Research Report

An Assessment for Learning System Called ACED: Designing for Learning Effectiveness and Accessibility

Valerie J. Shute

Eric G. Hansen

Russell G. Almond

Research & Development

June 2007

RR-07-26

An Assessment for Learning System Called ACED: Designing for Learning Effectiveness and Accessibility

Valerie J. Shute, Eric G. Hansen, and Russell G. Almond
ETS, Princeton, NJ

June 2007

Abstract

This paper reports on a 3-year, NSF-funded research and development project called ACED: Adaptive Content with Evidence-based Diagnosis. The purpose of the project was to design, develop, and evaluate an assessment for learning (AfL) system for diverse students, using Algebra I content related to geometric sequences (i.e., successive numbers linked by a common ratio). A key feature of the project was its dual focus on both learning effectiveness and accessibility. The project consisted of two studies. Study 1 (N = 268) experimentally evaluated the impact on learning of two ACED features: (a) elaborated task feedback and (b) adaptive sequencing of tasks. Study 2 examined the accessibility of ACED for students with visual disabilities (N = 4; two blind and two low-vision)—in a condition involving both adaptive sequencing and elaborated feedback—using a variety of accessibility features. Available features included voicing of test content (via synthesized and prerecorded audio), talking tactile graphics, font enlargement, screen magnification,and so on. Together these studies lay groundwork for the design of AfL systems that can enhance the learning of individuals with and without disabilities.

Key words: Accessibility, adaptivity/adaptation, Bayesian network, diagnostic assessment

Acknowledgments

We gratefully appreciate the support for ACED development and data collection by National Science Foundation Grant No. 0313202, and especially the support of John Cherniavsky, our program officer. We also want to acknowledge the various contributions to the development of ACED by the following: Edith Aurora Graf, Jody Underwood, Peggy Redman, Malcolm Bauer, Buz Hunt, Diego Zapata-Rivera, Irvin Katz, and Dan Eignor. In addition, we are grateful for the contributions from the New Jersey middle school mathematics teachers who assisted us with content issues, and to Angelique Peterson (head mathematics teacher at the high school where we tested Study 1). She organized the Algebra I students for us to test, and Waverely VanWinkle provided excellent support during Study 1 testing and during preliminary data analyses. Finally, we acknowledge the invaluable contributions to the project by the staff at Touch Graphics: Steve Landau, Richard Holborow, Nicole Rittenour, Dan Gieschen, and Ellen Rubin, and by the Overbrook School for the Blind (Philadelphia, PA) for their assistance in pilot testing Study 2.

Table of Contents

List of Tables

Page

List of Figures

Introduction

Problem

An assessment *for* learning (AfL) approach to education involves weaving assessments directly into the fabric of the classroom and using results as the basis to adjust instruction to promote student learning in a timely manner. This type of assessment contrasts with the more traditional, summative approach (i.e., assessment *of* learning), which is administered much less frequently than AfL and is usually used for grading and accountability purposes. In the past decade or so, AfL has shown great potential for harnessing the power of assessments to support learning in different content areas (e.g., Black & Wiliam, 1998) and for diverse audiences. Unfortunately, while assessment *of* learning is currently well entrenched in our educational system (such as through the No Child Left Behind Act, or NCLB), assessment *for* learning is not.

In addition to providing teachers with information about how their students are learning so that they can revise instruction appropriately, AfL systems may directly involve students in the learning process, such as by providing feedback that will help students gain insight about how to improve (Shute, 2007), and by suggesting (or implementing) instructional adjustments based on assessment results (Stiggins, 2002). These promises, however, need controlled evaluations to determine which features are most effective in supporting learning in a range of settings (see Jameson, 2006). Among the AfL features that have the greatest potential for supporting student learning and which would be suitable for investigation are task-level feedback and adaptive sequencing of tasks. Furthermore, it is important to explore potentially beneficial features for feasibility of making them accessible to individuals with disabilities. Given the highly visual nature of mathematics content, it makes sense to give attention to the barriers faced by individuals with visual disabilities.

Task-Level Feedback

Task-level feedback appears right after a student has finished solving a problem or task and may be contrasted with (a) more general summary feedback, which follows the completion of the entire assessment; and (b) more specific step-level feedback, which may occur within a task (VanLehn, 2006), such as the feedback used with intelligent tutoring systems. Task-level feedback typically provides specific and timely (often real-time) information to the student about a particular response to a problem or task and may additionally take into account the student's current understanding and ability level. In general, feedback used in educational contexts is regarded as crucial to knowledge and skill acquisition (e.g., Azevedo & Bernard, 1995; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Moreno, 2004), and may also influence motivation (e.g., Lepper & Chabay, 1985; Narciss & Huth, 2004). Immediate feedback on students' solutions to individual tasks has generally been shown to support student learning (Corbett & Anderson, 2001; Shute, 2007), especially when a response or solution is wrong. That is, when a student solves a problem correctly, it usually suffices to simply provide verification of the accuracy of the response (e.g., "You are correct"). But in the case of incorrect answers, research has suggested that it is more beneficial to provide not only verification of the incorrectness but also an explanation of how one would determine the correct answer (e.g., Kluger & DeNisi, 1996; Mason & Bruning, 2001; Narciss & Huth, 2004). In this research, we focused on feedback for incorrect answers and evaluated the contribution to learning that such elaborated task-level feedback provides relative to simple verification ("correct" or "incorrect") feedback.

Adaptive Sequencing of Tasks

Adaptive sequencing of tasks within an assessment contrasts with the more typical fixed, linear sequencing of tasks or items. Adaptive sequencing usually entails making adjustments to the sequence of tasks based on determinations such as: (a) which task would be most informative for refining an estimate of the student's proficiency level, and (b) which task would be most helpful in supporting the student's progress to a higher proficiency level. The rationale for employing an adaptive assessment is that students come to any new learning task with differing profiles. As educators, we want to take what we already know about students and add to that an understanding of what they are doing in real time in the AfL environment. We can then combine that information with knowledge about strategies for bringing individuals to a higher level of knowledge and adapt content to carry out those strategies. According to Bass and Glaser (2004), taking full advantage of such assessments requires the use of adaptive techniques that yield information about the student's learning process and outcomes. This allows teachers to take appropriate instructional actions and make meaningful modifications to instruction. The idea of tailoring content to fit the needs of learners is intuitively appealing and is being incorporated into more and more learning support systems. However, the effectiveness of adaptivity still lacks strong empirical support (Jameson, 2006; Shute & Zapata-Rivera, in press). These factors motivated the current research. Specifically, we wanted to experimentally test the value that adaptive task sequencing adds to learning compared to fixed (linear) sequencing of tasks.

Accommodations for Individuals With Disabilities

Accommodations are alterations to typical task performance conditions and are intended to overcome accessibility barriers that a student may face when interacting with test materials. Accommodations may include timing (extended testing time, extra breaks), presentation format (font enlargement, audio, braille, raised-line graphics), and response format (mouse, keyboard, tactile tablet). Accommodations in the area of algebra and other math domains are diverse, due in part to the various ways of representing math concepts and data (e.g., figures, charts, tables, and equations). Some individuals with visual disabilities have enough residual sight to benefit from large print or font or screen enlargement. Some need modified colors or contrast. Many others rely on audio (e.g., human reader, screen reader, audiocassette, audio CD; Hansen, Forer, & Lee, 2004; Hansen, & Mislevy, 2005; Shute, Graf, & Hansen, 2005). For graphical material such as figures and charts, raised line (tactile) drawings are often helpful, since description of such content is often more fully and rapidly understood than descriptions delivered solely by audio or braille—presented as hard copy or refreshable via the computer. Some new accommodations combine audio with tactile graphics, such as the Touch Graphics *Talking Tactile Tablet* (TTT). This kind of solution can be very helpful for visually-impaired users who do not read braille. The use of braille labels for audio-tactile graphics helps maximize the utility of this approach for readers of braille. AfLs must be designed to ensure that, to the extent feasible, disabilities do not pose accessibility barriers.

A controlled study of AfL features such as elaborated feedback and adaptive sequencing of tasks is greatly needed. Research is also necessary to explore ways to use accommodations in the area of mathematics to make such benefits available for individuals with disabilities. Addressing issues of accessibility is critical for ensuring the validity of scores for individuals with disabilities. Accessibility features (e.g., accommodations) may be able to remove barriers, enabling valid assessment wherein the scores accurately reflect what a student knows and can do. Partly due to the great diversity of disability and the relatively low numbers of individuals fitting a given profile, it is challenging to develop a research base. However, a reasonable first step is to conduct a study of the usability of an AfL system that implements features to facilitate learning for all students (e.g., elaborated feedback and adaptive sequencing of tasks) as well as accessibility features designed to ensure that the benefits of the system are available to individuals with disabilities.[1] Once we are confident that the system features are usable by certain populations, we can begin to examine other issues contributing to validity and learning effectiveness for those populations.

Purpose

This paper describes the evaluation of an AfL system for use by students with and without visual disabilities. The system, called Adaptive Content with Evidence-based Diagnosis (ACED), provided a platform for investigating the key AfL features of task-level feedback and adaptive sequencing of tasks. The system uses content from secondary school algebra curricula, specifically the topic of *sequences* (e.g., geometric sequences). The general goal of evaluating ACED was to determine whether such an AfL system facilitates learning for individuals without visual disabilities ($N = 268$) and to assess whether the system is usable by individuals with visual disabilities ($N = 4$). The primary research questions include the following:

1. *Learning effectiveness.* (a) For incorrect responses, is elaborated task-level feedback more effective for student learning than simple (verification only) feedback, and (b) is adaptive sequencing of tasks more effective for learning than linear sequencing?
2. *Accessibility*. Is the system usable by individuals with visual disabilities?

General Methodology

Project Strategy

Learning effectiveness was the focus of Study 1 and accessibility was the focus of Study 2. Specifically, Study 1 used a controlled evaluation to examine the impact of task-level feedback and task sequencing on learning. Study 1 also addressed the reliability and validity of the AfL assessment system. Study 2 was a small-scale study of the usability of system features for individuals with visual disabilities.

Project Design and Development

In general, the project involved several major steps, most of which are relevant to both Studies 1 and 2.

Selection of Content

The topic of sequences (e.g., arithmetic, geometric, and other recursive sequences, such as the Fibonacci series) was selected for implementation based on interviews with school teachers in New Jersey, review of the National Council of Teachers of Mathematics (NCTM) standards, and state standards in mathematics. Both Study 1 and Study 2 used content on geometric sequences.

Assessment Design

We employed an evidence-centered design (ECD) approach to create the ACED system. ECD (Mislevy, Steinberg, & Almond, 2003) provides a systematic way of reasoning about assessment design as well as a way of reasoning about student performance. The basic idea of ECD is to specify the structures and supporting rationales for the evidentiary argument of an assessment. By making the evidentiary argument explicit, the argument becomes easier to examine, share, and refine. Argument structures encompass, among other things, the claims (inferences) one wants to make about a student, the observables (performance data) that provide support for those claims, the task performance situations that elicit the observables from the students, and the rationales for linking it all together. The three main models used in ECD are as follows:

- *Proficiency model*: Establishes claims about a particular piece of knowledge, skill, or ability. The proficiency model describes what is to be measured, the conditions under which the ability is demonstrated, and the range and relations of proficiencies in the content area.
- *Evidence model*: Defines the evidence needed to support the claims. Evidence models describe what is to be scored, how to score it, and how to combine scores to support claims. These models thus establish the boundaries of performance and identify observable actions that are within those boundaries.
- *Task model*: Identifies tasks (and characteristics of tasks) that are able to elicit that evidence. Task models specify the inputs required to perform the observable actions as well as the outputs (work products) that result from performing the observable actions.

These models were developed during the first year of the project, while the second year focused on building the system's infrastructure, algorithms, and so on. Details on the system's features and functionality may be found in Shute et al. (2005). This paper focuses on our third year efforts, which involved testing the system on two populations of students: sighted (Study 1) and visually disabled (Study 2). For both groups of students, the evaluation of ACED utilized a subset of the overall proficiency model—*geometric sequences* (i.e., successive numbers linked by a common ratio). This set of proficiencies, about one third of the full set, is sufficiently rich yet not unwieldy, thus satisfying the constraints for a 2-hour testing session. Figure 1 illustrates the main concepts (or nodes) in the proficiency model, and the appendix contains a full description of the proficiencies. Each node in the proficiency model was linked to at least six tasks, with three levels of difficulty (easy, medium, and hard) and two parallel tasks per level.

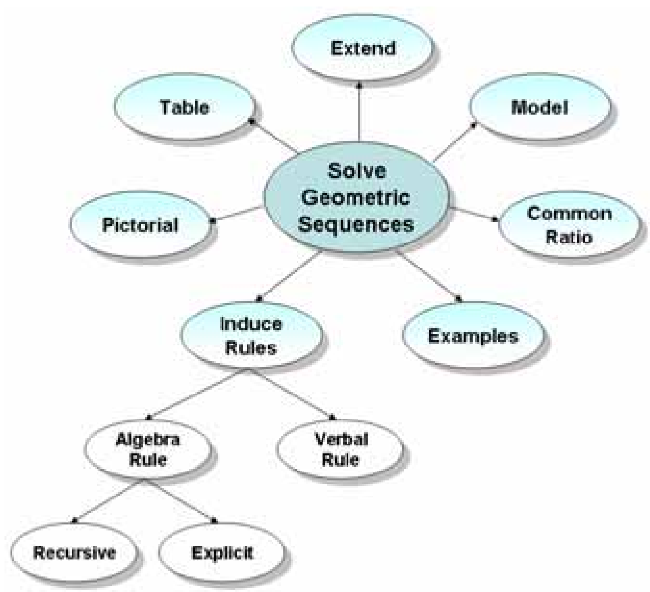


***Figure 1*. ACED proficiencies included in Study 1—geometric sequences.**

Estimation of knowledge and skills. To estimate students' knowledge and skills, we used StatShop (Almond, Yan, Matukhin, & Chang, 2006), software that estimates students' proficiency levels (e.g., high, medium, and low) for the concepts and skills defined in the proficiency model. StatShop uses a Bayesian network to integrate the evidence that is provided by students' solutions to the ACED tasks. This evidence is then automatically propagated throughout the network (Almond & Mislevy, 1999).

Adaptive algorithm. We integrated an adaptive algorithm into ACED to determine which task to present next to maximize measurement accuracy. This involved calculating the expected weight of evidence (Good & Card, 1971; Madigan & Almond, 1996) per task as the basis for selecting the subsequent task. For example, suppose we hypothesize that a given student's ability to solve geometric sequences (i.e., the parent node in the model) is at or above the *medium* level. The Bayes net can calculate the probability that this hypothesis holds, or equivalently, we can look at the log odds that the hypothesis holds. Now, suppose we observe a student's outcome from attempting to solve an ACED task. Entering this evidence into the Bayes net and updating the model will result in a change in the probabilities. The change in log odds is called the *weight of evidence* for a given hypothesis provided by evidence (Good, 1985). Typically, conclusions about a hypothesis will be based on the accumulation of many bits of evidence. Thus, weight of evidence can be used as the basis of graphical "explanations" of a hypothesis (Madigan, Mosurski, & Almond, 1997). For outcomes from tasks that have not been observed, one can calculate the *expected weight of evidence* under the hypothesis (Good & Card, 1971). The task that will be chosen maximizes the expected weight of evidence and is, in some sense, most informative about the hypothesis. This approach is related to the procedure commonly used in computer-adaptive testing of maximizing the Fisher information (Wainer et al., 2000). In ACED, computing the expected weight of evidence after each task response produces a new ordering of remaining tasks, unique for each student and based on the student's particular solution history.

Authoring tasks. A total of 174 tasks (i.e., ACED items) were authored for the full topic of sequences, including arithmetic, geometric, and other recursive sequences. A subset of 63 tasks comprised the geometric sequences branch of the proficiency model—the focus of this paper. Each task was statistically linked, via specific scoring rules, to relevant proficiencies. A little over half of the items were multiple-choice format, with four to five options from which to choose. The remaining items required the student to enter a short constructed response (a number or series of numbers). Task development entailed not only writing the items, but also crafting feedback for answer choices for multiple-choice items and for likely common responses to constructed-response items. Feedback for correct answers provided response verification (e.g., "You are correct!"). If the response was incorrect, the feedback provided elaboration—verification and explanation—to help the student understand how to solve the problem. An example task is shown in Figure 2.

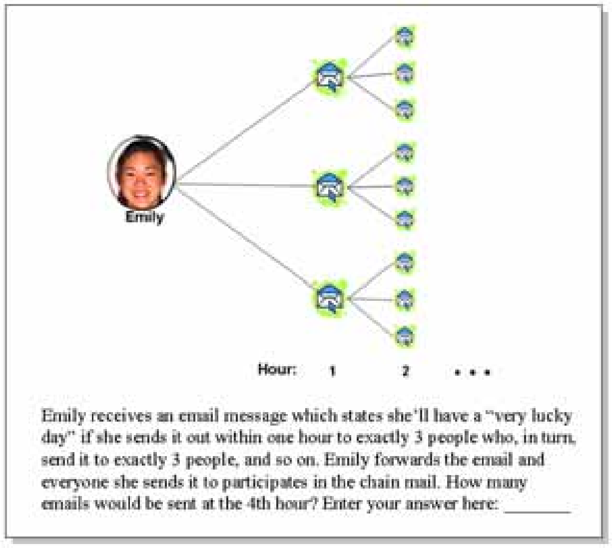


***Figure 2*. Example ACED geometric-sequence item with constructed response required.**

To the extent feasible, the elaborated feedback for an incorrect response was diagnostic in the sense of being crafted based on a diagnosis of the misconception or procedural bug suggested by the student's response. Feedback was written by two of the co-investigators and one other person, all of whom have doctoral degrees in cognitive psychology (or equivalent experience) and who have all worked professionally on mathematics projects.

If the learner entered an incorrect answer (e.g., "27") to the question shown in Figure 2, the elaborated feedback would indicate the following, "That's not correct. Three times as many e-mails go out every hour. That means 3 e-mails go out in the first hour, 9 go out in the second hour, and 27 e-mails (your response) go out in the third hour. The question asks about the number of e-mails in the fourth hour, which would be $3 \times 3 \times 3 \times 3 =$ **81**." Thus with elaborated feedback, the accuracy of the answer is indicated, along with an explanation about how to solve the problem and the correct answer.

Delivery platform. Originally, we designed and developed the ACED system with separate interfaces for sighted and visually disabled students. For the latter population, we included assistive technologies into the system for presenting math content. Once the components (models, accommodations, algorithms, etc.) of ACED were established and built near the end of the second year, we pilot tested ACED with students having visual disabilities using Touch Graphics TTT technology for audio-tactile graphics. The system provided the visually disabled users with the following accommodations: audio-tactile graphics, text-to-speech capability (providing basic "screen reader" functionality) for typing in constructed responses and hearing cues for navigation and editing, font enlargement, and color/contrast modification.[2] After pilot testing, we decided to integrate three interfaces—sighted, low-vision, and blind—into one system. The integrated ACED system was intended to have several benefits: (a) The adaptive item selection tools that were constructed for the mainstream (sighted students) version could also be seamlessly used for the blind and low-vision students; (b) the entire system could be integrated and loaded on a single computer, rather than distributed across a network with multiple servers providing individual functionality (as was the case in the first version of ACED); and (c) the comparison of performance across classes of users would be simplified, because a single set of data collection tools and database functionalities can be used for all students.

ACED features. The key features of the final delivery platform included the following:

- Audio-tactile graphics of assessment tasks: Includes an audio component via prerecorded and synthesized speech and a tactile component via raised line graphics and braille labels, which is helpful for individuals who are visually impaired or who have difficulty processing visual information.

- Visual display of assessment tasks: Tasks presented on the computer screen are suitable for fully or partially sighted individuals (e.g., students, teachers, or proctors).

- User-configured preferences: Students can modify font size and color of text or background. This is useful for students with some residual sight.

- Keyboard entry of constructed response answers: Students, whether visually impaired or not, can use a standard keyboard to type in answers to constructed response questions.

- Alternative interfaces: The system maintains two modes of feedback (elaboration and verification only) and two modes of task sequencing (adaptive and linear) and could administer tasks in any combination of modes, and with or without the TTT.

- Administration screen: This allowed the experimenter or proctor to set the feedback and task-sequencing modes. It further allowed configuration routines and walked participants through the selection of display options (e.g., with or without TTT; font size; color of font and background, etc.).

We now focus on Study 1—conducted with a sample of sighted students interacting with ACED.

Study 1. Feedback and Adaptive Sequencing: Effects on Learning

Method

Sample

A total of 268 Algebra I students participated in the study. These students attended the same mid-Atlantic state suburban high school. According to their teachers, for these students, geometric sequences were not explicitly instructed as part of the curriculum, although some geometric sequence problems may have been covered as part of other topics in algebra. Testing was conducted in sessions of about 20 students each, over a period of five days. Our full sample of students[3] was heterogeneous in ability, representing a full range of levels of math skills: honors (n = 38), academic (n = 165), regular (n = 27), remedial (n = 30),[4] and special education students (n = 8).

Procedure

Each 2-hour session consisted of the following activities. First, all students, at their desks, received a 10-minute introduction, which included the goals of our study and a description of upcoming activities. We also told them that their participation would not affect their math grade, and that it was important that they try their hardest. Finally, we reminded them that they were getting a reward for their participation.

After the introduction, all students took a 20-minute pretest. We created two test forms for the pre- and posttests—Forms A and B. The tests contained 25 items spanning the identified content (geometric sequences) and were administered in paper and pencil format. Items in each form were matched in relation to specific proficiency, difficulty level, and format. Calculators were permitted. After the pretest, students were randomly assigned to one of four conditions (see Design and Experimental Conditions) where they spent the next hour either at the computer in one of the three variants of ACED, or at their desks for the control condition. Following the 1-hour period, all students returned to their desks to complete the 20-minute posttest and a 10-minute paper-and-pencil survey.

Design and Experimental Conditions

For the evaluation in Study 1, we used a pretest-treatment-posttest design, with participating students randomly assigned to one of four conditions. All individuals regardless of condition received two forms of a multiple choice test—one form as a pretest and the other form as a posttest. The order of forms was randomly assigned so that half the students received forms in A-B order and the other half in B-A order. The control condition involved no treatment but only the pretest and posttest with an intervening 1-hour period (the duration of the ACED intervention) sitting at their desks reading content that was not related to math (e.g., other school work and magazines we obtained from the school library). Students assigned to the experimental conditions (Conditions 1, 2, and 3; see Table 1) took their assigned seats at one of the 26 networked computers in the laboratory where all testing occurred. After logging in, they spent the next hour solving geometric sequence problems presented on the screen.

Table 1

Four Experimental Conditions Used in Study 1

Condition	Code	Feedback for correct	Feedback for incorrect	Task sequencing
1. Elaborated feedback/adaptive sequencing	E/A	Verification	Verification + explanation	Adaptive
2. Simple feedback/ adaptive sequencing	S/A	Verification	Verification	Adaptive
3. Elaborated feedback/ linear sequencing	E/L	Verification	Verification + explanation	Linear
4. Control—no assessment, no instruction	CONTROL	N/A	N/A	N/A

For Research Question 1 ("Is elaborated feedback more effective for student learning than simple feedback?"), the main contrast of interest is between Conditions 1 and 2 (holding task-sequencing constant while varying type of feedback). Our hypothesis was that the elaborated feedback group (Condition 1) would experience greater learning than the simple feedback group (Condition 2). For Research Question 2 ("Is adaptive sequencing of tasks more effective for student learning than linear sequencing?"), the main contrast of interest is that between Conditions 1 and 3 (holding feedback type constant while varying task sequencing). Our hypothesis was that the adaptive sequencing group (Condition 1) would experience greater learning than the linear sequencing group (Condition 3).

While not used in a key research question, Condition 4, our control group, serves the useful function of establishing a base level of transfer from pretest to posttest. This condition thus provides a check on the overall impact of any of the three other conditions (1, 2, and 3). The sample sizes for the four conditions were as follows: Condition 1 (E/A; $n = 71$), Condition 2 (S/A; $n = 75$), Condition 3 (E/L; $n = 67$), and Condition 4 (Control; $n = 55$).[5]

Results for Learning Questions

Overall Learning via ACED

Because our two test forms (A and B) showed slight differences in difficulty (albeit, not significantly so), we scaled the scores for the pretest by producing z-scores for each pretest form (A and B) and then created a basic scale where 50 was the mean and 10 was the standard deviation. This served to equate the two forms of the pretest. Posttest scores (Forms A and B) were then placed on the same scale that was developed for the pretest via conversion tables we created from raw to scale scores for Form A and Form B (i.e., the raw score to z-score transformation was established on the basis of the pretest results only). As the Form A/B assignment was random, this implicitly produces random equivalent groups with equal mean and variance, equating between the two forms. Next, we examined the general question: Do students working with ACED (all 3 conditions, combined) show evidence of learning compared to the control group? An ANOVA was computed using combined ACED data (all three experimental conditions) in relation to the control group. First, there were no differences between the ACED and control groups in terms of their pretest scores (pretest: $F_{1, 266} = 0.23$; NS). However, the posttest scores of ACED students ($M = 56.5$, $SD = 10.7$, $n = 213$) were significantly higher than the posttest scores of the Control group ($M = 52.4$, $SD = 11.6$, $n = 55$); posttest $F_{1, 266} = 6.00$; $p < 0.02$. The effect size was 0.38. (Cohen's d).

Feedback and Adaptivity Effects on Learning

Research Questions 1 and 2 address the relationship to learning of (a) elaborated versus simple feedback, and (b) adaptive versus linear sequencing of tasks. Because our sample was so varied in ability level, we included two independent variables in the analysis: condition and academic level. An ANCOVA was computed with posttest score as the dependent variable, pretest score as the covariate, and condition (1-4) and academic level (1-5)[6] as the independent variables. The main effects of both condition ($F_{3, 247} = 3.41$, $p < 0.02$) and level ($F_{4, 247} = 11.28$, $p < 0.01$) were significant, but their interaction was not ($F_{12, 247} = 0.97$, NS). Figure 3 shows the main effect of condition in relation to posttest (collapsed across academic level) where the best posttest performance is demonstrated by students in the E/A condition, which also shows largest pretest-to-posttest improvement. Confidence intervals (95%) for the posttest data, per condition, are also depicted in the figure.

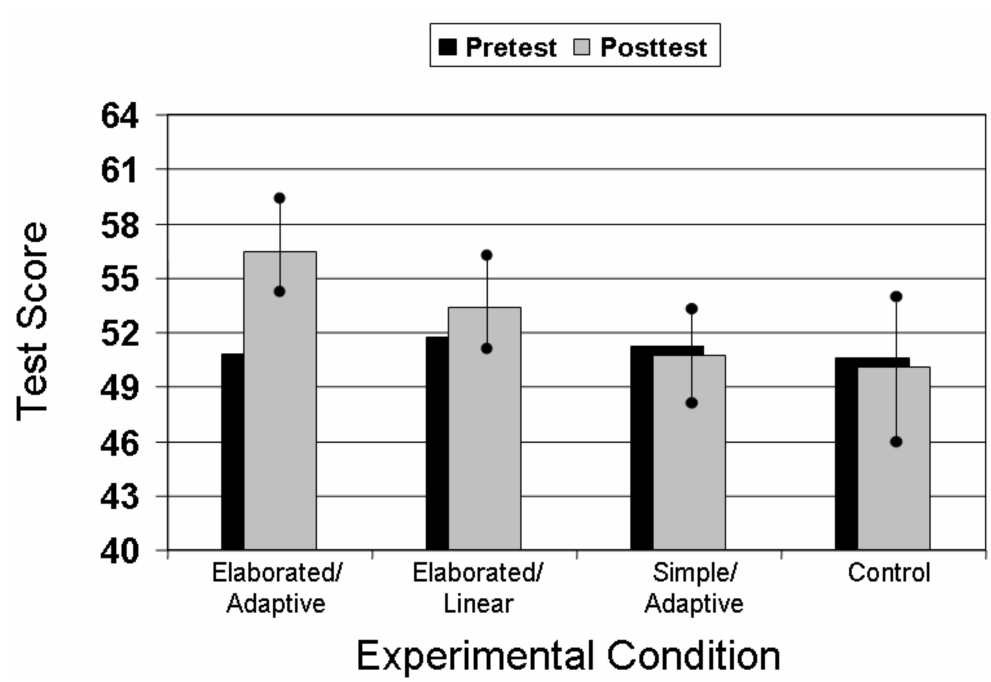


Figure 3. **Experimental condition by pre- and posttest scores.**

The general finding of a main effect of condition on learning prompted a more specific planned comparison (Bonferroni test) involving the three treatment conditions in relation to posttest data. Findings showed that the only significant difference was between the E/A and S/A conditions (i.e., task sequencing the same but type of feedback different). The mean difference =

5.62; $SE = 2.11$; $p < 0.05$. The difference between the E/A and E/L conditions (i.e., type of feedback the same but task sequencing different) was not significant. This suggests that the elaborated feedback feature was primarily responsible for the impact on learning.

In summary, regarding the learning questions: (a) The ACED AfL system did help students learn—mainly due to the effects of elaborated feedback; and (b) the sequencing of tasks (adaptive vs. linear) did not show significant effects on learning. However, as shown in Figure 3, the E/A versus E/L conditions show a slight advantage for adaptivity. Moreover, the adaptive algorithm used within ACED was intended to increase the accuracy of the assessment, not enhance the learning outcome. In the next section we examine assessment-related questions and present findings showing some benefits of adaptivity.

Results for Assessment Questions

To be worthwhile, an assessment must be valid and reliable. Validity refers to the extent to which the assessment accurately measures what it is supposed to measure. Reliability estimates the consistency of an assessment—the degree to which it rank orders students in the same way each time it is used. We start by examining the validity of the ACED measurement instruments. This is followed by analyses of reliability of our measures. Finally, we conclude this section with an analysis of efficiency. That is, while assessments should be both reliable and valid, to be useful they should also be efficient.

Validity

The first issue is whether the ACED proficiency estimates (i.e., within the Bayes net) predict outcome performance beyond that predicted by pretest scores. Each proficiency variable (see nodes in Figure 1) possesses a triplet of probabilities, reflecting the current estimate of performing high, medium, and low on that proficiency. To reduce the three numbers to a single number, we assigned the numeric values +1, 0, and -1 to the three proficiency states and took the expected value. This expected a posteriori (EAP) value can also be written as, $P(\theta_{ij} = \text{High}) - P(\theta_{ij} = \text{Low})$, where θ_{ij} is the value for Student *i* on Proficiency *j,* and $1*P(\text{High}) + 0*P(\text{Med}) + -1*P(\text{Low}) = P(\text{High}) - P(\text{Low})$. This results in a scale ranging from -1 to 1. As shown in Figure 1, all lower-level proficiency nodes (or subproficiencies) feed evidence into the main node—Solve Geometric Sequences (SGS). Thus higher values of EAP(SGS) should be associated with greater knowledge and skills overall on geometric sequence topics.

A regression analysis was computed with posttest score as the dependent variable and pretest score and EAP(SGS) as the independent variables.[7] Pretest score was entered into the equation first, followed by EAP(SGS). This regression analysis was intended to provide a general sense of whether the ACED estimates were valid or not and to indicate whether they accounted for any unique outcome variance beyond that attributable to pretest scores. Results showed that both of the independent variables significantly predicted student outcome data: Multiple $R = .71$, $F_{2, 210} = 106.57$, $p < 0.001$. Pretest score and the general estimate of proficiency accounted for 50% of the outcome variance, with EAP(SGS) accounting for 17% of the unique outcome variance over pretest score.

Reliability of ACED Tasks

All 63 tasks in the ACED pool of geometric items were statistically linked to relevant proficiencies. Students in all three ACED conditions were required to spend 1 hour on the computer solving the 63 items. This design decision was made in order to control for time and let outcomes vary; thus students in the two adaptive conditions spent the same amount of time on the program as those in the linear condition. Because students in all conditions had to complete the full set of 63 items, we obtained accuracy data (scored as 0/1 for incorrect/correct) per student, per item. These performance data were analyzed using a split-half reliability procedure (via SPSS). The Spearman Brown split-half reliability with unequal halves (i.e., 31 and 32) was equal to 0.84, while Cronbach's $\alpha = 0.88$, which is quite good.[8]

Reliability of ACED Proficiency Estimates

We analyzed proficiency estimates from the Bayes net, again using task performance data, which provided input to posterior probabilities, per node. These probabilities were then analyzed, making use of split-half reliabilities at the node level. The parent proficiency (EAP SGS) reliability was 0.88 (see Table 2). Moreover, the subproficiency estimates showed equally impressive reliabilities for their associated tasks. The same pattern held with data using probabilities—triplets per node—and MAPs (maximum a posteriori values). The reliabilities obtained for subproficiency estimates show how we can borrow strength and augment reliabilities in a meaningful way from indirect sources in the proficiency model, even though the tests are very small (i.e., the number of tasks per node is between 6–10 tasks). Furthermore, the high reliabilities of the subproficiencies suggest that they may be employed for diagnostic purposes.

Table 2

Spearman-Brown Split-Half Reliabilities

Proficiency (EAP)	Reliability
Solve geometric sequences (SGS)	0.88
Find common ratio	0.90
Generate examples	0.92
Extend sequence	0.86
Model sequence	0.80
Use table	0.82
Use pictures	0.82
Induce rules	0.78

Reliability of ACED Pretest and Posttest (Forms A and B)

Separate analyses were computed on the reliabilities of the pretest (Forms A/B) and posttest (Forms A/B) items. We computed Cronbach's α for each of the four tests, then used Spearman Brown's prophecy formula to increase the size of the tests to 63 items to render the tests comparable in length to the ACED assessment. These reliabilities are shown in Table 3.

Table 3

Reliabilities Across Pretests and Posttests

Test form	Pretest α	Posttest α
Adjusted Form A	0.84	0.82
Adjusted Form B	0.79	0.87
Adjusted avg. (A/B combined)	0.82	0.85

Efficiency of the ACED System

In Study 1, we required that students in all three ACED conditions spend 1 hour on the computer solving the set of 63 items. In fact, the majority (i.e., 80%) of the students completed all 63 tasks, and 95% of all students had 15 or fewer items remaining at the end of 1 hour. Those who completed the program early returned to their seats to read or rest until the hour was up.

As noted earlier, even though two of our ACED conditions employed our adaptive algorithm for task selection (i.e., Conditions 1 and 2—E/A and S/A), we still required students to complete the full set of 63 tasks. A typical rationale for using adaptive tests, however, relates to their efficiency capability. That is, adaptive algorithms rely on fewer tasks or items to determine proficiency level than more traditional approaches. The question here concerns what the data (proficiency estimates) would look like if we required fewer tasks to be solved (i.e., implementing a termination criterion into the adaptive algorithm). To answer this question, we selected the first N (where $N = 10, 15, 20, 25, 30, 40, 50$, and 63) tasks from the student records, and then calculated EAP values for the parent proficiency (solve geometric sequences) from each shortened test. Next, we computed correlations of each of these tests with the posttest score for the students.

What we expected to see was that the correlations, in general, should increase with test length until they reach an upper asymptote related to the reliability of the posttest. We hypothesized that the data from students in the linear condition (E/L) should reach that asymptote more slowly than the data from participants in the adaptive conditions. Figure 4 shows the results of the plot, confirming our hypothesis. The quick rise and asymptote of the two adaptive conditions shows that only 20–30 tasks are needed to reach the maximum correlation with the posttest. At that juncture, for those students, the next step would be instructional intervention by the teacher.

In the middle of the linear condition (E/L), there is a spike around Task 30, and then a subsequent drop, interrupting the gradual climb of the E/L curve. This is likely caused by a specific pattern of items (e.g., a collection of easy items that were good predictors followed by a collection of hard items that were poor predictors). In fact, when we reviewed the list of the 63 items in the order in which they appeared in the linear condition, Items 31–36 consisted of a group of very difficult items (determine the recursive rule for geometric sequences). An example of a task linked to the proficiency *recursive rule* (low-level node shown in Figure 1) is:

Let a_1, a_2, a_3, . . . , a_n, . . . be a sequence of integers. The first four terms of the sequence are: **-6, 24, -96, -384, . . .** Select the correct rule for finding the n+1th term from the nth term.

(a) $a_{n+1} = -6a_n$; (b) $a_{n+1} = 4a_n$; (c) $a_{n+1} = \frac{-a_n}{6}$; (d) $a_{n+1} = a_n \times 4^{n-1}$; (e) $a_{n+1} = -4a_n$.

Finally, the slight decline after Task 30 in the S/A condition could be a fatigue effect. Because many of the problems were unfamiliar and the feedback only indicated correct or incorrect (with no help or instruction), students may have stopped trying their best. Thus another potential benefit of the elaborated feedback could be student engagement. However, the noise associated with these data is large (e.g., the 95% confidence interval for the correlation coefficients ranges from .46 to .75) requiring another independent verification of this decline phenomenon.

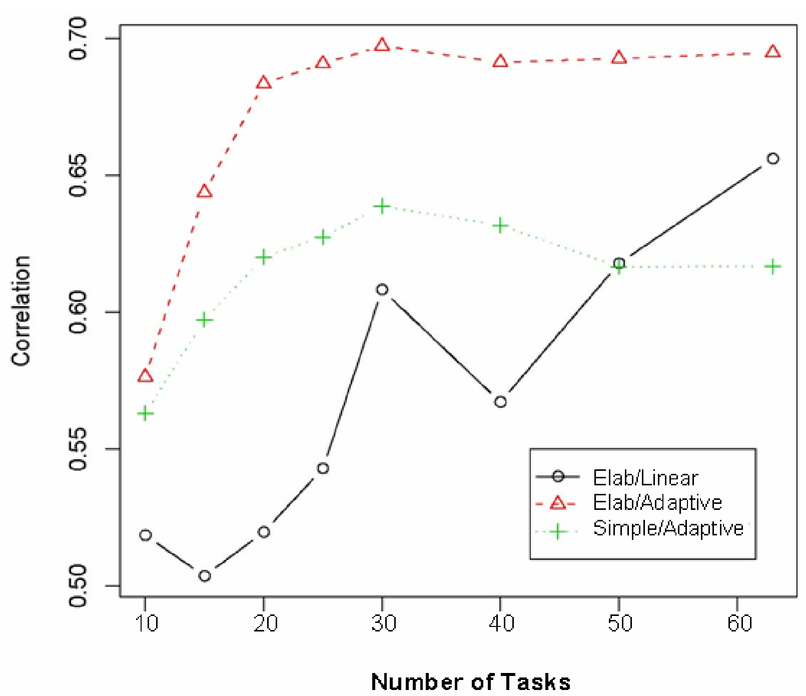


Figure 4**. Correlations of EAP (SGS) with posttest score by ACED condition.**

Summary of Study 1 Findings

In terms of learning from the ACED system, students in the three ACED conditions (combined) demonstrated significantly more learning of the topic geometric sequences than those assigned to the control condition. More importantly, we identified the source of the learning effects—elaborated feedback. That is, we found a significant difference in posttest scores between students in the E/A and S/A conditions. Both groups received adaptively-sequenced tasks, but they differed in the type of feedback received, with a clear benefit of task elaboration. There was no significant main effect of adaptivity (E/A vs. E/L), although we reiterate that the intention of the adaptive algorithm was more to enhance the precision of the assessment than to facilitate learning. The trend displayed in Figure 3, however, was in the right direction (i.e., students in the adaptive E/A condition showed slightly higher posttest scores than those in linear E/L condition).

Regarding predictive validity, using evidence-based assessment design with Bayes net technology facilitates estimation of proficiencies directly from performance data. Regression analysis showed that just a single proficiency estimate can significantly predict posttest performance, beyond that provided by the pretest. Next, the split-half reliability of the 63 ACED tasks was high (0.88), as was the parent proficiency (EAP SGS) reliability (0.88), and our adjusted pretest and posttest reliabilities (0.82 and 0.85, respectively). Finally, we saw that for students in the adaptive conditions (E/A and S/A), the ACED test could have terminated after approximately 20 items with no degradation in prediction of outcome. Because most of the students completed all 63 tasks, covering a range of geometric sequence topics, in 1 hour (or 1 task per minute, on average), administering a half-hour test yielding valid and reliable results would be much more efficient for students and teachers. We now turn our attention to the next population of students using ACED—those with visual disabilities.

Study 2. ACED Usability for Individuals With Visual Disabilities

There is a growing awareness of the need to ensure that the benefits of assessments are available to diverse test takers, including individuals with disabilities. For example, whereas student with disabilities were often excluded for state assessments, legislation has mandated inclusion in such assessments. According to the Individuals with Disability Education Act (Individuals with Disability Education Act [IDEA] of 2004), "All children with disabilities are included in all general State and districtwide assessment programs … with appropriate accommodations and alternate assessments where necessary and as indicated in their respective individualized education programs." (IDEA, 2004, Sec. 612(a)(16)(A)). Unfortunately assessments are often designed without taking accessibility into account from the beginning. This project sought to apply knowledge of accessibility as both a resource and a constraint during assessment design, to make the assessment useful to all students.

Specifically, Study 2 sought to explore the usability of accessibility features to enable access to mathematics content by individuals with visual disabilities. Thus the goal of this study was to explore the usability of ACED features with a small set of students who might reasonably benefit from either the low-vision or blind versions of the ACED system. As discussed earlier, accessibility may be regarded as a prerequisite to the validity of assessment results, since without accessibility, students with disabilities will tend to receive scores that are invalidly low. Furthermore, without valid assessment, is it difficult or impossible to demonstrate the effectiveness of a learning program or intervention.

ACED Accessibility Features

Following is a summary of the main features and functionality of the integrated ACED system, user-preference screens, and the administration utility.

ACED Functionality

To allow students (with and without visual impairments) complete access to ACED, we integrated into a single delivery system three distinct interfaces:

1. *Mainstream version*. This was for sighted students and used the computer monitor, keyboard, and optional mouse
2. *Screen-magnified* version. For this version, the student saw an enlarged image of the mainstream version. Parts that were off-screen could be accessed by sliding the image up, down, left or right to reveal hidden parts. The student could also set the magnification factor and color scheme, and choose to hear the items and answer choices read aloud.
3. *TTT version*. This version used the TTT from Touch Graphics, where the student can choose a voice, set the speech rate, and customize the TTT's touch sensitivity.

User Preferences

We created an application to allow students to interactively set up user preferences before starting a session—regardless of visual ability. This tool takes the form a series of screens that call for the student to make choices that allow the computer to deduce how to most effectively configure access features. For example, the first screen displays text and asks the student (via audio) to press the *Enter* key if they can read the text, and if not, then to press the space bar. If the space bar is selected, the computer then presents a series of enlarged font sizes asking the student if he/she can see any of those. If so, the student clicks on the preferred size. If not, he/she clicks the space bar, and so on. Once ACED assessment has begun, one can revise earlier settings.

Administration Utility

Administrators used a checkbox for activating the adaptive item selection feature. In the original version of ACED, the administrator was required to build a sequence of items that would be presented to the student by selecting from an item pool shown in one panel on the computer screen and dragging these into the sequence panel on the other side of the screen. Once the items were placed in the sequence panel, they could be rearranged as desired. In the integrated version, we implemented the adaptive item selection capability that can be used instead of preset sequences. If the adaptive item selection box was checked, the two panels and associated buttons used for arranging sequences were disabled, indicating that these functions were not available in the adaptive mode. Other features of the administration utility included: (a) tools for saving, opening, and deleting named test configurations; (b) a dialog box for setting one of the named sequences as "active" for the current administration; (c) a checkbox for activating elaborated task feedback that can be supplied in response to wrong answers; and (d) text boxes for setting the number of minutes for task administration, time-out, and the time-out warning message.

ACED uses its diagnostic front end to guide the student to one of its three interfaces to configure the system to user preferences in the areas of: (a) read-aloud or silent, (b) human[9] or synthetic voice, (c) speech rate for synthetic voice, (d) screen magnification level (non-TTT only), (e) text/background color scheme (non-TTT only), and (f) tactile sensitivity (TTT only).

The flow of events for TTT user sessions in ACED follows a general pattern. Once the student has logged in (which must be done for any version of ACED) by typing in a known user name or by creating a new user name, the student listens to a short description of how the system works. Once this is complete, the student is presented with the first question. At any time during the test, the student can use the raised-line graphic to access the main menu. A student scrolls through the menu options by pressing the up and down arrows, and selects an option by pressing the circle. During all Main Menu functions, the test duration timer pauses. The main menu options include:

- *Settings*. Students can adjust the narrator's speech rate, the rate of the screen reader's synthetic voice, and the TTT's touch sensitivity threshold (if applicable). Changes to settings are saved in the student's user profile.
- *Help*. A series of tutorials are presented that provide instructions and suggestions for how to read and interpret several categories of tactile graphic images.
- *Exit*. If the student wants to stop taking the test before completing all of the items, he or she can choose this Main Menu option.
- *Go to test.* This feature allows the student to return to the current item in the sequence after adjusting one of the settings or reviewing one of the help topics.

Method

The population of individuals with visual disabilities is very diverse, and a small-sample usability study such as this is only able to provide results that are suggestive of those that would be obtained with larger or more representative samples. However, according to Nielsen (2000), low numbers of users can be very helpful to a study like this one, because a study of 5-10 users is often sufficient to find 80% of usability problems.

Sample

The sample consisted of four students (S1, S2, S3, and S4) aged 17 to 20—two males and two females. All were high school students and had either an individualized education plan (IEP) or a Section 504 plan for visual impairment.[10] While all four students were considered legally blind, two had some useful vision (S1 and S2) and the other two had essentially no usable vision (S3 and S4). S1 and S2 reported using "large print or enlargement of print or pictures" as their primary method for study and learning; on the other hand, S3 and S4 reported using, "hearing, touch, and braille" to describe their primary methods of study. A summary of descriptive characteristics of the four participants is presented in Table 4.

Table 4

Characteristics of Participants (S1, S2, S3, S4)

Characteristic	S1	S2	S3	S4
ACED Mode used	*Low vision*	*Low vision*	*Blind*	*Blind*
Gender	Male	Female	Female	Male
Age	17	19	20	18
Born in the United States	Yes	No	No	Yes
Grade began attending school in United States	K	9th grade	9th grade	K
"I like mathematics as a subject area in school" (Yes, No)	No	No	Yes	No
Latest grade in math class	Satisfactory	80%	87%	80%
Use screen reader	No	No	Yes – JAWS	Yes – JAWS
Use large print test booklet in school	Yes	Yes	No	No
Had used a human reader in school	Yes	Yes	Yes	Yes
Use braille to take tests	No	No	Yes	Yes
Use raised-line drawings to take tests	No	No	Yes	Yes
Preferred media for math	Talking calculator	Calculator with enlarged font. CCTV (closed circuit television)	CCTV for diagrams, Nemeth braille code. Tactile graphics and scribe for testing.	Perkins Brailler, and math in Nemeth braille code.
Use of computer	2 times a week	2 times a month for 2 years	Use Braille Lite (electronic notetaker) every day; computer once a month	Daily, for 9 years.

Though not a criterion for selection, we learned that all participants had used a human reader in school. Moreover, all spoke English in the home, even though two of the participants (S2 and S3) were born in Latin America and began school in the United States in the ninth grade. All four students reported that at school they used audio (by machine or live reader) to take tests. Regarding the use of screen-reader software at school, S1 and S2 did not use this type of software, while both S3 and S4 did—citing JAWS (from Freedom Scientific) as their preferred screen reader program.[11]

Content

As with Study 1, the mathematics content for Study 2 was geometric sequences. Tasks used in this study consisted of a pool of 30 items, selected from the larger pool of 63 items from Study 1. Reducing the number of tasks from 63 to 30 was done to ensure that a significant proportion of items viewed by students in the study would involve graphics and tables, thereby providing a basis to evaluate the audio-tactile features of the TTT during Study 2. Also, because this was a usability study, we only required a representative subset of good tasks to administer. About half of the 30 items made use of graphics or tables on TTT overlay sheets, and the other half did not. As noted below, the four participants were able to access between 8 and 12 of the 30 items in the allotted time frame. For the student who accessed the most (i.e., 12 items): four were constructed response items and eight were single selection multiple choice items; four made use of graphics (presented as tactile graphics in the TTT mode) and eight did not; three made use of tables (again, with presentation as tactile graphics for the TTT mode) and nine did not.

Before the study began, all content was reviewed to avoid task designs in which the nature of the accommodation would conflict with the construct being measured (Hansen, & Mislevy, 2005; Mislevy, Steinberg, & Almond, 2003). For example, we reviewed adaptations for using audio-tactile mode, not only from the standpoint of accessibility, but also to avoid reducing demands for mathematics skills that the test was intended to measure.

Task Sequencing

The 30 tasks used in Study 2 were presented to each student based on the results from the same adaptive algorithm discussed earlier, so each participant received a different set of tasks during his or her time on the ACED system—differing as a function of each student's particular solution history.

Instruments

The ACED system was set up to administer up to 30 items in either of two modes—low-vision or blind. A series of setup screens guided the user in selecting between the low-vision and blind versions of ACED and in configuring the audio (e.g., pitch and speed), font size, color of text or background, and so forth. Both modes employed a laptop computer with audio output capabilities (e.g., speakers). Following is more detail about the two disability modes of the ACED system.

1. *Blind mode* (also called TTT mode). This mode used the TTT system from Touch Graphics as the main delivery mechanism. The TTT mode provided a rectangular touch-sensitive tablet, about 18 by 20 inches in dimension. A plastic overlay sheet with raised-line elements (graphics, text, and braille labeling) and color printing (for those with residual vision) was placed thereon and bound into place so that the sheet did not slide. The TTT acted under control of a laptop computer. The system guided the user via audio in how to set up successive items as required by the adaptive sequencing algorithm. When the student touched the tablet, the location of the touch was sent to the computer, which then invoked actions such as (a) playing audio, either prerecorded or synthesized speech that voiced the content; (b) providing audio instructions; or (c) other audio guidance. The blind mode also provided synthesized speech and a keyboard interface to guide the user in entering constructed responses. See Figure 5 for an illustration of a student from Study 2 using this mode. If you look closely at the figure, the right side of the tablet depicts the tactile version of the same chain e-mail task that was shown in Figure 2.

2. *Low-vision mode*. This mode is similar to an ordinary computer-based test (i.e., the regular mode), except that (a) the font size could be enlarged, and (b) it provided audio in the form of synthesized speech to read aloud the text and instructions. Students had a choice to use low-vision mode with or without audio. Low vision mode could be operated through keyboard alone or supplemented by the use of a mouse.

Pre-treatment survey. This survey posed questions about individuals' background (e.g., disability, home language), experience with assistive technologies, accommodations in school, and so forth.

Post-treatment survey. This instrument gathered information about reactions to the ACED experience.

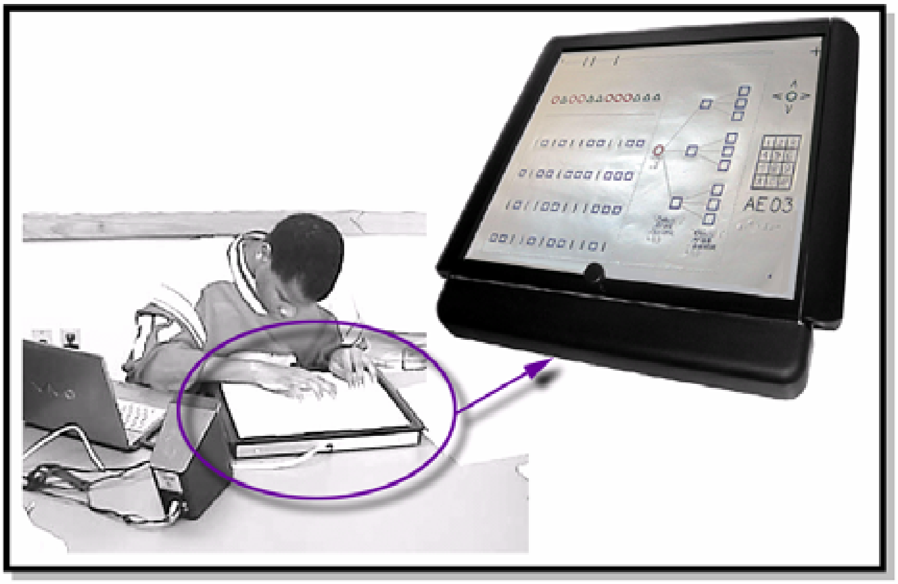


Figure 5. **An individual using the ACED system with the TTT feature.**

Procedure

Testing of our four participants took place at the Touch Graphics office in New York City. The first two authors of this paper were present at both sessions, and the tests were proctored by staff from Touch Graphics. Informed consent documents were read aloud and signed by all participants (or parents, for minors) prior to working on the ACED system. Participants worked individually during each session. Following is the basic procedure.

1. *Pretreatment survey and introduction*. The pretreatment survey questions were administered via interview. Students were informed that the purpose of the study was to evaluate the system rather than to evaluate their knowledge of mathematics. Nevertheless, they were asked to do their best in answering the questions in the ACED system.

2. *Familiarization and system configuration*. Each student received a computer-based tutorial on the use of the system and was guided by the system in selecting the system options. Students generally required very little help to use the system. Participants S1

and S2 used the low-vision version of ACED and participants S2 and S3 used the blind version of ACED.

3. *Assessment tasks and feedback.* Students used the ACED system with elaborated feedback and adaptive sequencing (referred to as Condition 1 in Study 1). The intent was to allow each student about an hour to become familiar with and use the ACED system. However, due to transportation and other logistical issues, the time was curtailed to a lesser amount of time (40 to 48 minutes) for three of the four participants, while the fourth person required a somewhat longer amount of time (72 minutes).
4. *Post-treatment survey*. Participants received a post-treatment survey administered by interview.

Results

We now present the results from the usability study in four parts. First, we examine the basic usage data describing what the students did on ACED and how they performed for both the low-vision and blind versions of ACED. Second, we examine the overall usability of the system in relation to important system features. Third, we describe the students' specific reactions to the low-vision and blind versions of ACED. Finally, we summarize the strengths and weaknesses of the system in relation to students' responses about what they particularly liked and did not like.

Usage of the ACED System

Table 5 shows basic ACED usage data for the four students with visual disabilities. Note that the time taken for familiarization and configuration of the system was longer for the blind students (47 minutes for S3, and 25 minutes for S4) than it was for the low-vision students (19 minutes for S1, and 12 minutes for S2). The blind students were using the TTT and therefore had to take time to become familiar with its interface and to configure it. Also notice that, as shown on the last line of the table, students took from 1.3 to 3.5 minutes per item completed (including feedback). This compares to roughly 1 minute per item for the nondisabled students using the regular version of ACED (i.e., about 60 minutes to complete 63 items). This difference is consistent with the expectation that students with disabilities accessing content with assistive technology generally need more time to access the content.

Table 5

Time, Items, and Percentage Correct

Participant	S1	S2	S3	S4
Modes used	Started in sighted mode then switched to *low-vision mode with audio*	*Low vision mode with audio* and large font, then switched to medium font	*Blind mode*	*Blind mode*
Time for familiarization and configuration (minutes)	19	12	47	25
Time answering items and receiving feedback (minutes)	29	28	25	16
Total time on system (minutes)	48	40	72	41
Items completed	9	8	9	12
Correct responses	1	5	4	8
Incorrect responses	8	3	5	4
Percentage correct (of items completed)	11%	63%	44%	67%
Average time per item completed (minutes)	3.2	3.5	2.8	1.3

Usage of the low-vision version of ACED. Participants S1 and S2 used the low-vision version of the ACED system (which employed the laptop but not the TTT). S1 initially entered the program as sighted (regular mode) but within 1–2 minutes, changed to the low-vision version with audio. S1 listened to the questions with synthesized speech as he read along and looked at the graphics on the screen. He used a hand-held calculator rather than the on-screen calculator available on the system. As shown in Table 5, S1 required 19 minutes for familiarization and configuration of the system, followed by 29 minutes answering items and receiving feedback. He completed nine items and answered only one correctly, which may be indicative of an access issue, lack of prerequisite knowledge of the system, or lack of knowledge in relation to math content. S2 used the low-vision mode with magnification, but due to her inexperience with screen enlargement software and computers in general, found it difficult to navigate. The student then decided to return the magnification level to normal-sized font, used the audio, and moved the screen close to examine the onscreen graphics. The student preferred to use the tab key instead of the mouse, but did use the mouse several times. In the beginning, the student expressed some frustration, but after several minutes, and as a result of the instructional feedback, she appeared to recall (or induce) the idea behind geometric sequences and was subsequently able to answer several questions correctly. She used a large calculator and scratch paper for problem solving. After the session, S2 asked, “How can I get this in school? It’s great!” As shown in Table 5, she used 12 minutes for familiarization and configuration of the system, followed by 28 minutes answering items and receiving feedback. She completed eight items and answered five correctly.

Usage of the blind version of ACED. Participants S3 and S4 used the blind version of the ACED system, which uses the TTT. These individuals pressed on the TTT to hear the meaning of various parts of the plastic overlay sheet. These individuals took longer on the familiarization and configuration of the system than did those using the low-vision mode (S1 and S2). This is understandable because of the greater number of special accessibility features on the blind version. S3 experienced some difficulty at the beginning of the session, being unfamiliar with geometric sequences as well as with the TTT. She answered the first 2 items incorrectly. She was apparently helped by the instructional feedback and answered 4 items correctly in a row. As shown in Table 5, she required 47 minutes for familiarization and configuration of the system, followed by 25 minutes answering items and receiving feedback. She completed 9 items and answered 4 correctly. S4 had been exposed to the TTT about 3 years earlier and therefore skipped some of the basic TTT introductory tutorial. He also selected a different synthesized voice (“Alice” as opposed to the default voice “David”) and increased the speech rate. As shown in Table 5, he used 25 minutes for familiarization and configuration of the system, followed by 16 minutes answering questions and receiving feedback. He completed 12 items and answered 8 correctly.

Overall Usability

The students all completed a survey at the end of their interaction with ACED, indicating the degree to which they agreed or disagreed with various statements that were read to them. Table 6 contains a summary of the students’ responses.

Table 6

Overall Usability of ACED Features

Statement	Participant (version used)			
	S1 (low vision)	S2 (low vision)	S3 (blind)	S4 (blind)
The [ACED] system was easy to use.	A	A	A	A
The synthesized speech in the screen reader was easy to understand.	A	A	A	A
I liked using the speech feature to have the test read aloud.	SA	SA	Not asked	Not asked
In was easy to use the [ACED] screen reader for questions requiring a keyboard.	A	A	A	A
It was easy to type in short numerical answers on the computer keyboard.	A	A	SA	SA
In was easy to understand the tables (row headings, column headings and cells) using the system.	SA	A	SA	SA
When my answer was wrong, the system helped me understand why my answer was wrong.	A	SA	SA	A
Having used the system during this session, I think that I now know how to use the system well enough to use it for an important test	N	SA	A	A

Note. Responses were coded as follows: SA = strongly agree; A = agree; N = neither agree nor disagree; D = disagree; and SD = strongly disagree.

As seen in Table 6, all four participants agreed that the ACED system was easy to use and that the synthesized speech was easy to understand. Understanding the audio was critical for individuals using the blind mode. The use of the audio feature was optional in the low-vision mode but was invoked by the two individuals using that mode, and both individuals strongly agreed that they liked using the speech feature to have the assessment read aloud.

All participants agreed that it was easy to use the screen reader for questions requiring a keyboard. Key screen-reader functionality included navigation commands and synthesized speech. And all agreed or strongly agreed that it was easy to type in short numerical answers on the computer keyboard.

The four participants agreed or strongly agreed that it was easy to understand the tables (row headings, column headings and cells) using the system. It seems significant that both individuals using the blind version strongly agreed, since access to tables is often a significant issue for individuals who are blind. Finally, four participants strongly agreed or agreed that when their answer was wrong, the system helped them to understand why the answer was wrong. This is a positive indication relative to the value of the approach for learning.

Specific Reactions to Low-Vision and Blind Versions of ACED

Table 7 shows reactions to some key features of the low-vision version of ACED. The voicing (speech) feature was especially liked by student S2.

Table 7

Agreement With Statements About Features of the Low-Vision Version

Statement	Participant	
	S1	S2
I liked the ability to change color schemes	N	SA
I liked the ability to magnify the screen.	N	SA
I liked using the speech feature to have the test read aloud.	SA	SA
I liked having the ability to change the speech rate.	N	SA

Note. Responses were coded as follows: SA = strongly agree; A = agree; N = neither agree nor disagree; D = disagree; and SD = strongly disagree.

Reactions to the blind mode. The main finding relating to the blind mode concerned the usability of the overlay sheets. Significantly, both individuals using the TTT's plastic overlay sheets found them easy to switch. S3 strongly agreed with the statement, "It is easy to switch the overlay sheets," while S4 agreed with the statement. Ease in switching the overlay sheets is a critical prerequisite in enabling students to function independently within a learning or assessment setting that uses a TTT. Students using the TTT (blind mode) were also asked to compare their preference for the TTT relative to other solutions. Table 8 shows that the TTT was preferred over each stated alternative by at least one of the two individuals. The TTT was least preferred (*disagree* and *neither agree nor disagree*) relative to the last option ("human reader and test booklet with braille and raised-line pictures"). This reaction suggests that the option which includes the use of a human reader (a relatively expensive solution) is fairly attractive. This result underscores, among other things, the diversity among individuals with disabilities. It also may reflect the relatively strong preference for braille and human readers among some individuals with visual disabilities, particularly total blindness. The TTT, while using braille labels, relies on audio more than braille.

Table 8

Preference for TTT Over Other Solutions

Statement	Participant	
	S3	S4
For understanding graphics, charts, tables, and math expressions, I prefer the TTT system over prerecorded audio alone	SA	N
For understanding graphics, charts, tables, and math expressions, I prefer the TTT system over raised-line graphics and prerecorded audio	A	A
For understanding graphics, charts, tables, and math expressions, I prefer the TTT system over raised-line graphics and a braille test booklet.	A	D
For taking important math tests, I prefer the TTT over a human reader and a test booklet with braille and raised-line pictures.	D	N

Note. Responses were coded as follows: SA = strongly agree; A = agree; N = neither agree nor disagree; D = disagree; and SD = strongly disagree.

Two final questions were asked of all four participants. In response to the statement "Having used the system during this session, I think that I now know how to use the system well enough to use it for an important test," 3 out of the 4 participants agreed (or strongly agreed). And in response to the statement "I enjoyed taking the test," the same 3 of the 4 participants agreed or strongly agreed.

Strengths and Weaknesses of ACED

Table 9 shows participants' comments about strengths of the system (features they liked) and its weaknesses (features they disliked). Note that all participants liked the audio capabilities and those using the TTT (S3 and S4) liked the audio-tactile capabilities. Regarding areas for improvement, individuals using the low-vision version (S1 and S2) found limitations in color modification and enlargement and one individual using the blind version (S4) would like improved navigation and ease of setup.

Table 9

Strengths and Weaknesses of Various ACED Features

Issue	Participant			
	S1	S2	S3	S4
What features did you like the most?	Voice (speech) along with the on-screen text. Voice is better than others used.	Pre-recorded audio and (synthesized) audio.	Audio-tactile interactivity, feedback, and confirmation. System and content is easy after some practice. Multi-modal approach is helpful. Easy to repeat question and voice is clear. Rephrasing of question is available. On-screen calculator. Booklet on holder.	Could press anywhere and it would tell you what you were pressing. Tactile parts were helpful. And narrator was clear and easy to understand.
What features did you really not like or think should be fixed?	Wanted higher color contrast options. Current color contrast options are too close to each other.	The enlargement (magnification) was too much, too tiring. Liked the voice better.	(None)	Need an easier way to repeat parts of questions (instead of using up and down arrows). You should be able to press the circle button to hear just the section you are on. Setup should require less pressing of things.

Summary of Study 2 Findings

Within the limitations due to the small size of the usability study, individuals with visual disabilities found both versions of the ACED system to be generally usable for a mathematics AfL system. Furthermore, the diversity of accessibility features—such as prerecorded and synthesized speech and audio-tactile graphics—appeared to contribute to the overall quite positive affect toward the approach.

By making diverse accessibility features available in a single system, ACED has enabled the benefits of elaborated feedback and adaptive sequencing to become available to individuals with blindness and low vision. As such the project represents a step forward in the quest to ensure equity and access to all individuals, regardless of disability.

Study 2 provides an example of how an AfL system (or testing, in general) can incorporate features that would make tests accessible to individuals with and without disabilities. It also illustrates the value that some users with visual disabilities find in test system features such as speech output (synthesized and prerecorded) and interactive audio-tactile graphics.

Summary and Discussion

We examined three main research questions tested within two studies relating to issues of learning effectiveness and accessibility of our ACED system. Regarding the learning questions explored in Study 1, we found that elaborated feedback (i.e., task-level feedback that provides both verification and explanation for incorrect responses) was, as hypothesized, more effective for student learning than simple feedback (verification only). More importantly, it contributed to significant pretest to posttest improvement on what was, for most of the students, a topic not explicitly instructed, according to their teachers. Study 1 did not show adaptive sequencing of tasks to be more effective for learning than linear sequencing. However, the adaptive condition did achieve its reliability maximum more efficiently than the linear test. That is, we saw that for students in the adaptive condition, the ACED assessment could have reasonably terminated after approximately 20 items with no degradation in prediction of outcome. Because most of the students completed all 63 tasks in 1 hour covering a range of geometric sequence proficiencies (or 1 task per minute, on average), administering a 20–30 minute test yielding comparable results would be much more efficient for students and teachers.

How did the ACED system fare in terms of validity and reliability? First, it was shown to be a reliable instrument. For instance, the split-half reliabilities of the ACED tasks, the proficiency estimates, and the pretests and posttests were all high. Second, in terms of validity and based on our findings, it appears that using an evidence-based assessment design with Bayes net technology facilitates valid estimation of proficiencies from performance data. Furthermore, regression analysis showed that just a single proficiency estimate—EAP(SGS)—significantly predicts posttest performance, beyond that predicted by pretest performance.

Study 2 provided us with valuable information regarding the usability of a system designed to accommodate visual disabilities for mathematics test taking. This finding provides some initial support for validity of the test results, though further research should be conducted to advance that argument. Also, because ACED was designed to be more than "just a test" (i.e., an assessment *for* learning), it was very encouraging that our visually-disabled participants also reported some learning from the elaborated feedback. In short, it seems that visually impaired students can use features that were embodied in ACED, both the low-vision and blind (TTT) versions, and they like the ACED system. Systems of this type might enable equal access to key learning-oriented features, such as instruction, feedback, and adaptive sequencing of tasks. One take-home message from this research concerns the importance of considering accessibility issues at the outset of designing an assessment system—particularly one that is intended to support learning.

Regarding next steps, we would like to enhance the current adaptive algorithm to better support student learning. The current adaptive algorithm was designed mainly to enhance assessment, but ACED—which incorporates elaborated feedback—was also intended to enhance learning. That is, while the expected weight of evidence algorithm has certain optimality properties for assessing student proficiencies, it does not necessarily have optimal properties for supporting growth in those proficiencies, i.e., student learning. The algorithm tends to select tasks that the student has an approximately 50% chance of getting right, but the idea that such tasks provide good learning opportunities is an untested hypothesis.

We would also like to improve our modeling of student learning. One limitation of the current approach is that the ACED scoring engine assumes that the student's proficiency does not change over the course of the ACED session, but that does not seem to be true. To illustrate the problem, suppose a student is struggling with one of the proficiencies represented in Figure 1 (e.g., "generate examples of geometric sequences"). The student receives one medium and two easy items linked to this proficiency and solves all three incorrectly. Later, something from the elaborated feedback provokes an “aha” moment and the student subsequently solves two easy items and one medium item correctly. ACED treats this as a cumulative set of six items: three right, three wrong, ignoring the temporal sequence, and assumes that the student’s proficiency is the same throughout the assessment period. The resulting probability distribution for this particular proficiency would likely be flat, and the EAP estimate would dip below zero and then come back up to zero, but rather slowly. But if ACED considered the temporal ordering of the data, the earlier observations would be discounted in some way, and the current EAP estimate would be positive. A more sophisticated engine needs to take student growth into account in a realistic way. Ideally, we should be able to use the inferences from the Bayes net as input to a planning system to help suggest next steps for the teacher or student. See Almond (in press) for a more detailed discussion of this topic.

In conclusion, we envision a role for the ACED program as part of a larger instructional support system. The adaptive AfL system would continue assessing the student until the best instructional options for that student become clear. Meanwhile, elaborated feedback would ensure that the assessment itself is an effective learning experience, supporting learning for students with a wide range of abilities and disabilities.

References

Almond, R. G. (in press). *Cognitive modeling to represent growth (learning) using Markov decision processes* (ETS Research Report Series). Princeton, NJ: ETS.

Almond, R. G., & Mislevy, R. J. (1999). Graphical models and computerized adaptive testing, *Applied Psychological Measurement, 23*, 223–238.

Almond, R. G., Yan, D., Matukhin, A., & Chang, D. (2006). *StatShop testing* (ETS Research Memorandum No. RM-06-04). Princeton, NJ: ETS.

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*(2), 111–127.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213–238.

Bass, K. M., & Glaser, R. (2004). *Developing assessments to inform teaching and learning* (CSE Rep. No. 628). Los Angeles, CA: UCLA, Graduate School of Education and Information Studies.

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

Chisholm, W., Vanderheiden, G., & Jacobs, I. (Eds.). (1999). *Web content accessibility guidelines (W3C recommendation)*. Retrieved February 12, 2007, from http://www.w3.org/TR/WAIWEBCONTENT

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309–319.

Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. *Proceedings of ACM CHI 2001 conference on human factors in computing systems* (pp. 245–252). New York: ACM Press.

Good, I. J. (1985): Weight of evidence: A brief survey. In J. Bernardo, M. DeGroot, D. Lindley, & A. Smith (Eds.), *Bayesian statistics 2* (pp. 249–269). Amsterdam: North Holland.

Good, I. J., & Card, W. (1971). The diagnostic process with special reference to errors. *Methods of Information in Medicine, 10*, 176–188.

Hansen, E. G., Forer, D. C., & Lee, M. J. (2004). *Toward accessible computer-based tests: Prototypes for visual and other disabilities* (TOEFL Research Rep. No. RR-78). Princeton, NJ: ETS.

Hansen, E. G., & Mislevy, R. J. (2005). Accessibility of computer-based testing for individuals with disabilities and English language learners within a validity framework. In M. Hricko & S. Howell (Eds.), *Online assessment and measurement: Foundation, challenges, and issues*. Hershey, PA: Idea Group Publishing, Inc.

Individuals with Disability Education Act of 2004, Sec. 612(a)(16)(A).

Jameson, A. (2006). Adaptive interfaces and agents. In J. A. Jacko & A. Sears (Eds.), *Human-computer interaction handbook* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.

Landau, S. (2005). *ACED for the TTT: Interim Report 1*. New York: Touch Graphics.

Lepper, M. R., & Chabay, R. W. (1985). Intrinsic motivation and instruction: Conflicting views on the role of motivational processes in computer-based education. *Educational Psychologist, 20*(4), 217–230.

Madigan, D., & Almond, R. G. (1996). On test selection strategies for belief networks. In D. Fisher & H. J. Lenz (Eds.), *Learning from data: AI and statistics IV* (pp. 89–98). New York: Springer-Verlag.

Madigan, D., Mosurski, K., & Almond, R. G. (1997). Graphical explanation in belief networks. *Journal of Computational and Graphical Statistics, 6*(2), 160–181.

Mason, B. J., & Bruning, R. (2001). *Providing feedback in computer-based instruction: What the research tells us.* Retrieved May 8, 2007, from http://dwb4.unl.edu/dwb/Research/MB/MasonBruning.html.

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

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*, 99–113.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multi-media learning. In H. M. Niegemann, D. Leutner, & R. Brunken (Ed.), *Instructional design for multimedia learning* (pp. 181–195). Munster, New York: Waxmann.

Nielsen, J. (2000, March). Why you only need to test with 5 users. *Jakob Nielsen's Alertbox Report.* Retrieved February 3, 2007, from http://www.useit.com/alertbox/20000319.html

Nunnally, J. (1978). *Psychometric theory.* New York: McGraw-Hill.

Shute, V. J. (2007). *Focus on formative feedback* (ETS Research Rep. No. RR-07-11). Princeton, NJ: ETS.

Shute, V. J., & Zapata-Rivera, D. (in press). Adaptive technologies. In J. M. Spector, D. Merrill, J. van Merriënboer, & M. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed.). Mahwah, NJ: Erlbaum Associates.

Shute, V. J., Graf, E. A., & Hansen, E. (2005). Designing adaptive, diagnostic math assessments for individuals with and without visual disabilities. In L. PytlikZillig, R. Bruning, & M. Bodvarsson (Eds.), *Technology-based education: Bringing researchers and practitioners together* (pp. 169–202). Greenwich, CT: Information Age Publishing.

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

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

Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., Mislevy, R. J., Steinberg, L., et al. (2000). *Computerized adaptive testing: A primer* (2nd ed.), Mahwah, NJ: Erlbaum Associates.

Notes

[1] Usability by individuals with disabilities is arguably the essence of accessibility. According to the Web Content Accessibility Guidelines (WCAG) 1.0, "Content is accessible when it may be used by someone with a disability" (Chisholm, Vanderheiden, & Jacobs, 1999, glossary entry: "Accessible").

[2] The pilot of the ACED system for individuals with visual disabilities involved four individuals with visual disabilities, three 14-year-old high school students and one 27-year-old adult. Individuals were presented 11 ACED items on the TTT in linear order. With regard to the statement that the system was "easy to use," three individuals strongly agreed and one agreed. The study also indicated that the participants preferred the TTT for graphical and mathematical material rather than for non-mathematical materials. The graphical nature of much of mathematics makes the TTT particularly useful for that subject area (Landau, 2005).

[3] Originally, we tested 290 students, which included a group of 22 English language learners (ELL) with very limited English proficiency. However, during the posttest, the bilingual teacher of the 22 ELL students assisted them not only in translating the problems, but, in many cases, in solving the problems. The teacher was not present during pretesting of the students. Consequently, the ELL students' data were removed from the database as their learning gains were artificially inflated. The final sample size for Study 1 is thus 268 students.

[4] These students generally require 4 years to complete a sequence of courses that others typically require 3 years to complete.

[5] The sample size for the control group was slightly less than that of the other groups for two reasons. First, our randomization procedure involved assigning students a number from 1–4 (starting at 1 in each class) which mapped to the four conditions. Because classes were not equal in size, we had fewer 4s (control) than other numbers. Second, because there was a disproportionate number of ELL students assigned to the control condition, when that group of students was removed from the study, it reduced the sample size of the control condition more than the other conditions.

[6] The five levels included: 1 (honors), 2 (academic), 3 (regular), 4 (remedial), and 5 (special education).

[7] Simple correlations among the three main variables—pretest, posttest, and EAP(SGS)—show that: (a) pretest × posttest, $r = .59$; (b) pretest × EAP(SGS), $r = .50$; and (c) posttest × EAP(SGS), $r = .65$, all significant at $p < 0.01$.

[8] A widely used rule of thumb of 0.70 has been suggested as a minimally accepted internal reliability score (e.g., Nunnally, 1978), and an ideal estimate of internal consistency is between 0.80 and 0.90, because estimates above .90 are suggestive of item redundancy or inordinate scale length (e.g., Clark & Watson, 1995).

[9] We included human-voice recordings of all spoken messages that were expected to be needed to permit a blind student to: log in to the system, set user preferences, move through items, enter responses, and listen to feedback. A voice-over actor with experience reading for blind audiences created the recordings, which were then edited to produce the necessary library of messages. The computer application played appropriate audio messages as needed to guide the student through the process of using the ACED for the TTT system and working through the items.

[10] Section 504 of the Rehabilitation Act of 1973 prohibits discrimination on the basis of disability in programs and activities that receive federal financial assistance. Any child covered under the Individuals with Disabilities Education Act has an IEP and is automatically covered under the Section 504 laws, but the opposite is not true.

[11] Screen reader software (e.g., JAWS) uses synthetic speech output and special navigation commands to allow users with visual disabilities to use applications such as Web pages, word processors, spreadsheets, etc.

Appendix

Proficiencies in ACED (Geometric Sequences)

Proficiency	Definition	Example task (difficulty = medium)
Common ratio	Find the common ratio in a geometric sequence of numbers.	Find the common ratio for the following geometric sequence: 5, -10, 20, -40, 80, . . . (Answer: -2)
Examples	Generate terms of a geometric sequence that satisfy the constraints provided by giving the first few terms of the geometric sequence.	Give the first 3 terms of a geometric sequence that satisfies the following: The first term is a negative integer between -6 and -4, inclusive. The common ratio is a positive integer between 4 and 6, inclusive. (Answer: There is more than one correct answer. Any example you choose that meets the above requirements will be scored as correct.)
Explicit rule	Generate an algebraic expression that represents the nth term for the given geometric sequence.	The first four terms of a geometric sequence are given below. Choose the algebraic expression that represents the n^{th} term in this sequence. 27, 81, 243, 729, . . . a. $3n + 27$ b. 5×3^n c. 3^{n+2} d. $3n$ e. 27^n (Answer: 3^{n+2})

(Table continues)

Table (continued)

Proficiency	Definition	Example task (difficulty = medium)
Model	Generate a geometric sequence that represents the given situation.	At a certain bakery chain, fresh baked goods are prepared every morning. The number of pounds of baked goods that the bakery chain sells per hour, starting with the hour the bakeries open, is shown in the chart below. Pounds of Baked Goods sold per hour Hour Pounds 1 384 2 192 3 96 4 48 If this pattern of sales continues, during which hour will the bakery chain first sell less than 15 pounds of baked goods? (Answer: 6)
Recursive rule	Generate or recognize a recursive formula for a given geometric sequence.	Let $a_1, a_2, a_3, \ldots a_n$, be a sequence of integers. The first four terms of the given sequence are: -3, -6, -12, -24. Select the correct rule for finding the $(n + 1)^{th}$ term from the n^{th} term. $a_{n+1} = -2a_n$ $a_{n+1} = 2a_n$ $a_{n+1} = a_n \times 2^{n-1}$ $a_{n+1} = a_n + 2^{n-1}$ $a_{n+1} = (-a_n)/3$ (Answer: $a_{n+1} = 2a_n$)

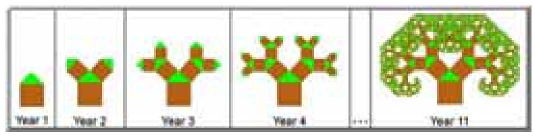
(Table continues)

Table (continued)

Proficiency	Definition	Example task (difficulty = medium)
Solve geometric	Identify which of the following sequences is not a geometric sequence.	Which of the following is not a geometric sequence? 6, 5½, 5, 4½, . . . 2, 1, ½, ¼, . . . 1, 3, 9, 27, . . . 3, 6, 12, 24, . . . (Answer: 6, 5½ , 5, 4½, . . .)
Table	Enter missing values in a table representing a geometric sequence.	The numbers in the table represent terms in a geometric sequence. Complete the table by filling in the values for A and B. Term number — Term value 1 — 1/2 2 — 1/8 3 — A 4 — B (Answer: A = 1/32, B = 1/128)
Verbal rule	Generate a verbal rule for a geometric sequence.	You have an excellent recipe for chocolate chip cookies. By the end of a week, you've shared the recipe with four of your friends. During the second week, each of your friends shared the recipe with four of their friends, so that sixteen new people know about the recipe. Assuming this pattern continues, how many new friends will get the recipe during the 10th week? 4^{10} 4^{9} 4^{8} 1,398,101 (Answer: 4^{10} , which equals 1,048,576.)

(Table continues)

Table (continued)

Proficiency	Definition	Example task (difficulty = medium)
Pictorial	Use geometric sequences to model or extend visual patterns provided.	Nicole receives a mysterious looking tree as a housewarming gift. She plants the tree in her garden. Each year, the tree grows new shoots. The pictures below show the pattern of growth across the years.  Observe the first 4 years of growth. Assuming the same pattern continues, how many new shoots appear in Year 11? (Answer: 1,024)
Extend	Extend the geometric sequence that has simple starting terms and common ratio.	Find the missing terms in the following geometric sequence: 2, __, 18, __, 162, (Answer: 6, 54)