

Designing online software for teaching the concept of variable that facilitates mental interaction with the material: systemic approach

Natalya A. Koehler · Ann D. Thompson · Ana-Paula Correia ·
Linda Serra Hagedorn

Published online: 11 October 2014

© Association for Educational Communications and Technology 2014

Abstract Our case study is a response to the need for research and reporting on specific strategies employed by software designers to produce effective multimedia instructional solutions. A systemic approach for identifying appropriate software features and conducting a formative evaluation that evaluates both the overall effectiveness of the multimedia instructional design and the effectiveness of specific features of the software is presented. The instructional software for teaching the concept of variable was designed as a research platform and tested with 90 undergraduate students at a Midwestern university. Behavior tracking and data collection instruments (pre-test, surveys, and delayed post-test) were embedded in the software. As a part of *design-engineering-develop* approach, two potential types of feedback, *single try* versus *two tries*, were tested in two experimental conditions. The results of the formative evaluation demonstrated preliminary evidence of the effectiveness of the designed software with either type of feedback for both high and low prior knowledge students. An innovative instructional strategy of helping the learner mindfully process the program feedback is described.

Keywords Formative evaluation · Instructional design · Web-based interactive multimedia · Computer-assisted instruction · Software development · Design experiment

N. A. Koehler (✉)

Instructional Design Faculty, International Institute for Innovative Instruction, Franklin University,
236 Town Engineering St., Columbus, OH 43215-3875, USA
e-mail: natalya.koehler@franklin.edu

A. D. Thompson

University Professor Emeritus, School of Education, N108 Lagomarcino Hall, Ames, IA 50011, USA

A.-P. Correia

School of Education, N165b Lagomarcino Hall, Ames, IA 50011-3198, USA

L. S. Hagedorn

Associate Dean of Undergraduate Programs, Educational Leadership & Policy Studies (ELPS),
Iowa State University, E262 Lagomarcino, Ames, IA 50011-3195, USA

Introduction

Need for computer assisted instruction for teaching basic statistics concepts

One of the key concepts to research methodology is the variable concept. Students' fail or demonstrate slow progress in research methods and basic statistics courses if they have problems identifying types of variables. The following remark comes from an experienced basic statistics course instructor: "I suspect that types of variables in statistics are tough for students to grasp. Rereading the standard definitions, they don't seem to make any sense unless they know first what a variable is." An effective low cost online instructional program that students could use at their convenience for bringing them up to speed with the variable concept would benefit them.

Historically, computer assisted statistics instruction (CAI) has a potential of increasing student engagement with course content because it allows interactivity and self-paced learning (Larreameydy-Joerns et al. 2005). The software for teaching statistics that is used in the classroom varies from software demonstrations and occasional interactive exercises to completely Web-based courses (Symanzik and Vulkasinovic 2006).

Despite the fact that the use of CAI in statistics education has grown to 80 % among universities, research about its effectiveness in teaching statistics shows controversial findings (Larwin and Larwin 2011). The authors found in their meta-analysis that CAI has a moderate impact on student achievement and that "it is not a panacea for all the ills that might plague statistics education" (p. 272), but it can serve as a very beneficial enhancement or supplement to small class sections (Larwin and Larwin 2011).

According to some scholars, technology needs to be used not only for computing numbers but for concept exploration, which would enhance student learning chance (Moore 1997; Friel 2007; Garfield et al. 2000). Chance et al. (2007) suggest that technology has a potential to expand the range of visualization techniques that can help learners understand concepts.

Sklar and Zwick (2009) in their paper provided recommendations for designing animated presentations in statistics and instructional approaches for teaching the material to non-statistical majors. One of their recommendations is segmenting animated presentations into scenes rather than using one single continuous presentation. Also, the authors emphasize the need for embedded questions that need to be provided during the presentation to assess viewer's understanding of the concept. Finally, the authors contend that more empirical research is needed to understand the best methods for teaching statistics to non-statistical majors in multimedia environments.

Wender and Muehlboeck (2003) investigated whether the computer-animated graphics were more effective than static images in teaching statistics. They found that student scores on retention and understanding of the concepts presented were significantly higher when the animated graphics were used compared to static graphics. Four statistical concepts were presented and explained to students in class. The presentations included graphics either in static or in animated form. The authors suggested that animations designed for teaching concepts should focus on functional relationships of sub-concepts. The interdependencies among sub-concepts need to be presented explicitly (Wender and Muehlboeck 2003).

"Given the increased emphasis on stand-alone tools and distance learning, research on the role of interactivity, engagement, and feedback takes on increased importance as educators continue work on improving the efficacy of technology-based statistics instruction." (Sosa et al. 2011, p. 122). To address this need, a low-cost online software for teaching the concept of variables (referred here as the Program) was designed and

evaluated for its overall effectiveness. The Program was designed as a research platform for testing and evaluation of the implemented program features and instructional design decisions.

Systemic approach to design engineering of the program

Many researchers noted the need for research on specific design features and strategies in computer assisted instruction. In their meta-analysis, Sosa et al. (2011) suggested that research on computer-assisted instruction in statistics would benefit from closer collaboration on measurement and design across researchers. They emphasized the importance of research focused on specific features of interest.

Ormel et al. (2012) stated the need for research on specific strategies that were used to facilitate the production of new theoretical understanding through design of instructional solutions. Bernard et al. (2009) in their meta-analysis underscored the importance of studies that focus on the specific features of interest. Interactions of design features with learner characteristics and more sensitive approaches to test the impact of learner-based features in computer-assisted instruction were identified as the key aspects of future research.

While conducting this study, we were working systematically and simultaneously toward the dual goals of solution development and theoretical understanding by studying what happened when those solutions came to life in real classrooms (McKenney and Reeves 2012). This case study uses a systemic approach for using existing knowledge to construct a web-based instructional solution and conducting a series of formative evaluations of the implemented solution, which will ensure the effectiveness of the developed product. The following systematic approach for incorporating the existing knowledge in the Program design was implemented.

- Identifying the benefits of multimedia and instructional methods that can potentially contribute to the effectiveness of stand-alone instructional software
- Defining potentially effective Program features and instructional design decisions in regards to the learning task (teaching the concept of variable) and the target population (low and high prior knowledge students in an Educational Psychology class and two Basic Statistics classes).
- Implementing the evidence-based research findings in the design of the Program.
- Building the Program as a research platform that allows online data collection about program use and student progress.

The following systematic approach to the evaluation of the Program was used.

Evaluation 1 (see Fig. 1).

- Evaluating the overall effectiveness of the Program by comparing student performance in each of the experimental conditions to no-treatment condition.
- Testing possible solutions for the key program features in experimental conditions.
- Collecting information about other potentially effective program features and instructional strategies through surveys. The findings serve as rationale for further design experiments in the second round of the evaluation series, if needed (Evaluation 2, 3 etc.)
- Collecting and analyzing the above data in regards to the student level of prior knowledge.

- Based on student perceptions of other program features (see the Evaluation 1) identifying the features (if any) to be tested in the experimental conditions in Evaluation 2.
- Collecting information about other potentially effective program features and instructional strategies through surveys. The findings serve as rationale for further design experiments in the third round of the series of evaluations, if any.
- Collecting and analyzing the above data in regards to students' level of prior knowledge.

Literature review

Knowledge sources informing the design of instructional strategies in the program

On one hand, for the software to be effective, instructional principles must be consistent with what is known about how people learn (Mayer 2008). On the other hand, multimedia have the potential of promoting meaningful learning by varying both the number of representations provided to students and the degree of student interactivity (Moreno and Valdez 2005). The literature review was conducted keeping the above aspects of multimedia in mind.

Compare and contrast strategy for teaching concepts

Concept learning can be facilitated by the instructor (or software) that presents examples and non-examples and through having students solve problems that facilitate comparison of the defining features of concepts. According to Richards and Godfarb (1986), "concept reasoning can be mastered through central tendency information, logical rules or single episodes depending upon which of these is activated in a particular task situation."

Tennyson and Cocchiarella (1986) in their instructional design model stated two basic relationships between concepts: successive and coordinate. They recommended that when teaching concepts defined as coordinate, an opportunity to compare and contrast from one coordinate concept with examples from another coordinate concept allows students to better develop the discrimination skill. Thus coordinate concepts need to be presented simultaneously (Tennyson and Cocchiarella 1986).

Litchfield (1987) recommends teaching a set of concepts presented simultaneously. Even though the attributes of different concepts are easily confused, this kind of presentation makes students compare and contrast similarities and differences between concepts and helps them clarify individual concepts (Litchfield 1987).

If the limited capacity of student working memory is taken into consideration, student can manipulate no more than four information elements at one time (Miller 1956). Thus, presenting no more than four concepts at one time can potentially benefit students.

Types of visuals

Previous research comparing the learning outcomes of student learning with text and illustrations versus animation and narration has produced inconsistent and controversial findings. On one hand, several studies have suggested that learning is enhanced in

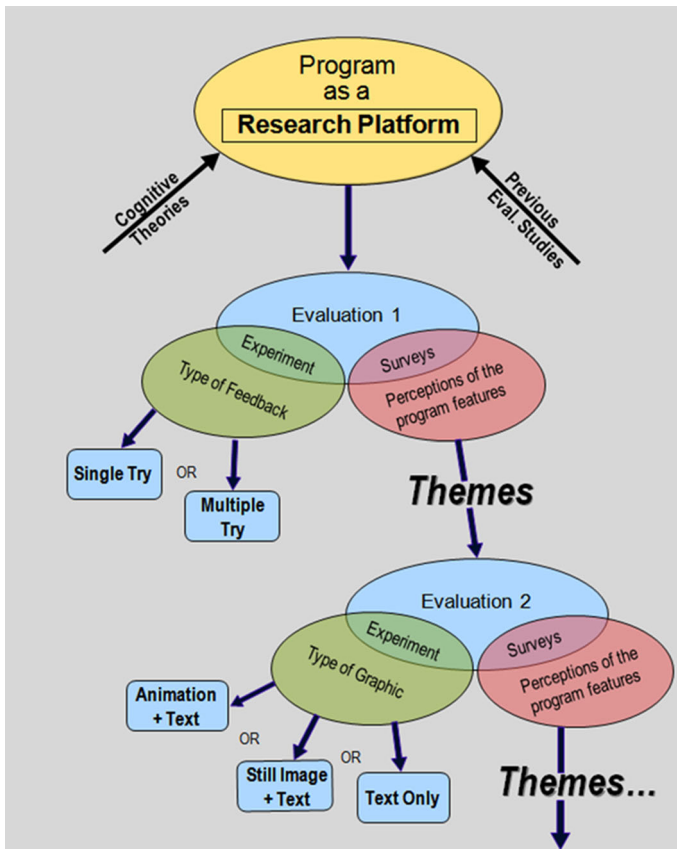


Fig. 1 Systemic approach for designing effective instructional multimedia. *Evaluation 2* (see Fig. 2)

computer-based animation environments (Park 1994; Tversky et al. 2002). For example, learners that have limitations in spatial ability benefit from animations because they may have problem mentally animate how a complex system works from a series of static diagrams (Hegarty and Sims 1994).

Some empirical results have indicated that animation is superior to static images in terms of retention and transfer of information (Mayer and Moreno 2002; Craig et al. 2002; Moreno et al. 2001). Mayer and Moreno (2002) stated the importance of principles of cognitive theory of multimedia learning to underlie animations design. Such design has a potential of promoting learner understanding. Gulz and Haake (2006) found that animations have the potential to make the learning experience more engaging. Animation appears to be most effective when presenting concepts of information students may have difficulty visualizing (Betrancourt 2005; Narayanan and Hegarty 2002).

On the other hand, in many studies dealing with abstract, scientific or technical content, animation did not turn out beneficial compared to static pictures (Tversky et al. 2002). Lowe (2003) showed that low prior knowledge students are often more focused on perceptually salient rather than thematically relevant features of animation. Clark and Mayer (2007) recommended using static illustrations unless there is a compelling instructional

rationale for animation. Hasler et al. (2007) noted possible disadvantages of animations if not designed properly. Animation may impose greater cognitive processing demands than static visuals when critical objects and their relations disappear during the animation.

The results obtained for learner-controlled pacing and segmentation of animation have produced inconsistent findings. Whereas some researchers observed positive effects of pacing and segmentation (Mayer et al. 2003), others found that system-paced instruction was more beneficial for learning (Tabbers et al. 2004). Plotzner and Lowe (2004) stated that animations are frequently used for attention gaining effect, and we know little about how animation needs to be designed in order to facilitate learning. One of the suggestions for building effective animations for instructional purposes is constructing them in ways that tap the positive features of static illustrations (Mayer et al. 2005).

Program feedback

Computer-based interactive learning environments support the learner engagement in an interactive learning exchange. Computer-generated feedback can potentially support this interaction because it can provide learners with information they may use to correct errors (Valdez 2012). The effects of different types and forms of informative feedback have been investigated in multiple instructional contexts and provided inconsistent findings (see the reviews by Azevedo and Bernard 1995; Bangert-Drowns et al. 1991; Butler and Winne 1995; Clariana 1993; Mason and Bruning 2001; Mory 1992, 1996, 2004). Verification (correct/incorrect) feedback has not shown much effect at promoting learning (Bangert-Drowns et al. 1991; Moreno 2004).

Most researchers believe that the program feedback that facilitates the greatest gains in learning must include both verification and elaboration. This combination strengthens student correct responses and is more effective than simply providing correct/incorrect feedback. (e.g., Bangert-Drowns et al. 1991).

As to the optimal feedback timing, there is substantial disagreement among researchers on whether the feedback should be given immediately after answering each test question (Keller 1983) or it should be delayed, which will give errors a chance to dissipate (Kulhavy 1977; Kulhavy and Anderson 1972; Kulhavy and Stock 1989). According to Keller's recommendations, formative (corrective) feedback needs to be provided when it will be immediately useful (Keller 1983).

In contrast, Butler and Roediger (2008) found that the delayed feedback benefited students when given soon after the test but not immediately after answering each item. A meta-analysis of 53 studies conducted by Kulik and Kulik (1988), showed that the studies that used actual classroom materials and quizzes revealed a better effectiveness of the immediate feedback compared to delayed feedback. On the contrary, in laboratory studies, delayed feedback demonstrated better effectiveness.

Another important aspect of feedback, the number of steps, was addressed by Spector et al. (2008). They emphasized the need for research on individual and situational conditions regarding the number of feedback steps and cycles. Clariana (1993) in his review of 30 studies compared single try feedback types (immediate knowledge of result, immediate knowledge of correct response, delayed feedback, no- feedback) to multiple try feedback and did not find any differences between single-try and multiple-try feedback. In regards to student prior knowledge differences, multiple tries are most effective with high prior knowledge students, and a single attempt with the correct answer feedback is the most effective with low prior knowledge students (Clariana 1993).

Program design based on the evidence-based findings from the previous research teaching the concept of variables

The Program was designed as a set of 20 scenarios, “extensive network of episodes involving the concept” (Richards and Godfarb 1986, p. 34). Students needed to identify independent variable (IV), dependent variable (DV), controlled variable or constant (CV), and the levels of independent variables (LIV). Also, theory explanation pop-ups (brief explanations of types of variables and examples of their use) were available on each screen for the learner to develop their concept reasoning based on “central tendency information, logical rules” (Richards and Godfarb 1986, p. 34).

The number of types of variables to discriminate from, increased gradually through the Program: 2 variables in scenario 1, 3 variables in scenarios 2–8, and 4 variables in scenarios 9–20. A screenshot of a problem scenario is presented in Fig. 2.

Visuals

The animations on the screens end as still images showing the completed state of the process. This strategy allows us to compensate for the fact that human perception is limited by the speed of neural networks while processing the temporarily changing information (Betrancourt 2005). The use of animations is not excessive. They emphasize the functional relationships between variables and make them explicit by portraying interdependencies (Wender and Muehlboeck 2003). Pointing arrows and highlights of important information are used to compensate for the fact that low prior knowledge students are often more focused on perceptually salient rather than thematically relevant feature of information (Lowe 2003).

In order for us to be able to compare and decide on the type of visuals to implement in the final version of the Program, we included three formats of presenting problem scenarios in the program design.

- Text scenario augmented with animation showing the functional relationships between variables (8 problems)
- Text scenarios augmented with still images (8 problems)
- Text only scenarios (4 problems)

Program feedback

The design of the program feedback (see Fig. 2) included the following features:

- (1) The Program uses response-contingent feedback to provide both verification (knowledge of result-KR- feedback presented as smiley faces) and item-specific elaboration (elaborative feedback-EF- presented as text) that explains the answer. The decision was made based on recommendations to use immediate feedback for retention of procedural or conceptual knowledge (Shute 2008; Kulik and Kulik 1988).
- (2) Students receive feedback regardless of the correctness of their response so that the learner can develop understanding when a correct answer was simply a guess.
- (3) The strategy of engaging the learner in the mindful processing of program feedback was implemented.

LearnScat Training Problem 7 of 20

Rule of Thumb
 People often confuse independent variables with dependent variables. This sentence will help you avoid confusion.
 Try to put your variables in this sentence.
 Independent variable(s) causes a change in (Dependent Variable) and it isn't possible that (Dependent Variable) could cause a change in (Independent Variable).
 For example: (Time Spent Studying) causes a change in (Student GPA), and it isn't possible that (Student GPA) could cause a change in (Time Spent Studying).

Choose One
 Dependent (DV)
 Independent (IV)
 Controlled (CV)
 Level of IV
 I want to know

Variable Type

Variable	Units	Variable Type
1. Type of plant - Tulips vs. Daisies		Dependent (DV)
2. Amount of water	Gallons	Controlled (CV)
3. Amount of sunlight	Foot candles	Controlled (CV)
4. Plant growth	Inches	Dependent (DV)

In this study, the plant growth depends on the type of plant (tulips vs. daisies). The other two variables, the amount of sunlight and the amount of watering (2 cans), are held constant for the left and right side of the flower bed.
 Answers: 1-independent variable (IV), 2-controlled variable (CV), 3-controlled variable (CV), 4-dependent variable (DV).

Fig. 2 A screenshot of a problem scenario

The types of variables in the training were manipulated in such a way, that an item classified as an independent variable in problem scenario A became a controlled variable (constant) in scenario B. This strategy put students in a situation where the probability of their correct answers was low. According to Kulhavy and Stock's certitude model of feedback (Kulhavy and Stock 1989), the learners who are informed that their answer is wrong when they are confident that their answer is correct would "exert much effort to find out what was remiss in their thinking" (Mory 2004, p. 749). This strategy makes the learner mindfully engage in the feedback content and pay more attention to the defining versus non-defining features of variables.

Research hypotheses

Hypotheses regarding overall effectiveness of the program

Hypothesis 1a The Program will facilitate retention of the concept of variable, and thus the average knowledge gain in the control condition will be statistically significantly lower than in each of the experimental conditions (*single try* condition or *two tries* condition).

Hypothesis 1b The average knowledge gain difference between *single try* condition and *two tries* condition will be statistically significant. Since the only difference between the experimental conditions is feedback type, the higher knowledge gains would indicate the more effective type of feedback for the Program.

Hypotheses regarding the effect of feedback type and student level of prior knowledge on knowledge retention

Hypothesis 2a There will be a statistically significant difference in knowledge gain between low prior knowledge (LPK) students in the *single try* condition and the *two tries* condition.

Hypothesis 2b There will be statistically significant difference in knowledge gain between high prior knowledge (HPK) students in the *single try* condition and the *two tries* condition.

Hypotheses regarding the effect of student level of prior knowledge on student perceptions of program helpfulness

Hypothesis 3a There will be statistically significant difference in student ratings of their satisfaction with the Program between the LPK students in the *single try* condition and the *two tries* condition.

Hypothesis 3b There will be statistically significant difference in student ratings of their satisfaction with the Program between the HPK students in *single try* condition (C 2) and *two tries* condition.

Hypotheses regarding the effect of problem presentation format on student perceptions of program helpfulness

Hypothesis 4a, 4b, and 4c There will be statistically significant difference in student survey ratings on how the formats of problem presentation (animation plus text, still image plus text, text-only) helped them remember the key concepts (hypothesis 4a), understand them (hypothesis 4b), and maintain their attention (hypothesis 4c).

Student perceptions of instructional strategies

Which instructional strategies did the students find the most and the least helpful? For this question, descriptive statistics were used for the analysis of the student survey ratings.

Methods

Participants

Participants were 90 undergraduate students (27 male, 63 female) of a Midwest university, USA (average age: males: 20.0; SD = 2.00, females: 21.0; SD = 7.13). The groups came from two Basic Statistics courses for non-statistics majors and an Educational Psychology course. Students were not required to perform at any criterion level on any measure given during the experiment in order to receive course credit.

Treatments

Hypotheses 1a and 1b Three comparison groups (Condition 1, Condition 2, and Condition 3) consisted of 30, 29, and 31 students correspondingly. Condition 1 was a control group (no-treatment condition); Condition 2 and Condition 3 were experimental conditions. The dependent variable was student knowledge gain measured as a difference between the pre-test and delayed post-test (5 days after the training). The independent variable was program feedback type: *single try* in Condition 2 and *two tries* in Condition 3. The only difference between experimental conditions was feedback type, *single try* versus *two tries*. *Single try* feedback consisted of the knowledge of result-KR-feedback presented

as green or red smiley faces (correct or incorrect response correspondently). *Two tries* feedback presented to the student the correct/incorrect status on the first try, and both the correct/incorrect status and the explanatory feedback (explanation of the correct answer) on the second try (See Fig. 2).

The comparison of student knowledge gain between the Condition 1 and Condition 2, and between the Condition 1 and the Condition 3 was done for testing the overall effectiveness of the Program (Hypothesis 1a). The comparison of student knowledge gain between two experimental conditions was done to determine the most effective type of feedback, *single try* or *two tries* (Hypothesis 1b).

Hypotheses 2a and 2b We conducted a factorial experiment that followed a 2×2 design. The between subjects factor was the type of feedback, *single try* or *two tries*. The within subjects factor was the level of student domain specific prior knowledge, low or high. We compared the LPK student knowledge gain and the HPK student knowledge gain in two experimental conditions, Condition 2 (LPK, $n = 15$; HPK, $n = 14$) and Condition 3 (LPK, $n = 17$; HPK, $n = 14$).

Hypotheses 3a and 3b The factorial experiment that followed a 2×2 design was conducted. The between subjects factor was the type of feedback, *single try* or *two tries*. The within subjects factor was the level of student domain specific prior knowledge, low or high. We compared the average LPK student ratings of their satisfaction with the program with the average HPK student ratings in two experimental conditions, *single try* condition (LPK, $n = 15$; HPK, $n = 14$) and *two tries* condition (LPK, $n = 17$; HPK, $n = 14$).

Hypotheses 4a, 4b, and 4c Each of the three comparison groups consisted of the LPK and the HPK students. The independent variable was the format of a problem scenario presentation: animation plus text, still image plus text, text-only. The dependent variable was the student ratings of the following survey items: how the Program helped them remember the key concepts (Hypothesis 4a), understand the key concepts (Hypothesis 4b), and maintain their attention (Hypothesis 4c).

Student perceptions of instructional strategies

The helpfulness of the instructional strategies implemented in the Program (learning from theory explanation pop-up and learning from program feedback) was rated by both the LPK and HPK students in both experimental conditions ($n = 60$). Descriptive statistics were used for the analysis of the obtained data.

Instruments

The data collection instruments for each condition are presented in Table 1.

Tests

Student learning performance was assessed with a pre-test taken before the training and delayed post-test taken on the 5th day after the trading. The same test was used for the pre-test and post-test. Each test item was designed as a problem scenario. In each scenario, participants were asked to select five choices from five dropdown menus: independent variable (IV), dependent variable (DV), controlled variable (CV), level of independent variable (LIV), and “I want to know”. The fifth choice “I want to know” was added to

Table 1 Experimental conditions and data collection instruments

Condition 1: no-treatment condition	Condition 2: experimental condition	Condition 3: experimental condition
<ul style="list-style-type: none"> •Pre-test •Delayed post-test (5 days delay) 	<ul style="list-style-type: none"> •Pre-test •Training episode with <i>single try</i> feedback •Survey 1 (taken immediately after the training) •Delayed post-test (5 days delay). The same as pre-test. •Survey 2 taken immediately after the post-test. It consisted of the first two items of survey 1. 	<ul style="list-style-type: none"> •Pre-test •Training episode with <i>two tries</i> feedback •Survey 1 (taken immediately after the training) •Delayed post-test (5 days delay). The same as pre-test. •Survey 2 taken immediately after the post-test. It consisted of the first two items of survey 1.

avoid random answers (see Fig. 3). Each correct answer was scored as one point. The maximum score was 50 points.

Each test item was designed to test student knowledge of defining features of independent, dependent, controlled (constant) variables and the level of independent variable. In each test item, students were asked to differentiate between the independent, dependent, controlled (constant) variables and the level of independent variable. The problem scenarios came from various fields: science, education, health care, nutrition, business, engineering, cooking etc. This strategy was used to support future knowledge transfer of the acquired concepts to a variety of contexts. The test was found to be highly reliable (10 test items, 5 questions in each test item; $\alpha = .84$).

Surveys

Students' perceptions on the overall effectiveness of the Program and individual program features were collected through Survey 1 (Post-training Survey) taken immediately after the training (see Fig. 4 in Appendix).

Survey 1 consisted of 11 items that measured student opinion on program features and overall experiences with the program. The rating scale of 5 (5- strongly agree, 1-strongly disagree) was used. The survey was found to be reliable (11 survey items, 3–6 questions in each survey item; $\alpha = .77$).

Survey items 1–3 allowed the researchers to collect data about how specific instructional strategies (learning through explanatory feedback, learning by solving problems, learning through theory explanations) and specific formats of the problem scenario presentations (animation plus text, still image plus text, text only) helped students learn about types of variables. The data were collected in regards to the following types of information processing:

- Helped the student recall the information learned through the Program (remembering)
- Helped the student identify the variable type (understanding)
- Helped the student maintain their attention

Survey items 4–6 allowed the researchers to collect student perceptions of the overall effectiveness of the Program. Survey items 7–11 helped the researchers understand how individual program features were perceived by students to make sure that none of the features were redundant (received low ratings).

Survey 2 (Follow up Survey) was administered immediately after the post-test (5 days after the training episode). The Survey 2 was a shortened version of the Survey 1 with only two first items included. The assumption was that the difference in the ratings between the surveys would indicate how well the program features helped students retain the information. For example, if the compared survey ratings went up or stayed the same, this could be an indication that the Program helped the students retain the information.

Procedure

The whole evaluation was done online, including the sign up process, pretest, assignment to conditions, actual training episode, post-training survey (Survey 1) administered immediately after the training, delayed post-test (5 days after the training), follow up survey (Survey 2) administered immediately after the delayed post-test. Students were

LearnStat Pre/Post Test (DEMO mode) to Menu Problem 4 of 10
(DEMO mode) Next Problem

Identify the variables by selecting the answers from the drop-down menus.

Can blueberries slow down aging? In this study, 15-month, 18 month and 20-month-old rats were assigned to three groups respectively. There were 100 rats in each group. They were fed their standard diet or a diet supplemented with blueberry, strawberry, or spinach powder. After 10 weeks, the rats were given memory and motor tests.

184

Variable	Units	Variable Type
1. Diet type (Standard diet-SD/SD+blueberries/SD+strawberries/SD+spinach powder)		Choose One
2. Age of rats	Months	Choose One
3. 20-month-old rats		Choose One
4. Memory test	Test score	Choose One
5. Motor skills test	Test score	Choose One

Save Answers

- Choose One
- Dependent (DV)
- Independent (IV)
- Controlled (CV)
- Level of IV
- I want to know

Fig. 3 A screenshot of a test item

allowed to participate in the study at their convenience over the time period of 3 weeks. The only requirement was to follow the timeline.

The instructors in three undergraduate courses asked their volunteer students to participate. As part of the recruitment process, students were asked not to use any instructional materials for studying types of variable other than the Program training (20–40 min) before and during the research study.

There was no formal training on the concept of variable for the participants neither before the training nor during the time period between the training and the delayed post-test. The participants were reminded by the researchers through emails to take the delayed post-test on the 5th day after the training.

The participants signed into the Program in a random order at their convenience. The following schema was used to assign students to conditions with a goal of having the same proportion of high and low prior knowledge students in each condition. Based on the results of the pre-test, the participants were assigned by the Program the low prior knowledge (LPK) or high prior knowledge (HPK) status. The level of students' prior knowledge was determined by the results of the pre-test. If their total pre-test score was 25 of 50 or lower, they were considered LPK students. If their total pre-test score was 26 of 50 or higher, they were considered HPK students.

As the participants were signing in, they were being assigned sequentially to the Condition 1, 2, 3, 1, 2, 3 and so on. The LPK student who signed in first was assigned to Condition 1, the next LPK student who signed in second was assigned to Condition 2 and so on. The same procedure was used for the students identified by the Program as HPK students. Thus, regardless of the status, LPK or HPK, the participants were assigned to the Condition 1, 2, 3 sequentially.

The participants accessed the Program at home at their convenience. The time on task was collected. The average time spent on the pre-test was 14.0 min (HPK: 14.6 min; LPK: 13.6 min) and 15.0 min (HPK: 17.5 min; LPK: 12.7 min) in the *single try* condition and the *two tries* condition correspondingly.

The average time spent on the *post-test* was 11.3 min (HPK: 10.6 min; LPK: 11.8 min) and 11.9 min (HPK: 12.3 min; LPK: 11.6 min) in the *single try* condition and the two tries condition correspondingly. The average time spent on the training episode was 23.8 min (HPK: 22.4 min; LPK: 24.8 min) and 28.2 min (HPK: 25.5 min; LPK: 30.5 min) in the *single try* condition and the two tries condition correspondingly.

Evaluation 1 results

Data analysis

The data from the ten students were excluded from our statistical analysis because the students did not follow the timeline. Three students took the post-test on the third and fourth day, four students did not do the post-test, and three students took the post-test 6–9 days after the training.

In this study, first we determined the overall effectiveness of the Program in regard to each feedback type. Second, we investigated the overall effectiveness of the Program both in regard to feedback type and student prior knowledge. Next, student perceptions of the overall program design in regard to student level of prior knowledge were analyzed. Third, student perceptions on how the format of problem presentation (animation plus text, still image plus text, text-only) helped them remember and understand the key concepts and maintain their attention were analyzed. Finally, average student ratings of the implemented instructional strategies were determined.

Statistical results

Hypotheses 1a & 1b A one-way between subjects ANOVA was conducted to compare the effect of overall program training on students' knowledge gain between the pre-test and delayed post-test for Condition 1 (C1), Condition 2 (C2), and Condition 3 (C3). Tests of the hypotheses were conducted using Bonferroni adjusted alpha levels of .017 per test (.05/3). There was a significant effect on student knowledge gain [$F(2, 87) = 34.2, p < 0.001$]. The results indicated that the knowledge gain was significantly lower in C1 ($M = 1.54, SD = 2.52$), than it was in both C2 ($M = 14.9, SD = 9.16, d = -1.99$) and C3 ($M = 16.06, SD = 8.58, d = -2.30$). Hypothesis 1a was confirmed. The pairwise comparison of the knowledge gain in C2 and C3 was non-significant [$F(2, 87) = 1.13, p = 0.838, d = -0.128$]. Therefore, Hypothesis 1b was rejected. The knowledge gain between the pre-test and delayed post-test was 30.8 % in C2 and 30.0 % in C3.

Hypotheses 2a & 2b Because of the small number of the LPK students in each experimental condition (15 in Condition 2 and 17 in Condition 3) and the HPK participants (14 in Condition 2 and 14 in Condition 3), non-parametric tests were used to analyze the data about the effects of feedback type within each prior knowledge level. Two two-sample Wilcoxon rank-sum (Mann–Whitney) tests were computed within each prior knowledge level. The results suggested that there was no statistically significant difference between the underlying distributions of the knowledge gain scores between the LPK students in the *single try* condition ($M = 19.67, SD = 9.78$) and in the *two tries* condition C ($M = 19.2, SD = 9.11$) ($Z = 0.189, p = 0.850, d = 0.052$).

As to the HPK students in Condition 2 ($M = 9.86$, $SD = 4.88$) and Condition 3 ($M = 12.29$, $SD = 6.29$), their knowledge gain scores were not significantly different either ($Z = -0.761$, $p = 0.447$, $d = -0.431$).

Hypotheses 3a & 3b Two two-sample Wilcoxon rank-sum (Mann–Whitney) tests were computed within each student prior knowledge level. The results suggested that there was statistically significant difference between the underlying distributions of the average student ratings of their satisfaction with the Program between the LPK students in the *single try* condition ($M = 4.07$, $SD = 0.41$) and in the *two tries* condition ($M = 4.38$, $SD = 0.32$) ($Z = 2.1714$, $p = 0.030$), $d = -0.837$). On average, The LPK students in the *two tries* condition were more satisfied with the program than in the *single try* condition.

As to the HPK students, there was no statistically significant difference in the student ratings between the *single try* condition ($M = 4.29$, $SD = 0.43$) and in the *two tries* condition ($M = 4.54$, $SD = 0.45$) ($Z = -1.447$; $p = 0.147$, $d = -0.588$). In both conditions, the HPK participants were highly satisfied with the Program. The average student ratings of their overall satisfaction with the Program are presented in Appendix Table 3.

Hypotheses 4a, 4b, & 4c A one-way ANOVA was conducted to compare the effects of three formats of problem presentation (animation plus text, still image plus text, text-only) on each of the following dimensions: helping students remember and understand the concept of variables and helping students maintain their attention.

There was a significant effect of a problem presentation format on students ability to remember the information learned through the Program [$F(2, 57) = 17.1$, $p = 0.001$]. The results indicated that the ratings were significantly lower for text-only format ($M = 3.15$, $SD = 0.81$) compared to the text augmented with a still image ($M = 3.74$, $SD = 0.82$) and to the text augmented with animation ($M = 4.08$, $SD = 1.00$). Therefore, Hypothesis 4a was confirmed.

There was a significant effect of a problem presentation format on student ability to understand the information learned through the Program [$F(2, 57) = 16.76$, $p = 0.001$]. The results indicated that the ratings were significantly lower for text-only format ($M = 3.16$, $SD = 0.88$) compared to the text augmented with a still image ($M = 3.77$, $SD = 0.89$) and to the text augmented with animation ($M = 4.13$, $SD = 1.01$). Therefore, Hypothesis 4b was confirmed.

There was a significant effect of a problem presentation format on students ability to maintain their attention while learning through the Program [$F(2, 57) = 51.00$, $p = 0.001$]. The results indicated that the ratings were significantly lower for text-only format ($M = 2.53$, $SD = 1.11$) compared to the text augmented with a still image ($M = 3.70$, $SD = 1.05$) and to the text augmented with animation ($M = 4.48$, $SD = 1.03$). Therefore, Hypothesis 4c was confirmed. The average student survey ratings of problem scenario presentation format by type of cognitive processing are presented in Appendix Table 4.

Student perceptions of instructional strategies

Descriptive statistics were applied for the analysis of the data of student perceptions of the most and least helpful program features. The results are presented in Table 2. The student ratings of the implemented instructional strategies were high regardless of student level of prior knowledge.

Table 2 Means (and SD) of students' survey ratings of their perceptions of the effectiveness of the instructional strategies implemented in the Program (1-strongly disagree, 5-strongly agree)

Instructional strategies	Student level of prior knowledge	
	LPK M (SD)	HPK M (SD)
Feedback	4.35 (0.68)	4.39 (0.63)
Theory explanations	4.01 (0.82)	3.93 (0.83)

Rationale for conducting evaluation 2

The results of the Evaluation 1 indicated the statistically significant differences between student ratings of different formats of problem scenario presentations (animation plus text, still images plus text, and text-only). The differences were found on the following dimensions: how the Program helped students remember the key concepts (hypothesis 4a), understand the key concepts (hypothesis 4b), and maintain student attention (hypothesis 4c). The conclusions were made about the need for further testing of this effect in Evaluation 2 (see Fig. 1). The format of problem scenario presentation will be tested in three experimental conditions in Evaluation 2. The only difference between the experimental conditions will be the presentation format. It will allow the designers to make the final decision on the most effective type of problem scenario presentation. This type will be implemented in the final version of the software. The second round of the evaluation (Evaluation 2) will be conducted according to the plan described in the Introduction section of this manuscript (See Fig. 1).

Discussion

This research paper describes the design and evaluation of the Program for teaching types of variables. The Program was designed as a research platform. Careful consideration of the elements that made the teaching principles used in the Program effective was done by using a systemic approach for the Program design and evaluation.

The results of this evaluation can be used for guiding program design and further evaluation. They cannot be generalized to larger populations because of the small number of participants in the sub-conditions (Condition 2—LPK: 15 students, Condition 2—HPK: 14 students, Condition 3—LPK:17 students, Condition 3—HPK: 14 students) and thus insufficient power to detect small differences.

Based on the results of the Evaluation, the Program can facilitate retention of the variable concept regardless of the feedback type. Five days after the training, students demonstrated a significant improvement in knowledge gain. The results showed that the number of attempts in the program feedback did not have a significant effect on either the LPK or HPK student knowledge gain. In contrast, according to the review of studies by Clariana (1999) that compared AUC (answer until correct feedback) with STF (single try feedback), “For low prior knowledge students, single-try feedback was more effective than multiple-try feedback, $ES = 0.11$ ” (p. 88).

Interestingly, the LPK students liked having *two tries* feedback better even though they did not show higher knowledge gain. This phenomenon needs further investigation. As to the HPK students, there was no statistically significant difference in student ratings of their

satisfaction with the Program between the *single try* condition and the *two tries* condition. In contrast, according to Clariana (1999), “for high prior knowledge learners, multiple-try feedback was better, $ES = 0.39$ ” (p. 88).

Student perceptions on how the format of problem presentation (animation plus text, still image plus text, text-only) helped them remember and understand the key concepts and maintain their attention, indicated the need for the next stage of the evaluation (Evaluation 2). Students gave higher ratings to problem scenarios augmented with animation compared to the ones augmented with still images and text-only. There was a significant effect of a problem presentation format on students ability to remember the information learned through the Program [$F(2, 57) = 17.1, p = 0.001$], understand the information [$F(2, 57) = 16.7, p = 0.001$], and maintain student attention [$F(2, 57) = 51.00, p = 0.001$]. These findings contradict the results obtained in other studies dealing with abstract, scientific or technical content (Tversky et al. 2002). The possible explanation for this phenomenon could be the fact that the animations in the Program were built in ways that included the positive features of static images (Mayer et al. 2005).

As to the instructional strategies used in the program, they received high ratings from both LPK and HPK students, which may indicate that direct instruction through theory explanation and immediate explanatory program feedback were important for students’ success. Redundant (low-rated) program features were not detected during the evaluation.

Surprisingly, when student ratings in Survey 1 (administered after the training) and Survey 2 (administered after the post-test 5 days after the training) were compared, the ratings for Survey 2 were higher (see Appendix Tables 5, 6). Since the Survey 2 was a shortened version of Survey 1, and the Survey 2 was taken immediately after the post-test, the findings could serve as an indication that the Program helped students retain the concept of variables.

Conclusion

The systemic approach used for the evaluation of the Program allowed the designer to limit the design space of potential features to the small subset of the space that needed to be explored. It helped the designers to decide on the inclusion, exclusion, and further investigation of the potential candidates of program features. The most effective ones will be included in the final version of the Program.

Based on the results of Evaluation 1 presented in this paper, the overall effectiveness of the program was determined. As indicated by significant improvement in knowledge gain, the Program can facilitate retention of the variable concept regardless of the type of feedback, single try and two tries, and the level of student prior knowledge, low and high. Students found all the program features helpful. Redundant program features were not detected. If this would not be the case, the systemic approach would help us determine the features that did not contribute to student success.

Also, systemic approach was helpful in deciphering specific features that contributed to student knowledge gain. Based on the results of Evaluation 1, we can conclude that it is not the feedback type that made the difference. Also, the results of Evaluation 1 demonstrated that students regardless of their level of prior knowledge, found animated problem scenarios more helpful than scenarios augmented with still images and text only scenarios. The effect of the problem presentation format on

students' knowledge gain will be tested in Evaluation 2. After that, the final decisions will be made on which problem scenario format will be used in the final version of the Program.

Limitations and implications for further research

The major limitation of this study is insufficient power to detect small differences that might underlie the phenomenon. For this reason, the findings about the lack of significant differences in students' knowledge gain between LPK students in Condition 2 and 3 and HPK students in Condition 2 and 3 should be treated with caution. The findings from this research cannot be generalized to larger groups.

The effect size for the analysis of student knowledge gain ($d = 1.99$ for single try condition and $d = 2.30$ for multiple try condition) was found to exceed Cohen's (1988) convention for a large effect ($d = .80$). It might be the combination of instructional methods and program features that contributed to this fact. Future research is needed to disentangle which features of the computer treatments contributed to differences in test performance.

Other factors worthy of future study include the format of problem scenario presentation (text augmented with animation, still images and text-only). Also, further research is needed to determine the effectiveness of the Program on higher learning criterion measure such as transfer tasks.

Another limitation of the study is gender issue. More than two-thirds of the participants were females. Even though we diversified the selection of problem scenarios, supporting animations and images and ensured their high quality, the differences in the male's and females' perceptions of the used visuals remained a possibility.

The contribution of this research

This research demonstrates a holistic method of conducting a formative evaluation of newly designed software while using this software as a research platform. It is common knowledge that conducting repeated formative evaluations of instructional software can be resource intensive. The initial formative evaluation described here demonstrates the use of the framework that allows the designer to limit the number of formative evaluation studies by collecting the information not only on the student knowledge gain, but also on specific program features of interest and on the interaction of design features in regard to student level of prior knowledge. This formative evaluation demonstrates how to conduct design experiments in the background and refine instructional software design in a non-intrusive fashion.

Appendix

See Fig. 4 and Tables 3, 4, 5 and 6

LearnStat Menu Next Post-Training Survey

1. Rate each of the following items regarding the extent that it helped you to **recall the information learned through the program**:

Very helpful Not helpful

5 4 3 2 1

	• The feedback with answers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
221	• Problems with <i>animated</i> (moving) images	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Problems with <i>still</i> (not moving) images	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Problems with <i>text only</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Solving problems by yourself	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Theory explanations in the pop-ups	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

LearnStat Menu Next Post-Training Survey

2. Rate each of the following items regarding the extent that it helped you to **identify the variables**:

Very helpful Not helpful

5 4 3 2 1

	• The feedback with answers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
222	• Problems with <i>animated</i> (moving) images	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Problems with <i>still</i> (not moving) images	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Problems with <i>text only</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Solving problems by yourself	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Theory explanations in the pop-ups	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 4 The example of a post-training survey (Survey 1) taken by the students. *Note* The short follow up survey (Survey 2) taken after the delayed post-test (5 days after the training) consists of the first two items from the post-training survey (Survey 1)

LearnStat Menu Next **Post-Training Survey**

3. Rate each of the following items regarding the extent that it helped you to maintain your attention:

		Very helpful			Not helpful	
		5	4	3	2	1
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• The feedback with answers					
223	• Problems with <i>animated</i> (moving) images	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Problems with <i>still</i> (not moving) images	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Problems with <i>text only</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Solving problems by yourself	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Theory explanations in the pop-ups	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

LearnStat Menu Next **Post-Training Survey**

4. Learning through the program was:

		Strongly Agree	Agree	OK	Disagree	Strongly Disagree
	• Easy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Interesting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Fun	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
224	• Helped me to understand the difference between variables	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5. The program:

		Strongly Agree	Agree	OK	Disagree	Strongly Disagree
	• Provides a meaningful environment to learn	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Provides enough practice	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Makes me think	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

6. I would recommend the program to other students if they have a problem understanding variables:

		Strongly Agree	Agree	OK	Disagree	Strongly Disagree
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 4 continued

LearnStat Menu Next Post-Training Survey

7. The **feedback** provided by the program was:

	Strongly Agree	Agree	OK	Disagree	Strongly Disagree
• Sufficient	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Helpful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

225 8. I like the idea of red and green **smiley faces** to indicate right/wrong answers:

	Strongly Agree	Agree	OK	Disagree	Strongly Disagree
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

9. The **animation** (moving pictures) in the program was:

	Strongly Agree	Agree	OK	Disagree	Strongly Disagree
• Helpful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Interesting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Sufficient	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

LearnStat Menu Next Post-Training Survey

10. The **still images** (not moving pictures) in the program were:

	Strongly Agree	Agree	OK	Disagree	Strongly Disagree
• Helpful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Interesting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Sufficient	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
226 • Necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

11. The **theory explanations** in the pop-ups were:

	Strongly Agree	Agree	OK	Disagree	Strongly Disagree
• Helpful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Interesting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Sufficient	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 4 continued

Table 3 Means (and SD) of students' survey ratings (1-strongly disagree, 5-strongly agree) of overall satisfaction with the Program

Categories of student satisfaction with the Program	Condition 2 (C2)		Condition 3 (C3)	
	LPK, n = 15 M (SD)	HPK, n = 14 M (SD)	LPK, n = 17 M (SD)	HPK, n = 14 M (SD)
1. I would recommend this Program to others	4.33 (0.72)	4.29 (0.61)	4.59 (0.51)	4.57 (0.51)
2. The Program made me think	4.20 (0.68)	4.29 (0.61)	4.47 (0.51)	4.50 (0.65)
3. Learning through the Program was interesting	3.73 (0.80)	4.14 (0.36)	4.18 (0.64)	4.36 (0.74)
4. Learning through the Program was easy	3.80 (0.56)	4.21 (0.70)	3.88 (0.60)	4.64 (0.50)
5. The Program helped me understand the difference between variables	4.27 (0.70)	4.50 (0.65)	4.76 (0.44)	4.54 (0.50)

LPK stands for Low prior knowledge students, HPK stands for high prior knowledge students, n stands for the number of students

Table 4 Means (and SD) of students' survey ratings (1-strongly disagree, 5-strongly agree) of problem scenario presentation format by type of cognitive processing

Types of cognitive processing	Condition 2 (C2)		Condition 3(C3)	
	LPK, n = 15	HPK, n = 14	LPK, n = 17	HPK, n = 14
	M (SD)	M (SD)	M (SD)	M (SD)
Helped me recall the concept of variables				
1. Learning by using theory explanation popups (deductive reasoning use)	3.73 (0.96)	4.21 (0.58)	3.88 (0.99)	3.86 (1.10)
2. Learning by solving problems (inductive reasoning use)	3.20 (0.94)	3.64 (1.01)	3.47 (1.18)	4.07 (1.27)
3. Learning through feedback	4.53 (0.52)	4.50 (0.65)	4.53 (0.62)	4.43 (0.65)
4. Problems as text with still images	3.67 (0.98)	3.21 (0.97)	3.94 (0.90)	4.21 (0.70)
5. Problems as text only	2.93 (0.96)	3.07 (1.00)	3.00 (0.96)	3.73 (0.96)
6. Problems as text with animation	3.73 (0.96)	3.73 (0.96)	3.73 (0.71)	3.50 (0.94)
Helped me identify variables				
1. Learning by using theory explanation popups (deductive reasoning use)	3.87 (0.83)	4.14 (0.66)	3.76 (1.08)	3.79 (1.25)
2. Learning by solving problems (inductive reasoning use)	2.87 (1.06)	3.64 (0.84)	3.71 (1.05)	3.79 (1.31)
3. Learning through feedback	4.53 (0.64)	4.57 (0.65)	4.41 (0.71)	4.29 (0.73)
4. Problems as text with still images	3.80 (1.08)	3.64 (1.01)	3.76 (1.15)	4.07 (0.73)
5. Problems as text only	2.73 (0.96)	3.21 (0.97)	3.12 (0.86)	3.50 (1.34)
6. Problems as text with animation	4.47 (0.64)	4.07 (1.27)	4.06 (1.09)	3.93 (1.27)
Helped me maintain attention				
1. Learning by using theory explanation popups (deductive reasoning use)	3.00 (1.13)	3.64 (0.50)	3.53 (1.01)	3.36 (1.60)
2. Learning by solving problems (inductive reasoning use)	2.53 (1.19)	3.57 (0.94)	3.47 (1.33)	3.64 (1.28)
3. Learning through feedback	3.80 (1.01)	4.50 (0.94)	4.18 (0.95)	4.07 (0.83)
4. Problems as text with still images	3.20 (1.08)	3.93 (1.00)	3.94 (1.09)	3.71 (0.91)
5. Problems as text only	2.13 (1.36)	2.64 (0.93)	2.82 (1.19)	2.50 (0.85)
6. Problems as text with animation	4.67 (0.72)	4.50 (1.16)	4.35 (1.11)	4.43 (1.16)

Table 5 Means (and SD) of students' survey ratings (1-strongly disagree, 5-strongly agree) at the post-training survey and follow-up survey for a *single try* condition

	Survey 1 (post-training survey taken after the training)		Survey 2 (follow up survey taken after the delayed post-test, 5 days after the training)	
	LPK	HPK	LPK	HPK
The Program helped me recall the Information (remembering)				
1. The feedback with answers	4.53 (0.52)	4.5 (0.65)	4.58 (0.51)	4.5 (0.67)
2. Problems with animations	4.2 (0.94)	3.93 (1.14)	4.25 (0.87)	4.08 (1.31)
3. Problems with still images	3.67 (0.98)	3.21 (0.97)	3.75 (0.87)	3.83 (0.83)
4. Problems with text only	2.93 (0.96)	3.07 (1)	3.08 (0.67)	3.5 (0.8)
5. Solving problems by yourself	3.2 (0.94)	3.64 (1.01)	3.67 (0.89)	4 (0.85)
6. Theory explanations	3.73 (0.96)	4.21 (0.58)	4 (0.85)	4.08 (0.67)
The Program helped me learn how to identify type of variable (understanding)				
1. The feedback with answers	4.53 (0.64)	4.57 (0.65)	4.58 (0.51)	4.42 (0.67)
2. Problems with animations	4.47 (0.64)	4.07 (1.27)	4.33 (0.78)	4.08 (1.31)
3. Problems with still images	3.8 (1.08)	3.64 (1.01)	3.67 (0.65)	3.67 (1.15)
4. Problems with text only	2.73 (0.96)	3.21 (0.97)	3.25 (0.87)	3.5 (0.9)
5. Solving problems by yourself	2.87 (1.06)	3.64 (0.84)	3.67 (0.78)	3.92 (0.9)
6. Theory explanations	3.87 (0.83)	4.14 (0.66)	4.08 (0.79)	4.17 (0.72)

Table 6 Means (and SD) of students' survey ratings (1-strongly disagree, 5-strongly agree) at the post-training survey and follow-up survey for a *two tries* condition

	Survey 1 (post-training survey taken after the training)		Survey 2 (follow up survey taken after the delayed post-test, 5 days after the training)	
	LPK	HPK	LPK	HPK
The Program helped me recall the information (remembering)				
1. The feedback with answers	4.53 (0.51)	4.43 (0.65)	4.38 (0.77)	4.17 (0.94)
2. Problems with animations	4 (0.87)	4.14 (1.29)	4.15 (0.9)	3.92 (1.38)
3. Problems with still images	3.94 (0.87)	4.21 (0.7)	3.85 (0.8)	3.58 (0.9)
4. Problems with text only	3 (0.67)	3.5 (0.94)	3 (0.58)	3.17 (1.11)
5. Solving problems by yourself	3.47 (0.89)	4.07 (1.27)	3.77 (0.93)	4.17 (0.72)
6. Theory explanations	3.88 (0.85)	3.86 (1.1)	4.08 (0.76)	4 (0.85)
The Program helped me learn how to identify types of variables (understanding)				
1. The feedback with answers	4.41 (0.51)	4.41 (0.51)	4.38 (0.77)	4.33 (0.89)
2. Problems with animations	4.06 (0.78)	3.93 (1.27)	4.31 (0.95)	3.83 (1.34)
3. Problems with still images	3.76 (0.65)	4.07 (0.73)	4 (0.71)	3.67 (0.89)
4. Problems with text only	3.12 (0.87)	3.5 (1.34)	3.08 (0.49)	3.08 (0.9)
5. Solving problems by yourself	3.71 (0.78)	3.79 (1.31)	3.85 (1.14)	4.33 (0.49)
6. Theory explanations	3.76 (0.79)	3.79 (1.25)	3.85 (0.99)	4.08 (0.9)

References

- 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. doi:10.2190/9LMD-3U28-3A0G-FTQT.
- 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, 213–238. doi:10.2307/1170535.
- Bernard, M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., et al. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, 79, 1243–1289. doi:10.3102/0034654309333844.
- Betrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In R. E. Mayer (Ed.), *The cambridge handbook of multimedia learning*. Cambridge University Press: New York. doi:10.1017/CBO9780511816819.019.
- Butler, A. C., & Roediger, H. L. I. I. I. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, 36, 604–616. doi:10.3758/MC.36.3.604.
- Butler, D., & Winne, P. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245–281. doi:10.2307/1170684.
- Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, 1(1), 50.
- Clariana, R. B. (1993). A review of multiple-try feedback in traditional and computer-based instruction. *Journal of Computer-Based Instruction*, 20(3), 67–74.
- Clariana, R. B. (1999). *CBT design: A feedback aptitude treatment interaction*. In Proceedings of Selected Research and Development Papers Presented at the National Convention of the Association for Educational Communications and Technology (vol. 21, pp. 87–91). Houston, TX, February 10–14, 1999.
- Clark R. C., & Mayer, R. E. (2007). *E-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (2nd ed.). San Francisco, CA: Pfeiffer & Company. doi:10.1002/pfi.4930420510.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Craig, S., Gholson, B., & Driscoll, D. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, 94(2), 428–434.
- Friel, S. (2007). The researcher frontier: Where technology interacts with the teaching and learning of data analysis and statistics. In G. W. Blume & M. K. Heid (Eds.), *Research on technology and the teaching and learning of mathematics: Cases and Perspectives* (Vol. 2, pp. 279–331). Greenwich, CT: Information Age Publishing Inc.
- Garfield, J., Chance, B., & Snell, J. L. (2000). Technology in college statistics courses. In D. Holton et al. (Eds.), *The teaching and learning of mathematics at university level: AnICMI study* (pp. 357–370). Dordrecht: Kluwer Academic Publishers. Retrieved from http://www.dartmouth.edu/~chance/teaching_aids/books_articles/technology.html. Accessed 7 Oct 2014.
- Gulz, A., & Haake, M. (2006). Design of animated pedagogical agents—A look at their look. *International Journal of Human Computer Studies*, 64(4), 322–339.
- Hasler, B. S., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load, and instructional animation. *Applied Cognitive Psychology*, 21(6), 713–729. doi:10.1002/acp.1345.
- Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory and Cognition*, 22, 411–430.
- Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status* (pp. 383–434). Hillsdale, NJ: Lawrence Erlbaum.
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, 47(1), 211–232.
- Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology*, 63(5), 505–512.
- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1(4), 279–308.
- Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, 58(1), 79–97. doi:10.2307/1170349.
- Larreamendy-Joerns, J., Leinhardt, G., & Corredor, J. (2005). Six online statistics courses: Examination and review. *American Statistician*, 59, 240–251. doi:10.1198/000313005X54162.

- Larwin, K., & Larwin, D. (2011). A meta-analysis examining the impact of computer-assisted instruction on postsecondary statistics education: 40 years of research. *Journal of Research on Technology in Education*, 43(3), 253–278.
- Litchfield, B. C. (1987). *The effect of presentation sequence and generalization formulae on retention of coordinate and successive concepts and rules in computer-based instruction* (Doctoral dissertation, Florida State University, 1987). *Dissertation Abstracts International*, 49, 486A.
- Lowe, R. K. (2003). Animation and learning: Selective processing of information in dynamic graphics. *Learning and Instruction*, 13, 247–262. doi:10.1016/S0959-4752(02)00018-X.
- Mason, B. J. & Bruning, R. (2001). Providing feedback in computer-based instruction: What the research tells us. Retrieved from: <http://dwb.unl.edu/Edit/MB/MasonBruning.html>. Accessed 7 Oct 2014.
- Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *The American Psychologist*, 63(8), 760–769. doi:10.1037/0003-066X.63.8.760.
- Mayer, R. E., Dow, G. T., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds. *Journal of Educational Psychology*, 95, 806–813.
- Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instructions. *Journal of Experimental Psychology: Applied*, 11, 256–265. doi:10.1037/1076-898X.11.4.256.
- Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review*, 14, 87–99.
- McKenney, S., & Reeves, T. (2012). *Conducting educational design research*. London: Routledge.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97. doi:10.1037//0033-295X.101.2.343.
- Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, 635, 123–165.
- Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, 32, 99–113. doi:10.1023/B:TRUC.0000021811.66966.1d.
- Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching. Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 19(2), 177–213.
- Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research and Development*, 53(3), 35–45.
- Mory, E. (1992). The use of informational feedback in instruction: Implications for future research. *Educational Technology Research and Development*, 40(3), 5–20. doi:10.1007/BF02296839.
- Mory, E. (1996). Feedback research. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 919–956). New York: Simon & Schuster Macmillan.
- Mory, E. H. (2004) Feedback research revisited, In D. H. Jonassen, (Ed.), *Handbook of research on educational communications and technology* (pp. 745–783, 2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Narayanan, H. N., & Hegarty, M. (2002). Multimedia Design for communication of dynamic information. *International Journal of Human-Computer Studies*, 57, 279–315.
- Ormel, B., Pareja Roblin, N., McKenney, S., Voogt, J., & Pieters, J. (2012). Research-practice interactions as reported in recent design research studies: Still promising, still hazy. *Educational Technology Research & Development*, doi:10.1007/s11423-012-9261-6.
- Park, O. (1994). Dynamic visual displays in media-based instruction. *Educational Technology*, 34(4), 21–25.
- Plotzner, R., & Lowe, R. (2004). Dynamic visualizations and learning. *Learning and Instruction*, 14, 235–240.
- Richards, D. D., & Godfarb, J. (1986). The episodic memory model of conceptual development: An integrative viewpoint. *Cognitive Development*, 1, 183–219.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. doi:10.3102/0034654307313795.
- Sklar, J. C., & Zwick R. (2009). Multimedia presentations in educational measurement and statistics: Design considerations and instructional approaches. *Journal of Statistics Education*, 17(3), 3. Retrieved from <http://www.amstat.org/publications/jse/v17n3/sklar.html>. Accessed 7 Oct 2014.
- Sosa, G., Berger, D. E., Saw, A. T., & Mary, J. C. (2011). Effectiveness of computer-assisted instruction in statistics: A meta-analysis. *Review of Educational Research*, 81, 97–128. doi:10.3102/0034654310378174.

- Spector, M., Merrill, M. D., Merrienboer, J. V., & Driscoll, M. (2008). *Handbook of research on educational communications and technology* (3rd ed.). New York: Routledge.
- Symanzik, J., & Vulkasinovic, N. (2006). Teaching an introductory statistics course with CyberStats, an electronic textbook. *Journal of Statistics Education*, 14(1), 1. Retrieved from www.amstat.org/publications/jse/v14n1/symanzik.html. Accessed 7 Oct 2014.
- Tabbers, H., Martens, R., & van Merrienboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, 74, 71–81.
- Tennyson, R. D., & Cocchiarella, M. J. (1986). An empirically based instructional design theory for teaching concepts. *Review of Educational Research*, 56(1), 40–71.
- Tversky, B., Bauer-Morrison, J., & Bétrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies*, 57, 247–262. doi:10.1006/ijhc.2002.1017.
- Valdez, A. (2012). Computer-based feedback and goal intervention: learning effects. *Educational Technology Research and Development*, 60, 769–784. doi:10.1007/s11423-012-9252-7.
- Wender, K. F., & Muehlboeck, J. S. (2003). Animated diagrams in teaching statistics. *Behavior Research Methods, Instruments, & Computers*, 35, 255–258.

Natalya A. Koehler is an Instructional Design Faculty at Franklin University. She specializes in instructional design of online, hybrid and face to courses. Her research combines design research, design experiments, rapid prototyping, design, development, and evaluation of Web-based interactive multimedia.

Ann D. Thompson is a University Professor Emeritus at the School of Education of Iowa State University. She specializes in instructional technology in the Department of Curriculum and Instruction. For her work, research, and innovative programs at Iowa State University, she received the Upton Sinclair Award for 2011.

Ana-Paula Correia is Associate Professor at the School of Education of Iowa State University. Her research combines online learning, curriculum development in Educational Technology, and collaborative learning. She is a past president for the Association for Educational Communications and Technology Research and Theory Division (2012–2013).

Linda Serra Hagedorn is Associate Dean of Undergraduate Programs, and Professor at Iowa State University. Her research focuses on college student success. She is currently President of the Association for the Study of Higher Education (ASHE) and a past vice-resident of Division J of the American Educational Research Association (AERA).