Visualizing Academic Assessment Data

Elizabeth SKLAR [a,b], Ilknur ICKE [b], Christopher CAMACHO [c], William LIU [c], Jordan SALVIT [a], and Valerie ANDREWLEVICH [a]

[a] *Dept of Computer and Information Science, Brooklyn College, City University of New York, 2900 Bedford Ave, Brooklyn, NY, USA 11210, sklar@sci.brooklyn.cuny.edu*

[b] *Dept of Computer Science, Graduate Center, City University of New York, 365 5th Avenue, New York, NY, USA 10016, iicke@gc.cuny.edu*

[c] *Children's Progress, Inc., 41 East 11th Street, Suite 1100, New York City, NY 10003 USA, chrisc,bill.liu@childrensprogress.com*

Abstract. We explore AI-based methods for analyzing and visualizing data collected from on-line assessments of young students' numeracy and literacy skills. Traditionally, educational assessment data is reported in the form of test scores, often two numbers that each represent a student's achievement or ability in mathematics and language. However, state-of-the-art, adaptive on-line environments are gaining favor for student assessment and these types of systems obtain a wealth of data. Distilling these robust data sets down to two numbers results in great loss of information; and we have been exploring richer ways of presenting and analyzing these data to allow deeper understanding of student performance by teachers, parents and education researchers.

Keywords. Artificial Intelligence, Educational Assessment

Introduction

State-of-the-art, adaptive on-line environments are gaining popularity as an effective and engaging method of student assessment. These types of systems gather robust, multi-dimensional, time-tagged data sets, yet still report results most frequently in terms of one or two statistical values indicating the number of questions a student answered correctly and how that score compares within the student's peer group. We have been exploring richer ways of presenting and analyzing these data sets, to allow deeper understanding of student performance by teachers, parents and education researchers.

The work discussed here involves data collected by an on-line multi-dimensional assessment tool called the Children's Progress[1] Academic Assessment (CPAA) which covers concepts that are essential to early childhood development. It is grouped around *core concepts* in language arts and mathematics: listening, pre-reading, alphabet knowledge, phonemic awareness, reading, writing mechanics, numbers and quantities, numeracy, operations, measurement, and patterns. These concepts were chosen to reflect US national and state academic standards for language arts and mathematics for pre-kindergarten (age 4) through second grade (age 8). In addition, the scoring rubrics for these core con-

[1] `http://www.childrensprogress.com`

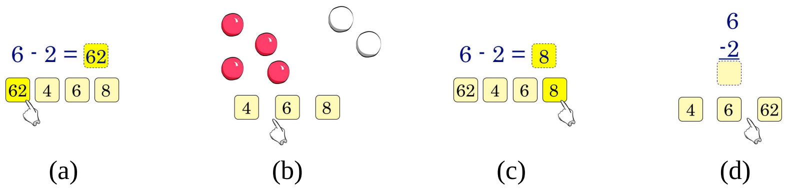


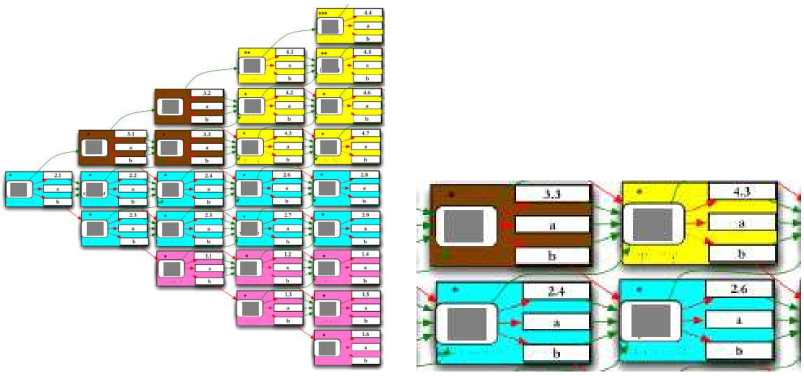
Figure 1. Example from the CPAA assessment tool. The initial question is a simple subtraction problem: $6-2=?$ If the child answers the question correctly, she moves on to more advanced subtraction scenarios. Otherwise, each incorrect response provides hints to guide the child to the correct answer. For example, if the child answers (a) $6-2=62$, then she is presented with a concrete hint: (b) "Here are six balls. If I take away two balls, how many balls would you have?" On the other hand, if the child responds (c) $6-2=8$, then she is presented with (d) a particular hint that is designed to direct her attention to the fact that this is a subtraction problem.

cepts are calibrated to individual state standards regarding end-of-year expectations for each grade. The core concepts are divided into *prime questions* which address specific concept components. For example, phonemic awareness is comprised of prime questions related to rhyming, initial sound, blending, and syllable counting; numeracy consists of individual questions related to correct order, ordinality, and numerical comparison.

The prime questions are organized within the assessment in an adaptive manner. That is, if a child answers a particular prime question correctly, then she would receive a more difficult prime question. For example, if a child is able to correctly answer that $3+2=5$, then she might be presented with a single-digit addition question resulting in a double-digit sum (e.g., $8+5=?$). On the other hand, if a child has difficulty answering any question, then she receives a *hint*—the same question again, presented in a different format. For example, a child might see the question $3+2\ =?$ presented in a vertical format. If the child continues to have difficulty with the question, then the question might be presented with concrete examples, e.g., "Here are three balls. If I give you two more balls, how many balls would you have altogether?" An example is shown in figure 1.

The content of each assessment is organized using a *lattice* data structure (see figure 2) [9]. This is a schematic of the underlying organization of the assessment's questions. The lattices are created manually by the assessment designers and are stored in a hierarchical database. When a student takes the assessment, learner interaction data is collected: user-identified and time-tagged journal entries indicating which question was asked and the answer provided by the student.

Assessment results are reported to teachers in several ways. Detailed *narrative reports* outline the performance of each student, divided into sections corresponding to each core concept covered by the assessment. Summary numeric data on an individual student and a whole class of students can also be viewed by the teacher. However, the amount of information detailed in the narrative reports can be overwhelming, particularly when a teacher has many students in her class. In addition, the narrative reports do not give comparative information indicating how well (or poorly) a student is performing in relation to the rest of her peer group. While classroom teachers are interested in accessing and understanding the wide range of data collected by the CPAA, they lack the tools or skills or experience to manage all the data that has been collected. The work described in this paper outlines our efforts to explore new AI-based methods for analyzing the data and communicating the results of the analysis to teachers.



(a) the overall structure of the lattice (b) a zoomed-in view

Figure 2. Sample lattice. Each node represents a question in the assessment (the actual data is blacked out for privacy). The arrows (links) indicate the direction in the lattice that the student moves if she answers a question correctly or incorrectly. The links pointing up or horizontally are taken if the student answers a question right. The links pointing down are followed if the student answers incorrectly, in which case the "hinting structure" inherent in the definition of the lattice kicks in and the student is asked the same question in a scaffolded form.

Data Analysis

We take an AI-based approach, viewing each student as if she were a *mobile agent* wandering around a *landscape* that is defined by the lattice structure. Using this perspective, we can analyze and visualize the data in a number of ways, as described below.

On top of the lattice structure, we overlay a *concept graph*, as described in [2]. A "concept" is defined as an atomic bit of knowledge within a domain, and theoretically, an entire domain can be represented as a graph of concepts, where each concept is illustrated by a node in the graph. Links connect the nodes and have real-valued weights associated with them, where the weights can indicate the strength of the relationship between the concepts and/or the probability (or possibiity) that a student's path might follow that link. Directionality of the links provides a "curriculum", i.e., an ordering for the presentation of concepts by a human teacher or an automated tutoring system.

Figure 3a illustrates the same lattice shown in figure 2a, as if it were a concept graph. Each question, roughly corresponding to the inner white rectangles labeled "a" and "b" in each colored box in figure 2b, is a node in the concept graph. Each red and green arrow in the lattice is represented by a link in the concept graph, correspondingly colored red and green. We consider this to be the landscape for each agent (i.e., student), where each agent begins at the leftmost node in the graph and ends at one of the nodes on the right edge of the graph. Figures 3b-d illustrate the paths taken through the concept graph by three sample students. At a glance, it is easy to see that they each take very different paths; currently, a teacher would have to read the text of three narrative reports in order to get this information. Still, while it is not difficult to compare two or three of these plots at a time, in order for a teacher to compare the paths of every student in her class, she would need to look at these plots for every student. Again, this could be overwhelming. Further, all the information collected about each student's performance is not shown in these plots. For example, the timing of each question is not illustrated.

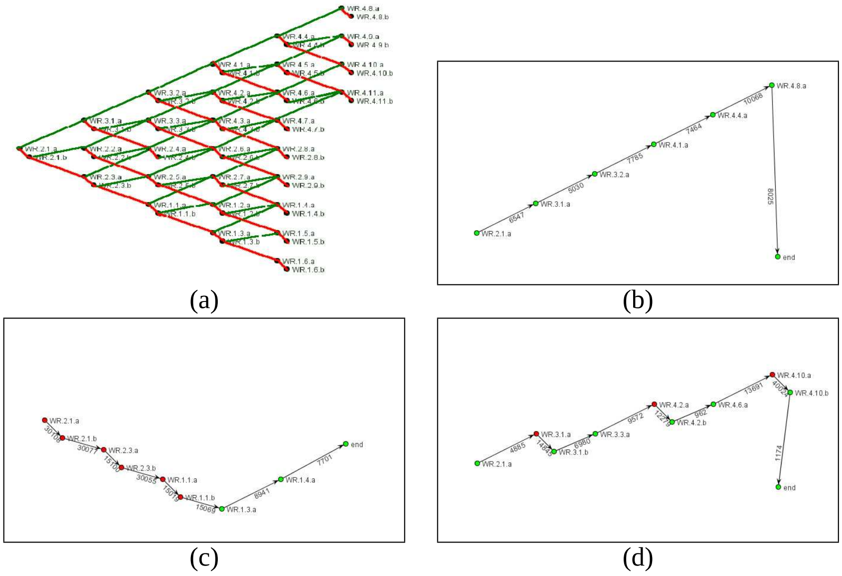


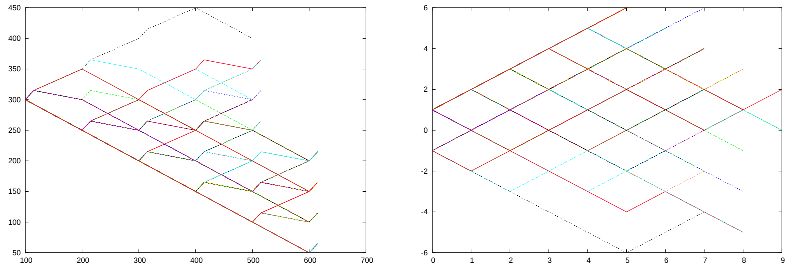
Figure 3. CPAA map for one portion of the lattice and paths taken by three students

We analyzed the student log data for a particular instance of the CPAA assessment given to first grade (age 7) students in Fall 2006. There were 117 students in our sample. Note that the plots shown here are only for the portion of the assessment described above; but these are representative of the entire data set. Figure 4a overlays the same type of plots shown in figure 3 for all 117 students in our sample. This shows the variation in paths taken by the students and indicates that almost every possible path is taken by at least one student. In the plot, the two axes represent the (x, y) locations of each node, as defined in the concept graph (figure 3a). All the students start at the same node, but rapidly diverge, which is interesting given that the group of students are all the same age and all attend the same school.

Figure 4b illustrates the cumulative scores for each student in the sample set. The vertical axis represents score. The horizontal axis indicates "time": each time a student answers a question and moves on to another question, the time is incremented. For easy comparision, we started all students with a score of 0. If a student gets the answer to a question right, then her score increases by 1; if a student gets the wrong answer, her score decreases by 1. Note that this is not how scoring is computed within the actual CPAA assessment, but was adopted to simplify the illustrations here. Again the disparity amongst the students—all of the same age from the same school district—is marked.

It is desirable to pick out which students take similar paths. Currently, we are exploring the use of clustering algorithms in order to group paths with similar features. The results of this work will be able to provide a teacher with clusters of students in her class with skill levels in common for each core concept.

The analyses described thus far focus on the paths taken by students through the assessment's landscape. However, these ignore timing data. In order to get a more complete picture of a student's performance, individually but particularly comparatively within a group, we have developed an animated "video" reporting tool that allows teachers to re-



(a) paths taken through the lattice (b) cumulative scores

Figure 4. Data for 117 sample students

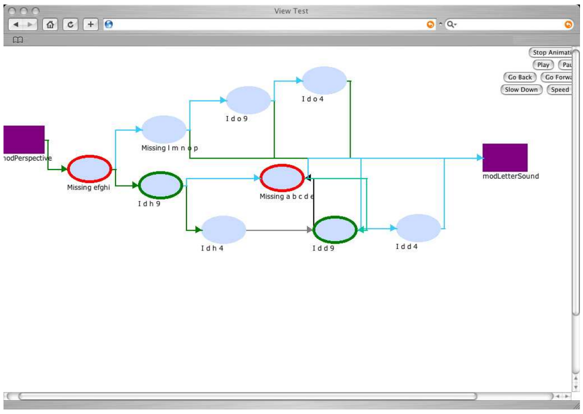


Figure 5. Visualization of the CPAA: Structure-based Video Report.

play the assessment of any of her students, alone or in a group. The tool is interactive in order to allow maximum flexibility for examining individual, or groups of, students. An example screen shot is shown in figure 5. A set of VCR-like controls (located in the upper right corner of the screen), allow the user to "play", "stop", "pause" and even "fast forward." When "play" is pressed, an animation begins that highlights, over time, which nodes a user has visited. Each oval in the figure represents a "prime question" in the assessment tool. The border of an oval is drawn in green when the student has gotten the answer to a question right, and red otherwise. In the figure, the ovals without borders are those which the student never visited.

We take advantage of the fact that the underlying structure of the assessment is hierarchical in order to display the content of the entire assessment, logically split over several screens. The structure represented in figure 5 contains all nodes within the prime question layer corresponding to one of the core concepts. The dark rectangles at the far left and right of the diagram indicate entry and departure points for the concept covered, leading from and to, respectively, the previous and next core concepts in the assessment.

An informal focus group held with in-practice classroom teachers provided feedback that the visual representation of nodes and transition links was too abstract for the typical, technically challenged early elementary school teacher. This visualization works

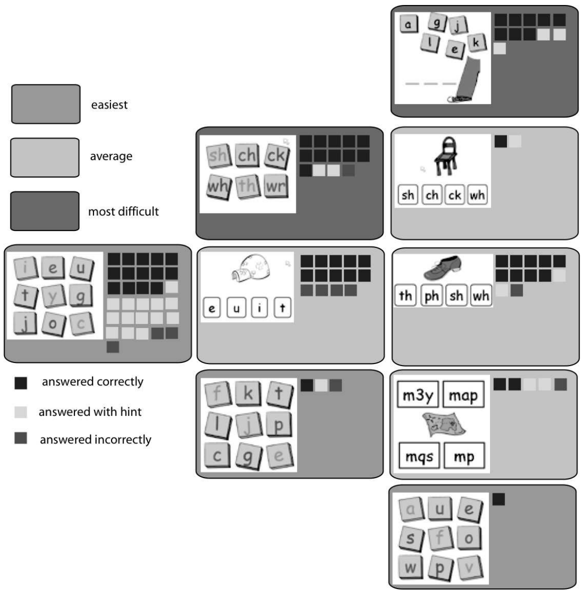


Figure 6. Visualization of the CPAA: Content-based Video Report.

well for showing the underlying structure of the CPAA assessment and demonstrating how a single student progresses from one sub-concept to the next. The mobile agent-based approach, where we explained to teachers that each of their students was being represented as an invidual agent and their paths through the landscape represented the sequences of questions they were asked, worked very well. The teachers saw that as a natural way to connect their students to the large data set they were exploring with our tool. However, for viewing the progress of multiple students at a time, or even an entire class, the visualization was limited.

A second visualization was designed and is illustrated in figure 6. This drawing illustrates a set of nodes relating to one sub-concept in the assessment. Each rectangle corresponds to a question node in the lattice; the content of the rectangles includes a representative portion of the screen (or animation) that appears when the student is being assessed on the concepts in that particular node. In order to overcome the problem of how to represent multiple students progressing through the assessment at the same time, the right half of each rectangle contains a schematic of boxes, where each box represents a student. A color scheme highlights how well the students have performed on each question asked. Another focus group is planned for evaluating this new interface design.

Summary

The intersection of AI techniques and educational applications is broad, from traditional intelligent tutoring systems to the development of intelligent pedagogical agents [11, 10] to the use of data-mining techniques for building student models from interaction data [13] to the use of machine learning techniques for building student models [1] to the use of Bayesian networks for analyzing standardized test scores [15]. Here we have presented a number of data-backed, AI-based methods for examining, analyzing and visualizing data collected in an on-line academic assessment environment. Our goal is to develop techniques that allow teachers to take advantage of the wealth of data that is now available to them from state-of-the-art computerized assessment systems. We have explored several types of data analyses, taking an agent perspective to view student actions as paths through a landscape. Animated visualizations help teachers observe the timing with which students traverse the paths, either individually or in groups. Current work is pursuing the use of simulated learners as a means to predict student outcomes and as the basis for peer tutors in an on-line tutoring and assessment environment.

Acknowledgements

This work was partially supported by NSF ITR #02-19347, NSF PFI #03-32596, and US Dept of Education SBIR #ED-06-PO-0895.

References

[1] Beck, J. E., and Woolf, B. P. 2000. High-level student modeling with machine learning. In *Proceedings of Fifth International Conference on Intelligent Tutoring Systems*, 584–593.

[2] Sklar, E., and Davies, M. 2005. Multiagent simulation of learning environments. In *Fourth International Conference on Autonomous Agents and Multi Ag ent Systems (AAMAS-2005)*.

[3] Sklar, E., and Parsons, S. 2004. Towards the Application of Argumentation-based Dialogues for Education, In *Third International Conference of Autonomous Agents and Multi Agent Systems (AAMAS)*, 1420–1421.

[4] Sklar, E., Parsons, S., and Davies, M. 2004. When is it okay to lie? A simple model of contradiction in agent-based dialogues. In *Proceedings of the Workshop on Argumentation in Multiagent Systems (ArgMAS) at Autonomous Agents and MultiAgent Systems (AAMAS-2004)*.

[5] Sklar, E. 2000. CEL: A Framework for Enabling an Internet Learning Community, PhD thesis, Department of Computer Science, Brandeis University.

[6] Sklar, E., and Salvit, J., and Camacho, C., and Liu, W., and Andrewlevich, V. 2007. An agent-based methodology for analyzing and visualizing educational assessment data, In *Sixth International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2007)*.

[7] Brown, A. H. 1999. Simulated Classrooms and Artificial Students: The Potential Effects of New Technologies on Teacher Education. *Journal of Research on Technology Education (formerly Journal of Research on Computing in Education)* 32(2):307–318.

[8] Daum, S. C. 1997. Using simulated students to improve residents' teaching. *Acad Med.* 72(5).

[9] Galanter, E., and Galanter, M. 2003. Adaptive evaluation method and adaptive evaluation apparatus. United States Patent, no. 6,511,326 B1.

[10] Johnson, W. L.; Rickel, J. W.; and Lester, J. C. 2000. Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education* 11.

[11] Johnson, W. 1995. Pedagogical agents for virtual learning environments. In *International Conference on Computers in Education*.

[12] Mark, M. A., and Greer, J. E. 1993. Evaluation methodologies for intelligent tutoring systems. *Journal of Artificial Intelligence and Education* 4:129–153.

[13] Mavrikis, M. 2005. Logging, replaying and analysing students' interactions in a web-based ile to improve student modeling. In *The 12th International Conference on Artificial Intelligence in Education, AIED 2005, Young Researcher Track Proceedings*, 101–106.

[14] Ohlsson, S. 1996. Learning from performance errors. *Psychological Review* 103:241–262.

[15] Pardos, Z. A.; Heffernan, N. T.; Anderson, B.; and Heffernan, C. L. 2006. Using fine grained skill models to fit student performance with bayesian networks. In *Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems (ITS 2006)*, 5–12.

[16] Spoelstra, M. 2006. Simulating the Effects of Goal Structures in Human Learning Environments. Master's thesis, Artificial Intelligence Section, Department of Computer Science, Faculty of Sciences, Vrije Universiteit Amsterdam, The Netherlands.

[17] VanLehn, K.; Ohlsson, S.; and Nason, R. 1996. Applications of simulated students: An exploration. *Journal of Artificial Intelligence in Education* 5(2):135–175.

[18] Virvou, M., and Manos, K. 2003. A Simulated Student-Player in Support of the Authoring Process in a Knowledge-Based Authoring Tool for Educational Games. In *Third IEEE International Conference on Advanced Learning Technologies (ICALT'03)*.

[19] Vizcaino, A., and du Boulay, B. 2002. Using a simulated student to repair difficulties in collaborative learning. In *Proceedings of the International Conference on Computers in Education*.