Assessment and Learning in Intelligent Educational Systems: A Peek into the Future

Valerie SHUTE[a], Roy LEVY[b], Ryan BAKER[c], Diego ZAPATA[d], and Joseph BECK[e]
[a] *Florida State University,* [b] *Arizona State University,* [c] *Carnegie Mellon University,*
[d] *Educational Testing Service,* [e] *Worcester Polytechnic Institute*

Abstract. New demands associated with living in a highly-technological and globally-competitive world require today's students to develop a very different set of competencies than previous generations of students needed. The general goal of education is to prepare young people to live independent and productive lives. Unfortunately, our current educational system is not keeping pace with these changes and new demands. The world is becoming increasingly complex and to make progress toward fixing educational woes, we need to have a good sense of bearings—where we are, and where we're heading. This paper is intended to provide such bearings, specifically in s of a fresh vision for education. We envision new modes of learning teaching using stimulating online environments such as games and simulations, coupled with an assessment infrastructure covering a broad set of competencies and other attributes to support learning. This represents a long view of the field to inform current R&D efforts.

Keywords. Assessment, data mining, diagnosis, games, learning, modeling.

Introduction

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. 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; lifelong learning is a reality for most citizens. Similarly, there is a shift in how instructional content is delivered. No longer is most content learned from an instructor at the front of a room; more and more content is being delivered electronically via software—in school and outside of school.

This paper peeks into the future of education, 20 years hence. We describe a future of assessment and learning that is intended to inform current and near-future R&D efforts. For example, we envision new modes of learning and teaching using rich online environments coupled with an assessment infrastructure covering a broad set of competencies and other attributes to support learning. Toward this end, we need to identify ways to fully engage all students through learning environments that meet their diverse needs and interests. Online games and simulations are potentially optimal venues. One big problem is that most online games currently lack an assessment infrastructure, especially in relation to educationally-valuable knowledge and skills. However, assessment results can and should have important implications for learning.

Assessing generally refers to the process of gathering important information about competencies and attributes, either in formal or informal learning contexts. This should lead to valid and reliable inferences, both diagnostic and predictive. Too often, classroom and other high-stakes assessments are used for purposes of grading, promotion, and placement, but not to engender learning (i.e., a typical educational cycle is: Teach. Stop. Administer test. Next loop, with new content). The stance we take on assessment is that it should: (a) encourage and support, not undermine, the learning process for learners and teachers (as well as online agents); (b) provide formative information whenever possible (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 vision of assessment has its primary goal to improve learning [4, 14, 18] which 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," or assessment *for* learning, in contrast to "summative assessment" (or assessment *of* learning).

Consider the following metaphor representing an important shift that occurred in the world of retail outlets (from small businesses to supermarkets to department stores, as suggested by Pellegrino, Chudhowsky, & Glaser [11]). 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 (cf. [15]). This comprises a main feature of our vision of assessment to support learning as embedded in rich online environments, 20 years hence.

There are two other key aspects of our vision that can inform a research agenda. 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, but aren't being assessed today. Specifically, we propose assessing general cognitive competencies such as problem-solving ability and systems thinking, as well as non-cognitive competencies and attributes such as teamwork, motivation, frustration, and open-mindedness. Since these learner characteristics can affect learning, it is essential to start thinking about how we can assess them in valid and reliable ways. 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: (1) comprehensive profiles/models (*what* and *how* to assess), (2) seamless and ubiquitous assessment (*when* and *where* to assess), and (3) assessment information for decision making (*who* is using the assessment data). Each of these will be briefly explained.

Three Themes for Assessing Learning in Online Environments

Theme 1: Comprehensive Models

For the first theme, 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 attributes refers to determining which sets of learner characteristics help to direct or support a learner's education. In addition to domain-specific knowledge and skills (e.g., reading, math, and computer literacy 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 other personal attributes include boredom, frustration, and excitement, which can have an impact on students' 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 [6, 7], 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, etc.

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. However, we need an explicit research goal to develop powerful new tools and modeling 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 (when relevant), 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., socio-cultural and instructional environment data).

Assessing students yields information that can 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. Additional research relating to longitudinal assessment is also necessary to support this dynamic perspective of the role of assessment in student learning and progress. Currently, 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 [9] and analytic techniques that support inferences about students along multiple dimensions and at multiple grain sizes [8]. 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.

Theme 2: Seamless and Ubiquitous Assessment

For the second theme, 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 normally crossed by teachers. 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 and evaluation of the design, implementation, and interpretation of the ensuing data from seamless and ubiquitous assessment and learning systems. Some existing examples of systems that implement embedded assessment are as accurate as some of the best instruments available. For example, the accuracy of the Reading Tutor [10], which listens to students read stories aloud, and uses automated speech recognition to assess their reading proficiency, is comparable to the one produced by the Woodcock Reading Mastery Test [3]. The Assistment system for mathematics education can predict future performance on standardized mathematics exams as well as the test itself can [16]. Another example of these systems is assessing creative problem solving while a player is immersed in a game environment, such as Oblivion [13]. Finally, ongoing research at the Naval Air Warfare Center Training Systems Division (NAWCTSD) includes embedded assessment within various simulations. 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.

Theme 3: Assessment Information for Decision Making

Our third theme relates to the needs of various stakeholders (e.g., students, parents, teachers, administrators, policy makers, researchers) with regard to assessment information to enable informed evidence-based decisions. Important research relating to this theme concerns the determination of particular stakeholder requirements with respect to assessment information or score reports.

Recognizing that each stakeholder has different information needs, assessment models should provide each stakeholder with access to assessment information in meaningful forms. Current reports tend to convey assessment information in a one-size-fits-all manner rather than conveying the information stakeholders need to make a decision. In addition, 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 [5]. 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
Policy makers, adminis-trators	• Validity and reliability of inferences based on assessments • Information collected on the students • Types of assessment claims supported	Policy makers need evidence to decide whether the current educational policies are effective and appropriate. Assessment data can provide this evidence, although information has to be summarized to be useful.
Teachers, mentors, tutors	• What is being learned (content, competencies, other attributes) ○ Relative to other students ○ Relative to student ○ Relative to standards	Teachers have a diverse set of needs (e.g., individual progress, sub-groups, the whole class). Progress can be measured in relation to the learner, a group, or criterion. Teachers can use assessment results to determine what works and what does not to inform future teaching.
Students	• Strengths and weaknesses of valued competencies. • Levels/types of other attributes.	Students can use assessment results to learn content, hone skills, and learn about learning.
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 the teacher? Should we switch schools?

Having briefly defined our three themes comprising our vision of assessment to support learning, we now focus on the benefits of and barriers to this vision.

Benefits and Barriers of Implementing This Vision

The envisioned shifts in content types and delivery mechanisms described in the Introduction of this paper present many challenges. For instance, 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.

Benefits of this 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 from 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. 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 toward a model where they will retain more of what they learn, we are enabling them to better succeed in cumulative domains, such as mathematics.

The third direct benefit is that this shift in assessment mirrors the 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 U.S. states are requiring students to pass a high-stakes final exam in order to graduate from 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 effectively is fundamentally limited by our ability to assess them. If students have varying degrees of proficiency 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 (e.g., intelligent educational games, specific instructional modules) it is necessary to have precise 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 the 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 suggest suitable educational objects to students as well as accurately evaluate those objects' efficacy. In addition to the various benefits described above, there are other educational issues that our proposed vision can help to address or resolve.

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's 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. Again, research is needed to identify attributes as well as their relation to learning outcomes and processes. Given this perceived increase in importance, it is important that we research good methods to assess students in these capabilities and to figure out how to take into account various cognitive and non-cognitive abilities when designing instruction. Moreover, 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. Finally, 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. In all cases, 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.

Broader 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 amount of investment required to develop such a model—for a single construct for a single system—could easily require hiring a full time graduate student for an entire year. Thus, the current approach does not scale to the increasing numbers of electronic learning environments. Our vision and associated research agenda 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 educational objects that would have been created, our vision will also increase the number of such artifacts that are built since a broader set of content creators will be able to participate.

Structuring the Data Deluge. Given a world where learners are using a variety of electronic educational objects, and those objects are continuously assessing learner progress on a variety of measures, it is possible to become drowned in details. Therefore, we recommend 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 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 IEEE, ISO, IMS, or SCORM.

By making assessment information available to a broader variety of members of the educational establishment, the likelihood that the learner will succeed is improved. For example, young learners could benefit from their parents being informed of learning deficiencies and providing additional help or motivation. Teachers could benefit from seeing a summary of areas of weakness 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 is desired by and meaningful to stakeholders. So, by considering the social processes of learning outside of software, the assessment technologies described herein are intended to enhance the learner's experience and support network, resulting in effective, efficient, and enjoyable instruction.

Challenges and Barriers to this Vision

Though the vision and agenda of research discussed herein have 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.

Generalizable Educational Models. The vision we have outlined depends to a significant degree on the success of models in generalizing among educational 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") [cf. 1], 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 (e.g., to new environments), 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 educational objects and how the competencies and constructs advocated here manifest themselves within such 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.

Considering the Stakeholder. We have presented a need for more focused research on the 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) 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 paper 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.

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 [2] 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.

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.

Summary

What are the critical research and development questions that can begin to move us toward the vision we have described? The principal goals will be to figure out (a) which 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 systems (e.g., educational games). Modeling, assessing, and supporting students in relation to an expanded set of competencies and attributes is intended to allow students to grow in 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. Such research is 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, controlled evaluations need to be conducted on advanced, online educational systems to determine what works, for whom, and under what conditions.

In conclusion, we foresee three major funding targets to move these ideas (and the field) forward. This includes 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? and (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?

References

[1] 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.

[2] 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.

[3] 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.

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

[5] 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.

[6] 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.

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

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

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

[10] 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.

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

[12] 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.

[13] 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.

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

[15] Van Merriënboer, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID* model, Educational Technology, Research and Development, 50(2), pp. 39-64.

[16] Feng, M., Beck, J., Heffernan, N. & Koedinger, K. (2008) Can an Intelligent Tutoring System Predict Math Proficiency as Well as a Standardized Test? In Baker & Beck (Eds.). Proceedings of the 1st International Conference on Education Data Mining. pp.107-116. Montréal 2008.