students are no longer able to "cram" for an exam. Although cramming provides excellent short-term recall, and is a viable strategy for passing an exam, it is a poor route to long-term retention and transfer of the material. Thus, standard educational policy is to assess students in a manner that is in conflict with their long-term success. By providing a continuous assessment model, the best way for students to do well is to do well every day; although this statement sounds tautological, it is not how most classes are structured. By moving students towards a model where they will retain more of what they learn, we are enabling them to better succeed in cumulative domains, such as mathematics, which are essential to our nation's economic health.

The third direct benefit is that this shift in assessment mirrors the national shift from evaluating students based on the number of years they have sat at a desk to evaluating students on the basis of acquired competencies. A growing number of states are requiring students to pass a high-stakes final exam in order to graduate high school. While we do not especially resonate with the model of a pencil and paper, high-stakes test for which students must prepare, this shift toward ensuring students have acquired "essential" skills fits with our proposal of continuous assessment. Many educators would argue that certain milestone assessments are needed to ensure quality across larger populations, and thus such assessments would have to be based on the same principles. In line with our proposed vision, the next steps would entail broadening the set of educationally valuable competencies and attributes to be more aligned with current (and near future) educational needs.

In addition to the direct benefits to education, there are substantial indirect benefits as well. First, our ability to instruct students is fundamentally limited by our ability to assess them. If students have varying degrees of proficiency and ability at cognitive and non-cognitive attributes, and this varying background predicts how well a student will respond to a given intervention, then in order to provide optimal instruction it is necessary to accurately assess students. Furthermore, for us to understand the efficacy of different educational objects (writ large) it is necessary to have a crisp understanding of the student's knowledge before and after being exposed to the intervention. Thus, an ability to assess students will also enable us to better evaluate, presumably numerous, educational objects with which future students will spend much of their educational time. Second, our current capacity to assess students is often limited in that it is based on a relatively small number of test items. As we move to a seamless assessment model, we will be able to more accurately assess students since we will have access to a much broader collection of the student's learning data. More accurate assessments enable us to both better suggest learning objects to students as well as more accurately evaluate those objects' efficacy.

In addition to the various benefits described above, there are other educational issues that our proposed assessment vision can help to address or resolve.

### 3.1.1. 21st Century Skills

Students will need to develop a different set of competencies than those in the current schools. The issue is not that the 21st century is that different from the 20th, it is that it is different from the 19th century upon which much of "modern" schooling is based. In particular, so-called "soft skills" (e.g., teamwork, computer literacy, and presentation skills) are expected to become more important in education than they are now. Given this increase in importance, it is important that we find good methods to assess students in these capabilities. Furthermore, given the growing importance of lifelong learning we must find methods of assessing those cognitive and non-cognitive factors that are likely to be predictive of learner success so as to best guide the learner. As we are envisioning seamless and ubiquitous assessment in the context of lifelong learning, this vision can readily lead us to seamless and ubiquitous learning integrated with job performance support systems.

Of course, we would like to help learners develop skills such as self-regulation or self-explanation. If we hope to develop methods for improving those skills, we must have a means of knowing whether the student has improved. Therefore, our approach of building comprehensive models of learner competencies and attributes, and then developing assessment techniques to infer levels of those constructs is necessary in a shifting educational landscape.

### 3.1.2. The Emergence of Educational Technology

Currently, assessment within educational software is typically handled on a system-by-system basis. To measure a specific construct (e.g., persistence, help-seeking) requires a substantial amount of effort to construct a model that is particular to the system in question. The construction of such a model—for a single construct for a single system—costs approximately one year's time for a graduate student. Thus, the current approach does not scale to the increasing numbers of electronic learning environments.

Our proposal of building comprehensive models of general learner characteristics, and constructing them in such a way as to transfer across systems, avoids this problem. Aside from reducing the costs of electronic learning objects that would have been created, our proposal will also increase the number of such artifacts that are built since a broader set of content creators will be able to participate.

### 3.1.3. Harvesting the Data Deluge

Given a world where learners are using a variety of electronic learning objects, and those objects are continuously assessing learner progress on a variety of measures, it is possible to become drowned in details. Therefore, we have recommended that assessment designers think about who the potential consumers are of this knowledge, and determine how they can distill the assessment content down to be of use to each stakeholder. If this is the responsibility of individual designers, it would be very helpful to provide them with a framework for orientation – a shared data dictionary that prevents duplication of efforts and streamlines nomenclature and categorization. Otherwise it will be extremely difficult to aggregate information across individual contributions. As we described earlier, our envisioned taxonomy would first have to be established by corresponding research and then disseminated (and perhaps governed) by a body similar to other shared standards as coordinated by the IEEE or ISO.

So, by making assessment information available to a broader variety of members of the educational establishment, the odds the learner will succeed are improved. For example, young learners could benefit from their parents being informed of learning deficiencies and providing additional help or motivation. Teachers would (probably) benefit from seeing a summary of areas of weakness of several students in the class above and beyond a report for each student; such a report would enable an immediate alteration of teaching methods. This highlights the importance of mechanisms that facilitate the communication of data in a way that this desired by and meaningful to stakeholders. Thus, by considering the social processes of learning outside of software, the assessment technologies described in this proposal are able to enhance the learner’s experience and support network, resulting in more effective, efficient, and enjoyable instruction.

## *3.2. Challenges and Barriers*

Though the agenda of research discussed herein has a number of direct and indirect benefits for the science and practices of learning, there are several challenges that will need to be addressed for the agenda to be successfully completed, and for it to achieve its full potential influence on the scientific community and on educational practitioners. These are now described below.

### *3.2.1. Generalizable Educational Models*

The vision we have outlined depends to a significant degree on the success of models in generalizing between learning objects, competencies and other learner attributes. However, the study of generalizability of these types of models is still in its infancy. There exist examples of the study of the generalization of models between learning objects, involving stratified cross-validation ("leave-out-one-learning-object-cross-validation") (Baker et al, 2008), but the methodology used is generally overly simplified, and does not explain why models can generalize in some cases but not others. The scientific literature on transfer learning, from the machine learning community, is a valuable resource for understanding the transfer of models, but has not generally been applied to the types of models developed by the educational research community. Selecting from and applying this literature to the type of educational models proposed here is likely to increase the success of the proposed research agenda.

In addition, there is significant variation in the design of learning objects and how the competencies and constructs advocated here manifest themselves in learning objects. This may lead to the need for meta-models – drawing from the cognitive modeling literature – that express the competencies and constructs at higher levels that can be automatically translated to the low-level features of the environment often found in machine learned models of educational constructs.

### *3.2.2. Considering the Stakeholder*

In the vision proposed in this paper, we have argued that there is a need for dissemination of information to a wide variety of stakeholder communities. We cannot expect that school administrators, teachers, or parents will become experts in complex data analysis. Hence, we will have to develop tools for communicating assessment information to these stakeholders in their own language, and tools that these stakeholders can use to explore the deluge of data available. This effort will require deploying methods from the interaction design and human-computer interaction communities, in order to develop reporting and communication tools that are useful, usable, and desirable to members of these communities.

In general, the educational practice community (including teachers and administrators) absolutely must be included in the development of this vision and the research agenda proposed here. Traditional assessment methodologies such as tests have a long and rich history of usage by these communities, and have been successful at addressing specific summative assessment needs. It is our view that the methods proposed in this document have the potential to be more efficient, less disruptive, and better able to assess more complex competencies and educational constructs than existing educational measurement methods that are widely used today. However, they must be designed in collaboration with these partner communities in order to achieve wide usage and high effectiveness. In particular, these communities must be involved in the choice of competencies and constructs to assess, or the models may make little impact on educational practice.

### *3.2.3. Walled Gardens*

At the moment, developers of educational software have little incentive to cooperate in making the educational content interoperable. For instance, consider a student using learning objects A and B developed by two different software companies. The content in learning object A is prerequisite to the content in learning object B, hence students predominantly experience learning object A first. Learning object A distills information about the student that can make learning object B more effective, resulting in learning object B being highly effective. What is the incentive for the developer of learning object

A to share that information to learning object B? In a competitive commercial environment, sharing information has the potential for asymmetrical impact, where the information-receiving learning object appears significantly more effective than the information-donating learning object. Additional issues to be resolved relate to intellectual property; e.g., the question of revenue generation, compensation, and possibly royalties.

This problem can be addressed in part through using the educational data mining method of learning decomposition (Beck, 2006) to infer when learning object B's effectiveness was likely enhanced by receiving information from learning object A (i.e., by analyzing object B's effectiveness both when object A's information is present and missing). This may increase the incentive for sharing information, as the developers of learning object A can point to their software's benefits on students' future learning.

### *3.2.4. Privacy*

As with any large-scale data management project, incorporating data that can potentially identify individuals and which gives a broad range of information about individuals, privacy concerns must be accounted for. The positive intentions of educational practitioners and technology experts notwithstanding, any large quantity of data provides risks that inadvertent errors or intentional abuses can lead to privacy violations. Hence, all efforts must be taken to ensure that data is as anonymous as possible, including removal of obvious personal information such as names and birthdays (and its replacement with unique personal identifiers which cannot be reverse-engineered to link to a person), and scrubbing of potential identifying information. Such practices are already standard in large public educational data repositories such as the Pittsburgh Science of Learning Center DataShop, TalkBank, and the Kingsbury Center. All research supported by this initiative should study these existing examples and attempt to match or improve on the privacy practices used by these repositories.

## *4. Summary*

What are the critical research and development questions that can begin to move us toward this vision? The principal goals will be to figure out (a) what attributes to value, assess, and support for 21st century success, and (b) how to accomplish the design and development of robust assessments, which would ultimately be embedded within online educational systems (see Figure 1).

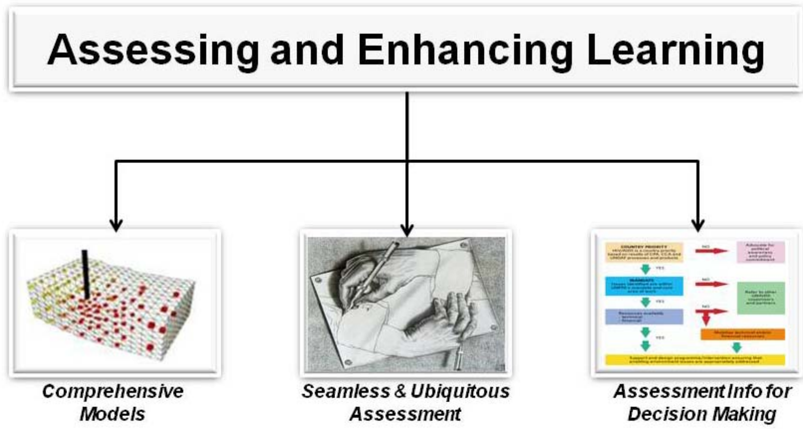

Figure 1. Three strands of assessing and enhancing learning research

Modeling, assessing, and supporting students in relation to an expanded set of competencies and attributes is intended to allow students to grow in a number of important new areas, function productively within multidisciplinary teams, identify and solve problems (with innovative solutions), and communicate effectively. Critical research and development issues include those related to assessment and modeling—particularly in support of student learning and also that can be delivered in a cost-effective way. Good formative assessment research is thus needed given changes in (a) the types of learning we are valuing today (and in the near future), as well as (b) the new, broader set of contexts in which learning is taking place.

Additional research and development will need to be done in terms of effective and efficient adaptive technologies that are closely coupled with valid diagnostics, and research is needed that facilitates the linking of results from various forms of assessment, e.g., to support the creation of developmental or vertical scales. Tools for critically evaluating theories and models will also be required. Finally, more controlled evaluations need to be conducted on advanced, online educational systems to determine what works, for whom, and under what conditions.

## *Major Funding Targets*

In short, the principle steps forward include research on:

1. Understanding the full complement of characteristics that are brought to bear in learning - what are they, how do they relate, how do we get evidence about them, and how do we take that evidence to inform learning?

2. Fusing assessment and learning - what are the new sources of assessment, how do they flow to, from, and with learning, and how can we tear down conceptual and practical barriers between assessment and learning?

3. Rendering assessments useful to all parties - who makes what decisions, what information do they need, how does assessment provide evidence for those decisions, and how to best communicate the complicated results of assessment to each party?

Table 2 summarizes our 3 themes (and sub-themes) in terms of where they should be two- and 20-years down the road.

**Table 2**. Themes over time

| Theme | 2 yr | 20 yr |
|---|---|---|
| **Comprehensive Model: ID'ing key comps** | Proof of concept with some extensive lit reviews and conceptual models of cog & non cog variables (e.g., ST, persistence) within and across domains. | Wide range of important skills identified and justified, such as: inquiry skills, causal reasoning, collaboration , communication, critical thinking , reflexivity, self-regulation, time management skills, cognitive flexibility/adaptability , indexing skills/info mgt, focused attention, systems thinking , interdisciplinary thinking, creativity, self efficacy, problem solving (structured & ill-structured), domain expertise, tolerance/compassion, identity management , visual literacy skills. |
| **Comprehensive Model: Modeling** | Demonstrate ability of a single model (or close variants) to function effectively across multiple LEs. Demo capability to model multiple competencies in single LE. | Robust (conceptual and computational representation) models of all above comps, as well as their interactions. Vertical scales for key competencies. |
| **Comprehensive Model: Assessing** | Proof of concept for embedded asst for JIT learning support and/or use DM (e.g., "discovery with models") analysis for increasing scientific understanding of construct. | Most assts are intimately combined within the learning experience thus rarely need summative asst (e.g., only for auditing purposes). Assts can be accumulated into lifelong student portfolios and profiles. |
| **Seamless assessment** | Proof of concept of actual stealth asst approaches. Asst activities provide for more authentic learning & continuous trajectories | All s/w that students are using (not only games, learning envirs, but also applications like Word) are employing seamless asst. |
| **Ubiquitous assessment** | Proof of concept: Data from one LE improves accuracy of model in other environment (in both formal & informal envirs). | Diff interoperable platforms for the delivery & analysis & support of key comps, relevant data, and the inferences linking them altogether. |
| **Assessment information for decision making** | Demonstrate that the system can produce actionable recommendations for some subset of the stakeholders who don't normally get this type of info. | All school principals, teachers, and most parents can understand clearly and act upon data, easily making the kinds of inferences that are currently made by educational data miners. |

## References


Baker, R. S. J. d., Corbett, A. T., Roll, I., Koedinger, K. R. (2008) Developing a generalizable detector of when students game the system. User Modeling and User-Adapted Interaction, 18, 3, 287-314.

Beck, J.E. (2006). Using learning decomposition to analyze student fluency development. *Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems* (pp. 21-28). Jhongli, Taiwan.

Beck, J. E., & Sison, J. (2006). Using knowledge tracing in a noisy environment to measure student reading proficiencies. *International Journal of Artificial Intelligence in Education, 16,* 129-143.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.

Brna, P., Self, J., Bull, S., & Pain, H. (1999). Negotiated collaborative assessment through collaborative student modeling. *Proceedings of the workshop Open, Interactive, and other Overt Approaches to Learner Modelling at AIED99*. Le Mans, France, pp. 35-44.

Bull, S., & Pain, H. (1995). Did I say what I think I said, and do you agree with me?: Inspecting and questioning the student model, *Proceedings of World Conference on Artificial Intelligence in Education (AACE)*, Charlottesville, VA, pp. 501-508.

Coburn, C. E. & Talbert, J. E. (2006). Conceptions of evidence-based practice in school districts: Mapping the terrain. *American Journal of Education, 112*(4), 469-495.

Dimitrova, V. (2003). StyLE-OLM: Interactive open learner modelling. *International Journal of Artificial Intelligence in Education* 13(1), 35-78.

Hartley, D. & Mitrovic, A. (2002) Supporting Learning by opening the student model. In *Proceedings of ITS 2002*. pp. 453-462.

Honig, M. I. & Coburn, C. E. (2008). Evidence-based decision making in school district central offices: Toward a research agenda. *Educational Policy, 22*(4), 578-608.

Kapoor, A., & Horvitz, E. (2007). Principles of lifelong learning for predictive user modeling. In *Proceedings of the Eleventh Conference on User Modeling* (UM 2007), 37-46.

Kay, J. (1998). A Scrutable User Modelling Shell for User-Adapted Interaction. Ph.D. Thesis, Basser. Department of Computer Science, University of Sydney, Sydney, Australia.

Kay, J. (2008). Lifelong learner modeling for lifelong personalized pervasive learning. *IEEE Trans on Learning Technologies, 1*(4), 215-228.

Leighton, J. P., & Gierl, M. J. (Eds.) (2007). *Cognitive diagnostic assessment for education: Theory and practices*. Cambridge University Press.

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

Mostow, J., & Aist, G. (1997). The sounds of silence: Towards automated evaluation of student learning in a reading tutor that listens. *Proceedings of the Fourteenth National Conference on Artificial Intelligence* (AAAI-97) (pp. 355-361). Providence, RI.

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

Shute, V. J. (2007). Tensions, trends, tools, and technologies: Time for an educational sea change. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 139-187). New York, NY: Lawrence Erlbaum Associates, Taylor & Francis Group.

Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. J. Cody, & P. Vorderer (Eds.), *The Social Science of Serious Games: Theories and Applications*. Philadelphia, PA: Routledge/LEA.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning, *Phi Delta Kappan Professional Journal, 83*(10), pp. 758–765.

Stiggins, R. (2006). Assessment for learning: A key to motivation and achievement. *Edge: The Latest Information for the Education Practitioner, 2*(2), 1-19.

Zapata-Rivera, D. & Greer, J. (2004) Inspectable Bayesian student modeling servers in multi-agent tutoring systems, *International Journal of Human-Computer Studies, 61*(4), 535–563.