Adaptive Technology for E-Learning: Principles and Case Studies of an Emerging Field

Kathleen Scalise
College of Education, Educational Leadership (Applied Measurement), University of Oregon, Eugene, OR 97403. E-mail: kscalise@uoregon.edu

Diana J. Bernbaum
Graduate School of Education–Quantitative Measurement and Evaluation, University of California, Berkeley, CA 94720

Mike Timms
WestEd, 300 Lakeside Dr., 25th Floor, Oakland, CA 94612

S. Veeragoudar Harrell, Kristen Burmester, Cathleen A. Kennedy and Mark Wilson
Graduate School of Education—Quantitative Measurement and Evaluation, University of California, Berkeley, CA 94720

This article discusses the rapidly emerging field of computer-based assessment for adaptive content in e-learning (National Research Council, 2002), which we call *differentiated e-learning*. In e-learning products, a variety of assessment approaches are being used for such diverse purposes as adaptive delivery of content, individualizing learning materials, dynamic feedback, cognitive diagnosis, score reporting, and course placement (Gifford, 2001). A recent paper at the General Teaching Council Conference in London, England, on teaching, learning, and accountability described assessment for personalized learning through e-learning products as a "quiet revolution" taking place in education (Hopkins, 2004). In our study, we examine approaches for the use of assessment evidence in e-learning in four case studies. The products in the case studies were selected for exhibiting at least one exemplary aspect regarding assessment and measurement. The principles of the Berkeley Evaluation & Assessment Research Center Assessment System (Wilson & Sloane, 2000) are used as a framework of analysis for these products with respect to key measurement principles.

Introduction

Software for differentiated e-learning that anticipates what the user wants or needs and makes suggestions or delivers a personalized product is rapidly emerging in many areas. The basic logic to this type of software is similar across applications: Take what is known about a person and use statistical models or other approaches to guess, or infer, what should happen next. In other words, such software is "assessing" the user and trying to adjust the information each person receives to fit his or her needs.

As in other assessment contexts, the basics of good measurement practices can help to make inferences more accurate and useful. While technology is certainly available to deliver many personalized products to people, good evidence and good conclusions drawn from the evidence can make the difference in how appropriate the fit is for personalized delivery. For instance, Wal-Mart in January had to issue a public apology and take down their personalization system after it made embarrassing mistakes about who should get what product recommendations. As reported in the *New York Times* (Flynn, 2006), their system suggested that "customers who looked at a boxed set of movies that included 'Martin Luther King: I Have a Dream' also might appreciate a 'Planet of the Apes' DVD collection as well as 'Ace Ventura: Pet Detective' and other irrelevant titles." Wal-Mart admitted that some of the combinations the system came up with were offensive, and the trouble with the inferences was attributed to assessment and data-mining technology that needed to do a better job in the decision-making process.

While this example shows the potential pitfalls of making questionable inferences, the promise of personalization also is great. In the world of education, a recent paper delivered

Received June 26, 2006; revised March 5, 2007; accepted March 5, 2007

 • Published online 00 xxxxxx 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20701

at the General Teaching Council Conference in London, England, on teaching, learning, and accountability described assessment for personalized learning as a "quiet revolution" taking place in education (Hopkins, 2004). The paper said that personalization was one of the five principles informing the
AQ1 Government's Five Year Strategy for Children and Learners. Personalized learning was described as (a) tailoring educational products, assessments, and approaches to meet the needs and aspirations of individual learners, and (b) designing teaching, curriculum, and school strategies to create a coherent learning system tailored to the individual pupil.

In light of the potential importance of this emerging area of technology, in this article we present four case studies of e-learning products with assessment approaches that allow for personalized learning and dynamic delivery of content. In other words, the assessment data drive determination of what an individual needs as well as the specific content to deliver to meet those needs (for more details, see later section on "Dynamically Delivered Content and Personalized Learning"). The products are examined for the soundness of their assessment strategies according to a framework that describes four principles of good assessment and measurement practice. The framework is part of the Berkeley Evaluation & Assessment Research Center (BEAR) Assessment System (BAS; Wilson & Sloane, 2000), which describes techniques used in the construction of high-quality assessments. We begin with a description of the four principles, and then consider each case study in light of one of the principles. Each case study was selected as an exemplary model of the application of one of the principles. Developers of software who have a component of customization or personalization can use this information to consider assessment approaches in light of good measurement practice, to better understand the range of decision-making approaches available to technology platforms, and to optimize some of these strategies in regard to drawing inferences about what students know and can do. In addition, those involved in the selection of e-learning products can use such examples as those provided to help evaluate e-learning alternatives in light of the quality of assessment data that is produced from different products and how that data are used, either within the product or subsequently by teachers and students.

The four e-learning products considered in our case studies are:

- NetPASS, a research prototype for the Cisco Learning Institute, that uses Evidence Centered Design (ECD) to develop curriculum and online assessments for students working toward certification in computer networking.
- Quantum Tutors, educational tutoring software developed by Quantum Simulations, Inc., based on approaches in cognitive psychology and learning.
- FOSS Self-Assessment System, a hinting and feedback approach to personalization using Rasch family item-response models.
- ALEKS, which assesses student "knowledge states" and attempts to provide individual and class reports on mastery to teachers and students.

Principles of Assessment

Four principles that any assessment system or approach must address to be useful in learning settings, according to the BEAR Assessment Principles are:

- Assessments should be based on a developmental perspective of student learning.
- Assessments in e-learning should be clearly aligned with the goals of instruction.
- Assessments must produce valid and reliable evidence of what students know and can do.
- Assessment data should provide information that is useful to teachers and students to improve learning outcomes.

Principle 1, a developmental perspective of student learning, means assessing the development of student understanding of particular concepts and skills over time as opposed to, for instance, making a single measurement at some final or supposedly significant time point. A developmental perspective requires clear definitions of what students are expected to learn, and a theoretical framework of how that learning is expected to unfold as the student progresses through the instructional material. Traditional classroom assessment strongly supports a developmental perspective. Here, we affirm what is perhaps the obvious: For diagnostic information to be diagnostic, it must be collected in relationship to some set of goals about what is to be learned.

Principle 2, establishing a good match between what is taught and what is assessed, means that the goals of learning and the measurements and inferences made regarding learning should be related. Reports abound of teachers interrupting their regular curricular materials to "teach the material" students will encounter on district- or statewide tests. Resnick and Resnick (1992) argued that "Assessments must be designed so that when teachers do the natural thing—that is, prepare their students to perform well—they will exercise the kinds of abilities and develop the kinds of skill and knowledge that are the real goals of educational reform" (p.) AQ2
Diagnostic assessment approaches that do not match the goals of instruction fail this test.

Principle 3, quality evidence, addresses issues of technical quality in assessments. By making inferences about students that are reliable, valid, and supported by evidence, e-learning assessment procedures are beginning to gain "currency" in the educational community. Reliability concerns the reproducibility of results whereas validity relates to whether an assessment measures what it is intended to measure. To ensure comparability of results across time and context, these issues must be addressed in any serious attempt at developmental assessment systems to drive personalized delivery in e-learning products.

The fourth principle examined in these case studies, the value of assessment data to teachers and students, is perhaps the most critical: E-learning assessment systems must provide information and approaches that are useful for improving learning outcomes. Teachers must have the tools to use systems efficiently and to explain resulting data effectively and appropriately. Students also should be able to participate

DOI: 10.1002/asi

in the assessment process, and they should be encouraged to develop essential metacognitive skills that will further the learning process. If teachers and students are to be held accountable for performance, they need a good understanding of what students are expected to learn and of what counts as adequate evidence of student learning. Teachers are then in a better position, and a more central and responsible position, for presenting, explaining, analyzing, and defending their students' performances and outcomes of their instruction. Students are better able to develop their own metacognitive skills and to bring them to bear in the learning process. In addition, e-learning assessment procedures should be accessible to teachers to avoid a climate of "black box" assessment, in which the logic of the assessments and personalization are known only to the software developers.

These four principles also relate to the Assessment Triangle developed by the National Research Council Committee on the Foundations of Assessment and published in their report, "Knowing What Students Know" (National Research Council, 2001). The Assessment Triangle, shown in Figure 1, is a model of the essential connections in a coherent and useful assessment system. In this triangle, assessment activities (i.e., the observation vertex) must be aligned with the knowledge and cognitive processes (i.e., the cognition vertex) through the instructional process, and the scoring and interpretation of student work (i.e., the interpretation vertex) must reflect measures of the same knowledge and cognitive processes. Meaningful connections among the three vertices—cognition, observation, and interpretation—are deemed essential for assessment to have an optimal impact on learning.

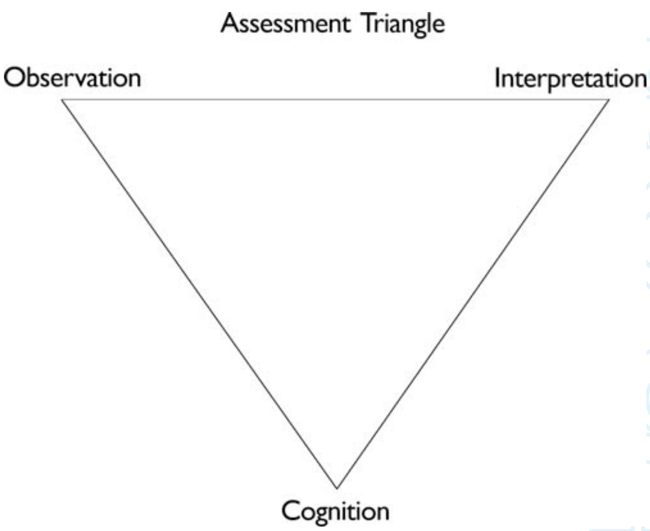


FIG. 1. Assessment Triangle, from the National Research Council Report.

Dynamically Delivered Content and Personalized Learning

Dynamically delivered content, or use of the computer to provide different learning experiences to different students, is of course only one possible form of personalized learning that might be made available to students through e-learning. Obviously, where a person can be involved in differentiating instruction (e.g., a teacher, tutor, or parent), he or she can use information observed or collected, either online or offline, to change instruction and meet the needs of different students. This is a longstanding practice in the very important field of differentiating instruction. Here, in an e-learning context, differentiating instruction can still take place with a person involved; for instance, when assessment reports are generated and teachers, tutors, parents, or students themselves use the reports to reflect on what instruction would be best, or when teachers score performance or other types of complex assessments that are offered over the computer. However, to be clear on our definitions, the term *dynamically delivered content*, sometimes called *data-driven content*, is usually used to mean relatively real-time streaming of different flows of content to different students, using some kind of preprogrammed computer algorithm or back-end database process involving embedded or stand-alone assessments in the e-learning product. Note that the preprogramming is usually not so-called "machine learning," in which the computer itself decides on the meaning of the scores and interprets the outcome space, but this logic is almost always established by people such as teachers and subject-matter experts and then described in terms the software can use to score students. This is the meaning we will use here when the term is mentioned.

Overview of the Principles and Products

Table 1 shows an outline of the products to be examined in the case studies, and the BEAR principle that they were selected to illustrate. As discussed previously, each product

TABLE 1. Overview of products and BEAR principles.

Product	BEAR Principle
1. NetPASS, Cisco Learning Institute	Assessments should be based on a developmental perspective of student learning.
2. Quantum Tutors, Quantum Simulations, Inc.	Assessments in e-learning should be clearly aligned with the goals of instruction.
3. FOSS Self-Assessment System, Univ. of California, Berkeley	Assessments must produce valid and reliable evidence of what students know and can do.
4. ALEKS Products, ALEKS/Mcgraw-Hill	Assessment data should provide information that is useful to teachers and students to improve learning outcomes.

DOI: 10.1002/asi

was selected because it showed some good practices toward the principle it illustrates; however, there could be other good approaches to achieving these principles, and not all of these products achieve all principles, by any means. The conclusions and implications section of this article discusses why it is important to achieve these principles. As the field of adaptive e-learning matures, we expect to see more products establishing a sound basis in these areas.

Principle 1. Developmental Perspective in E-Learning Products

Case Study: Networking Performance Skill System (NetPASS)

The first case study will consider the NetPASS, an assessment of computer networking skills from the Cisco Learning Institute (http://www.education.umd.edu/EDMS/mislevy/
AQ3 CLI/). NetPASS is a performance-based e-learning product for assessment in which students encounter simulations and live interactions in computer network design, implementation, and troubleshooting. Note that NetPASS is used in conjunction with physical classrooms that offer hands-on opportunities for using networking equipment. According to Dennis Frezzo, a Cisco employee who has worked on NetPASS for the past 10 years, "Cognitively, students don't
AQ4 develop the mental model without the hands-on experience." Thus, the developmental perspective used to measure students' networking ability in NetPASS also incorporates a hands-on skills exam with real equipment.

We will consider NetPASS as an example of the use of sound principles for instructional design and assessment with a developmental perspective. Learning goals are clearly defined, and assessments are based on the evidence-centered design approach. NetPASS also is designed to serve students' needs as they progress through a series of increasingly complex performance levels, lending support to the idea of developing proficiency over time.

The Cisco Networking Academy Program (CNAP) teaches students about computer technologies and computer networking. Recognized as a philanthropic educational outreach program, the Academy has been supporting schools with their program since 1997. In just 7 years, more than 1.6 million students in 10,000 Academies have participated
AQ3 in the program (http://cisco.netacad.net/public/index.html). Academies are located in high schools, technical schools, colleges, universities, and community-based organizations in 150 countries.

The Academies rely on instructor-facilitated learning environments in thousands of classrooms around the world. Components of these learning environments include access to online curriculum, access to real networking equipment in the classroom, a trained instructor, and an online assessment system. Part of that online assessment system uses NetPASS to give summative skills-based (i.e., performance) exams in the middle and end of the terms. NetPASS provides diagnostic feedback to students in the Academy program by indicating topics in the online curriculum that students should relearn.

Combining this online Web-based course with hands-on labs and a three-tiered instructor training and support system, NetPASS prepares its students for industry standard certifications in networking. After passing the Certification examinations, students are awarded a certificate to work in the following positions: Cisco Certified Network Associate, Cisco Certified Network Professional, and Cisco Certified
Network Expert (http://cisco.netacad.net/public/index.html). AQ3

NetPASS is an example of an interactive e-learning product that guides students through the curriculum using assessment and formative feedback. It also provides a summative assessment of their knowledge at several points along the way, and makes inferences about a wide range of student knowledge, skills, and abilities in the targeted domains.

Developmental perspective in NetPASS. NetPASS is able to provide a rich, interactive, and personalized learning environment for its students because it is based on a systematic approach to learning and assessment: Evidence Centered Design (ECD). ECD is a four-process approach to measuring students' knowledge and ability:

- Student model, which represents a construct (i.e., knowledge, skills, and abilities) of the domain measured by the assessment.
- Task model, which may include items with many complex features appropriate enough to make evidentiary claims about a student's abilities in the domain.
- Evidence model, a specific argument (based on both quantitative and qualitative measures) about how to interpret students' work given the specifications of the first two models.
- Assembly model, the component of ECD which specifies how the task should be presented to the student (Behrens, Mislevy, Bauer, Williamson, & Levy, 2004; Williamson et al., 2004).

The conception of students' knowledge, skills, and abilities in the NetPASS student model can be considered to be one type of developmental trajectory, describing a progression of student learning. For example, in their third semester of the CNAP curriculum, students are expected to:

- First, use TCP/IP utilities to troubleshoot network connectivity problems,
- Next, identify the cause of connectivity problems at the physical, data link, and network layers of the OSI model. AQ5
- Finally, use a systematic approach to identify and solve network problems.

This hierarchy of knowledge (an example of the student model) forms a basis for the tasks chosen. The students' responses on these tasks provide evidence about the student ability along this trajectory. This developmental model of understanding suggests, for example, that to solve a problem involving networking connectivity, one would first be expected to know how to check for correct Web addresses

DOI: 10.1002/asi

> using TCP/IP utilities. The next level of troubleshooting complexity requires an understanding of the local aspects of networking. This involves physical components that allow the computer to connect to the Internet, such as cables, switches, routers, Ethernet cards, DSL lines, and transporters. The last level of complexity in this model requires knowledge of programming and the use of Web language (e.g., HyperTextTransportProtocol), so that students can troubleshoot the problem when it is beyond the physical space in which they are working.

This perspective of increasing complexity is typical in the computer networking industry, which often refers to "networking in layers." Seven "layers" were created by the
AQ6 Department of Defense to troubleshoot computer networking problems. Cisco employees typically work within the arena of the first four layers, involving cables, Ethernet cards, routers, and transporters. The programming stage layers are typically reserved for the expert network engineers.

Measuring the developmental trajectory. Subject-matter experts worked with the design team at Cisco to identify the skills and knowledge necessary for proficiency in achieving the many certificates offered by Cisco (Williamson et al., 2004). Once a final list was identified, the design team and subject experts also discussed ways in which this knowledge could best be represented. Choosing representations such as log files, configuration files, worksheets, network diagrams, and essays to be used as tasks, the subject experts also described the features that they expected to see on the final work products. These observable features included the steps that students take to identify problems and the problems that they are able to actually locate, as well as relevant connections between the steps followed and problems identified, and the problem-solving logic used for determining the ultimate cause of the problem.

Using experts in the field helps to establish "content validity," or how well the inferences relate to the content area of interest, and helps ensure that the appropriate material is included in the tasks and in the expected responses. Such content validity is one of the many pieces of evidence that help assure that the students have had an opportunity to learn the subjects that will be tested on the certification exams. Like standard setting in the world of large-scale tests, experts help establish criteria for what and when assessments should be presented to students.

Although the focus of this case study is on developmental perspective and not the measurement model used, we mention here that the measurement model used by NetPASS is based primarily on Bayesian networks, which represent beliefs about student proficiencies as joint probability distributions over the proficiency variables (Levy & Mislevy, 2004). NetPASS gathers responses from students and applies the measurement model to ultimately provide students with instructionally relevant diagnostic feedback, which is expected to help them develop their knowledge, skills, and abilities in computer networking.

Research and evaluation in the development of NetPASS. As an evaluative measure, the design team at Cisco used cognitive task analysis (CTA; Newell & Simon, 1972) to help identify the knowledge and strategies that students use while working on a NetPASS assessment. According to Williamson et al. (2004), CTA

> seeks to expose (a) essential features of task situations for eliciting certain behaviors, (b) internal representations of task situations, (c) the relation between problem-solving behavior and internal representation, (d) processes used to solve problems, and (e) task characteristics that impact problem-solving processes and task difficulty. (p.) AQ7

CTA helps to inform the content of the test as well as the construct being measured.

To take an example using the hierarchy of knowledge associated with the third semester of the CISCO Academy curriculum, CTA was used to study students' abilities in designing, implementing, and troubleshooting computer networks. Twenty-four students from high schools and community colleges participated in the study from different geographic locations. The participants were selected to represent a range of abilities described by teachers as high, medium, and low. Teachers from the Academy chose low-, medium-, and high-ability students to participate in the CTA. As a validating measure, a pretest at the beginning of the CTA confirmed the spread of abilities across the group (Williamson et al., 2004).

Four scenarios were given to each of the students, based on their ability level, and the students were asked to think aloud as they solved each of the problems (Williamson et al., 2004). Their solutions were recorded and transcribed for analysis. Ten researchers, computer network instructors, and subject-matter experts examined the students' responses to identify patterns within each of the three ability groups. After examining and discussing students' processes for troubleshooting, they categorized student actions into four categories: gathering information about router configuration, making changes to fix problem with router, testing network after changes, and getting information about commands. This classification system was used to describe patterns of behaviors in performance and identified a hierarchical model within two main categories of behavior: Correctness of Procedure and Correctness of Outcome. Results of these analyses informed the design and implementation of the four models of evidence-centered design (i.e., student, task, evidence, assembly).

Principle 2. Matching Instructional Goals With Assessment in E-Learning

Case Study: Quantum Tutors

Our second case study involves the product Quantum Tutors, developed by Quantum Simulations, Inc. (http://www. AQ3
quantumsimulations.com/). This for-profit company develops artificial intelligence tutoring, assessment, and professional

DOI: 10.1002/asi

development software. The Quantum Tutors are e-learning products designed for students from middle school through college to improve their knowledge and appreciation for the sciences, and are funded by the National Science Foundation, National Institutes of Health, and the U.S. Department of Education. Quantum Tutors is considered a good model of an e-learning product that aligns assessment with instructional goals.

Currently, there are nine tutors developed in the following areas of science: Measurement, The Elements, Ionic Compound Formulas, Mathematics of Chemical Formulas, Equation Balancing, Oxidation Numbers, Chemical Bonding, Chemical Reactions, and Stoichiometry. There also are five math tutors on Measurement, Ratio and Proportion, Percentages, Scientific Notation, and Metric Units.

Quantum Tutors are Internet-delivered with a text-based, dialogue-driven interface that takes a conversational
AQ3 approach (for a sample screen shot, see Figure 2). A demonstration of the Quantum Tutors is available at www.quantumsimulations.com/demo.html. Students enter their own work on the problem, and the Tutor interprets the work and provides coaching and feedback based on the student's responses. Because the Quantum Tutors can accept any problem that the student wants to work on and use artificial intelligence technology to tutor the student one-on-one based on knowledge level and learning style, the software is intended to be compatible with a variety of textbooks, curricula, or state/accreditation standards at both the high-school and college levels.

Pedagogical foundations of Quantum Tutors. To provide students with the right amount of help, the system uses a form of student scaffolding called *cognitive apprenticeship*. This involves modeling how an expert would perform the task so that the students can observe and build a conceptual model of the processes that are required. The approach also involves coaching that consists of observing students while they carry out the task and offering hints, scaffolding, and feedback.

This approach is implemented in the Quantum Tutors by creating a worked-out solution with detailed explanations for numerous problems entered by the student or teacher. While some example problems are provided with the Tutors for convenience, no problems are "prestored" in the system, and students and teachers can enter any problem they create or encounter. One of the most important ways that the Tutors support student inquiry is through recognizing the role of the student's prior knowledge. Thus, the tutoring is created dynamically as students enter their own problem.

Tutors also are designed to allow a student to direct his or her own learning or inquiry by asking questions from a menu that matches his or her current understanding. In other words, the menus are dynamically generated based on the system's assessment of the student's current knowledge. This is where the Tutors begin—where the student has a level of understanding. The Tutors use further prompts to extend a student's thinking beyond his or her current level of understanding, which assists the cognitive apprenticeship.

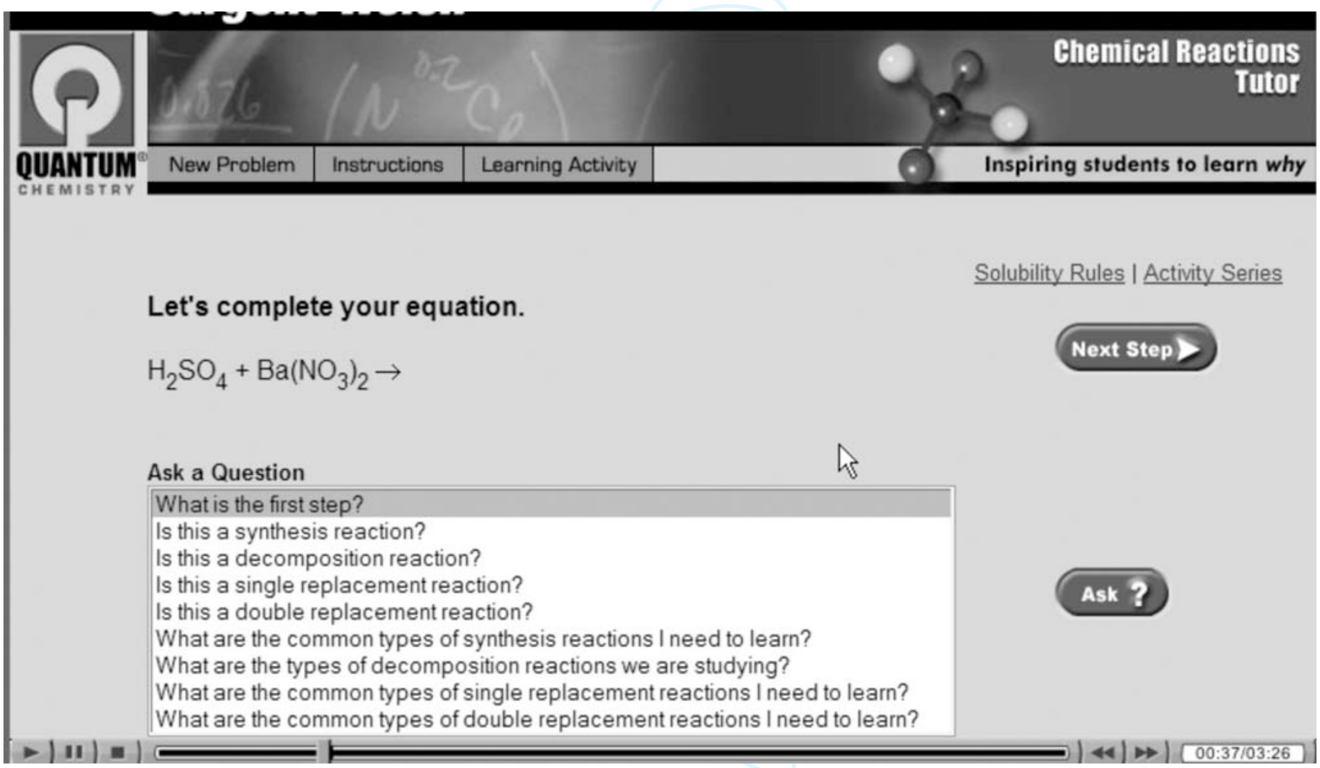


FIG. 2. Sample screen from the stoichiometry Quantum Tutor.

Measurement foundations of Quantum Tutors. Quantum Tutors use the "model tracing" approach of the cognitive tutors developed by the Pittsburgh Advanced Cognitive Tutors (PACT) center at Carnegie Mellon University. Model tracing works by comparing the student's solution of a problem to an expert system for the domain of interest. Production rules, or rules about knowledge and skills in a given domain, are in this system based on an approach from the work of cognitive scientist John Anderson's ACT-R model representing skill-based knowledge (Anderson, 1993; Anderson & Lebiere, 1998).

As the student progresses through the problem solving, the model tracing system generates at each step the set of all possible "next steps" by referring to the production rules. These possible next steps are not displayed to students but are used by the computer to evaluate the quality of the student's next step in problem solving. The computer-generated set of possible steps is called the *conflict set*, and the decision as to which is the best next step to take from the entire set of possible steps is called *resolution of the conflict set*. The computer assesses each of the possible next steps in the conflict set and decides if it is productive, counterproductive, or illegal (i.e., one that violates a fundamental principle). It is the group of productive solutions which the tutor then evaluates as to which is most teachable and presents those options to the student.

What makes the Quantum Tutors different from the traditional Carnegie Mellon tutor approach on which they are based is that the Quantum Tutor generates a list of questions from which the student selects to obtain an answer. This allows for some student control over the system while still providing targeted and scripted feedback. The generation of the questions is implemented in the expert system in a similar way to the productions rules.

Research and evaluation. In February 2005, Quantum Simulations, Inc. released the results of research conducted at Duquesne University in Pittsburgh, PA, where 235 science majors participated in a study as part of a national research effort to measure the effectiveness of the Quantum Tutors (see Table 2). The study randomly divided 14 sections of the Duquesne general chemistry course into two groups: a treatment group ($n = 97$) who used the Quantum Tutors for at-home study and a control group ($n = 138$) who received the same course instruction, but did not use the Tutors.

After both groups received a lecture discussing how to assign oxidation numbers, an important skill needed to understand introductory chemistry concepts, students were given a pretest to assess baseline performance. Both groups were then given identical homework assignments on oxidation numbers to prepare for the posttest, which was given 1 week after the lecture. To accurately measure student mastery and comprehension, students were required to write all steps in solving each problem rather than simply selecting an answer from a multiple-choice list. As an objective measure of performance, only completely correct solutions were accepted; no partial credit was awarded.

The Tutor-using group improved by 45% in number of correct solutions reached, outperforming the control group by an average of nearly 21% more problems solved correctly on the posttest. In addition, 41% of students in the Tutor group solved 80% or more of the posttest problems correctly, compared to only 16% of students in the control group. This evidence was cited as demonstrating the Tutor's ability to help more students gain full mastery of the concepts and skills required to assign oxidation numbers.

Principle 3. Obtaining Quality Evidence

Case Study: FOSS Self-Assessment System

Thus far, this article has considered generating a developmental perspective and arriving at a good match between instruction and assessment. Issues of studying and reporting technical qualities of assessment evidence in e-learning systems are illustrated in our third case study on the Full Option Science System (FOSS) Self-Assessment System, which uses an automated hints system designed by Dr. Michael Timms of WestEd and the University of California, Berkeley.

FOSS is a hands-on approach to teaching science that uses kits and materials to bring inquiry-based science education into classrooms. Established about 20 years ago at the Lawrence Hall of Science at the University of California, Berkeley, FOSS now delivers curricular materials to nearly 1 million students worldwide. The FOSS Self-Assessment System is an e-learning product that has been used to provide supplementary assistance to a small number of students using the FOSS science curriculum on force and motion. The system was developed as part of the NSF-funded Principled Assessment Designs for Inquiry Project, a collaborative

TABLE 2. Results of the Duquesne University Study of the Quantum Tutors.

Duquesne University Study	Tutor Group N=97	Control Group N=138
Student Improvement		
Students Solving 80% or More Post-Test Problems Correctly	41.2%	15.9%
Improvement in Correct Solutions Reached from Pre-Test to Post-Test	+45.0%	+28.0%

DOI: 10.1002/asi

project led by the University of Maryland and SRI International. The FOSS Self-Assessment Systems customizes hints to be delivered to students as they work on problems in which they apply the equation for calculating speed and variations of the equation.

Prior to the construction of the self-assessment system, a literature review on assistance by human and computer tutors was conducted. From the review, a set of principles was developed that guided the decision-making process as the system was built. The self-assessment system has three components: a pretest, practice problems, and a series of "quick-checks." Students first take the pretest so the system can determine their initial ability. The pretest includes one problem from each of the 10 levels in the application of the speed equation. In this first stage of the system, students do not receive any type of assistance from the system. Students then work on practice problems during which they receive error feedback on their answers as well as hints from the system. Students then choose when they feel ready for the next level of problems, at which time they take the quick-check problems. The quick-check includes up to three problems from the student's current problem level and one from each of the preceding two levels to provide sufficient measurement information and an "overlap" of scores between levels.

Decision making in the giving of hints. The heart of the FOSS Self-Assessment System is the ability to measure the gap between student ability and the difficulty of the assessment tasks and activities as a way to decide what hints to give students. Elements involved in the inference-making process include an analysis of both student ability and task difficulty. Ability estimates and item difficulties were calculated from a multivariate item-response model, the multidimensional random coefficients multinomial logit model
AQ8 (Adams, Wilson, & W., 1997). Proficiency on one or more variables can be estimated for each student using this item-response model.

The assessment tasks were chosen to elicit evidence of the features of student proficiency. The key link for reliability and validity evidence in this approach is that estimates of the difficulty of the task can be made with item-response models, such that the task difficulty can be directly compared to the student-proficiency estimates, in the same metric. The tasks can be seen as representing one set of markers or milestones along the learning progression, and student proficiency can be measured by the success rates on these "calibrated" tasks, for which the difficulties are estimated statistically by prior analysis. The tasks themselves, however, do not become a locked-down "map" of what is measured because new tasks can be developed and added into or changed within the system. The "map" is the theory itself—or construct (Wilson, 2005)—of what is being measured whereas the tasks or assessment activities are a sample of many possible ways that student proficiency might be measured, or one set of markers that might be substituted for other markers in the system at any time.

Developing the FOSS system. The FOSS system contains 10 levels of problems and uses item shells, which are prototypes of assessment tasks that share general features, to generate new, similar problems. The items measure knowledge of speed and mathematical ability in the study of force and motion. Knowledge of speed is defined as (a) understanding that speed is a relationship between distance and time, and (b) being able to use velocity equations to calculate speed. Mathematical ability is defined as selecting and using appropriate formulas and equations to solve science problems, understanding the meaning of symbols used in equations, and performing required mathematical computations.

Difficulties of the items in the pretest, the tutorial, and the quick-checks were calculated using a type of generalized Rasch family item-response model called a *two-dimensional partial credit item response analysis* (Wu, Adams, & Wilson, 1998). The task specification connected student responses with measures of knowledge of speed and mathematical ability. During the tutorial phase, the ability estimates of students were calculated by a "scoring engine" based upon the calibrated difficulty of each item in logits. The scoring engine returned an ability estimate for knowledge of speed as well as mathematical ability. The student ability gap was then based on the knowledge of speed ability estimates and item difficulties. It was this calculated gap that determined the types of hints that students should receive.

Research and evaluation. Research on the Self-Assessment System was conducted in three parts over a period of approximately 18 months. Part 1, in May of 2004, was an exploration with 22 students in Grade 8 from a middle school in Berkeley, California, and 4 older students/adults. From this data, the item, or task, difficulty estimates were obtained. In Part 2 of the study, in July 2004, 4 students participated in individual "think alouds" using an alpha version of the software. This part of the study used response processes to build evidence for the validity of the system. Part 3 was a field study in two classrooms in 2005, each at different schools in California, as well as a randomized design study involving the comparison of three different implementations of the Self-Assessment System to determine if students had learned more effectively using the full version of the tutor. One school in California and three schools in Tennessee participated in this randomized design study. The randomized design study was the focus of further analysis on the basis of reliability and validity.

The first implementation of the Self-Assessment System contained the full tutoring system. The second type contained the tutor that only gave error feedback and not hints. The third type offered no help. Results showed that both the fully implemented tutoring system and the error-analysis-only implementation produced statistically significant higher learning gains over the implementation with no personalized assessment feedback. The effect size was about 0.70 *SD*s.

One reason for the similar performance of the partial system with the full system may be an interesting "ceiling

DOI: 10.1002/asi

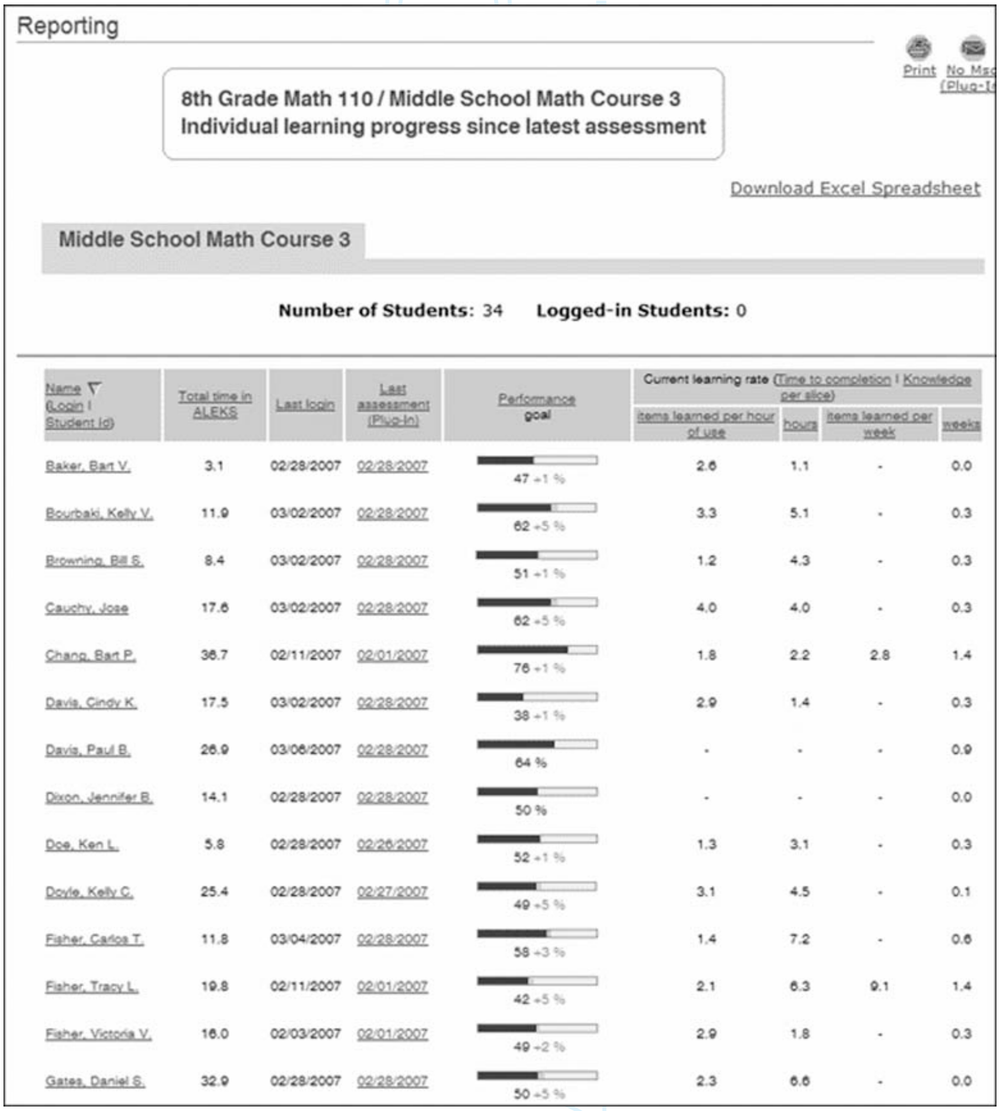
effect” in the instructional design of the hints. Students using the full tutoring system who performed well were quickly classified into mastery learning, where the system offered the least amount of hinting. This level of hinting was essentially like that offered in the error-feedback implementation of the tutor, so it is not surprising that those two groups performed similarly. This shows that it may be important to carefully consider challenge-level implications in the instructional interventions assigned to personalized learning to make sure that at all challenge levels, even for that of exceeding mastery, there are opportunities for growth and new learning.

Principle 4. Making E-Learning Assessment Reports Useful in Classroom Instruction

Case Study: ALEKS

The final case study considers e-learning diagnostic approaches that are useful for teachers and students, beyond just the adaptivity of the software. A comprehensive assessment program not only issues assessments to students and determines their ability level but also synthesizes relevant information into reports that are easily understood and used by those who make instructional decisions. The fourth BEAR assessment principle focuses on the usability of information generated by assessments, and although many programs attempt to assess student knowledge, few provide insight into what the results of the assessment mean in instructional terms.

This case considers ALEKS (http://www.aleks.com/), a math tutoring system, which uses computer-based, authentic AQ3 student-input item responses to determine what ALEKS describes as the student’s “knowledge state.” Examples of well-designed reports that are useful for teachers, students, and other stakeholders can be found in ALEKS. It is an interactive Web-based math tutoring system that uses “authentic” student input—meaning in this case, no multiple-choice response items—to calculate and determine individual proficiency levels within various content areas. It was developed by a collaborative group of cognitive scientists and software


Reporting
Print No Msg (Plug-In)
8th Grade Math 110 / Middle School Math Course 3
Individual learning progress since latest assessment
Download Excel Spreadsheet
Middle School Math Course 3
Number of Students: 34 Logged-in Students: 0

Name (Login I Student Id)	Total time in ALEKS	Last login	Last assessment (Plug-In)	Performance goal	Current learning rate (Time to completion I Knowledge per slice): items learned per hour of use	hours	items learned per week	weeks
Baker, Bart V.	3.1	02/28/2007	02/28/2007	47 +1 %	2.6	1.1	-	0.0
Bourbaki, Kelly V.	11.9	03/02/2007	02/28/2007	62 +5 %	3.3	5.1	-	0.3
Browning, Bill S.	8.4	03/02/2007	02/28/2007	51 +1 %	1.2	4.3	-	0.3
Cauchy, Jose	17.6	03/02/2007	02/28/2007	62 +5 %	4.0	4.0	-	0.3
Chang, Bart P.	36.7	02/11/2007	02/01/2007	76 +1 %	1.8	2.2	2.8	1.4
Davis, Cindy K.	17.5	03/02/2007	02/28/2007	38 +1 %	2.9	1.4	-	0.3
Davis, Paul B.	26.9	03/06/2007	02/28/2007	64 %	-	-	-	0.9
Dixon, Jennifer B.	14.1	02/28/2007	02/28/2007	50 %	-	-	-	0.0
Doe, Ken L.	5.8	02/28/2007	02/26/2007	52 +1 %	1.3	3.1	-	0.3
Doyle, Kelly C.	25.4	02/28/2007	02/27/2007	49 +5 %	3.1	4.5	-	0.1
Fisher, Carlos T.	11.8	03/04/2007	02/28/2007	58 +3 %	1.4	7.2	-	0.6
Fisher, Tracy L.	19.8	02/11/2007	02/01/2007	42 +5 %	2.1	6.3	9.1	1.4
Fisher, Victoria V.	16.0	02/03/2007	02/01/2007	49 +2 %	2.9	1.8	-	0.3
Gates, Daniel S.	32.9	02/28/2007	02/28/2007	50 +5 %	2.3	6.6	-	0.0

FIG. 3. ALEKS Teacher Report: Overall view of the class.

DOI: 10.1002/asi

engineers at the University of California and has been funded by the National Science Foundation. The ALEKS system covers most middle- and high-school math curriculum content organized into units by topic. The content of each tutoring unit was developed based on educational standards, and content coverage is proportionally weighted within each content area.

The tutoring program collects the student responses and estimates student mastery level for specific standards-based skills. The student's proficiency level, or supposed "knowledge state" within a content domain, is determined in a single 40-min ALEKS session, but often includes a more robust set of longitudinal information on each student. The ALEKS assessments use adaptive item administration to quickly and programmatically hone in on the student's ability level. Using the targeted item responses, ALEKS determines the student's level and produces individualized progress reports. The set of achieved and unachieved skills is graphically represented so that each student's report highlights his or her overall progress, progress on specific skills, distance to a short-term goal, and distance to mastery within the area.

Reporting on student ability. ALEKS reports provide immediate and detailed graphical representations of the student's knowledge state through class list tables, progress bars, and pie charts. Reporting information is viewable by the student, teacher, and parent. Class lists and progress bars are used to provide teachers with a quick reference on the distance each student has moved toward obtaining the content goal (see Figure 3).

A variety of pie charts are generated according to the user. One example pie chart (see Figure 4) provides the student with a more detailed look at his or her mastery of each skill. Each slice of the pie chart corresponds to a content area and an educational standard. The slice is shaded according to the level of content mastered by the student. If the slice is completely shaded, like the green section in the example, the student has mastered most of the content from that area. If there is less shading in an area, more information has yet to be mastered. While the dark shading indicates mastery of a portion of the content, the light section indicates unmet skills. Clicking on the light sections of the pie chart opens a list of the skills that the student has yet to master. Some of the more developed units within ALEKS include more than one pie chart corresponding to each major content area. Each of the major content areas have subcontent areas.

Students may use the reports and feedback information to self-direct their own learning while teachers and parents may use the reports to make decisions about what instructional direction to provide the student. For simplicity, a

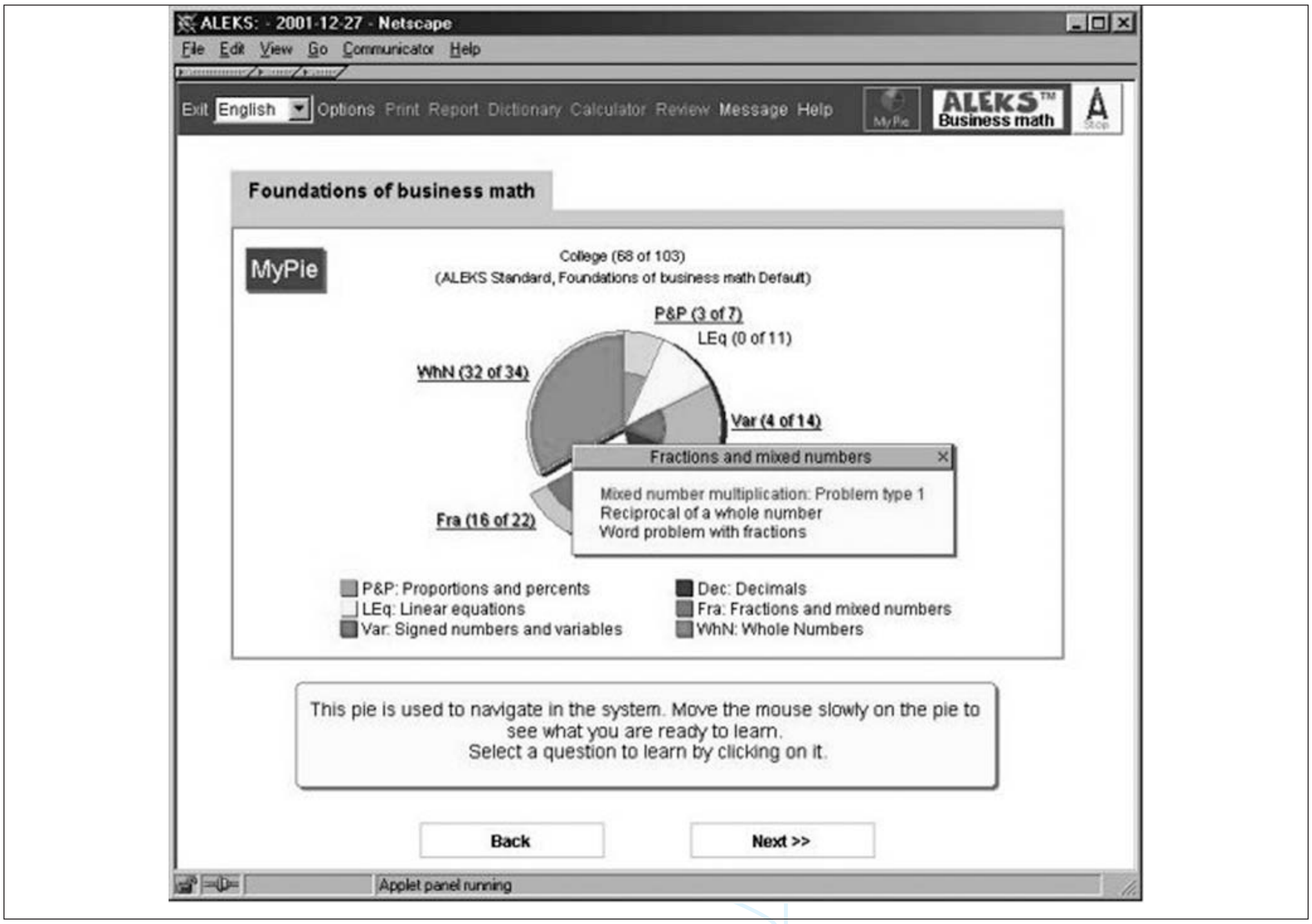


FIG. 4. ALEKS Student Report.

DOI: 10.1002/asi

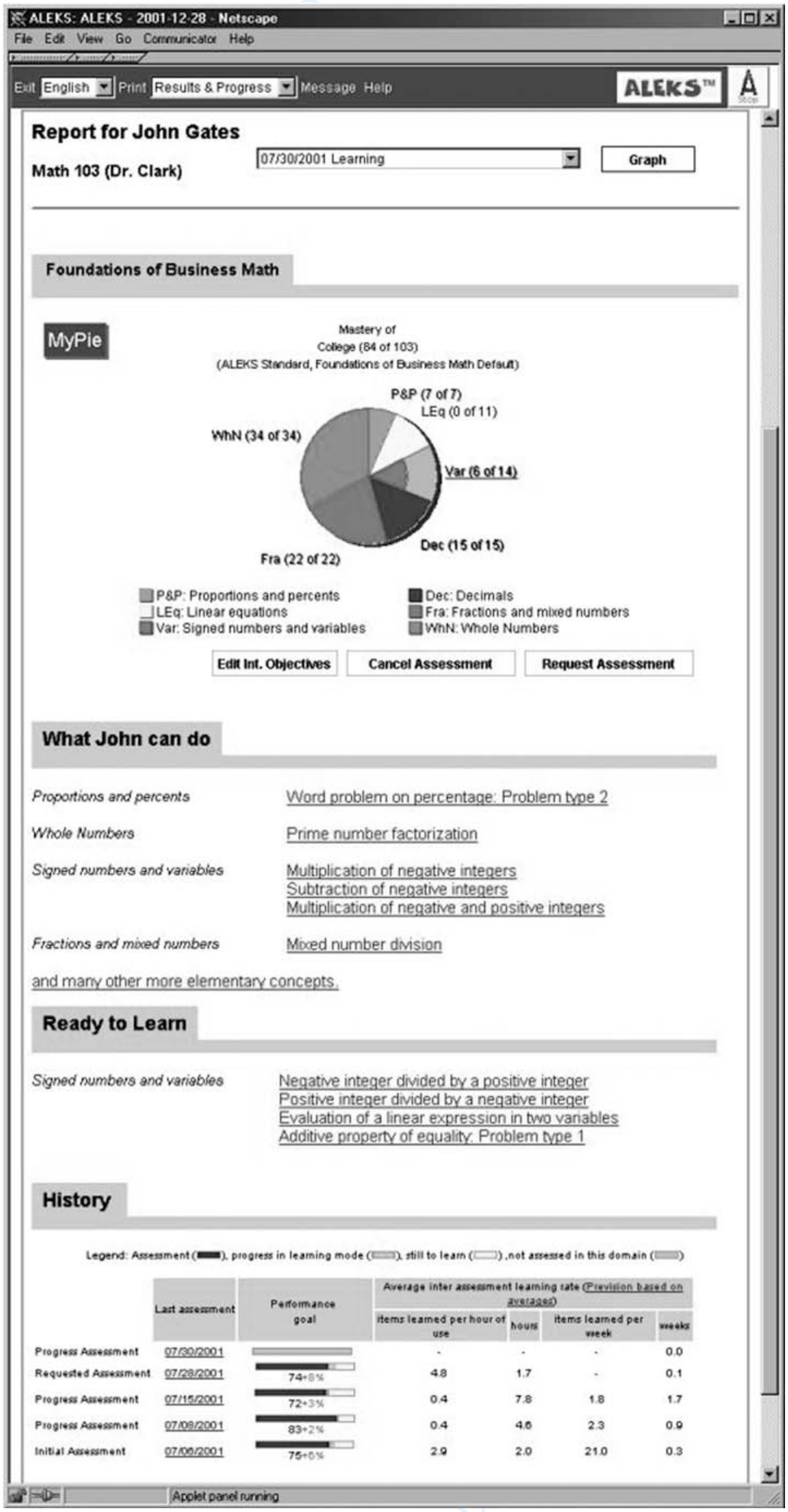


FIG. 5. Teacher Report: A detailed view of one student.

DOI: 10.1002/asi

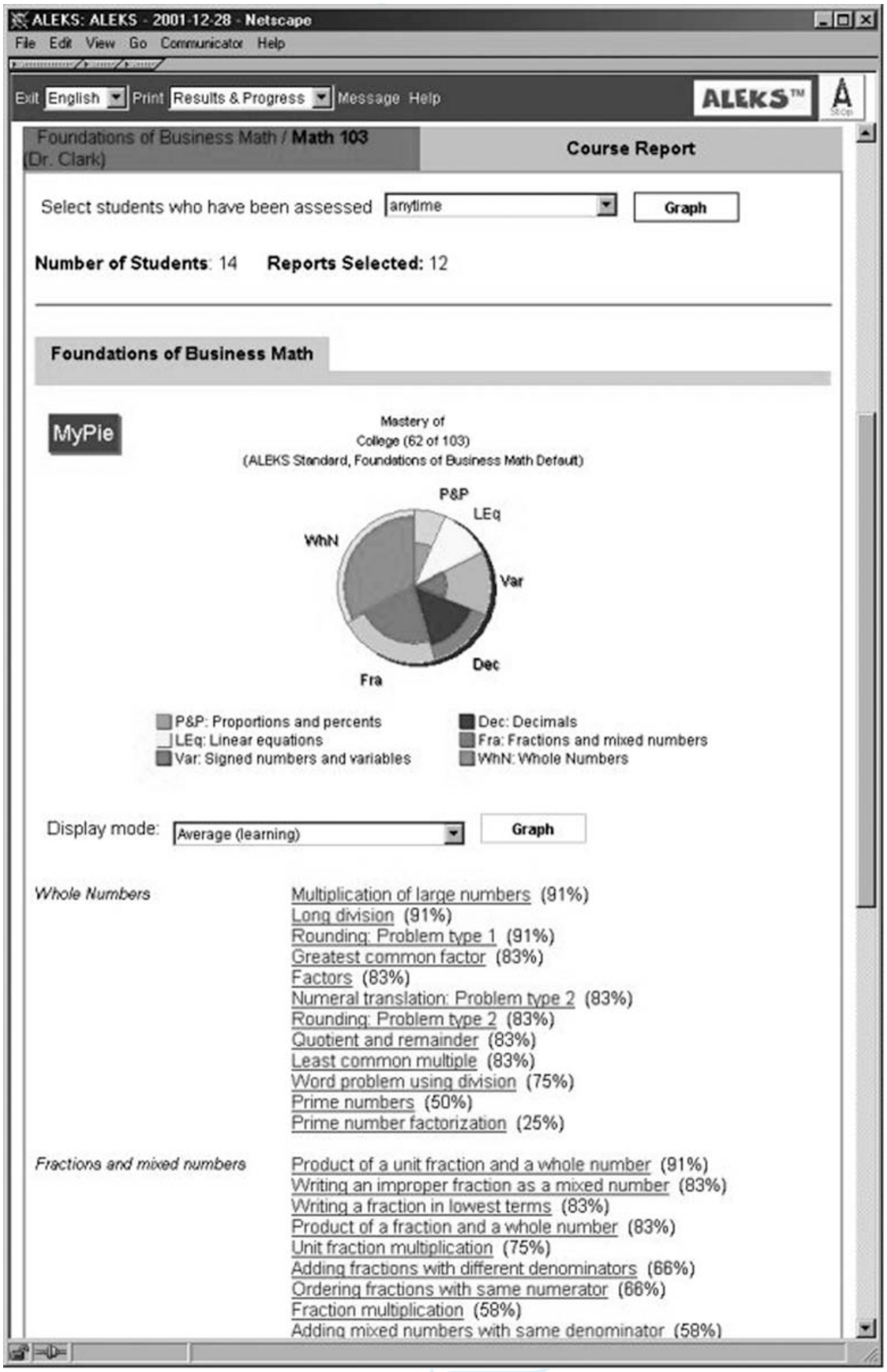


FIG. 6. Teacher Report: A cumulative view of one class.

parent or teacher has the ability to assess all, some, or one of his or her children/students at a time (see Figures 3, 5, and 6). This flexibility allows the teacher a quick view of the whole class, a detailed view of a student, or a detailed overview of the class's strengths and weaknesses. In all reports, the instruction and follow-up has been closely woven into the display. ALEKS is an example of a tightly coupled reporting and instructional system.

Incorporating reporting and instruction. ALEKS not only displays reports to be viewed but also aids in the interpretation of reports and makes precise recommendations

DOI: 10.1002/asi

concerning further instructional study. For the student, the interactive feedback links to relevant practice, and suggests subsequent instruction that the student is most ready to learn based on his or her set of skills. For example, when clicking on a slice of a pie chart, the list of concepts appears with a distinction between concepts that the student has mastered and those which he or she is "ready to learn." The student may click on any of the skills from the list and transition out of the reports, into the relevant instructional content. In the instructional-content section, called the "learning module," ALEKS displays an explanation of the concept, a variety of tools with which to work through the concept, and immediate feedback on the student's performance.

Maintaining a tight link between instruction and assessment, the learning module issues instructional content and assessment items in tandem. Clicking on the skill first presents the student with a short explanation of the skill and an assessment item. The student may attempt to solve the new problem or may seek further explanation on the concept. After further explanation, an item is again offered to test the student's understanding. If the student attempts to answer the question and is unsuccessful, the system analyzes the answer and attempts to evaluate the nature of the error, with advice for continued work on this concept. If the student attempts to solve the problem and is successful, a new problem embodying the same concept is generated. Because of the immediacy of the instructional content and assessment reporting, ALEKS is able to produce content and questions that are close to the student's actual ability.

The teacher reviews the individual or class reports and then has a number of options. First, he or she may investigate which skills have yet to be mastered, either by the individual or by the group, by clicking to obtain a list of the skills that the student or group of students is ready to learn. The instructor may continue to investigate the ready-to-learn skills by reading through the proposed methods of instruction on that topic. Rather than having a static set of instructional pages, standards, and items, the teacher also can use the "instructor module" to edit the items and standards that were used in the generation of assessment reports.

These examples illustrate that for personalized learning, not only is it important for software to have effective means of assessing student understanding but that the interpretation stage also is key. One approach is to synthesize relevant information into reports that are easily understood and used by students for their own reflection and by teachers and others who make instructional decisions.

Conclusions, Implications, and Future Directions

E-learning products that offer personalized learning for students have been described as part of a "quiet revolution" taking place in education (Hopkins, 2004). In this article, we presented four case studies of e-learning products with different assessment approaches that allow for personalized learning and dynamic delivery of content. The products are examined for the soundness of their assessment strategies according to the BEAR Assessment Principles: a developmental perspective of learning, a match between instructional goals and assessment, the generation of quality evidence, and providing information to teachers and students that is useful to improve learning outcomes.

Principle 1, a developmental perspective of student learning, means assessing the development of student understanding of particular concepts and skills over time, as opposed to, for instance, making a single measurement at some final or supposedly significant time point. A developmental perspective requires clear definitions of what students are expected to learn, and a theoretical framework of how that learning is expected to unfold as the student progresses through the instructional materials. In Case Study 1, we examined the Cisco NetPASS product and its approach to establishing a developmental perspective. Theoretical learning trajectories and evidence for their validity are established through a systematic approach involving multiple steps that are both sequential and iterative. Subject-matter experts lend evidence to content validity, and research and evaluation are used to help inform future iterations of the instrument. NetPASS promotes the use of instructionally relevant feedback and uses a model that allows for a variety of content in the assessments used. The evidence-centered design approach in NetPASS helps provide the individualized, diagnostic feedback that is rarely found in most classrooms today.

Principle 2, establishing a good match between what is taught and what is assessed, is illustrated in Case Study 2. Reports abound of teachers interrupting their regular curricular materials to "teach the material" that students will encounter on district- or statewide tests. Personalized e-learning products can avoid this problem if they are matched closely enough to the goals of instruction so that they do not become a source of teachers teaching to a test in which there is no alignment between the assessment and the goals of instruction. The Quantum Tutors example in Case Study 2 illustrates attention to the match between instruction and assessment. Using a "model tracing" approach developed by Carnegie Mellon University's PACT group, the systems compare the student's solution of a problem to an expert system for the domain of interest, mapping both assessment and instruction to the same representation of this domain. Production rules are used to represent the domain knowledge in this system, an approach based on the work of cognitive scientist John Anderson's ACT-R model (Anderson, 1993; Anderson & Lebiere, 1998). By mapping the domain in this way, the software makes inferences about student strengths and weaknesses, and can suggest additional work. This alignment between instruction and assessment helps ensure that assessment data are relevant to the learning process, and is of real use in classroom products.

To make inferences about students that can be supported by evidence, issues of technical quality, Principle 3, have to be addressed. Technical studies on reliability and validity are important in designing quality assessments and also help e-learning assessment procedures to gain "currency" in the educational community. Reliability refers to how reproducible assessment results are, and validity concerns whether we are

DOI: 10.1002/asi

measuring what we say we are measuring. In the FOSS Self-Assessment System of Case Study 3, careful attention to the use of assessment evidence is seen, with Rasch family item-response models helping to provide reliability and validity evidence. The heart of the FOSS Self-Assessment System is measuring the size of the gap between student ability and the difficulty of the problems as a way to decide what hints to give a student. Key to making the system work is that item-response models can be used to estimate the difficulty of each task, and task difficulty can be directly compared to student proficiency. The gap measured is the difference between the two estimates. In FOSS, three tutoring approaches were compared for how much they helped the students: (a) no help; (b) error feedback, but not hints or other instruction; and (c) full tutoring with hints. The mean scores for the full tutor and error-only groups were similar, and both were considerably higher than those of the no-help group. An interesting "ceiling effect" of the full tutor system was that students who performed well were quickly classified into mastery learning, where the system offered the least amount of hinting. This illustrates that it may be important to carefully consider challenge-level implications in the instructional interventions assigned to personalized learning to make sure that at all challenge levels, even exceeding mastery, there are opportunities available for growth and new learning.

The final principle we considered in these case studies also is critical: E-learning assessment products must be interpretable by teachers and students. Teachers must have the tools to use the system efficiently and to explain resulting data effectively and appropriately. Students also should be able to participate in the assessment process, and they should be encouraged to develop essential metacognitive skills that will further the learning process. Case Study 4 considers the ALEKS reporting system. The designers of ALEKS took care to create a reporting system intended to be both easy to understand and easy to use for students, teachers, and parents. They also made efforts toward a tight and immediate linkage between the instructional content and the reporting system. These factors have contributed to the success of ALEKS' e-learning assessment in classroom instruction and the popularity of the product among some of its dedicated users. This example illustrates that attention to reporting functions should be a key aspect of usability analysis for computer interfaces with assessment functions.

The implications of adaptive e-learning products are potentially large. If it indeed turns out that personalized learning is a quiet revolution taking place in education, many students could be affected by adaptive products. Tailoring educational products, assessments, and approaches to meet the needs and aspirations of an individual learner could be a powerful approach to improving learning outcomes, although that is still very much an open question. Little research is available to understand how these products function, much less to describe best practices or fully understanding learning outcomes. It cannot be denied, however, that designing teaching, curriculum, and school strategies to create a coherent learning system tailored to the individual pupil would be a huge change in how we go about teaching and learning.

We can say that the motivation for differentiated instruction (Tomlinson & McTighe, 2006), whether differentiated through teacher intervention or use of other strategies such as computer-adaptive technology, includes that traditional curricular materials and assessments can lead to the production of inert learning activities, sometimes marginally responsive to where the student is in the knowledge acquisition cycle (Gifford, 1999; Hopkins, 2004). By comparison, differentiated instruction approaches are seen as moving teaching and learning activities toward the needs of the student. Technology can help teachers lower the resource barrier for differentiated instruction and also marry potentially powerful assessment tools with new information technologies to capture and analyze student data, rapidly deploy new media, facilitate collaboration, and provide other e-learning amenities such as asynchronous learning (Gifford, 2001; Parshall, Davey, & Pashley, 2000).

Technology to deliver differentiated instruction is now readily available, with back-end databases and a variety of multimedia-rich streaming techniques for which the flow of content to students can be adjusted in near real-time (Turker, Görgün, & Conlan, 2006); however, the inferential machinery necessary to decide who should get what, and the techniques by which such inferences will be made, are mainly lacking or show limited development in most products (Scalise et al., 2006; Timms, 2000). The usual measurement concerns of high-quality data and inferences can quickly derail efforts to make such inferences in an accurate and speedy fashion (Osterlind, 1998; Wilson & Scalise, 2003), threatening to undermine the usefulness of dynamically personalized learning objects and products in the e-learning marketplace.

The size of this potential market for these products also is an open question. How many students and teachers will want to use some form of a differentiated e-learning product? Will school systems and educational leaders choose to adopt these products? Will parents and other stakeholders in education increasingly expect more personalized approaches for learning needs? What about the home market and the business professional development market: Will they move toward embracing some of these products?

Again, it is too early to answer these questions. We can say that the emergence of this market has been suggested for several years now. Surveys such as EPIC (Clark & Hooley, 2003) found that the main benefit educational organizations using online learning hope to achieve is no longer cost savings or course management, but greater access to learning and greater flexibility in learning approaches. This suggests that investments will be made in learning products that do a good job of offering such flexibility. The main factor that organizations felt would determine the success of online learning was good content. This was ahead even of leadership buy-in, implementation practices, and IT improvements. EPIC described how as the e-learning industry matures, the quality of learning design is going up. They described this as the quality tide rising.

DOI: 10.1002/asi

As quality rises and investments are made in e-learning design, best principles and sound practices are important to consider. We hope that developers of software using a component of customization or personalization will incorporate sound principles into their products. Developers should understand the range of decision-making approaches available in assessment and should plan their e-learning products to optimize some good measurement and assessment strategies. This would help fulfill the promise of the emerging field of personalized e-learning and might bring to fruition some new tools that could substantially help instructors and students in the teaching and learning process.

Alternatively, if the e-learning field moves toward black-box assessment, where the logic, scoring, and match to instruction are not transparent to teachers, students, and other stakeholders and quality evidence about inferences is not collected and/or made available, we believe this could set back or undermine educational efforts and learning outcomes. Also note that these products may have increasing impact in education and assessment as they gain market share and enter many educational settings.

It also is important that those who select e-learning products have the information available to become knowledgeable about what e-learning assessments offer, and that they know how evidence and inferences are being used in the systems. Not only are these systems being used for "high-stakes" decisions such as course placement, identifying who has access to honors, and advanced-placement opportunities, and in some cases licensing and accreditation, but we believe that whenever assessment is used to affect learning in the classroom, these are truly the high stakes.

Some additional standards for evidence and reporting are probably necessary in this field. The four principles described here give us a possible lens for what some of these standards might be. We invite others to join this conversation and help shape this important emerging field in educational measurement.

Acknowledgments

This article and the associated e-learning case studies are based on work of the Technology and Assessment Group (TAG) of the University of California, Berkeley.

References

AQ8 Adams, R.J., Wilson, M., & W., W. (1997). The multidimensional random coefficients multinomial logit model. Applied Psychological Measurement, 21(1), 1–23.

Anderson, J.R. (1993). Rules of the mind. Hillsdale, NJ: Erlbaum.

Anderson, J.R., & Lebiere, C. (1998). The atomic components of thought. Mahwah, NJ: Erlbaum.

Behrens, J., Mislevy, R., Bauer, M., Williamson, D., & Levy, R. (2004). Introduction to evidence centered design and lessons learned from its application in a global e-learning program. International Journal of Testing, 4(4), 295–301.

Clark, D., & Hooley, A. (2003). Epic Survey 2003: The future of e-learning. Brighton, England: Epic Group PLC.

Flynn, L.J. (2006, January 23). "Like this? You'll hate that (Not all Web AQ9 recommendations are welcome.) New York Times.

Gifford, B.R. (1999). Computer-mediated instruction and learning in higher AQ10 education: Bridging the gap between promise and practice. Paper presented at the 104th annual meeting of the North Central Association of Colleges and Schools, Commission on Institutions of Higher Education, Chicago. Available: http://ishi.lib.berkeley.edu/sche/projects/university/april12gifford.html AQ3

Gifford, B.R. (2001). Transformational instructional materials, settings and economics. In the case for the distributed learning workshop. Minneapolis, MN: The Distributed Learning Workshop.

Hopkins, D. (2004,). Assessment for personalised learning: The quiet revo- AQ11 lution. Paper presented at the Perspectives on Pupil Assessment, New Relationships: Teaching, Learning and Accountability, General Teaching Council Conference, London.

National Research Council. (2001). Knowing what students know. Washington, DC: National Academy Press.

National Research Council. (Ed.). (2002). Technology and assessment: Thinking ahead—Proceedings from a workshop. Washington, DC: National Academy Press.

Newell, A., & Simon, H.A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Osterlind, S.J. (1998). Constructing test items: Multiple-choice, constructed-response, performance, and other formats. Norwell, MA: Kluwer.

Parshall, C.G., Davey, T., & Pashley, P.J. (2000). Innovative item types for computerized testing. In W. Van der Linden & C.A.W. Glas (Eds.), Computerized adaptive testing: Theory and practice (pp. 129–148). Norwell, MA: Kluwer.

Resnick, L.B., & Resnick, D.P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B.R. Gifford & M.C. O'Connor (Eds.), Changing assessments: Alternative views of aptitude, achievement and instruction (pp. 37–76). Boston, MA: Kluwer.

Scalise, K., Bernbaum, D.J., Timms, M.J., Harrell, S.V., Burmester, K., Kennedy, C.A., et al. (2006). Assessment for e-learning: Case studies of an emerging field [Online]. 13th International Objective Measurement Workshop. AQ12 Retrieved April 5, 200x, from http://bearcenter.berkeley.edu/publications/Scalise_eLearning.pdf

Timms, M.J. (2000). Measurement issues in student modeling for intelligent tutoring systems. Berkeley: University of California, Berkeley POME Position Paper.

Tomlinson, C.A., & McTighe, J. (2006). Integrating differentiated instruction + understanding by design: Connecting content and kids. Alexandria, VA: Association for Supervision and Curriculum Development.

Turker, A., Görgün, I., & Conlan, O. (2006). The challenge of content creation to facilitate personalized e-learning experiences. International Journal on E-Learning, 5(1), 11–17.

Williamson, D., Bauer, M., Steinberg, L., Mislevy, R., Behrens, J., & DeMark, S. (2004). Design rationale for a complex performance assessment. International Journal of Testing, 4(4), 303–330.

Wilson, M. (2005). Constructing measures: An item response modeling approach. Mahwah, NJ: Erlbaum.

Wilson, M., & Scalise, K. (2003). Reporting progress to parents and others: Beyond grades. In J.M. Atkin & J.E. Coffey (Eds.), Everyday assessment in the science classroom (pp. 89–108). Arlington, VA: NSTA Press.

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. Applied Measurement in Education, 13(2), 181–208. AQ13

Wu, M., Adams, R.J., & Wilson, M. (1998). The generalised Rasch model. In ACER ConQuest. Hawthorn, Australia: ACER.

Author Queries

AQ1: Please add a complete entry to the References for the "Government's Five Year Strategy for Children and Learners."
AQ2: Please provide the page number(s) for the quote.
AQ3: Please check the Web address to make sure the link still works.
AQ4: Please indicate if this quote was taken from a printed/published source. If so, list the year in the text and add a complete entry to the References.
AQ5: Please spell out OSI.
AQ6: Please add the year in the text as well as a complete entry to the References for the Department of Defense.
AQ7: Please provide the page number(s) for the quote.
AQ8: Please check the surname of the third author: "W."
AQ9: Please provide the page number(s).
AQ10: Can you provide the date of the meeting? Please also check the Web address to make sure the link still works.
AQ11: Can you provide the date of the meeting?
AQ12: I assume this was part of the proceedings of the workshop, correct? Where was the workshop held? Please add the year of the retrieval. Finally, please check the Web address to make sure the link still works.
AQ13: Please provide the name(s) of the editor(s), if applicable