Intelligent Systems

V J Shute, Florida State University, Tallahassee, FL, USA
D Zapata-Rivera, Educational Testing Service, Princeton, NJ, USA

Introduction

Different models of educational measurement engender different instructional practices in the classroom or within intelligent instructional systems, and thus have different effects on student learning. Historically, the main aim of measuring students' educational progress was to identify differences among students in order to rank order them by achievement. This type of measurement model makes heavy use of summative assessment, which is useful for accountability purposes but only marginally useful for guiding day-to-day instruction and supporting student learning. In contrast, student-centered measurement models rely mostly on formative assessment, which can be very useful in guiding instruction and supporting student learning, but provide an inadequate basis for accountability purposes. Now, instead of considering these as disjunctive measurement models, it may be possible to effectively combine them within new, enhanced intelligent systems that use both kinds of assessments – summative and formative – and which leverage computer technology, educational measurement, and cognitive science to address problems that face education today.

A unified approach to educational measurement may be justified based on the following assumptions: (1) individual differences among students have powerful effects on learning, (2) these effects can be quantified and predicted, and (3) technology can capitalize on these effects to the benefit of teachers and students (as well as others, such as administrators and parents). The goal is to figure out how to effectively and efficiently integrate appropriate assessment and instruction to improve student learning and education in general. This is substantially easier said than done, although significant advances have been made in both educational-measurement technologies and intelligent systems.

We begin this article by broadly defining educational measurement, specifically in terms of the critical role assessment plays in education. Summative and formative assessment both have their place in education and we examine how they may be more tightly and strategically coupled. Next, we define intelligent systems, focusing on the usage of assessment data by computers to make inferences about students' cognitive and other attributes. This provides the foundation for rendering decisions about what content to present next to the student. Measuring what students know and do not know (and to what degree) is the first step in remediating students or advancing them to higher levels. This requires the development and use of valid and reliable assessment tools and results to make accurate diagnoses and guide learning. We conclude with an outline of an approach to incorporating evidence-based assessments into intelligent systems to improve learning.

Definitions

Educational Measurement

Educational measurement may be broadly defined as the application of a standard scale or measuring tool to determine the degree to which educationally valuable knowledge, skills, and abilities have been acquired. According to the website of the National Council on Measurement in Education, this includes theory, techniques, and instrumentation available for measurement of educationally relevant human, institutional, and social characteristics. We measure to obtain information, and such information may or may not be useful, depending on the accuracy of the instruments and the skillful manner with which they are used.

Assessment is a general term that includes testing. Progress toward educational goals is typically assessed through testing. Assessment is both an instrument and a process by which information is obtained relative to a known objective or goal. Since inferences are made about what a person knows on the basis of his/her responses to a limited number of assessment tasks or items, there is always some uncertainty in inferences made on the basis of assessments. The goal in educational measurement is to minimize uncertainty or error; thus, key aspects of assessment quality are validity and reliability. Reliability refers to the consistency of assessment results – the degree to which they rank order students in the same way. Validity refers to the extent to which the assessment accurately measures what it is supposed to measure, and the accuracy of the inferences made from task or test results.

Types of Assessment

We consider here two main types of assessment: summative and formative. Summative assessment reflects the traditional approach used to assess educational outcomes. This involves using assessment information for high

International Encyclopedia of Education (2010), vol. 4, pp. 75-80

stakes, cumulative purposes, such as promotion, certification, and so on. It is usually administered after some major event, like the end of the school year or marking period or before a big event, like college. Benefits of this approach include the following: (1) it allows for comparing student performances across diverse populations on clearly defined educational objectives and standards; (2) it provides reliable data (e.g., scores) that can be used for accountability purposes at various levels (e.g., classroom, school, district, state, and national) and for various stakeholders (e.g., students, teachers, and administrators); and (3) it can inform educational policy (e.g., curriculum or funding decisions).

Formative assessment reflects a more progressive approach in education. This involves using assessments to support teaching and learning. Formative assessment is tied directly into the fabric of the classroom and uses results from students' activities as the basis on which to adjust instruction to promote learning in a timely manner. This type of assessment is administered much more frequently than summative assessment, and has shown great potential for harnessing the power of assessments to support learning in different content areas and for diverse audiences. In addition to providing teachers with evidence about how their students are learning so that they can revise instruction appropriately, formative assessment may directly involve students in the learning process, such as by providing feedback that will help students gain insight about how to improve, and by suggesting (or implementing) instructional adjustments based on assessment results.

We now turn our attention to intelligent computer-based systems which have been around for several decades, but have yet to be fully embraced by education. Their primary goal is to enhance student learning; so assessment should, in theory, play a key role in these systems.

Intelligent Systems

Intelligent systems (also known as intelligent tutoring systems) refer to educational software containing an artificial-intelligence component. The software tracks students' work, adjusting feedback and providing hints along the way. By collecting information on a particular student's performance as well as other cognitive and noncognitive variables, the software can make inferences about strengths and weaknesses, and can suggest additional work.

A summary of requirements for intelligent systems was presented by Hartley and Sleeman in the early 1970s. They argued that these systems must possess: (1) knowledge of the learner (student model), (2) knowledge of the domain (expert model), and (3) knowledge of teaching strategies (pedagogical model). It is interesting to note that this simple list has not changed in more than three decades; however, advances have been made in each of the three areas. All of the computer-resident knowledge marks a radical shift from earlier knowledge-free, computer-assisted instructional programs. Furthermore, the ability to diagnose students' errors and adapt instruction based on the diagnosis represents a key difference between intelligent versus other computer-based systems, such as simulations. Intelligent systems are also aligned with the features and goals of formative assessment. The three main components of intelligent systems – student, expert, and pedagogical models – are now briefly described.

A student learns from an intelligent system primarily by solving problems – ones that are appropriately selected or tailormade, and that serve as learning experiences for that student. The system may start by assessing what the student already knows. Information about the student is maintained within what is called the student model, which is updated during the course of learning. The system then must consider what the student needs to know. This information is embodied in the domain-expert model. Finally, the system must decide what unit of content (e.g., assessment task or instructional element) ought to be presented next, and how it should be presented. This is achieved by the pedagogical model (or tutor). From all of these considerations, the system selects or generates a problem, then either works out a solution to the problem (via the domain-expert model), or retrieves a prepared solution. The intelligent system compares its solution to the one the student has prepared and performs a diagnosis based on differences between the two as well as other information available in the student model. Feedback is offered by the system based on considerations such as how long it has been since feedback was last provided, whether the student already received some particular advice, and so on. After this, the program updates the student model, and the entire cycle is repeated, starting with selecting or generating a new problem.

Despite the great promises of intelligent systems, they are currently not widely used in classrooms, partly because of their cost, and also because of measurement limitations. We now focus on the latter in more detail, describing how assessments differ between traditional intelligent systems and newer, enhanced intelligent systems. This is intended to provide the foundation on which to consider a new view of educational measurement within intelligent systems.

Assessments' Role in Intelligent Systems

For the most part, traditional intelligent systems use a formative assessment model, where different student actions invoke different instructional decisions or paths. This comprises the basis for adaptive instruction. New enhanced intelligent systems extend the assessment capabilities of traditional systems. Some of these enhancements include the use of: evidence-based assessment data, explicit links to state curriculum standards, formative and summative sources of assessment information, new measurement techniques from educational psychology and cognitive science, and an explicit and strong role for teachers. Both types of intelligent systems are now discussed in turn.

Traditional Intelligent Systems

As noted earlier, formative assessment is explicitly intended to support student learning, defining the role of the student as an active, creative, and reflective participant in the learning process. Learning environments that make use of formative assessment typically include individualized instruction, along with hands-on, authentic learning activities. Assessments are used primarily to inform teaching and improve student learning.

One major downside of this model is that formative assessment is often implemented in a nonstandardized and hence less-rigorous manner than summative assessment. This can hamper the validity and reliability of the assessment tools and data. As the validity and reliability of the assessment data affect the accuracy of the student diagnosis, and the diagnosis informs instructional support, if the first part of the chain is weak, the rest (i.e., diagnostic accuracy and effective instructional support) would consequently be compromised. In other words, the effectiveness of an intelligent system in achieving its goal hinges on the goodness of information in the student model (i.e., the inferences about what the student knows and can do).

Traditional intelligent systems that employ formative assessment utilize a rich source of student data from which to draw inferences. For example, evidence is captured from all past and current student–system interactions, and may differ in type and grain size. Thus, in addition to the nonstandardization of methods for implementing formative assessment in traditional intelligent systems, there are also problems with accurately modeling student knowledge within such multifaceted environments. This poses a number of psychometric challenges (e.g., modeling of multiple abilities, capabilities, and other learner characteristics) regardless of the measurement model employed.

We now take a closer look at new intelligent systems that are starting to integrate formative and summative sources of assessment information, and are employed within real classroom settings.

Enhanced Intelligent Systems

Most current intelligent systems reside primarily in the laboratory. This isolation from real classrooms explains why their designs have not been overly concerned with summative types of assessment, and also explains to some extent why they have not been widely adopted. That is, learning systems deployed within laboratory-based environments do not have to comply with the same high standards (e.g., accountability requirements), as in real-classroom environments. However, as these systems move out of the laboratory and into the classroom, the need for accountability (e.g., standards and norm-referenced assessments) increases.

Summative assessments are explicitly designed for accountability purposes. They represent a source of valid and reliable evidence of student knowledge. Due to national and international accountability requirements and interests, summative assessments are widely used in schools. For example, in the US, summative assessments have received increased attention after the US Congress passed the No Child Left Behind (NCLB) initiative in 2001. And on the international front, the Programme for International Student Assessment (PISA) is being used to compare student achievement in countries all over the world. The measurement community has made important advances in the development of psychometric models (e.g., Rasch, item-response theory) that provide reliable and valid assessment information, typically presented as a single measure of ability at a particular point in time for any given student. These data, however, have limited use for formative purposes. And one often-cited downside of this emphasis on accountability is that teachers tend to view testing as time taken away from valuable instruction and learning.

Over a decade ago, Snow and Mandinach called for the development of principles for creating valid and useful instructional-assessment systems. We are, only now, beginning to see intelligent systems entering classrooms that integrate sound assessment and instruction. These systems are characterized by (1) a strong presence of teachers in all phases of the project, (2) a cognitive model that is used to drive instructional and assessment interactions, and (3) explicit connections to state standards and standardized state tests.

An example of a successfully deployed intelligent system can be seen in the web-based cognitive tutors designed and developed by Anderson, Koedinger, and colleagues at Carnegie Mellon University. A derivation of their cognitive-tutor approach is called assistments – the merging together of robust assessment with instructional assistance into one system. Assistments use real (i.e., released) items from the Massachusetts comprehensive assessment system (MCAS) state exams within the system for both assessment and instructional purposes.

Table 1 summarizes the main features that separate traditional from enhanced intelligent systems with regard to the role assessments play.

While assessments provide a good example of joining formative and summative models within an intelligent system, this important blending is still uncommon despite calls for their union. A few other systems demonstrating similar capabilities include: Sistema de Evaluacion Inteligente mediante Tests (SIETTE) by Conejo and colleagues, adaptive content for evidence-based diagnosis (ACED) by Shute and colleagues, and English assessment-based learning environments (ABLE) by Zapata-Rivera and colleagues. The next section presents an evidence-based approach designed to create valid assessments for summative or formative purposes, and which may be implemented as part of an intelligent system.

Evidence-Centered Design and Intelligent Systems

An intelligent system that includes valid assessments, for formative and/or summative purposes, must elicit behavior from the student that bears evidence about key skills and knowledge. In addition, the system must provide principled interpretations of that evidence in terms that suit the purpose of the assessment. Figure 1 sketches the basic structures of an evidence-centered approach to assessment design, formulated by Mislevy and colleagues at Educational Testing Service (ETS).

Table 1 Assessments' role in traditional vs. enhanced intelligent systems

Issue	*Traditional Systems*	*Enhanced systems*
Design methods based on evidentiary argument	Mostly absent	Present
Assessment focus	Mostly formative assessment	Formative and summative assessment
Links to standards	Mostly absent	Present
Measurement models	Largely *ad hoc*	Variegated, informed by advances in educational measurement and cognitive science
Evaluations	Mostly laboratory based	Classroom based
Role of teacher	Very limited or absent	Strong

Working out these variables and models and their interrelationships is a way to answer a series of questions posed by Messick in the early 1990s that get at the very heart of assessment design:

- What complex of knowledge, skills, or other attributes should be assessed? A given assessment – formative or summative – is meant to support inferences for some purpose, such as a licensing decision, provision of diagnostic feedback, guidance for further instruction, or some combination. Variables in the competency model describe the knowledge, skills, and abilities on which the inferences are to be based. The term student model is often used to denote a student-instantiated version of the competency model. That is, values in the student model express the assessor's current belief about a student's level on variables within the competency model.
- What behaviors or performances should reveal those constructs? An evidence model expresses how the student's interactions with, and responses to, a given problem constitute evidence about student-model variables. Observable variables summarize aspects of specific task performances, and may come from either formative or summative sources. Then depending on the type and origin of the source(s) of evidence, different parameters are used to update the student model.
- What tasks or situations should elicit those behaviors? Task-model variables describe features of tasks or situations that will be used to elicit performance. A task model provides a framework for characterizing and constructing situations with which a student will interact to provide evidence about targeted aspects of knowledge. The task models will vary in line with the purpose of the assessment and its administration.

Within intelligent systems employing such evidence-based assessment, the student model would accumulate and represent belief about the targeted aspects of skill. These beliefs are often expressed as probability distributions

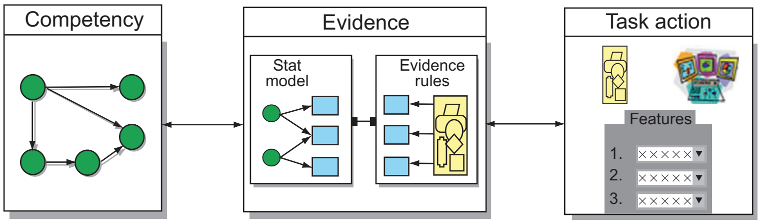

Figure 1 The three central models of an evidence-centered assessment design. Reproduced from Mislevy, R. J., Steinberg, L. S., and Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspective*, **1**(1), 3–62.

for student-model variables. The way that this works in practice is that the evidence model extracts observables (e.g., scores) from student work (i.e., what the student says or does), and provides, a way to aggregate scores that are then used to update the student model. In other words, the evidence model describes how evidence about a set of skills is connected to the competency-model variables using a psychometric model. Task models express situations that can evoke required evidence.

Based on the information in the student model, a viable approach to select and deliver content to the learner is needed – one that fits his or her needs at the time. This would provide context and coherence for delivering adaptive instruction, one of the main goals of intelligent systems. Following is an example of a model to support the select-and-deliver goal. It has been extended from the simpler two-process model that resides at the core of intelligent systems – diagnosis and prescription – and from a process model to support assessment, by Mislevy, Almond, and colleagues at ETS.

Four-Process Adaptive Cycle

The success of any intelligent system to promote learning requires accurate diagnosis of student characteristics (e.g., algebra knowledge, troubleshooting skill, and engagement). The collection of student information then can be used formatively for the prescription of optimal content, such as hints, explanations, hypertext links, practice problems, encouragement, metacognitive support, and so forth. Student information can also be used in a summative manner, such as providing reports on student achievement. The framework, described in this section, involves a four-process cycle connecting the student to appropriate educational materials and other resources (e.g., learning objects, peers, applications, and pedagogical agents) through the use of a student model, shown as the small human icon at the top of Figure 2.

The main components of this four-process cycle are (1) capture, (2) analyze, (3) select, and (4) present. The solid arrows in the figure show a normal, complete loop used in many intelligent systems, while the dashed arrows show variations of the cycle that have been used in other kinds of systems. For example, the dashed line that goes upward from the student (i.e., the large human icon in at the bottom of Figure 2) to the student model depicts an intelligent system where the student is allowed to interact directly with the student model. The nature of this interaction and the effects on the student model can vary, such as negotiating the value of a particular variable with the system or the teacher. The four processes are now briefly defined.

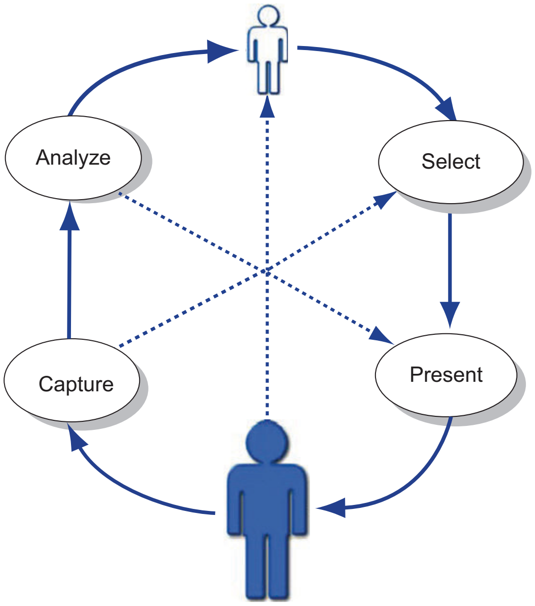

Figure 2 Four-process adaptive cycle. Reproduced from Shute, V. J. and Zapata-Rivera, D. (2008). Adaptive technologies. In Spector, J. M., Merrill, D., van Merriënboer, J., and Driscoll, M. (eds.) *Handbook of Research on Educational Communications and Technology*, 3rd edn, pp 277–294. New York: Lawrence Erlbaum Associates, Taylor & Francis Group.

- *Capture.* The capture process entails gathering personal information about the student as he or she interacts with the environment. Relevant information, obtained through formative or summative assessment, can include cognitive as well as noncognitive aspects of the learner. This information is used to update the internal student model maintained by the system.
- *Analyze.* The analyze process requires the creation and maintenance of a student model by properly integrating evidence sources from student performance in the environment. This usually involves representing information in the student model through inference mechanisms in relation to students' proficiency states based on specific performance data.
- *Select.* Information (i.e., content in the broadest sense) is selected according to the model of the student maintained by the system and the goals of the system (e.g., next learning object or test item). This process is often required to determine how and when to intervene.
- *Present.* Based on results from the select process, specific content is presented to the learner. This entails appropriate use of different media, devices, and technologies to effectively and efficiently convey information to the learner.

Discussion

Educational measurement involves obtaining observations from students' responses in order to make inferences about their knowledge and skills. Reliable instruments are critical in this effort. The way in which information from assessments is captured, analyzed, and ultimately used, strongly influences the student model. First, the granularity of the assessment information influences the kinds of assessment claims that the system can make at any point in time (e.g., general ability vs. component-skill level). Second, the type of evidence that is available from assessment tasks influences the reliability of the claims that the system can make about the student (e.g., data from calibrated test items are more reliable than data from, say, homework assignments). Third, the type of evidence available from assessment tasks influences the validity of the claims that the system can make about the student (e.g., the student is fully proficient at a particular skill vs. more evidence is needed to support the claim). Finally, the complexity of the assessment information – or the manner in which assessment information is interpreted – can range from very simple methods to quite sophisticated ones, such as using probability-based models. To effectively handle the diverse types of assessment information, from either formative or summative sources, an evidence-based assessment framework (e.g., evidence-centered design) is recommended.

Merging instruction with valid formative and summative assessment information in the form of intelligent systems opens up new possibilities and challenges that, if surmounted, could result in improved student learning and educational outcomes. New tools and methodologies are needed to get a more complete profile of the learner. These tools should focus on functions such as: (1) helping create student models that go beyond skills (e.g., conceptual understanding, social aspects of learning, emotional states, and other noncognitive attributes); (2) creating a framework to analyze incoming student work and update student model(s) properly (e.g., evidence-centered design (ECD), along with the four-process adaptive cycle); (3) informing the teacher, in easy-to-understand language, what to do next with the class and/or student and how to understand the data that are derived from the system; and (4) encouraging students to become more active and accountable for their own learning.

Early intelligent system design and development was typically accomplished without much input from psychometricians and/or assessment specialists. This may be partially explained by the chasm between traditional and emergent models of educational measurement. Current advances in intelligent systems, cognitive science, and educational measurement provide the opportunity for integrating assessment and instruction into powerful new intelligent systems. These new systems have the potential to improve student learning and also standardized test scores. Additional benefits of integrating formative and summative assessments with instruction include: (1) re-using rigorous assessment tasks and items from summative tests directly in the service of learning, and (2) increasing the chances of successful adoption of intelligent systems into mainstream education.

Further Reading

Black, P. and Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice* **5**(1), 7–74.

Black, P. and Wiliam, D. (1998b). *Inside the Black Box: Raising Standards through Classroom Assessment*. London: School of Education, King's College.

Conejo, R., Guzman, E., Millán, E., *et al.* (2004). SIETTE: A web-based tool for adaptive testing. *International*. *Journal of Artificial Intelligence in Education* **14**(1), 29–61.

Linn, R. L., Baker, E. L., and Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher* **20**(8), 15–21.

Martin, J. D. and VanLehn, K. (1995). A Bayesian approach to cognitive assessment. In Nichols, P., Chipman, S., and Brennan, R. (eds.) *Cognitively Diagnostic Assessment*, pp 141–165. Hillsdale, NJ: Erlbaum.

Messick, S. (1992). The interplay of evidence and consequences in the validation of performance assessments. *ETS Research Report, RR-92-39*, pp 1–42. Princeton, NJ: Educational Testing Service.

Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In Brennan, R. L. (ed.) *Educational Measurement,* 4th edn., pp 257–305. Westport, CT: Greenwood.

Mislevy, R. J. and Gitomer, D. H. (1996). The role of probability-based inference in an intelligent tutoring system. *User-Modeling and User-Adapted Interaction* **5**, 253–282.

Mislevy, R. J., Steinberg, L. S., and Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspective* **1**(1), 3–62.

National Research Council (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*, Committee on the Foundations of Assessment, Pellegrino, J., Glaser, R., and Chudowsky, N. (eds.). Washington, DC: National Academy Press.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Kaufmann.

Razzaq, L., Feng, M., Heffernan, N. T., *et al.* (2007). Blending assessment and instructional assistance. In Nedjah, N., deMacedo Mourelle, L., Borges, M. N., and Almeida, N. N. (eds.) *Intelligent Educational Machines within the Intelligent Systems Engineering Book Series*, pp 23–49. Berlin: Springer.

Shute, V. J. (1995). SMART: Student modeling approach for responsive tutoring. *User-Modeling and User-Adaptive Interaction* **5**(1), 1–44.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research* **78**(1), 153–189.

Shute, V. J. (2007). Tensions, trends, tools, and technologies: Time for an educational sea change. In Dwyer, C. A. (ed.) *The Future of Assessment: Shaping Teaching and Learning*, pp 139–187. New York: Erlbaum/Taylor and Francis.

Shute, V. J., Hansen, E. G., and Almond, R. G. (2007). An assessment for learning system called ACED: Designing for learning effectiveness and accessibility. *ETS Research Report.* Princeton, NJ: Educational Testing Service.

Shute, V. J. and Psotka, J. (1996). Intelligent tutoring systems: Past, present, and future. In Jonassen, D. (ed.) *Handbook of Research for Educational Communications and Technology*, pp 570–600. New York: Macmillan.

Shute, V. J. and Zapata-Rivera, D. (2008). Adaptive technologies. In Spector, J. M., Merrill, D., van Merriënboer, J., and Driscoll, M. (eds.) *Handbook of Research on Educational Communications and Technology,* 3rd edn, pp 277–294. Mahwah, NJ: Erlbaum.

Snow, R. E. and Mandinach, E. B. (1991). *Integrating Assessment and Instruction: A Research and Development Agenda*. Princeton, NJ: Educational Testing Services.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education* **16**(3), 227–265.

Zapata-Rivera, D., VanWinkle, W., Shute, V. J., Underwood, J. S., and Bauer, M. I. (2007). English ABLE. *Proceedings of the 13th International Conference of Artificial Intelligence in Education*, 8p.