UCLA | CRESST NATIONAL CENTER FOR RESEARCH ON EVALUATION, STANDARDS, AND STUDENT TESTING

TOWARDS INDIVIDUALIZED INSTRUCTION WITH TECHNOLOGY-ENABLED TOOLS AND METHODS: AN EXPLORATORY STUDY

Gregory K. W. K. Chung
Girlie C. Delacruz
Gary B. Dionne
Eva L. Baker
John J. Lee
Ellen Osmundson

JUNE 2016

CRESST REPORT **854**

Towards Individualized Instruction
With Technology-Enabled Tools and Methods:
An Exploratory Study

CRESST Report 854

Gregory K. W. K. Chung, Girlie C. Delacruz, Gary B. Dionne, Eva L. Baker, John J. Lee,
and Ellen Osmundson
CRESST/University of California, Los Angeles

June 2016

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Building, Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported by a grant from the Institute of Educational Sciences under the Educational Research and Development Centers Program, PR/Award Number R305A050004, with funding to the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

An earlier version of this paper was presented at the 2007 annual meeting of the American Educational Research Association, Chicago, IL.

To cite from this report, please use the following as your APA reference: Chung, G. K. W. K., Delacruz, G. C., Dionne, G. B., Baker, E. L., Lee, J. J., & Osmundson, E. (2016). *Towards individualized instruction with technology-enabled tools and methods: An exploratory study* (CRESST Report 854). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

TOWARDS INDIVIDUALIZED INSTRUCTION WITH TECHNOLOGY-ENABLED TOOLS AND METHODS: AN EXPLORATORY STUDY[1]

Gregory K. W. K. Chung, Girlie C. Delacruz, Gary B. Dionne, Eva L. Baker, John J. Lee, and Ellen Osmundson
CRESST/University of California, Los Angeles

Abstract

This report addresses a renewed interest in individualized instruction, driven in part by advances in technology and assessment as well as a persistent desire to increase the access, efficiency, and cost effectiveness of training and education. Using computer-based instruction we delivered extremely efficient instruction targeted to low knowledge learners in sixth-grade algebra readiness classes and eighth-grade Algebra 1A classes. Our research questions were the following: (1) To what extent can very brief exposure to instruction result in learning? and (2) How effective is the instruction compared to no exposure? We found that computer-based "instructional parcels" can be developed to provide very brief instruction that results in learning of mathematics content.

The current interest in personalization and individualized instruction is driven in part by advances in technology (e.g., Advanced Distributed Learning, 2006; IEEE Learning Technology Standards Committee, 2006), advances in assessment (e.g., National Research Council, 2001; Williamson, Behar, & Mislevy, 2006), and a persistent desire to increase the access, efficiency, and cost effectiveness of training and education (e.g., Fletcher, Tobias, & Wisher, 2006). While the idea of individualized instruction has existed for some time (Courtis, 1938), what is remarkable are the striking similarities of desired goals and methods between current research in training and education and work beginning almost a century ago (e.g., teaching machines, Pressey, 1926, 1927; Skinner, 1958; Thorndike, 1912; programmed instruction, Lumsdaine & Glaser, 1960; McDonald, Yanchar, & Osguthorpe, 2005; mastery learning, Bloom, 1968; domain-referenced testing, Baker, 1974; Hively, 1974; Hively, Patterson, & Page, 1968; criterion-referenced testing, Glaser, 1963; CAI, Atkinson, 1968; Suppes & Morningstar, 1969; intrinsic programming, Crowder, 1960; and hypertext, Engelbart, 1962). What differs today, however, is the availability of technology to make practical many of the ideas central to individualizing instruction. First, the means now exist to deliver tasks that can implement and streamline many of the capabilities that were cumbersome in the original formulations of programmed instruction, for example, embedded and dynamic testing, immediate feedback, active participation during instruction, and instructional branching. All were

[1]We would like to thank Donna Morris, Kristin Fairfield, Whitney Wall and the students of Culver City Middle School for participating in this study; Jenny Maguire for providing expert review of our materials; Patty Augenstein for assistance with the computer lab; and Michelle Chaldu and Long Nguyen for help with data collection. We would like to thank Joanne Michiuye of UCLA/CRESST for editorial help with this manuscript.

recognized as important capabilities to increase learning, but their implementation was limited by the available technology of the time. Second and perhaps more important, advances in the science and technology of the assessment of student learning have enabled cost-effective methods to embed assessments into digital environments to support diagnosis, feedback, and the selection and delivery of appropriate "instructional parcels" tailored to a learner's particular level of knowledge.

In this report, we describe our research and development efforts at individualizing instruction for low knowledge students learning pre-algebra. Our effort was part engineering (drawing on the strongest empirical evidence to design the instruction), part instructional design (developing instruction on several pre-algebra concepts), part efficacy testing (testing whether very brief exposure to pre-algebra instruction could result in learning), part assessment research (testing a novel "next step" assessment for rapid and precise diagnosis), and part technology development (testing automated reasoning techniques with Bayesian networks). Our long-term goal is to develop an approach to support "faster, cheaper, and better" ways of delivering individualized instruction in a distributed learning context. Our research questions for the current study were as follows: (1) To what extent can very brief exposure to instruction result in learning? and (2) How effective is the instruction compared to no exposure? Implications for education, training, and online systems are discussed.

In the remainder of this report, we first describe our instructional design of what we term "instructional parcels"—brief, theory-based, multimedia instruction and practice designed to rapidly provide conceptual instruction. Then we describe the use of an innovative and efficient approach to the design of measures of pre-algebra knowledge. We then describe the study method and results of a test of the approach with middle school students.

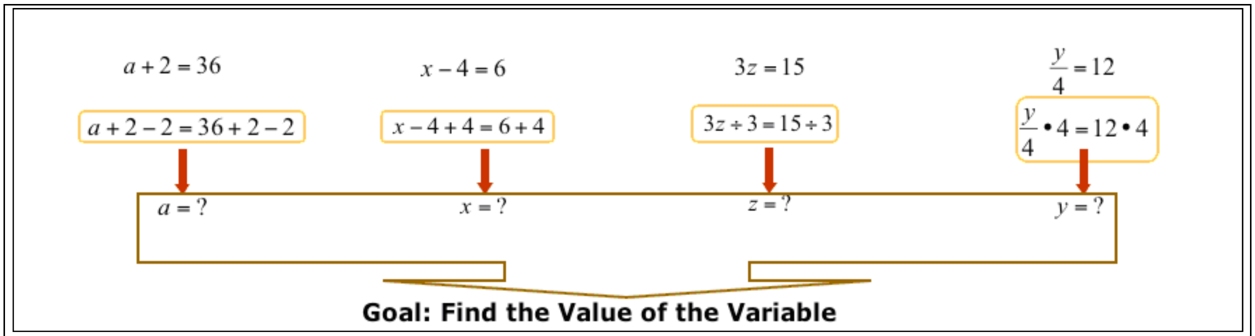
Designing Instruction for Efficient Learning

A basic assumption underlying our instructional design was that for instruction to be maximally effective, particularly when brief, the instructional design should incorporate the features known to promote learning. We drew extensively on the work related to multimedia learning and cognitive load, analogical reasoning, and feedback (e.g., Black, Harrison, Lee, Marshall, & Wiliam, 2003; Catrambone, 1998; Chandler & Sweller, 1991; Kluger & DeNisi, 1996; Mayer, 2001, 2005b; Novick & Holyoak, 1991; Sweller, van Merriënboer, & Paas, 1998). Our goal was to implement in computer-based instruction the properties with strong empirical evidence of effectiveness to deliver extremely efficient instruction targeted to low knowledge learners.

Conceptual Instruction

There were two broad instructional design objectives we set for the actual software. The first objective addressed the overall structure of the instruction—how should the to-be-learned information be structured to facilitate understanding in content that is typically “brittle” (application of math procedures under very specific conditions and tied to the surface form of the math expression)? We addressed this objective by providing instruction that conveyed both the concepts and procedures of the underlying math concept. Importantly, instruction was based on multiple examples applied across different math problems. The desired learning outcome was for students to understand that the same solution method or reasoning applied across problems that differed in surface features. Instruction was chunked into three areas: what (i.e., the goal underlying the math procedure), how (i.e., the procedure), and why (i.e., an explanation of why the particular math concept or procedure “works”).

We designed the instruction to emphasize the relational correspondences among examples. Gick and Holyoak (1983) found that performance on transfer items was improved if instruction made explicit the comparisons among examples to promote the recognition of similarities. One technique is to require learners to explicitly compare examples for similarities (Catrambone & Holyoak, 1989; Cummins, 1992; Novick & Holyoak, 1991). Such explicit comparisons promote performance on transfer tasks by guiding the learner’s attention to more abstract generalizations that might not be obvious in the surface features of the examples. This issue may be particularly important with low prior knowledge learners, where directing learners’ attention to superficial but semantically related aspects of the problems appears key to learning the underlying structure (Catrambone, 1998). Figure 1 shows sample instruction of introducing the goal behind the concept of *transformations*. *Transformations* refers to the idea of isolating variables by using inverse operations.



Narration: What are we trying to do? Even though they all look different, we have the same goal. We’re trying to find the value of the variable. We want to know what *a*, *x*, *z*, and *y* are equal to.

Figure 1. Example of instruction that emphasizes the underlying math concept across different surface examples. The actual on-screen instruction unfolds in steps and the visual highlighting occurs in coordination with the narration.

Multimedia-Based Instruction

The second instructional design objective addressed the delivery of the instruction—what techniques could be used to facilitate the communication of the content? We focused on techniques that specifically addressed limitations of human cognition (e.g., limited working memory capacity), that exploited human sensory channels (visual, auditory), that would be appropriate for math, and that were within the capabilities of the available technology. Where practical, we also adopted techniques with large effect sizes. We addressed the second objective by adopting many of the guidelines derived from research on multimedia and cognitive load (e.g., Clark, Nguyen, & Sweller, 2006; Mayer, 2001, 2005a). Figure 2 shows an example of what the screen looked like. Note that students could control pacing via the control buttons on the lower right of the screen.

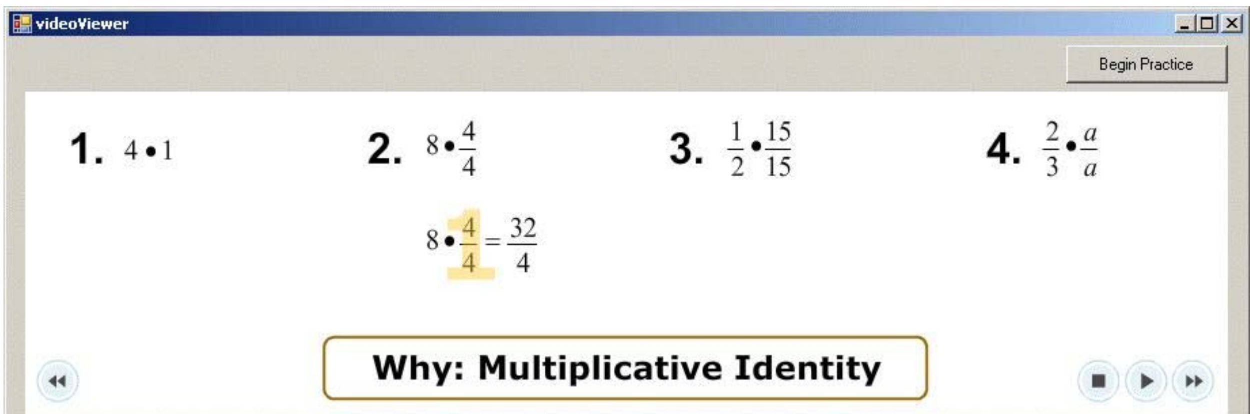


Figure 2. Sample screen shot illustrating (a) multiple examples, (b) cuing, (c) segmenting, and (d) split attention.

Worked-Example Practice

Similar to the instruction, practice was with multiple examples with different surface forms but the same underlying math concept. The practice stage involved students "solving" three math problems in parallel using a "next step" approach. We attempted to implement a simplified worked example in three stages. First, students were required to identify the appropriate next step of solving the problems. Next, given that students were successful, they were asked to identify the expression (i.e., if one carried out the next step procedure, what would be the resultant expression?). Finally, given that students were successful, they were required to identify the underlying math concept. Figure 3 shows a sample screen. Note that in the actual application, the practice is administered in three stages, and only the current and previous stages are shown on the screen. The student advances to the next stage only after getting all the questions correct on the current stage.

The purpose of this approach was to provide practice that emphasized the important first steps involved in solving math problems, to expose students to math language usage when applying math operations, to provide students with practice applying a math concept on different surface forms of the problem, to have the technical means of capturing students' online problem solving processes, and to gather information on whether students commit common errors.

The use of the "next step" was inspired by the work of Kalyuga and Sweller (2004) and Kalyuga (2006). They hypothesized that what students wrote down as the first step in solving an equation reflected their level of expertise in solving equations. High correlations (*r*s between .7 and .9) were found between performance on specifying the first step of a solution and performance on specifying the entire solution. Their finding suggests that the first step is the critical event, particularly for low prior knowledge students. Thus our instruction attempted to support students in identifying the critical first step.

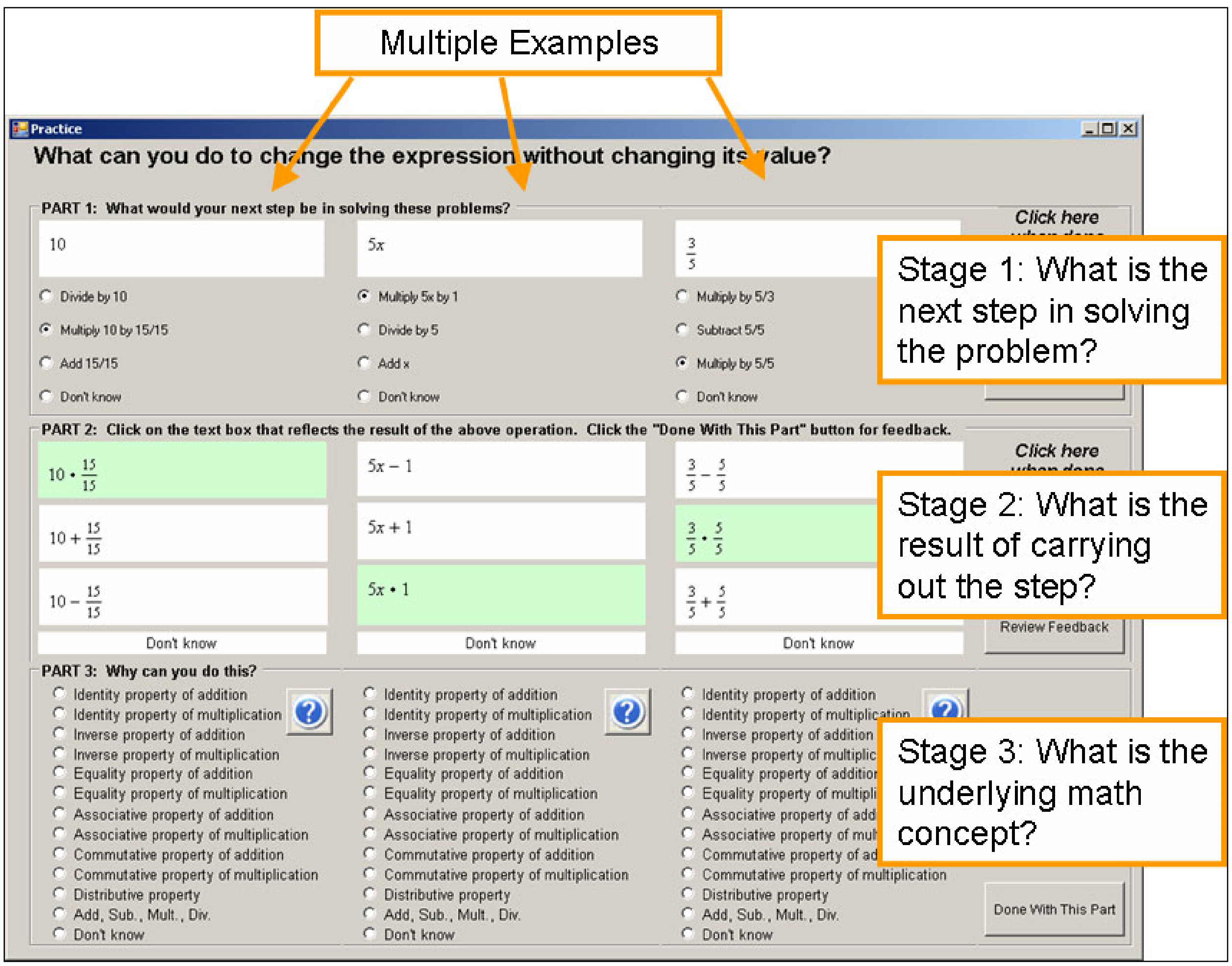


Figure 3. Screen shot showing the practice.

Tailored Feedback

After attempting to solve the problems at each stage, feedback was provided to students as shown in Figure 4. The feedback included (a) knowledge of results (whether the student got the problem correct or incorrect), and (b) explanatory feedback that provided guidance to learners on what they should focus on to solve the problem. These two techniques have been found to be effective feedback methods, particularly for learners with low knowledge of the domain (e.g., Azevedo & Bernard, 1995; Bangert-Drowns, Kulick, & Morgan, 1991; Black et al., 2003; Kluger & DeNisi, 1996).

If the student specified an incorrect response, the explanatory feedback stated what was wrong with the response and a hint about what the student should think about to correctly solve the problem. If the student specified a correct response, the feedback explained why the response was correct. If the student specified "don't know," the feedback provided guidance on what the student should consider. Finally, access to segments of the original instruction was provided. Students could play the video that related to the goal, why, or how instructions that were relevant to the particular practice problem. In addition, students could view instruction on related common errors. Such feedback was provided for all three stages of the practice. The general timing of the feedback was based on findings from Kester, Kirschner, and van Merriënboer (2005), who found that procedural information presented prior to the practice task and explanatory feedback provided during the practice task led to the most efficient learning.

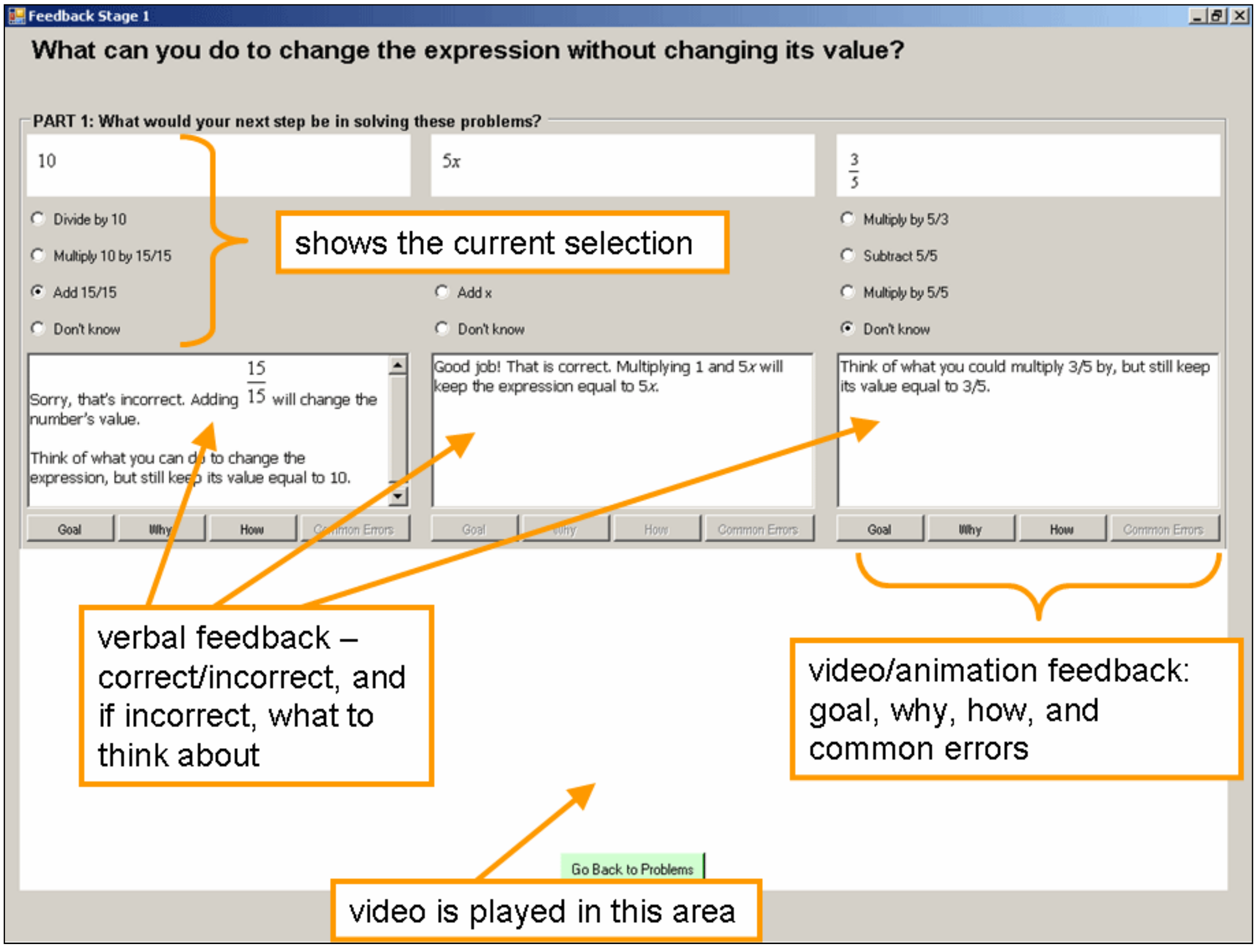


Figure 4. Sample feedback screen.

Task Sequence

The general format of instruction was an initial animation-based instruction followed by practice opportunities with tailored feedback. Figure 5 shows the flowchart of the task sequence.

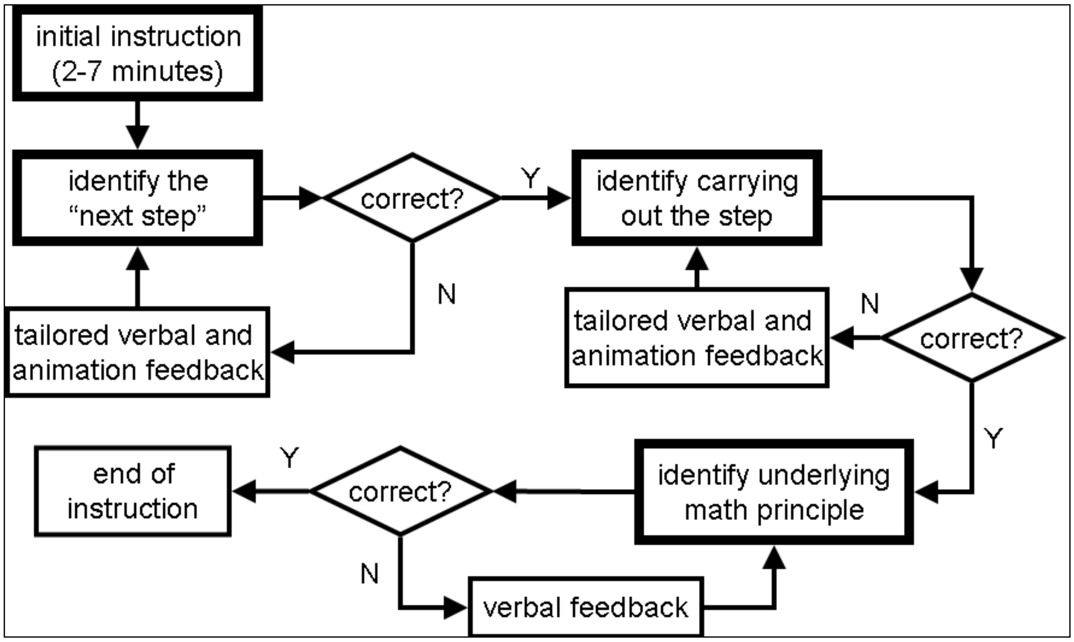


Figure 5. Flow chart of general task sequence. There are four general stages: (a) initial instruction, (b) identifying the "next step" in solving a problem, (c) identifying the equation or expression that results from carrying out the step, and (d) identifying the common underlying math property.

Summary

Thus, our overall instructional design strategy was pragmatic and focused on addressing the question: To maximize the chances of learning, which instructional features have been shown to be very effective, relevant to math, and traceable to an empirical research base? Table 1 summarizes the design properties of the software, our implementation, and the research underlying the specific math property. A fuller description is given in Appendix A.

Table 1

Summary of Instructional Design Features Adopted

Instructional design property	Research base
Use of coordinated graphics and human narration	Modality principle: Learning is higher with graphics and narration than from graphics and printed text (Mayer, 2011). Mean effect size is 0.7.
Complementary sources of information—graphics and audio	Coherence principle: Learning is better when the same information is not presented in more than one format and when extraneous words, pictures, and sounds are excluded (Mayer, 2011). Mean effect size is 1.0. Split attention principle: Materials should be physically and temporally integrated (Ayres & Sweller, 2005).
Learner-controlled pacing	Segmenting principle: Learning is greater when a multimedia message is presented in user-paced segments rather than as a continuous unit (Mayer, 2005b). Effect sizes in the range of 1.0.
Visual annotations	Signaling principle: Learning is deeper from a multimedia message when cues are added that highlight the organization of the essential material (Mayer, 2011). Mean effect size is 0.5. Temporal contiguity principle: Learning is deeper from a multimedia message when corresponding picture or animation and narration or words are presented simultaneously rather than successively (Mayer, 2011). Mean effect size is 1.3.
Use of lay language, first and second person references, and use of math-specific language	Personalization principle: Learning is deeper when the words in a multimedia presentation are in conversational style rather than formal style (Mayer, 2005d) and using a human rather than computer-generated voice (Mayer, 2011). Effect sizes in the range of 0.8 to 1.3.
Instruction centered around worked examples	Worked example: Consists of a problem formulation, solution steps, and the final solution (Clark & Mayer, 2011; Renkl, 2005).
Target low knowledge learners	Prior knowledge principle: Instructional strategies that help low knowledge individuals may not help or may hinder high knowledge learners (Mayer, 2001). Effect sizes in the range of 0.6.
Subgoal chunking and labeling, multiple examples	Structure emphasizing instruction: Instruction that emphasizes the structural features that are relevant to the correct solution procedure (Renkl, 2005). Grouping solution steps by goals and methods, and explicitly labeling the chunks as *goals* and *method*, particularly across different problems with different surface features but with the same underlying solution structure, promotes problem solving in low knowledge learners (Catrambone, 1998).
Knowledge of results during practice Explanatory feedback tailored to participants' selection	Knowledge of results and explanatory feedback promote learning (Bangert-Drowns et al., 1991; Hattie, 2009; Kluger & DeNisi, 1996). Feedback is particularly effective when used to illuminate goals for students, progress toward goals, and determine next steps. Examples of effective feedback can be used for corrective purposes, to provide information about past attempts, to point to alternative strategies, and to increase effort and motivation (Hattie, 2009).

Designing Assessments for Rapid and Precise Diagnosis

To support diagnoses of student knowledge gaps, items were developed to sample the following topics: (a) 12 properties of algebra, (b) transformations and related operations, (c) arithmetic, and (d) fractions. We used a novel item-generation process intended to provide precise information about students' pre-algebra knowledge. Our assumption is that solving a problem requires the successful application of different concepts across multiple steps. Table 2 shows a step-by-step derivation for solving the equation $7x - (3x - 2) = 38$.

Table 2

Item Generation Technique Based on Problem Derivation

Step	Equation	Transition	Operation used from previous step to the current step
0	$7x - (3x - 2) = 38$	step 0:	given, solve for *x*
1	$7x + (-1)3x - (-1)2 = 38$	step 0 → step 1:	distributive property
2	$7x - 3x + 2 = 38$	step 1 → step 2:	multiplication
3	$x(7 - 3) + 2 = 38$	step 2 → step 3:	factor
4	$x(4) + 2 = 38$	step 3 → step 4:	subtraction
5	$4x + 2 = 38$	step 4 → step 5:	commutative property of multiplication
6	$4x + 2 - 2 = 38 - 2$	step 5 → step 6:	addition property of equality (transformation)
7	$4x + 0 = 36$	step 6 → step 7:	subtraction
8	$4x = 36$	step 7 → step 8:	additive identity
9	$\left(\frac{1}{4}\right)4x = \left(\frac{1}{4}\right)36$	step 8 → step 9:	multiplication property of equality (transformation)
10	$(1)x = \left(\frac{1}{4}\right)36$	step 9 → step 10:	multiplicative inverse
11	$x = \left(\frac{1}{4}\right)36$	step 10 → step 11:	multiplicative identity
12	$x = \frac{36}{4}$	step 11 → step 12:	multiplication
13	$x = 9$	step 12 → step 13:	division

Each step-to-step transition requires the use of a single or small number of algebra properties or arithmetic. The transitions were coded to capture the algebra knowledge required in each transition. Note that the decomposition process yielded simpler equations and terms, which we used as a source for items. Thus our items were generally of the form *identify a valid next step* instead of *solve an entire equation*.

Deriving items from a single equation yields several benefits: (a) an efficient, simple, and precise item-generation scheme that is inherently coherent; (b) a set of items that students can solve quickly; (c) items amenable to a selected-response format, particularly the multiple true-false (MTF) format (Frisbie, 1992); and most interestingly, (d) a built-in transfer test—solving the original equation (or sub-steps within the equation). Example items are shown in Figure 6.

The use of the *next step* was inspired by intriguing findings by Kalyuga and Sweller (2004) and Kalyuga (2006). The authors gave students equations, but instead of asking students to fully solve an equation, the authors asked students to simply write down the first step of their solution. Kalyuga and Sweller found this technique to be highly predictive of performance on tasks that required students to solve the entire equation. High correlations were found between the two methods across different studies, ranging from .7 to .9. While the authors were interested in identifying ways of measuring expertise (with more skipped steps indicative of higher expertise), the strong relation between the two forms suggested a very efficient assessment.

An equation is given below. For each statement below, circle whether the statement is T (true) or F (false). <u>Circle DK (don't know) if you are unsure.</u>

$$7x-(3x-2)=38 \quad \text{(step 1)}$$

DP (OPER)	72.	$7x-3x+2=38$ is a possible NEXT STEP after step 1.	●T	F	DK
CP-ADD (CE)	73.	$7x-(3x-2)=38$ has the same value as $7x-(2-3x)=38$.	T	F●	DK
AP-ADD (CE)	74.	$7x-(3x-2)=38$ has the same value as $(7x-3x)-2=38$.	T	F●	DK
DP (CE)	75.	$7x-3x-2=38$ is a possible NEXT STEP after step 1.	T	F●	DK
DP (CE)	76.	$7x-5x=38$ is a possible NEXT STEP after step 1.	T	F●	DK
DP (CE)	77.	$7x-(-1)3x-(-1)2=38$ is a possible NEXT STEP after step 1.	T	F●	DK
DP (OPER)	78.	$7x+(-1)3x-(-1)2=38$ is a possible NEXT STEP after step 1.	●T	F	DK

DP = distributive property
CP-ADD = commutative property of addition
AP-ADD = associative property of addition
CE = common error
OPER = operation

Figure 6. Sample multiple true-false item. Each item is mapped to a single math concept, allowing for precise diagnoses of math knowledge. Dots denote the correct answer.

Research Questions

We had two research questions related to whether our instructional technique was effective:

- To what extent can very brief exposure to instruction result in learning?

- How effective is the instruction compared to no exposure?

The first question addresses the basic question of whether our implementation of the instructional techniques can be effective. Can even very brief exposure to instruction on pre-algebra be effective? We know of no other work that has, as a main goal, explicitly designed instruction for brevity. Thus, it is an open question on whether such an approach could be effective.

The second research question was comparative. If our approach demonstrated learning, how effective was the instruction? In our case, we used the most straightforward comparison—the no instruction condition. Because we were not manipulating instructional variables, we were very early in the research, and given that comparisons between instructional techniques are problematic (particularly technology vs. "traditional" instruction), we assumed the most appropriate and interpretable contrast at this stage of the research would be between students receiving our instruction and students receiving no instruction.

Method

Participants

Background. One hundred fifteen participants were recruited from three teachers at an urban middle school in southern California. At the end of the first semester participants were drawn from two sixth-grade algebra readiness classes (n = 54) and three eighth-grade Algebra 1A classes (n = 52). The sample was split about evenly by gender and students' ethnicity was diverse, including 31% Latino, 25% Asian or Pacific Islander, 17% White, and 13% African American. About three fourths of students reported receiving A's or B's in math, and nearly all students agreed or strongly agreed that they were able to understand their teacher's explanations in math class, and nearly all students agreed or strongly agreed that they were able to read and understand most of the problems and explanations in their math textbook. When asked how knowledgeable they were about pre-algebra, 83% of students reported being moderately or very knowledgeable.

Correlations among self-reported math grades, knowledge of pre-algebra, and pretest are shown in Table 3. In general, achievement variables were significantly related with affective variables and language variables. We interpret these relations in light of the student background information as suggesting that our sample was typical of middle-school pre-algebra students in urban settings. Appendix B contains a more detailed description of the sample and Appendix C contains a more detailed description of the measures.

Table 3

Descriptive Statistics and Intercorrelations on Background Variables

Variable	*n*	*M*	*SD*	1	2	3	4	5	6	7
Achievement										
1. Pretest[a]	113	54.2	15.1	–						
2. Self-reported grades in math[b]	109	2.9	5.0	-.20	–					
3. Self-reported pre-algebra knowledge[c]	102	3.0	0.6	.43**	-.31**	–				
Affect										
4. Interest in math[d]	78	2.9	0.7	.31**	-.07	.22*	–			
5. Self-concept in math[d]	78	3.1	0.8	.37**	-.19	.52**	.49**	–		
6. Belief in "math myths"[e]	78	3.2	0.4	.42**	.00	.02	.19	.30**	–	
Language										
7. Understand teacher's explanations[e]	107	1.6	0.5	-.39**	.13	-.42**	-.29**	-.47**	-.19	–
8. Understand problems and explanations in textbook[e]	108	1.6	0.6	-.47**	.25*	-.34**	-.31**	-.32**	-.35**	.56**

[a]Maximum score = 84. [b]1 = *Mostly A's*, 2 = *Mostly B's*, 3 = *Mostly C's*, 4 = *Mostly D's*, 5 = *Mostly F's*. [c]1 = Not *knowledgeable at all*, 2 = *Somewhat knowledgeable*, 3 = *Moderately knowledgeable*, 4 = *Very knowledgeable*. [d]1 = *Disagree*, 2 = *Disagree somewhat*, 3 = *Agree somewhat*, 4 = *Agree*. [e]1 = *Strongly agree*, 2 = *Agree*, 3 = *Disagree*, 4 = *Strongly disagree*.

Design

A pretest, posttest two-condition design was used. Participants in both conditions received a pretest one instructional day before the start of the intervention. On the day of the computer-based activity, participants in the experimental condition received the individualized remediation followed by a posttest. Participants in the control condition received the posttest followed by the treatment. This design was used for the control condition to allow those students to participate in the computer activity.

Measures

Pretest scales and transfer scales. The pretest consisted of 84 items that spanned a range of mathematical knowledge and included simple identification of math facts (e.g., whether 0/3 = 0) to more complex knowledge (e.g., equations with variables on both sides of the equations). About 40% of the items used the MTF format as shown in Figure 6. MTF items are highly efficient for gathering achievement data and they tend to yield higher reliabilities than other selected-response formats (Frisbie, 1992). Appendix D contains example items from the pretest.

Twelve scales were developed and tested in the pretest. Scales were dropped from the analyses for the following reasons: (a) reliability less than .60 (three scales), (b) no instruction delivered that directly mapped to the scale (two scales), or (c) the scale had less than four items regardless of reliability (three scales).

Two complex tasks were created, one related to fractions and the other scale related to rational number equivalence. The fractions scale was made up of constructed-response items that required students to evaluate or solve a problem related to adding or multiplying fractions. The rational number equivalence scale was an explanation task, which required students to explain how to find equivalent fractions and to provide numeric examples. Appendix E contains the transfer items and the scoring rubric.

Table 4 shows the scale information for the pretest scales and the complex task scales. Overall, the pretest subscales retained for analyses had moderate reliability and the transfer scales had acceptable reliabilities.

Table 4

Scale Reliability

Scale	*n*	No. of items	Alpha
Pretest	81	84	.89
Pretest subscales			
Adding fractions	101	8	.68
Distributive property	97	8	.70
Transformations	104	6	.63
Multiplicative identity	93	7	.61
Complex task scales			
Multiplying and adding fractions	96	6	.71
Rational number equivalence	113	2	.75

Diagnosing Knowledge Gaps

Bayesian network (BN) of pre-algebra knowledge. We developed a BN to represent the domain of pre-algebra. Each hypothesis node could assume the state *understands* or *does not understand*, and each observable node could assume the state *correct* or *incorrect*. Participants' performance (correct or incorrect) on a test item was used as inputs to the observable nodes. Relations captured the presumed knowledge dependencies among the different concepts. For example, understanding the concept of *addition* implies understanding of the concept *commutative* *property of addition*. The BN was constructed by an electrical engineering graduate. In general, nodes higher in the network tended to be more conceptual and abstract (e.g., *simplifying*) while leaf nodes tended to be skill-oriented (e.g., *multiplicative identity*). The BN included 28 concepts related to the properties of algebra, arithmetic, equality, evaluation, fractions, simplifying, rational numbers, and factoring.

A student *understanding* a concept was operationalized as a fusing of test item performance data across the items that required use of the particular concept. The items sampled a variety of conditions (e.g., identifying the "next step" in a solution process; use of multiple problems that incorporated fractions, integers, variables; correctly solving a problem).

Given a student's performance on the items of the type shown in Figure 6, the BN was used to compute the probability of a student understanding various pre-algebra concepts. The probabilities were used to classify a student's understanding on a particular concept as high, medium, or low. This classification was taken as the diagnosis of a student's knowledge and used as the basis for determining what concepts each student would receive.

Procedure

A pretest was administered to students via the classroom teacher. The pretest was a self-contained paper booklet, with all the directions contained in the booklet. Teachers were instructed not to help students with the content. Students had 60 minutes to complete the pretest. The teachers reported that students finished the pretest within a range of 45–60 minutes.

The following week, which represented a gap of one instructional day, students were administered pre-algebra instruction on a computer in a computer lab. Students were randomly assigned to computer stations. Students received individualized instruction in the following sense: Based on students' pretest performance, students received instruction that they were predicted to need help on. In general, students received instruction on four to six concepts within the allotted 45-minute period. Participants in the instruction condition received the instruction before the posttest, and participants in the control condition completed the posttest before receiving instruction.

The posttest was tailored to each individual and included only the concepts the student received instruction on. The items on the posttest were identical to the items on the pretest. After completing the posttest, students filled out a demographic survey, a satisfaction survey, and a "math myths" survey.

Results

Preliminary Analyses

Check for accuracy of diagnoses. Analyses were conducted to evaluate the quality of the classification of students into high, medium, and low knowledge categories. We checked for group differences on the pretest by classification. There were significant differences by knowledge level for the *adding fractions* and *transformations* scales, with students classified as low performing the lowest, followed by students classified as medium and high. For the *distributive property* scale, there was no difference between the medium and high groups, and for the *multiplicative identity* scale, there was no difference between the low and medium groups. Because we were only interested in participants who demonstrated knowledge gaps, we collapsed the low and medium groups. Overall, with the collapsing, the groups appeared differentiated.

Check for preexisting group differences. A t test was performed to test for a preexisting difference in group means between the control ($M = 57.7$, $SD = 16.1$) and experimental ($M = 53.0$, $SD = 14.6$) conditions. The difference between the two conditions was close to significant ($t(111) = 1.45$, $p = .15$). Because the t test suggested a potential difference between groups on pre-algebra knowledge, subsequent between-group analyses included the pretest as a covariate.

Check for implementation of the treatment. The proportion of time spent on the instruction relative to the actual running time of the video instruction was computed for each student. The computation was based on students' log files that contained event markers indicating the start and end times related to the instruction. A value less than 1 suggests that the student had skipped through the video and a value greater than 1 suggests the student replayed portions of the video.

Table 5 shows the distribution of students with respect to the time spent on different instruction for different concepts. As Table 5 shows, a number of students did skip the instruction. We used the proportion measure as an indicator of treatment implementation and reasoned that if students did not receive sufficient exposure to the instruction, then they could not profit from it. Thus, we excluded from subsequent analyses students who spent less than two thirds of the time viewing the instruction.

Table 5

Distribution of Students Spending Different Amounts of Time on the Initial Instruction by Initial Skill Classification

Concept	Instructional time (mm:ss)	Proportion of time spent on the initial instruction				
		≤ 0.33	0.33 – 0.67	0.67 – 1.00	1.00 – 1.33	≥ 1.33
Adding fractions	10:58	10	5	15	12	2
Distributive property	4:09	12	11	11	12	2
Transformations	4:03	7	9	14	7	0
Multiplicative identity	3:04	4	3	10	13	2

Main Analyses

Overview of analyses. The main analyses addressed the two research questions: (a) To what extent can very brief exposure to instruction result in learning? and (b) How effective is the instruction compared to no exposure? Note that separate analyses were conducted by scale. For example, participants who received instruction on *adding fractions* were tested for learning by comparing posttest to pretest performance on the *adding fractions* scale. The items in the scale were identical on both occasions. Because participants received instruction on multiple concepts the same participant was included in analysis related to those concepts. That is, if a participant received instruction on *adding fractions* and *multiplicative identity*, then that participant was included both in the *adding fractions* analysis and in the *multiplicative identity* analysis. Thus, caution is warranted on interpreting the results as there may be spillover effects.

Between-group comparisons included the pretest as a covariate, as the random assignment resulted in pretest performance favoring the control condition (p = .15). The pretest included the same items on all posttest scales, in addition to more basic math facts. The pretest was taken as a much broader measure of pre-algebra knowledge than any one scale. Support for this interpretation is seen in the significant correlations between the pretest and nearly all achievement measures, attitudinal, and language measures (see Table 3).

Research question 1: To what extent can very brief exposure to instruction result in learning? To address the first question, we examined within-subjects learning from pretest to posttest, by condition. Table 6 shows the results of the analyses. In the no-instruction condition, there were no significant differences between pretest and posttest on the scales or on the transfer tasks. This result strongly suggests opportunity-to-learn effects due to instruction and (re)testing effects were not strong enough to affect posttest performance. In the instruction condition, students in general performed higher on the posttest compared to the pretest, on nearly all the scales. Posttest results for the complex task are tentative.

Table 6

Within-Subjects Comparison of No-Instruction and Instruction Conditions

	No instruction							Instruction						
		Posttest	Pretest	Test of group differences (paired *t* test)					Posttest	Pretest	Test of group differences (paired *t* test)			
	n	*M (SD)*	*M (SD)*	*t*	*df*	*p*	*d*	*n*	*M (SD)*	*M (SD)*	*t*	*df*	*p*	*d*
Scale														
Adding fractions[a]	9	4.67 (2.29)	3.67 (1.66)	1.50	8	.17	--	22	4.95 (2.13)	3.91 (2.00)	2.63	21	.02	0.50
Distributive property[a]	13	4.46 (1.45)	4.62 (2.43)	-0.23	12	.78	--	18	5.44 (1.50)	4.44 (2.12)	1.80	17	.09	0.54
Transformations[b]	13	4.23 (1.79)	3.62 (1.80)	1.98	12	.07	0.34	21	4.76 (1.09)	3.81 (1.91)	3.21	20	<.01	0.61
Multiplicative identity[c]	7	5.29 (1.25)	5.00 (1.83)	0.60	6	.57	--	21	5.81 (0.98)	4.00 (1.61)	5.29	20	<.01	1.36
Complex tasks														
Multiplying and adding fractions[b]	12	3.00 (1.28)	2.67 (1.72)	0.74	11	.47	--	11	4.09 (1.58)	3.36 (2.25)	1.62	10	.14	0.37
Rational number equivalence[b]	13	1.46 (2.18)	1.08 (1.89)	0.81	12	.43	--	12	4.25 (2.56)	3.17 (2.48)	1.78	11	.10	0.43

[a]Maximum score = 8. [b]Maximum score = 6. [c]Maximum score = 7.

The within-subjects results provide evidence that our computer-based instruction was effective on posttest scales and the results are suggestive of a learning effect on the complex tasks. While the general finding that learning occurred from instruction is not surprising, we note with emphasis that significant learning occurred with *very brief* exposure to instructional activities (effect sizes from .50 to 1.4).

Research question 2: How effective is the instruction compared to no exposure to instruction? To address the second question, we examined differences on the posttest between the instruction and no-instruction condition. Table 7 shows the results of the analyses. With the exception of the *adding fractions* scale, there were significant differences ($p < .06$) on the other posttest scales and on the transfer tasks. The effect sizes ranged from 0.5 to 0.9, suggesting the instruction was effective.

Table 7

Time on Task, Adjusted Means, Standard Errors, and Analysis of Covariance (ANCOVA) for Pre-Algebra Concepts

	Time spent on task components (instruction condition)		Comparison of treatment effects									
	Instruction	Practice	Instruction			No instruction			ANCOVA[a]			
	M (SD)	*M (SD)*	*n*	Adj. *M*	*SE*	*n*	Adj. *M*	*SE*	*df*	*F* ratio	*p*	effect size[b]
Scale												
Adding fractions[c]	10:46 (2:37)	6:30 (3:32)	22	5.02	0.40	9	4.52	0.62	28	0.46	.50	--
Distributive property[c]	4:09 (0:57)	1:55 (0:40)	18	5.39	0.36	14	4.22	0.41	29	4.46	.04	0.76
Transformations[d]	3:50 (0:25)	3:14 (2:19)	21	4.90	0.26	13	4.00	0.33	31	4.58	.04	0.77
Multiplicative identity[e]	3:33 (2:21)	2:31 (1:43)	21	5.90	0.21	7	5.03	0.38	25	3.91	.06	0.91
Complex tasks	19:17 (3:37)	13:17 (6:08)										
Multiplying and adding fractions[d]	--	--	12	3.98	.31	13	2.94	.30	22	5.74	.03	0.91
Rational number equivalence[d]	--	--	12	4.09	.65	13	1.61	.62	22	7.51	.01	0.50

[a]Between-groups *df* = 1. [b]Hedge's *g*. [c]Maximum score = 8. [d]Maximum score = 6. [e]Maximum score = 7.

The between-subjects results provide evidence that our computer-based instruction was effective, compared to having no instruction. Not only did participants learn in general as suggested in the within-subjects analysis, but the learning attributable to the instruction (taking differences in pre-algebra knowledge into account) resulted in participants in the instruction condition performing much higher than participants in the no-instruction condition (effect sizes range from .8 to .9).

The most intriguing finding was an effect on the students' performance on the transfer tasks. The complex tasks were constructed requiring students to generate a solution (vs. the MTF format). In addition, the rational number equivalence task required participants to first explain how to find fractions that are equivalent to 2/3, and then provide examples. Student responses were judged for conceptual understanding (vs. procedural). Participants in the instruction condition outperformed participants in the no-instruction condition on the posttest complex tasks, with effect sizes of .9 (multiplying and adding fractions) and .5 (rational number equivalence).

Discussion

This study investigated whether an integrated instructional and assessment system, drawing on empirically supported instructional design techniques and assessment approaches, could result in low prior knowledge students learning pre-algebra topics after very brief exposure to instructional activities—typically less than 5 minutes of direct instruction and less than 3 minutes of practice. We found evidence of instructional effectiveness. Participants who received instruction outperformed participants who did not receive instruction on nearly all measures of performance. This finding held for posttest scales (that were identical to pretest scales) and for near transfer items that required solving fraction problems or required explanations related to rational number equivalence. While such differences are unremarkable in and of itself, what is remarkable is that such differences were observed given the brevity of the instructional treatment.

Limitations

This study has several limitations. First, about one third of the sample did not view the instructional videos despite their brevity. This result may be because the videos were not engaging, not understandable, or other reasons. A second limitation arises from the individualization of instruction and assessment. Individuals were given instruction only on topics they were predicted to need remediation on, and they were tested only on those particular topics. Thus, for each topic, only a subset of the sample experienced the intervention. This reduced the sample size used for the analysis of effects; thus, these results should be taken as tentative.

Implications

The most important finding is that we have tentative evidence that computer-based "instructional parcels" can be developed to provide very brief instruction that results in learning of pre-algebra content. The most important implication of this finding is that our approach is practical and works in field settings. These instructional parcels were designed to be self-contained, stand-alone content with embedded practice. While we did not implement a fully automated system, every aspect of the work has been designed to support such an implementation. For example, the pretest items (MTF format) were paper-based but they were designed to be implemented on a computer and designed to work with Bayesian networks (e.g., dichotomous scoring simplifies the integration with Bayesian networks).

A persistent general question that underlies our work is to what extent can "dumb" technology using smart methods be used to result in significant learning outcomes? Is the only way to meet the "2-sigma" challenge (Bloom, 1984) with intelligent tutoring systems (ITS)? Are there alternative instructional methods and technologies that can be used to replicate many of the desirable techniques associated with human tutors and ITSs, but in a way that does not rely on highly sophisticated artificial intelligence approaches? ITSs require extremely detailed student models and sophisticated inference engines. Development of ITSs is expensive and proprietary, and requires very specific skills.

As suggested by the results of this study, the simple approach we adopted may be able to support student learning and fill many classroom capability gaps. We think we have developed a general approach for targeted instructional methods that is simple, flexible, easy to use, and workable under nominal classroom conditions. We believe that the combination of judicious design of multimedia messages, tailored feedback addressing specific misconceptions, and multiple-example practice opportunities is key to promoting learning in general, and perhaps critical when learning is expected to occur over a brief period of time.

Next Steps

There are several next steps of this work. First, the measures need to be refined so that more scales can be included in the analyses. Several scales were dropped due to poor reliability and low number of items. A second area is to examine retention: While effects were demonstrated, they were based on testing immediately following the intervention. Demonstrating longer term retention would bolster claims of effectiveness considerably. Another area to investigate is the accuracy of the initial diagnoses. While we used a Bayesian network to compute probabilities that a student understood different concepts, we did not conduct an extensive examination of the accuracy of our diagnosis. This is clearly an area in need of further investigation, as individualization is in large part based on accurate diagnosis of knowledge capabilities and gaps. Finally, our initial objective was to design the most effective instruction by integrating a variety of techniques. The drawback of this approach is that it is not possible to identify the relative effectiveness of the various techniques. Future work that addresses this issue may yield even more efficient instruction by providing information that would inform decisions about tradeoffs between development time, development cost, degree of learning, length of instruction, and practice time.

The signature properties of our instructional parcels is that they are brief, stand alone, and combine assessment, instruction, and practice. They are not intended to replace teachers or large curricular units, but rather to provide instruction on a single topic. We believe our approach is feasible and applicable to a variety of contexts ranging from K-16 instruction to military training, providing supplementary help to students on an as needed basis. To realize the promise of individualized instruction, systems need to address the practical constraints of educational and training environments. Instructors in school settings typically have well over 100 students and are often limited in what they can do to meet individual student needs—whether it is one-on-one tutoring, keeping track of students' progress, or even correcting homework. Flexible instructional parcels may be one solution.

References

Advanced Distributed Learning. (2006). *Sharable Content Object Reference Model (SCORM)*, Version 2004 3rd Edition. Alexandria, VA.

Amdahl, K., & Loats, J. (1995). *Algebra unplugged*. Broomfield, CO: Clearwater Publishing.

Atkinson, R. C. (1968). Computerized instruction and the learning process. *American Psychologist, 23*, 225–239.

Ayres, P., & Sweller, J. (2005). The split-attention principles in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 135–146). New York, NY: Cambridge University Press.

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback computer-based instruction. *Journal of Educational Computing Research, 13*, 109–125.

Baker, E. L. (1974). Beyond objectives: Domain-referenced tests for evaluation and instructional improvement. *Educational Technology, 14*(6), 10–16.

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

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice.* Buckingham, England: Open University Press.

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1*(2).

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General, 127*, 355–376.

Catrambone, R., & Holyoak, K. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 1147–1176.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293–332.

Clark, R. C., & Mayer, R. E. (2011). *e-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. New York, NY: John Wiley & Sons.

Clark, R. C., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. San Francisco, CA: Pfeiffer.

Courtis, S. A. (1938). Contributions of research to the individualization of instruction. *National Society for the Study of Education, 37*, 201–210.

Crowder, N. A. (1960). Automatic teaching by intrinsic programming. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning: A source book* (pp. 286–298). Washington, DC: National Education Association of the United States.

Cummins, D. D. (1992). Role of analogical reasoning in the induction of problem categories. *Journal of Educational Psychology, 18*(5), 1103–1124.

Engelbart, D. C. (1962). *Augmenting human intellect: A conceptual framework*, AFOSR-3233 Summary Report, SRI Project No. 357. Menlo Park, CA: Stanford Research Institute.

Fletcher, J. D., Tobias, S., & Wisher, R. A. (2006). Learning anytime, anywhere: Advanced distributed learning and the changing face of education. *Educational Researcher, 36*(2), 96–102.

Frisbie, D. A. (1992). The multiple true-false item format: A status review. *Educational Measurement: Issues and Practice, 5*(4), 21–26.

Gick, M. L., & Holyoak, K. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist, 18*, 519–521.

Hattie, J. (2009). *Visible learning*. New York, NY: Routledge.

Hively, W. (1974). Introduction to domain-referenced testing. *Educational Technology, 14*(6), 5–10.

Hively, W., Patterson, H. L., & Page, S. H. (1968). A "universe defined" system of arithmetic achievement tests. *Journal of Educational Measurement, 5*, 275–290.

IEEE Learning Technology Standards Committee (LTSC) P1484. (2006). IEEE P1484.12.3., draft 8, extensible markup language (XML) schema definition language binding for learning object metadata. Retrieved June 1, 2006, from http://ieeeltsc.org.

Kalyuga, S. (2006). Rapid cognitive assessment of learners' knowledge structures. *Learning and Instruction, 16*, 1–16.

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96*, 558–568.

Kester, L., Kirschner, P. A., & van Merriënboer, J. G. (2005). Timing of information presentation in learning statistics. *Instructional Science, 32*, 233–252.

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*, 254–284.

Lumsdaine, A. A., & Glaser, R. (1960). *Teaching machines and programmed learning: A source book*. Washington, DC: National Education Association of the United States.

Marsh, H. W., Hau, K.-T., Artelt, C., & Baumert, J. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing, 6*, 311–360.

Mayer, R. E. (2001). *Multimedia learning*. New York, NY: Cambridge University Press.

Mayer, R. E. (Ed.). (2005a). *Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.

Mayer, R. E. (2005b). Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 169–182). New York, NY: Cambridge University Press.

Mayer, R. E. (2005c). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 183–200). New York, NY: Cambridge University Press.

Mayer, R. E. (2005d). Principles of multimedia learning based on social cues: Personalization, voice, and image principles. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 201–212). New York, NY: Cambridge University Press.

Mayer, R. E. (2011). Applying the science of learning to multimedia instruction. In J. Mestre & B. Ross (Eds.), *Psychology of learning and motivation: Cognition in instruction* (pp. 77–108). New York, NY: Oxford University Press.

McDonald, J. K., Yanchar, S. C., & Osguthorpe, R. T. (2005). Learning and programmed instruction: Examining implications for modern instructional technology. *Educational Technology Research & Development, 53*, 84–98.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment. J. Pellegrino, N. Chudowsky & R. Glaser (Eds.). Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 398–415.

Pressey, S. L. (1926). A simple apparatus which gives tests and scores—and teaches. *School and Society, 23*(586), 373–376.

Pressey, S. L. (1927). A machine for automatic teaching of drill material. *School and Society, 25*(645), 549–552.

Renkl, A. (2005). The worked-out example principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 229–245). New York, NY: Cambridge University Press.

Skinner, B. F. (1958, October 24). Teaching machines. *Science, 128*, 969–977.

Suppes, P., & Morningstar, M. (1969, October 17). Computer-assisted instruction. *Science, 166*, 343–350.

Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.

Thorndike, E. L. (1912). *Education: A first book*. New York, NY: MacMillan.

Williamson, D. M., Behar, I. I., & Mislevy, R. J. (Eds.). (2006). *Automated scoring of complex tasks in computer-based testing*. Mahwah, NJ: Erlbaum.

Appendix A:
Summary of Design Guidelines Implemented in Instruction

Key instructional design property	Implementation	Multimedia principle
Use of coordinated graphics and human narration	The instruction combined graphics and narration. Text was used to show the math principle being covered and the associated type of explanation (i.e., Goal, Why, How). A narration was used to step the student through the instruction.	Modality principle: Learning is higher with graphics and narration than from graphics and printed text (Mayer, 2005b). Effect sizes in the range of 0.97.
Complementary sources of information—graphics and audio	We designed the instruction so that each channel (visual and aural) provided different information. Thus, one channel alone would be insufficient to understand the information. Graphics were used to show the form of the math expression or equation, and narration was used to provide the ongoing explanation that was coordinated with the graphics display. All the relevant material was coordinated physically and temporally. Equations and their decomposition appeared adjacent to each other, and the narration was coordinated with the animation.	Coherence principle: Learning is better when the same information is not presented in more than one format (Mayer, 2005c). Effect sizes in the range of 1.3. Split attention principle: Materials should be physically and temporally integrated (Ayres & Sweller, 2005).
Learner-controlled pacing	User controls were provided so that students could stop, replay, or fast-forward through the instruction. Instruction was segmented into major sections that explained what the goal of using the math operation was, why the math concept was important, and how to apply the operation.	Segmenting principle: Learning is greater when a multimedia message is presented in user-paced segments rather than as a continuous unit (Mayer, 2005b). Effect sizes in the range of 1.0.
Visual annotations	Visual annotations were added to the instruction to explicitly point out what the narration was referring to. Several cuing techniques were used: circling the element under discussion, highlighting or superimposing an image to demonstrate equivalency (see Figure 2), or using arrows to denote procedural flow. Color conventions were used such that the visual annotations were always the same color and different from the math equation or expression. The visual and aural information were always synchronized in delivery. That is, the narration was coordinated with the graphics and cuing. The cuing was intended to guide students' attention to the important information and the narration explained the related math procedure or concept.	Signaling principle: Learning is deeper from a multimedia message when cues are added that highlight the organization of the essential material (Mayer, 2005c). Effect sizes in the range of 0.6. Temporal contiguity principle: Learning is deeper from a multimedia message when corresponding animation and narration are presented simultaneously rather than successively (Mayer, 2005c). Effect sizes in the range of 1.3.
Use of lay language, first and second person references, and use of math-specific language	An informal delivery style was adopted rather than a technical style. The use of technical math language was minimal. Typically, the math terminology was introduced only after the math concept was presented in lay language. The presentation was of the form: "… and in math we call this the commutative property of addition." First and second person language was used as well.	Personalization principle: Learning is deeper when the words in a multimedia presentation are in conversational style rather than formal style (Mayer, 2005d). Effect sizes in the range of 1.3.

Key instructional design property	Implementation	Multimedia principle
Instruction centered around worked examples	All instruction was centered around the use of worked examples that demonstrated the application of the particular math concept. The initial expression or equation was given, and the intermediate steps of the solution were shown (with use of cues) and explanatory narration.	Worked-out example: Consists of a problem formulation, solution steps, and the final solution (Renkl, 2005).
Target low knowledge learners	Our target student population was low knowledge students.	Prior knowledge principle: Instructional strategies that help low knowledge individuals may not help or may hinder high knowledge learners (Mayer, 2001). Effect sizes in the range of 0.6.
Subgoal chunking and labeling	Instruction was chunked to emphasize conceptually meaningful chunks of a problem's solution or subgoals.	Grouping of solution steps by goals and methods, and explicitly labeling the chunks as *goals* and *method*, particularly across different problems with different surface features but with the same underlying solution structure, promotes problem solving in low knowledge learners (Catrambone, 1998).
Multiple examples	Instruction emphasized explicit comparison of problems, use of multiple examples with different surface forms but the same structural form. Instruction also directed attention to both abstract and concrete commonalities among the math problem examples.	Structure emphasizing guideline: Instruction that emphasizes the structural features that are relevant with respect to the selection of the correct solution procedure.
Knowledge of results during practice and explanatory feedback tailored to participants' selection	The content of the feedback differed depending on learners' performance. The feedback is whether the selection was correct or incorrect. If the response was incorrect, the feedback briefly explained why the response was incorrect and provided a suggestion on what to consider. Access to a portion of the original video was also provided to specifically address the goal of the problem, the reason why operations were performed, how to execute the procedure, and any common errors associated with the problem.	Knowledge of results and explanatory feedback promote learning (Bangert-Drowns, Kulick, & Morgan, 1991; Kluger & DeNisi, 1996).

Appendix B:
Student Background Detail

Demographics. One hundred fifteen participants were recruited from three teachers at an urban middle school in southern California. The participants were drawn from two sixth-grade algebra readiness classes (n = 54) and three eighth-grade Algebra 1A classes (n = 52). Fifty were males and 48 females, and there were 34 Latino, 27 Asian American, 19 White, 14 African American, and four biracial. Of the remaining participants, nine did not report grade information, 15 did not report gender information, and eight did not report ethnicity information.

Self-reported achievement. Ninety-seven percent of the students reported understanding spoken English well always or most of the time, and all students reported reading English well. Seventy-three percent of the students reported receiving A's or B's in math, 16% receiving C's, and 4% receiving D's. Ninety-eight percent of students reported being able to understand their teacher's explanations in math class, and 94% agreed that they were able to read and understand most of the problems and explanations in their math book. When asked how knowledgeable they were about pre-algebra, 83% of students reported being moderately or very knowledgeable.

Self-reported math myths. Students held a variety of positive and negative notions about math. For example, 83% of students disagreed with the idea that there is only one way to get an answer; 93% disagreed that if a math problem took longer than 10 minutes, it is impossible; 88% disagreed with the idea that math is too hard for most people to learn; 97% agreed that most math problems could be solved given enough time; and 97% disagreed with the idea that only geniuses are capable of understanding formulas and equations. On the other hand, 35% of students perceived math as just plugging numbers into formulas, 43% of students agreed that there is a math gene that some people have and others don't, and 50% of students agreed that math is mostly memorizing.

Appendix C:
Student Background Measures

Math myths. We adopted a scale from Amdahl and Loats (1995) to measure participants' beliefs about math. The items in this scale were:

When solving problems, there is only one way to get the answer.

If a math problem takes more than 5 or 10 minutes, it is impossible.

Math is just plugging numbers into formulas.

There is a math gene some people have and others don't.

Math is hard—too hard for most people to learn.

Math is mostly memorizing.

With enough time, most math problems can be solved.

Only geniuses are capable of understanding formulas and equations.

Only geniuses are capable of creating formulas and equations.

Participants were instructed to indicate for each item, on a 4-point Likert scale (1 = *strongly agree,* 2 = *agree,* 3 = *disagree,* and 4 = *strongly disagree*), how much they agreed with the statements about math. Cronbach's alpha was .72 (n = 94).

Interest in mathematics. We adopted a scale from Marsh, Hau, Artelt, and Baumert (2006) to measure participants' interest in math. The items in this scale were:

When I do mathematics, I sometimes get totally absorbed.

Because doing mathematics is fun, I wouldn't want to give it up.

Mathematics is important to me personally.

Participants were instructed to indicate for each item, on a 4-point Likert scale (1 = *disagree,* 2 = *disagree somewhat,* 3 = *agree somewhat,* and 4 = *agree*), how much they agreed with the statements about math. Cronbach's alpha was .63 (n = 104).

Self-concept in mathematics. We adopted a scale from Marsh et al. (2006) to measure participants' self-concept in math. The items in this scale were:

I get good marks in mathematics.

Mathematics is one of my best subjects.

I have always done well in mathematics.

Participants were instructed to indicate for each item, on a 4-point Likert scale (1 = *disagree*, 2 = *disagree somewhat*, 3 = *agree somewhat*, and 4 = *agree*), how much they agreed with the statements about math. Cronbach's alpha was .80 (n = 103).

Student survey. We administered a background survey to gather information on participants' age, grades in math, perceived level of knowledge in pre-algebra, and language. We also asked students about their experience with the software and asked them for comments about how the system could be improved.

Appendix D:
Sample of Pretest Items

For each statement below, circle whether the statement is T (true) or F (false). Circle DK (don't know) if you are unsure.

		T	F	DK
1.	To compute $\frac{3}{6}+\frac{4}{4}$, you need to make the denominators equal to 10	T	F	DK
2.	$\frac{12}{1}\times\frac{1}{12}$ has the same value as $\frac{12}{144}$	T	F	DK
3.	$a\times(2\times3)$ has the same value as $(a\times2)\times3$	T	F	DK
4.	$\frac{2b}{2b}\times\frac{1}{2}$ has the same value as $1\times\frac{1}{2}$	T	F	DK
5.	$12(3+4+1)$ has the same value as $12(1+3+4)$	T	F	DK
6.	$12\times(-1)$ has the same value as -12	T	F	DK
7.	$\frac{12}{3}-\frac{2}{3}$ has the same value as $\frac{12-2}{3}$	T	F	DK
8.	$\frac{12a}{12a^2}$ has the same value as $\frac{0}{a}$	T	F	DK
9.	To compute $\frac{12}{3}+\frac{2}{5}$, you need to make the denominators equal to 15	T	F	DK
10.	$-5\times(3\times8)$ has the same value as $(-5\times3)\times8$	T	F	DK
11.	$12\times\frac{3}{4}$ has the same value as $\frac{3}{4}\times12$	T	F	DK
12.	$12\times3\times4$ has the same value as $3\times12\times4$	T	F	DK
13.	$\frac{\frac{42}{70}}{\frac{70}{70}}$ has the same value as $\frac{42}{70}\times\frac{70}{70}$	T	F	DK

Figure D1. A sample of items related to basic math facts.

An equation is given below. For each statement below, circle whether the statement is T (true) or F (false). <u>Circle DK (don't know) if you are unsure.</u>

$$4x = 36 \quad \text{(step 1)}$$

51.	$\frac{4x}{4} = \frac{36}{4}$ is a possible NEXT STEP after step 1.	T	F	DK
52.	$\left(\frac{1}{4}\right)4x = \left(\frac{1}{4}\right)36$ is a possible NEXT STEP after step 1.	T	F	DK
53.	$\frac{4x}{4}$ has the same value as x .	T	F	DK
54.	$\frac{4x}{4}$ has the same value as $\left(\frac{1}{4}\right)4x$.	T	F	DK

Figure D2. Multiple true-false format using a single step equation.

Two steps in solving the equation below are given. For each statement below, circle whether the statement is T (true) or F (false). <u>Circle DK (don't know) if you are unsure.</u>

$$4x = 36 \quad \text{(step 1)}$$

$$\left(\frac{1}{4}\right)4x = \left(\frac{1}{4}\right)36 \quad \text{(step 2)}$$

61.	To solve the problem, you need to isolate x .	T	F	DK
62.	$1 \bullet x = \left(\frac{1}{4}\right)36$ has the same value as $x = \left(\frac{1}{4}\right)36$.	T	F	DK
63.	$1 \bullet x = \left(\frac{1}{4}\right)36$ is a RESULT of carrying out step 2.	T	F	DK
64.	$\left(\frac{1}{4}\right)4x = \left(\frac{1}{4}\right)36$ has the same value as $x\left(\frac{1}{4}\right)4 = \left(\frac{1}{4}\right)36$.	T	F	DK

Figure D3. Multiple true-false format involving fractions and multiple steps in the prompt.

Appendix E:

Transfer Test Items

Adding and multiplying fractions transfer items

Evaluate $\frac{2}{4}+\frac{3}{2}$

Evaluate $\frac{1}{10}+\frac{1}{2}+\frac{2}{5}$

Evaluate $\frac{8}{12}\times\frac{3}{4}$

Evaluate $\frac{a}{4}\times\frac{2b}{3}$

Reduce $\frac{6}{32}$ to its simplest form.

Find the prime factors of 56.

Rational number equivalence explanation transfer items

(a) Explain how to find fractions that are equivalent to $\frac{2}{3}$.

(b) Show an example of how to use your method to find three fractions that are equivalent to $\frac{2}{3}$.

Scoring rubric:

Part a:

Score point	Description
3	The student mentions all of the following: • you multiply (or divide) the fraction by a/a (as long as a ≠ 0)* or you multiply (or divide) the numerator and the denominator by the same # (as long as it is not 0)* • doing the above is the same thing as multiplying by 1 • and so the number does not change. *Note: The student should not be penalized if they do not include the statement about the number being non-zero. *Note: The student should not be penalized if they do not include the statement about the number being non-zero.
2	The student mentions that you multiply (or divide) the top and bottom number by the same non-zero* number. *Note: The student should not be penalized if they do not include the statement about the number being non-zero.
1	The student unclearly expresses the idea of manipulating the top and bottom numbers by the same number, or by something, e.g. student might say "multiply the fraction by 5." Or "you do the same thing to the top and bottom."
0	The student gives either no explanation or an incorrect explanation.

Part b:

The student scores a maximum of 4 points—one for each correct equivalent fraction given (no credit given for rewriting 2/3) and one point for showing how they found the equivalent fractions (1 point total for showing any work). If more than 3 equivalent fractions are given, the student does not score additional points. If no equivalent fractions are given, the student scores 0 for part B.