2012

Research-based design of interactive multimedia for solving instructional problems

Natalya A. Koehler
Iowa State University

Research-based design of interactive multimedia for solving instructional problems

by

Natalya A. Koehler

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Co-majors: Education (Curriculum and Instructional Technology); Human Computer Interaction

Program of Study Committee:
Ann D. Thompson, Major Professor
Gary D. Phye
Ana-Paula Correia
Connie Hargrave
John P. Boysen

Iowa State University

Ames, Iowa

2012

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I wish to express my gratitude to the people who have helped me make this dissertation and PhD journey a reality.

First and foremost, I would like to thank my mentors and teachers, Dr. Ann Thompson, Dr. Gary Phye, and Dr. Ana-Paula Correia for their continuous support, encouragement, wisdom, and trust in me. I also thank my committee members, Dr. Hargrave and Dr. Boysen for their support of my work. I also extend my thanks to Dr. Richard Clark for his support and appreciation of my research work and to the Department of Curriculum and Instruction for providing funding during my PhD study.

I would also like to thank Dr. Lance Wilhelm and Ms. Aileen Mahood Sullivan for their support from the Ames Community School District, Iowa State University, Des Moines Area Community College, and Ames High School students for their participation in the design studies and appreciation of my work.

CHAPTER 1: GENERAL INTRODUCTION

Background

As cited in the report commissioned by Cisco Systems (2008) "the complexity of teaching and learning becomes increasingly apparent as the psychological, cognitive, social, and emotional aspects of learning become known." A major accomplishment of psychology has been the development of a science of learning aimed at understanding how people learn.

A few decades ago the research of cognitive psychologists was far removed from teaching in real classrooms. Today, cognitive psychologists attempt to test and refine their theories in real instructional settings. Researchers want to see how different classroom settings and classroom interactions influence applications of their theories (Bransford, Brown, & Cocking, 2000). "In attempting to apply the science of learning, a central challenge of psychology and education is the development of a science of instruction aimed at understanding how to present material in ways that help people learn" (Mayer, 2008, p.760).

The design of instruction should be guided by a research-based theory of how instruction works. This theory should take into consideration instructional methods that affect learning processes. This is the central premise in the science of instruction (Clark and Mayer 2008; Mayer 2009). It is common knowledge that effective instruction can be conducted in a classroom and over the internet. It is the use of verified instructional theory that determines the effectiveness of the instruction not the environment where it is delivered (Merrill, 2007). Clark stated that the delivery system (multimedia) is merely the truck but it is what is on the truck that counts (Clark, 1983, 1994a, 1994b). "Choice of delivery system is a matter of economics and convenience not a matter of instructional effectiveness" (Merrill, 2007, p.9).

On the other hand, researchers suggest that, compared with classes with a traditional lecture-oriented approach, those using multimedia are better liked by students and yield slight but statistically significant improvements in student learning as measured by both student self-report and objective outcome testing (e.g., Deimann & Keller , 2006; De Westelinck, Valcke, De Craene, & Kirschner, 2005; Frey, 1994; Kulik & Kulik, 1987; Mayer, 1997; McNeil & Nelson, 1991; Petty & Rosen, 1990; Sekuler, 1996; Welsh & Null, 1991; Worthington, Welsh, Archer, Mindes, & Forsyth, 1996).

New Web 2.0 technologies continue adding more nuances to multimedia learning. The overall trend is that students engaged in learning that includes multimodal designs, on average, outperform students who learn using traditional approaches with single modes (Cisco Systems, 2008). One of the challenges in research on multimedia is that it is difficult to determine whether it is the effect of multimedia or instructional strategies used in the multimedia.

This dissertation presents the design studies of three multimedia applications that combine both the effect of multimedia and instructional strategies used in these applications. Moreover, the evaluation of these multimedia applications conducted in real classrooms allows the researcher to derive instructional principles that are both grounded in theory and supported by evidence from authentic tasks.

Organization of the Dissertation

Using a multiple-paper format, this dissertation includes three publishable papers. Design, development, and formative evaluations of three interactive multimedia programs are presented in the included papers. The three instructional design studies in this dissertation contribute to the body of knowledge about research-based design of effective multimedia instructional programs.

Chapter 1: General Introduction

The first chapter introduces the research topic, presents a statement of the problem investigated, outlines the main purpose of the dissertation, and describes the organization of the dissertation chapters.

Chapter 2: A Design Study of a Web-based Multimedia Instructional Program for Teaching Types of Variables

This article presents the design engineering effort of a multimedia instructional program for teaching types of variables. An evaluation study involving 90 undergraduate students from a Midwest university provided preliminary evidence of the program effectiveness. As to the expected knowledge gain differences between the two experimental conditions with the number of feedback steps as the only difference between the conditions, the results did not demonstrate the significant difference in students' knowledge gain. The knowledge gain was high in both conditions. Apparently, it was not the number of feedback steps that was a decisive factor that facilitated students' learning through the program.

Chapter 3: A Design Study: Development of Flash-based Interactive Formative Assessment Software for Teaching Gas Laws Using PhET Simulation in a Secondary Science Class

This article presents an example of a multimedia software program focused on specific concepts and skills identified as elements of twelfth grade science course. The students' and teacher's perceptions of students' overall educational experiences and specific program features were collected. The study provided preliminary evidence that the program was an important scaffolding and formative assessment activity that made students' learning of the target critical scientific concepts effective.

Chapter 4: A Design Study of a Multimedia Instructional Grammar Program with Embedded Tracking

This published article (Koehler, Thompson , & Phye, 2011) presents two design studies of multimedia software for teaching grammar within the context of history and geography of the USA. The program was tested with ten and four adult ESL students in a Midwest community college. Preliminary results demonstrated effectiveness of the software.

Chapter 5: General Conclusions

The final chapter of this dissertation summarizes the findings of Chapters 2, 3, and 4 and presents recommendations for further research and practical implications in the field of multimedia instructional design.

References

Bransford, J. D., A. L. Brown, & Cocking, R.R. (Eds.). (2000). *How People Learn*: *Brain, mind, experience, and school*. Washington, D.C., National Academy Press.

Cisco Systems. Inc. (2008). Multimodal learning through media: What the research says. Retrieved from Edutopia website: http://www.edutopia.org/multimodal-learning-teaching-methods-media

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445-459.

Clark, R. E. (1994a). Media and method. *Educational Technology Research and Development, 42*(3), 7-10.

Clark, R. E. (1994b). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21-29.

Clark, R. C., & Mayer, R. E. (2008). *E-Learning and the science of instruction*. San Francisco: Pfeiffer.

Deimann, M. & Keller, J. (2006). Volitional Aspects of Multimedia Learning. *Journal of Educational Multimedia and Hypermedia*, *15*(2), 137-158.

De Westelinck, K, Valcke, M., De Craene, B., & Kirschner, P.(2005). Multimedia learning in social sciences: Limitations of external graphical representations. *Computers in Human Behavior*, *2,* 555-573.

Frey, D. K. (1994). Analysis of students' perceptual styles and their use of multimedia. *Perceptual and Motor Skills, 79*, 643–649.

Koehler, N., Thompson, A., Phye, G. (2011). A design study of a multimedia instructional grammar program with embedded tracking. *Instructional Science*, *39*(6), 939-974. doi:10.1007/s11251-010-9161-2.

Kulik, J. A., & Kulik, C. L. C. (1987). Review of recent research literature on computer-based instruction. *Contemporary Educational Psychology, 12*, 222–230.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*, 1–19.

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.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

McNeil, B. J., & Nelson, K. R. (1991). Meta-analysis of interactive video instruction: A 10-year review of achievement effects. *Journal of Computer-Based Instruction, 18*, 1–6.

Merrill, M. D. (2007). The Proper Study of Instructional Design. In Reiser, R. A. and Dempsey, J. V. (Eds.), *Trends and Issues in Instructional Design and Technology,* 2nd ed. (336-341). Pearson Prentice Hall.

Petty, L. C., & Rosen, E. F. (1990). Increase in mastery levels using a computer-based tutorial/simulation in experimental psychology. *Behavior Research Methods, Instruments & Computers*, *22*, 216–218.

Sekuler, R. (1996). Teaching sensory processes with multimedia: One of my teaching assistants is a mouse. *Behavior Research Methods, Instruments & Computers, 28*, 282–285.

Welsh, J. A., & Null, C. H. (1991). The effects of computer-based instruction on college students' comprehension of classic research. *Behavior Research Methods, Instruments & Computers*, *23*, 301–305.

Worthington, E. L., Jr., Welsh, J. A., Archer, C. R., Mindes, E. J., & Forsyth, D. R. (1996). Computer-assisted instruction as a supplement to lectures in an introductory psychology class. *Teaching of Psychology*, *23,* 175–18.

CHAPTER 2: A DESIGN STUDY OF A WEB-BASED MULTIMEDIA INSTRUCTIONAL PROGRAM FOR TEACHING TYPES OF VARIABLES

Abstract

This article presents a design study of a web-based interactive multimedia program for teaching types of variables. Students demonstrate slow progress in research methods and basic statistics classes if they struggle with identifying types of variables. A web-based multimedia instructional program (referred to here as the Program) to bring students up to speed on this concept was designed to serve two functions: (1) a teaching tool, and (2) a design-engineering platform, since the *design-engineering-develop* approach to innovation (Bryk & Gomez, 2008) has been identified as an effective instructional design strategy. Given the fact that feedback has been extensively identified as an important instructional strategy (Mory, 2003), two types of feedback - "single try" and "multiple try" - were identified as potentially effective ways to test the effectiveness of the Program. Even though research has been conducted on comparing "single try" feedback to AUC (answer until correct) feedback (Clariana, 1993), this research study specifically compares "two tries" feedback with "single try" feedback.

Another critical condition for learning is prior knowledge (Clark & Mayer, 2007), defined here as the student's preexisting attitudes, skills, experiences, and knowledge of the concepts at hand, in this study types of variables (independent, dependent, controlled variables, and levels of independent variable).

The article highlights the theoretical foundations for the underlying instructional design decisions. Then it describes the program features and the program design as a research platform. The design study of the Program was conducted with 90 undergraduate

students at a Midwestern university. In addition to the evaluation of the overall effectiveness of the Program, the effectiveness of the two types of feedback was tested as a part of using the Program as a research platform. The results of the program evaluation demonstrate the effectiveness of the Program with either type of feedback.

Introduction

This study is an effort to demonstrate the design engineering of a multimedia instructional program for teaching types of variables. In order to make the design engineering process easy, the Program was designed both as a teaching tool and a research platform for testing potentially effective program features and instructional design strategies.

Behavior tracking and data collection instruments (pre-test, survey, and delayed post-test) were imbedded in the Program, which allowed the designers to collect data from three perspectives: users' (students' perceptions of program features measured by the survey), instructor's (knowledge gain between pre-test and delayed post-test), and multimedia instructional designer's (count of the use of implemented program features at different stages in the training episode). A design study of the Program was conducted in a Midwest university. The purpose of the design study, data collection instruments, and data analysis plan are described below.

Literature Review

Computer technology has altered our ability to manage information. Multimedia has shortened the distance between people and information because it allows the computing to move from text to natural presentation of information through graphics, sound, images and video. Using multimedia provides multi-sensory experience for the learner in online environments. The benefits of multi-mode instruction are highlighted by Jensen and Sandlin

(1991). Multimedia mirrors the way in which the human mind thinks, learns, and remembers by moving easily from words to images, to sound, stopping along the way for interpretation, analysis, and in-depth exploration (Jensen & Sandlin, 1991).

Web-based multimedia instructional programs have additional benefits such as self-paced learning at the convenient place and time. The programs can be used by students both for regular classroom instruction and at home for remedial purposes because the training can be done at the students' convenience and at their own pace. In addition, multimedia instructional programs that support interactivity and assist students in customizing instruction to their needs can provide additional benefits to learners. "The key features of multiple media, user control over the delivery of information and interactivity, can help learners come to a deeper understanding through supporting conceptualization and contextualization of the new material being presented"(Cairncross & Mannion, 2001, p.162).

One of the key concepts in teaching research methodology involves types of variables. Students' fail or demonstrate slow progress in research methods and basic statistics courses if they have problems identifying types of variables. An effective instructional program that students could use at their convenience for bringing them up to speed with variables concept would benefit them.

On one hand, for a program to be effective, instructional principles must be consistent with what is known about how people learn. "By maintaining overlapping theoretical and practical goals, researchers can derive instructional principles that are both grounded in theory and supported by evidence from authentic tasks" (Mayer, 2008).

On the other hand, effective instructional design is typically based on a design-engineering-develop approach to innovation (Bryk & Gomez, 2008). If researchers engage in classroom-based research, the observed learning behaviors can be sources of data that inform next steps in the project. In this way, attention to knowledge use could be incorporated into the early stages of their work. Since the level of students' prior knowledge of types of variables concepts varies in a classroom, the ultimate goal of the researchers will be to design a simple low-cost program with linear navigation and pre-determined feedback that could address the needs of students with different level of expertise of a variable type concept.

Multimedia applications offer a particular valuable opportunity for feedback because it provides opportunities for students' self-assessment. The effects of different types and forms of informative feedback have been investigated in multiple instructional contexts and provided inconsistent findings (see reviews by Azevedo & Bernanrd, 1995; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Butler & Winne, 1995; Clariana, 1993; Mory ,1992, 1996; Mason & Bruning, 2001).

Mory (2003) identified as one of the directions for future research the identification of variables that can reflect internal cognitive and affective processes of learners that might potentially affect how feedback is perceived. In this research, an attempt to design a pre-determined psychologically well-founded informative tutoring feedback that would be helpful for both LPK and HPK students in their learning of the concept of variable types through the Program was made.

As the leading researchers in multimedia instructional design suggest, there is a need "to know more about the best types of feedback to give" and about the techniques that can "ensure that learners reflect on feedback" (Clark & Mayer, 2007, p.252). The aforementioned needs guided the researchers' effort to design the pre-determined feedback that is effective for teaching concepts for both LPK and HPK students.

In this research, careful consideration of the elements that made the teaching principles effective was done by collecting empirical validation of the applied program features and instructional design methods. Behavior tracking and all the data collection instruments (pre-test, delayed post-test, post-training survey, delayed survey after the post-test) were imbedded in the Program, which allowed the designers to collect data from three perspectives: users' (students' perceptions of the features in the Program), instructor's (knowledge gain between pre-test and delayed post-test), and multimedia instructional designer's (which type of feedback better supports students' learning and how students' with different level of prior knowledge use different features in the Program). The data were analyzed for patterns to determine how the performance aids were used by students with different levels of prior knowledge of the variables concept.

Moreover, as part of program design engineering process, along with the evaluation of the overall effectiveness of the Program, the researchers tested two types of feedback that were identified as having a potential to benefit the students learning through the Program. The assumption was that the type of feedback that was the most helpful for both LPK and HPK students would be implemented in the final version of the Program.

Several studies have suggested that learning is enhanced in computer-based animation environments (Park, 1994; Tversky, Bauer-Morriso, & Betrancourt, 2002). Animation appears to be most effective when presenting concepts of information that students may have difficulty envisioning (Betrancourt, 2005). On the other hand, in many studies dealing with abstract, scientific or technical content, animation did not turn out beneficial compared to static pictures (Lowe, 1996; Tversky, Bauer-Morriso, & Betrancourt, 2002). Clark & Mayer (2007, p.72) recommended "using static illustrations unless there is a compelling instructional rationale for animation."

In this study, animation augmented eight text scenarios, still images augmented the next eight scenarios; the other four scenarios were text-only. In the first eight scenarios, the animations were used to show the concept of change in a dependent variable when independent variable was changed. These animations ended in still images showing the completed state of the process change. In addition, learners had an option to replay the animation. This strategy allowed the learner to perceive functional relations between variables by watching the animations. At the same time, watching a still image would compensate for the fact that "human perceptual equipment is not very efficacious regarding processing of temporally changing information" (Betrancourt, 2005, p.290).

Moreover, Lowe (2003) showed that low prior knowledge students are often more focused on perceptually salient rather than thematically relevant features of animation. To lower this tendency, arrows, highlighting, and labels were implemented to guide students' attention to important features of the animation. Other potential program features were compared during the design study experiment. The insights on how this approach contributed to the design process are provided in this dissertation.

Program Description

The Program consists of 20 scenarios for identifying different types of variables and takes students approximately half an hour to complete. Since the scenarios come from various contexts, the Program can be used both in basic statistics and various research methodology courses. All performance aids are incorporated in the Program as pop-ups and could be used when needed.

Purpose of the Study

The aim of this study was, first, to evaluate the effectiveness of the design of low cost multimedia instructional program with linear navigation and predetermined feedback, second, collect information about how different features in the Program helped students memorize, understand, retain the information**,** and how they helped students maintain their attention. Two potentially effective types of feedback were compared and the most effective type was chosen.

Students' perceptions on how the design of different features in the Program supported their cognitive and metacognitive processes were collected through two surveys. The obtained quantitative and qualitative data were considered in regards to students' prior knowledge. Data were collected from 3 perspectives:

1. Students' (how different program features and instructional design methods help them learn)
2. Multimedia instructional designer's (which features are used the most/the least, possible navigation and visual design problems)
3. Instructors' (students' knowledge gain between pre-test and delayed post-test)

Research Questions

The following research questions were addressed:

Research Question #1: How well did the Program facilitate retention of the acquired concepts depending on the feedback type?

In this study, *single try* feedback was a response-contingent feedback consisting of the Knowledge of Results (KR) and the pre-determined Elaborative Feedback (EF). The KR consisted of green and red smiley faces, and the EF was a text explanation of the correct answers. The *two tries* feedback consisted of two steps. The users were presented the KR on the first try and the KR combined with the EF after the second try (see Figure 2.1).

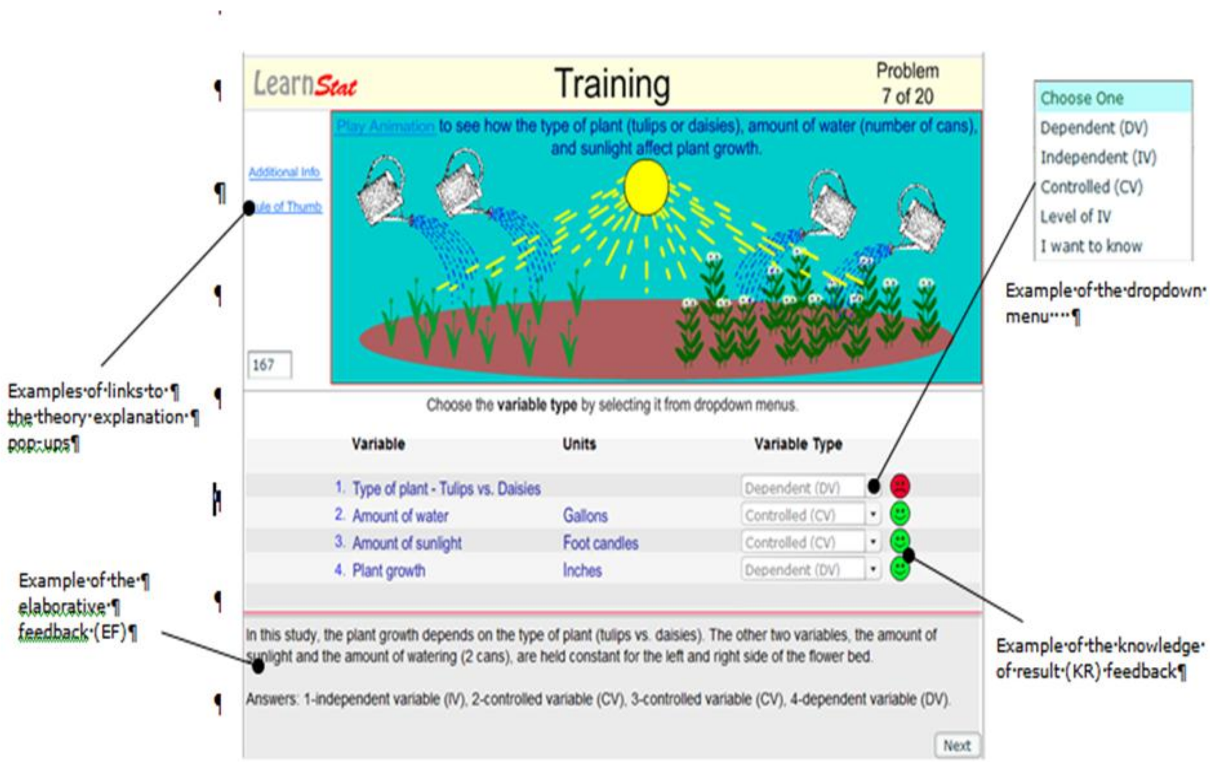


Figure 2.1 A screenshot of a problem scenario

It was hypothesized (Hypothesis 1a) that the Program would facilitate retention of the concept of variables, and the average knowledge gain in the control condition would be statistically significantly lower than in each of the experimental conditions. The only difference between experimental conditions was feedback type, *single try* or *two tries*. Also, it was expected that the average knowledge gain differences between the two experimental conditions would be statistically significant (Hypothesis 1b). Since the only difference between the experimental conditions was feedback type, the higher knowledge gains would indicate the more effective type of feedback for the Program.

Research Question #2: How different were students' experiences with the Program when their prior knowledge was considered?

It was expected that there would be a statistically significant difference in knowledge gain between the Low Prior Knowledge (LPK) students across two experimental conditions (Hypothesis 2a). Similarly, a statistically significant difference between the High Prior Knowledge (HPK) students across two experimental conditions was expected (Hypothesis 2b).

Methodology

Data were collected in undergraduate Basic Statistics courses for non-statistics majors and an Educational Psychology course participating in this study with 90 participants. The experiment followed a 2x2 design with the first factor as the between subjects factor and the last factor as the within subjects factor. The conditions tested included control condition and two experimental conditions, "single try" (ST) feedback and "two tries" (TT) feedback. Participants were randomly assigned to each of the three conditions, the equal number of students per condition. The between subjects factor was the type of feedback, ST and TT. The within subjects factor was the level of students' domain specific prior knowledge based on their pre-test score. The level of prior knowledge was determined according to the median split (Mdn=25) of participants' pre-test scores. Research questions and data collection instruments are presented in Table 2.1

Table 2.1 Research questions and data collection instruments

Research Questions and Hypotheses	Data Collection Instruments
How well did the Program facilitate retention of the acquired concepts depending on the feedback type?	1. Students' pre-test and delayed post-test scores stored in the database.
How different were students' experiences with the Program regarding their prior knowledge?	1. Students' pre-test and delayed post-test scores stored in the database. 2. Likert-scale survey

Data Collection

The participants received instruction through the Program at their convenience for 15-40 minutes without any help from the teachers. The pre-test, the training episode, and the post-test were done at the students' own pace. The data collection process is presented in Table 2.2

Students' perceptions on how the feedback helped them recall and understand the information along with how it helped them maintain their attention were collected in a likert-scale survey embedded in the Program and administered after the training.

Table 2.2 Data collection

Time Schedule	Procedures
Day1	Pre-test, training episode, likert survey
Day5	Post-test

Data Analysis

Students' Knowledge Gain

The knowledge gained during the training was assessed with a pre-test and delayed (5th day after the training) post-test. The pre-test and post-test were the same and consisted of ten scenarios. In each scenario, participants were asked to make five choices by selecting from five dropdown menus: independent variable, dependent variable, controlled variable, level of independent variable, and "I want to know". The fifth choice "I want to know" was added to avoid random answers. Making the correct choices required conceptual knowledge, that is, coherent mental models of types of variables. Each correct answer was scored as one point. The maximum score was 50 points.

Students' pre-test scores served as indicators of their prior knowledge of types of variables. The difference between delayed post-test scores and pre-test scores served as indicator of students' retention of the types of variables concept. All the tests as well as the survey were embedded in the Program and the students' responses were captured in the database. All the items in the tests were designed to check the students' ability to differentiate between independent, dependent & controlled variables as well as levels of independent variables.

Hypotheses 1a and 1b were validated by the pairwise comparison of the single try feedback condition and two tries feedback condition. As to the comparison of students' performance in each of the experimental conditions in regards to students' prior knowledge, non-parametric two-sample Wilcoxon rank-sum (Mann-Whitney) tests were computed within each prior knowledge level and the significance level was divided by two to avoid type I error (i.e., the value of the significance level was set at 0.025).

Students' Perceptions of the Program and their Use of Performance Aids

Students' ratings of different program features and instructional methods were examined in regards to their prior knowledge. Descriptive statistics was used for the analysis of the data. Means, medians and standard deviations of students' ratings of their overall experience and the program features were calculated for each experimental condition and level of prior knowledge (for LPK and HPK students by condition).

Results and Discussion

Research Question #1: How well did the Program facilitate retention of the acquired concepts depending on the feedback type?

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 (no training), Condition 2 (training with a single try feedback), and Condition 3 (training with a *two tries* feedback). Tests of the three a priori hypotheses were conducted using Bonferroni adjusted alpha levels of .017 per test (.05/3). There was a significant effect of the program training on students' knowledge gain [(F(2,87)=34.18, p=0.000]. Results indicated that the knowledge gain was significantly lower in control group condition (M = 1.9, SD = 2.52), than were those in both *single try* feedback condition (M = 14.93, SD = 9.16) and *two tries* feedback condition (M = 16.06, SD = 8.58). Hypothesis 1a was confirmed. The pairwise comparison of the *single try* feedback condition and *two tries* feedback condition was non-significant (0.838), [F (2, 87) =1.13, p=0.838]. Hypothesis 1b was rejected. The knowledge gain between the pre-test and delayed post-test (5 days after the training) in Condition 2 was 30.8% and in Condition 3 (30.0%).

Research Question #2: How different were students' experiences with the Program when their prior knowledge was considered?

Since the distribution of low prior knowledge (LPK) and high prior knowledge (HPK) students' scores per condition was not normal, 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 and the significance level was divided by two to avoid type I error (i.e., the value of the significance level was set at 0.025).

The results suggested that there is no statistically significant difference between the underlying distributions of the knowledge gain scores between low prior knowledge students in *single try* feedback condition (Condition 2) (M=19.67, SD= 9.78) and *two tries* feedback condition (Condition 3) (M=19.18, SD= 9.11) (z = 0.189, p = 0.8501).

As to high prior knowledge 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).

As to students 'overall satisfaction with the Program (see Table 2.3), it was higher for Condition 3 (C3) compared to Condition 2 (C2) for both LPK and HPK students even though there was no significant difference in knowledge gain. Learning through the Program was easier (LPK:3.84, HPK: 4.43) and more interesting (LPK: 3.97, HPK:4.25) for HPK students.

In contrast, the ratings for the survey statement "The Program helped me understand the difference between variables." were marginally the same (LPK:4.53, HPK: 4.57). The same is true about students' ratings of the statement "I would recommend the Program to others." (LPK: 4.47, HPK: 4.43). It allows the designer to assume that the Program was equally helpful for both LPK and HPK students and the program features address the needs of students with different levels of prior knowledge.

Table 2.3 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**		**Condition 3**	
	LPK, n=15	**HPK, n=14**	**LPK, n=17**	**HPK, n=14**
	M (SD)	**M (SD)**	**M (SD)**	**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)

Note. LPK stands for Low prior knowledge students; HPK stands for high prior knowledge students; n stands for the number of students.

As to students' ratings of program features, the program feature that received the highest rating was feedback across both conditions. Interestingly, both survey items *Learning by using theory explanation* popups (deductive reasoning use) and *Learning by solving problems* (inductive reasoning use) were rated higher by high prior-knowledge students. In contrast, the item *Learning through feedback* received almost the same ratings.

Also, in both conditions, text scenarios augmented with animation were consistently higher rated compared to the ones with still images and text only. Scenarios with animation were the most helpful for maintaining attention. On the other hand, the ranges of ratings for all three items (animation, still images, text-only) are large (min: 1, max: 5), which means that the preferences may be related to students' individual differences related neither to prior knowledge nor to conditions. Students' survey ratings by the kind of cognitive processing supported by the Program (the target concept recall, understanding, and maintaining attention during the training) are presented in Table 2.4.

Table 2.4 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**		**Condition 3**	
	LPK, n=15 M (SD)	**HPK, n=14 M (SD)**	**LPK, n=17 M (SD)**	**HPK, n=14 M (SD)**
1.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)
2.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)
3. 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)

Conclusions

The multilayered nature of this design-based research is an attempt to contribute both to the science of instruction (formative evaluation of the Program with the goal of identifying possible problems for further modifications and improvements) and the science of learning (the experimental testing of two types of feedback design grounded in cognitive theories and validating the feedback design using students' input on how the feedback and other features in the Program help them maintain attention, understand, and retain the target concepts). "By maintaining overlapping theoretical and practical goals, researchers can derive instructional principles that are both grounded in theory and supported by evidence from authentic tasks" (Mayer, 2008).

It might be argued that for both experimental conditions - "single try" (ST) and "two tries" (TT) feedback – the Program facilitated retention of knowledge. There was no significant difference between conditions, which is in tune with Clariana's study (1993).

This work allowed the researchers to make conclusions about the overall effectiveness of the Program by comparing the students' knowledge gain in the control group with the students' knowledge gain in each of the experimental conditions. The knowledge gain was compared for both LPK and HPK students across the conditions to make sure that the Program meets the needs of students with different levels of prior knowledge. In addition, the comparison of the students' knowledge in the two experimental conditions and by prior knowledge allowed the designers to choose the most effective feedback.

Students' perceptions on how the design of different features in the Program supports their cognitive and metacognitive processes provided information about the justification and helpfulness of the Program features. Also the data from this research study provided themes for the next stage of the Program formative evaluation. One of the topics of interest in the next stage that emerged from this research was the comparison of different formats of problem scenario presentations: problem scenario augmented with animation, problem scenarios augmented with still images, and text only scenarios.

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.

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

Bera, S. & Liu, M. (2006). Cognitive Tools, Individual Differences, and Group Processing as Mediating Factors in a Hypermedia Environment. *Computers in Human Behavior. 22*(2), 295-319.

Betrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In R. E. Mayer (Ed.). *The Cambridge Handbook of Multimedia Learning*. New York: Cambridge University Press.

Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisation. *Learning and Instruction, 14*(3), 325–341.

Bryk, A. S., & Gomez, L. M. (2008). Ruminations on reinventing an R&D capacity for educational improvement. In F. M. Hess (Ed.), *The future of educational entrepreneurship: Possibilities for school reform*. Cambridge: Harvard Education Press.

Butler, D. & Winne, P.(1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational research*, *65* (3), 245-281.

Cairncross, S & Mannion, M. (2001).Interactive Multimedia and Learning: Realizing the Benefits. *Innovations in Education & Teaching International, 38* (2), 156-164, doi: 10.1080/14703290110035428

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.

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.

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

Clark, R. & Mayer, R. E. (2008). *E-learning and the science of instruction* (2nd ed.). San Francisco: Jossey-Bass.

Eylon, B., Ronen, M., & Ganiel, U. (1996). Computer simulations as tools for teaching and learning: Using a simulation environment in optics. *Journal of Science Education and Technology, 5,* 93–110.

Jensen, R. E. & Sandlin, P.(1991). *Why do it? Advantages and dangers of new ways of computer-aided teaching/instruction*. San Antonio, TX: Department of Business Administration, Trinity University.

Klauer, K., & Phye, G. (2008). Inductive reasoning: A training approach. *Review of Educational Research, 78*(1), 85–123. doi:10.3102/0034654307313402.

Koehler, N., Thompson, A., Phye, G. (2011). A design study of a multimedia instructional grammar program with embedded tracking. *Instructional Science, 39*(6), 939-974. doi:10.1007/s11251-010-9161-2.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 989-998.

Lawson, A. E., Baker, W. P., DiDonato, L., Verdi, M., & Johnson, M. (1993). The role of hypothetico-deductive reasoning and physical analogies of molecular interaction in conceptual change. *Journal of Research in Science Teaching, 30*(7), 482-495.

Lewins, A., Taylor, C. & Gibbs, G. (2005). What is qualitative data analysis? School of Human & Health Sciences, University of Huddersfield, United Kingdom.

Limson, M., Witzlib, C. & Desharnais, R. (2007). Using web-based simulations to promote inquiry. *Science Scope, 30*(6), 36-42.

Lowe, R. K. (2003). Animation and learning: Selective processing of information in dynamic graphics. *Learning and Instruction*, *13*, 247-262.

Marzano, R. J. (2007*). The Art and Science of Teaching: A Comprehensive Framework for Effective Instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.

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.

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

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. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

Merrill, M. D. (2007). The Proper Study of Instructional Design. In Reiser, R. A. and Dempsey, J. V. (Eds.), *Trends and Issues in Instructional Design and Technology*, 2nd ed. (336-341). Pearson Prentice Hall.

Mory, E. (1992). The use of informational feedback in instruction: Implications for future research. *Educational Training Research and Development, 40*(3), 5-20.

Mory, E. (1996). Feedback research. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology*. New York: Simon & Schuster Macmillan.

Mory, E. H. (2003) Feedback research revisited. In J. H. Jonassen, (Ed.), *Handbook of Research on Educational Communications and Technology* (pp.745-783) New York, NY: MacMillian Library Reference.

Papanek, V. (1985). *Design for the Real World: Human Ecology and Social Change* (2nd ed.). Chicago, IL: Academy Chicago Publishers.

Park, O. (1994). Dynamic visual displays in media-based instruction. *Educational Technology, 34*(4), 21-25.

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210.

Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003). On the science of education design studies. *Educational Researcher, 32*(1), 25-28.

Spector, J. M., Merrill, M. D., van Merrienboer, J., Driscoll, M. P.(Eds.) (2008). *Handbook of research on educational communications and technology*. New York, NY: Lawrence Erlbaum Associates.

Sweller, J. (1999). *Instructional design in technical areas*. Melbourne: ACER.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp.19–30). Cambridge, UK: Cambridge University Press.

Tversky, B. Bauer-Morrison, J., & Bétrancourt, M. (2002) Animation: Can it facilitate? *International Journal of Human-Computer Studies*, *57*, 247-262.

CHAPTER 3: A DESIGN STUDY OF A FLASH-BASED INTERACTIVE FORMATIVE ASSESSMENT SOFTWARE FOR TEACHING GAS LAWS USING PhET SIMULATION

Abstract

This paper presents a design study of a Flash-based interactive Gas Laws Program. The article describes the instructional decisions underlying the design of the Gas Laws Program, program features, and a design study conducted in five high school science classes with a total of 114 students as a part of their science curriculum. The study was conducted over a two weeks period. Data on the students' and teacher's perceptions of the overall educational experience and specific program features were collected and analyzed.

Introduction

With the advent of curriculum reform, large scale evaluation efforts must be supported by standardized formative evaluation efforts that inform instruction and student learning. The following is an example of a multimedia software program designed to focus on specific concepts and skills identified as elements of a twelfth grade science course.

In this paper, first, the rationale behind the instructional design decisions implemented in the Gas Laws Program is described. Second, the data collection and analysis methods are presented. Students' and teacher's perceptions of their experience with the program confirmed the effectiveness of the implemented decisions.

Literature Review

Simulation is "computer- based model of natural process or phenomenon that reacts to changes in values of input variables by displaying the resulting values of output variables" (Spector, Merrill, van Merriënboer & Driscoll, 2008, p.457). Classroom activities that surround, support, and assess students' learning are critical parts of any simulation. As stated by Adams et al. (2008, p.12), the authors of PhET (Physics Education Technology project) Interactive Simulations (2006) implicitly assume that their simulations "will be used in the context of an educational setting where teachers will primarily provide the scaffolding and goals for the simulation use."

From a cognitive perspective, it is argued that the need to coordinate and assimilate concepts or elements into knowledge constructs is the primary generator of information complexity in a difficult subject such as chemistry (Sweller & Chandler, 1994). Simple tasks are said to have low element interactivity, and contain elements that can be learned in isolation, whereas complex tasks contain elements that must be learned in concert with one another. A subject is complex, not because of the number of elements to be learned, but the need to simultaneously assimilate the many elements before meaningful learning can occur (Sweller, 1999; Sweller & Chandler, 1994).

The topic of gas behavior is a regular component of many chemistry curricula that requires integrated understanding of various areas of introductory chemistry. Many students have considerable difficulties understanding the concepts and processes involved. It is common knowledge among educators that the gas laws topic is one of the most difficult for students to master in chemistry.

The central premise of this project was that students would achieve deeper conceptual understanding of gas laws on the microscopic level as a result of careful integration of a Flash-based interactive formative assessment program (referred to here as the Program) with the Gas Laws PhET simulation. The Program allowed students to perform hypothetico-deductive reasoning (Lawson, Baker, DiDonato, Verdi, & Johnson, 1993) and enabled translation between visual representations of gas laws and proportional reasoning. The Program was meant to facilitate and coordinate students' assimilation of gas laws concepts on the microscopic level. The specific objectives for the lesson during which the Program was used as a part of the curriculum were identified by the teacher. The lesson was expected to help students better understand the same concepts on the macroscopic level at a later stage.

There is no doubt that web-based simulations will be an ever-growing part of science education. At this point, the important question is how to design highly productive educational uses of web-based simulations. Even though Clark and Mayer (2007) described evidence for guidance in the form of explanation and reflection, there is still need for research "on the most appropriate format, source, timing, and type of guidance to use for different instructional goals at different learning stages" (Clark & Mayer, 2007, p.374).

The necessity of design research that tests new ideas in real classrooms has been stated by Shavelson, Phillips, Towne, & Feuer (2002). The authors believe that "coupling scientifically warranted knowledge and rich contextual information in some narrative form might lead to increased understanding and use of scientific research in practice."

To address the above challenges the formative assessment (the design study) of the Program was conducted. During the study, preliminary data were collected about what users (the students and the teacher) think about the effectiveness of the Program.

The purpose of this paper is to describe the instructional decisions underlying the Program and the results of the formative evaluation conducted in real classroom settings. The Program was used in five high school science classes for regular classroom instruction. The students' and teacher's perceptions of the effectiveness of the training were collected, and the correspondence of the instructional designer's and users' perceptions of the Program effectiveness was determined.

Research Questions

Research questions and data collection instruments are presented in Table 3.1.

Table 3.1 Research questions and data collection instruments

Research questions	Data collection instruments
1. What were the students' perceptions of how the Program supported their cognitive processing of the target concepts?	Survey consisting of four open-ended questions administered to students on paper during the last five minutes of the classroom period.
2. Which program features and instructional strategies implemented in the Program were the most and least helpful?	Students' survey ratings on the rating scale of 5 (5 – strongly agree, 1 – strongly disagree) for program features such as pop-up explanations, images, model explorations in the simulation lab, guiding questions, diagrams, and feedback.
3. What were the teacher's perceptions of how the objectives of the lesson were met?	A semi-structured interview with the instructor about how the Program helped her students learn about gas laws on the microscopic level and how easy the Program was to use.

The Instructional Design of the Program

The Objectives for the Lesson

The collaborating teacher identified the purpose of this lesson as introduction to gas behavior on the microscopic level with use of a visual tool that can be manipulated. The knowledge of common students' misconceptions of gas laws helped the teacher identify the objectives of the lesson. Such knowledge is a prerequisite for developing effective instructional approaches using the potential of new technologies (Eylon, Ronene, & Ganiel, 1996).

Students were expected to be able to answer the following questions as a result of working with the simulation:

1. How does a gas cause pressure?
2. What happens to gas pressure as its volume changes?
3. What happens to gas pressure as its temperature changes?
4. What happens to gas pressure as the number of gas particles changes?
5. Why do most gases (close to an ideal gas) behave almost in the same way regardless of their identity under the same conditions of volume, temperature, and pressure?

In our study, students' learning through the Gas Laws PhET simulation (see Fig 4) was guided by the Flash-based interactive formative assessment, the Program, (see Figures 3.1-3.3, 3.5-3.14) and handout materials (see Appendix 1-2). The Program was designed to address specific conceptual difficulties in the domain. For example, it is common knowledge that many students have a problem understanding that most gases (close to an ideal gas) behave almost in the same way regardless of their identity under the same conditions of volume, temperature, and pressure. In order to preserve the "exploratory" nature of the PhET simulation, a four step strategy was used in the design of the Program:

1. Prior Knowledge Activation/Predictions - students' select factors that may potentially affect gas pressure from multiple choice menu.
2. Hypotheses Validation- students check their predictions in the simulation lab.
3. Reflection - students' type an explanation of the findings into the Program text box.

4. Explanation - pre-determined elaborative feedback provided by the Program augmented with images and diagrams.

Moreover, the Program was designed to support important cognitive processes such as attention, activation of prior knowledge, elaboration –rehearsal, encoding, and retrieval (Clark, Nguyen, & Sweller, 2006). The description of instructional design features supporting these cognitive processes and illustrations are presented below.

Program Design Description

Step 1.

At the beginning of the Program, the interactive assignments were added to gain students' attention and activate their prior knowledge. In these assignments, students were asked to match the type of interaction between molecules and states of matter. The pre-determined feedback in a form of correct responses was presented to students when they clicked on the Check Answer button (see Figure 3.1). The integration of the students' new knowledge into their pre-existing cognitive schemas was done through overt linkages by the Program pointing out the connections "between content previously addressed in class and content that is about to be presented in a critical-input experience" (Marzano , 2007, p 41).

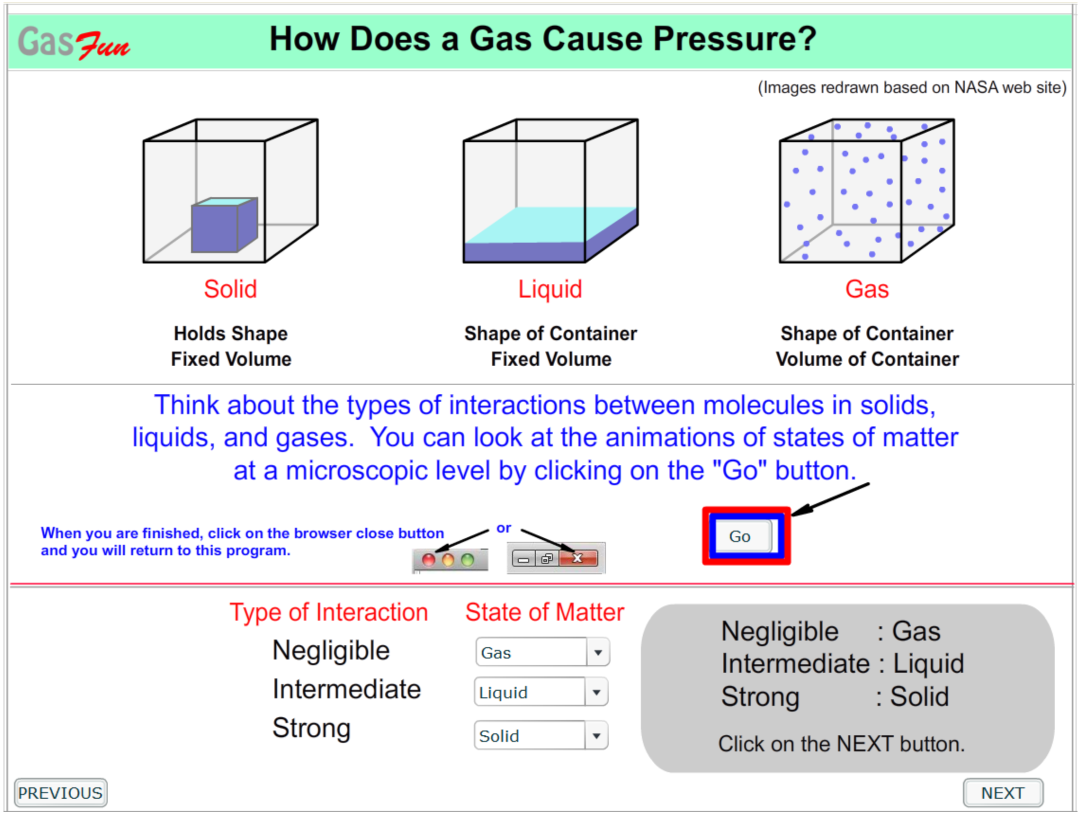


Figure 3.1 The screen shot illustrating the pre-determined feedback

The visual introduction of the nature of gas pressure was presented both on a small and large scale (see Figure 3.2). The assumption was that even if students had little or no prior knowledge relative to this specific topic, they would at least activate related knowledge that would allow them to make important linkages (Marzano, 2007).

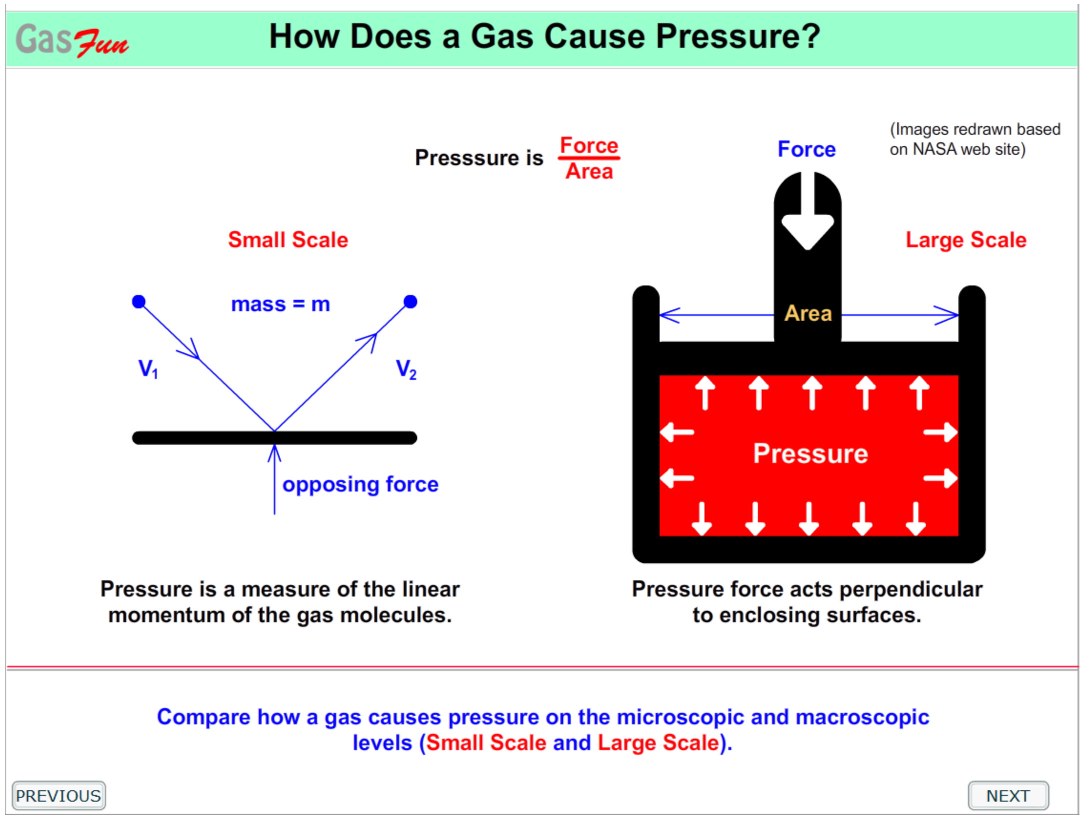


Figure 3.2 The screen shot illustrating the concept of gas pressure

After that, students were asked to make their predictions about what might affect gas pressure: type of gas, temperature of gas, mass of gas, volume of gas, speed of gas molecules, and number of gas moles (See Fig 3.3, 3.4). Lawson and colleagues found that students who were able to perform hypothetico-deductive reasoning were better in acquiring new concepts (Lawson, Baker, DiDonato, Verdi & Johnson, 1993).

Step 2.

Then students were taken to the PhET lab simulation to check their predictions by going through targeted exercises (see Fig 3.4).

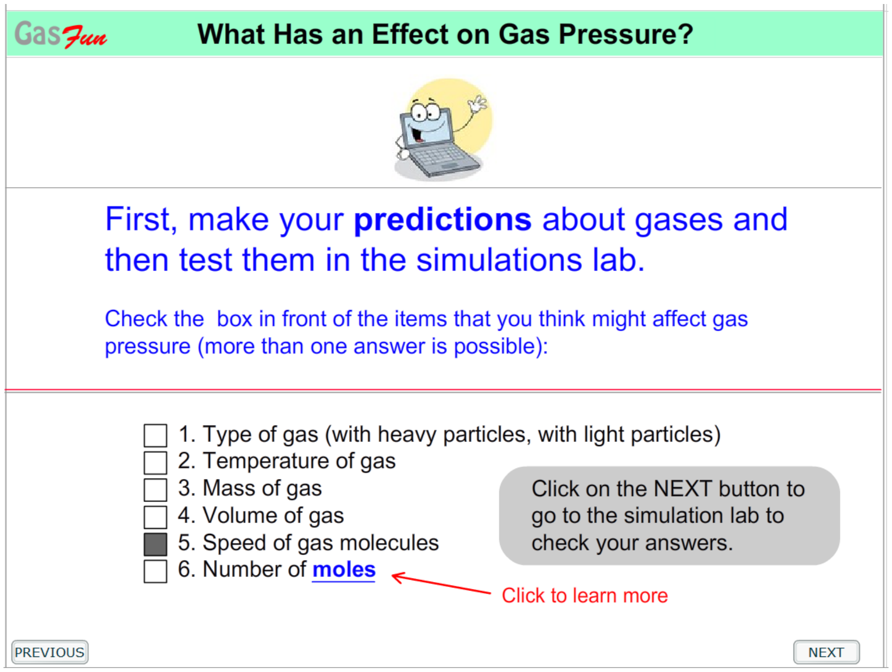


Figure 3.3 The screen shot illustrating the predictions that students need to make

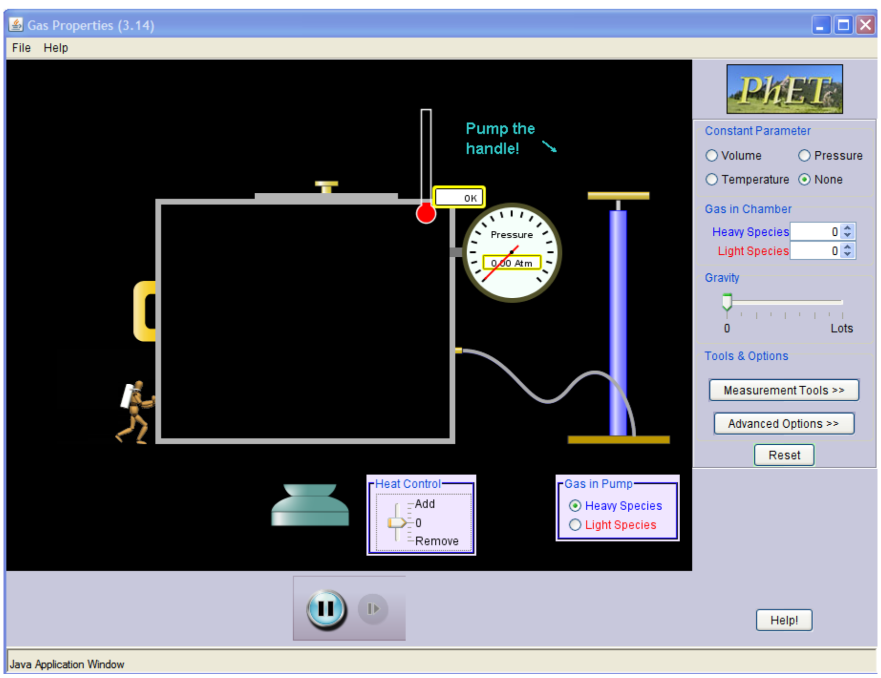


Figure 3.4 The screen shot illustrating the Gas Laws PhET simulation lab

The goal of Experiments 1 and 2 in the PhET simulation lab was to guide students' thinking in the direction that gases behave in very similar ways regardless of the type of gas.

Students were asked to go through the same procedure in the PhET simulation lab by changing gas mass both for “light species” (a lighter gas) and “heavy species” (a heavier gas) and recording gas pressure. The volume of the gas was kept constant. Providing detailed step by step instructions in the handouts (See Appendix A) allowed the students to work efficiently with the simulations and record their findings in the tables provided in the handouts.

Steps 3 & 4.

After conducting the experiments (Experiment 1 and 2) in the simulation lab, the elaboration rehearsal of the acquired concepts was supported by the guiding questions presented by the Program (see Figure 3.5).

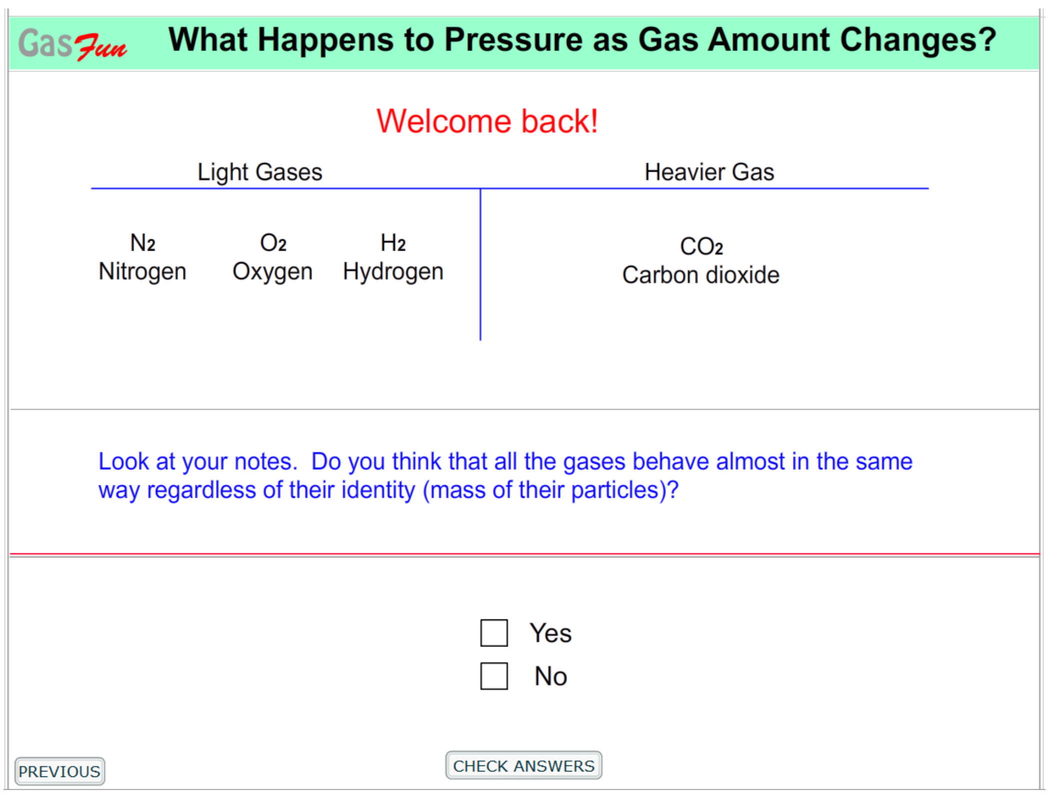


Figure 3.5 The screen shot illustrating the “compare and contrast” instructional strategy

These questions stimulated students' inductive reasoning thinking and helped them organize newly acquired information by making them use a "compare and contrast" strategy. Students were asked to compare the behavior of "light species" and "heavy species". Only after having students respond to the question (see Figure 3.5), Avogadro's law was presented as the elaborative feedback to the student's response (see Figure 3.6). This instructional strategy was used to promote the elaboration of the newly acquired information about gas behavior. "As a result of rehearsal and elaboration, incoming content from the instructional environment is transformed to result in expanded schemas stored in long-term memory." (Clark, Nguyen, & Sweller, 2006, p.37). Having the students investigate the behavior of two types of gases in the simulation lab and after that compare the behavior of the gases before the introduction of the new concept of Avogadro's law was expected to facilitate students' learning of this challenging concept.

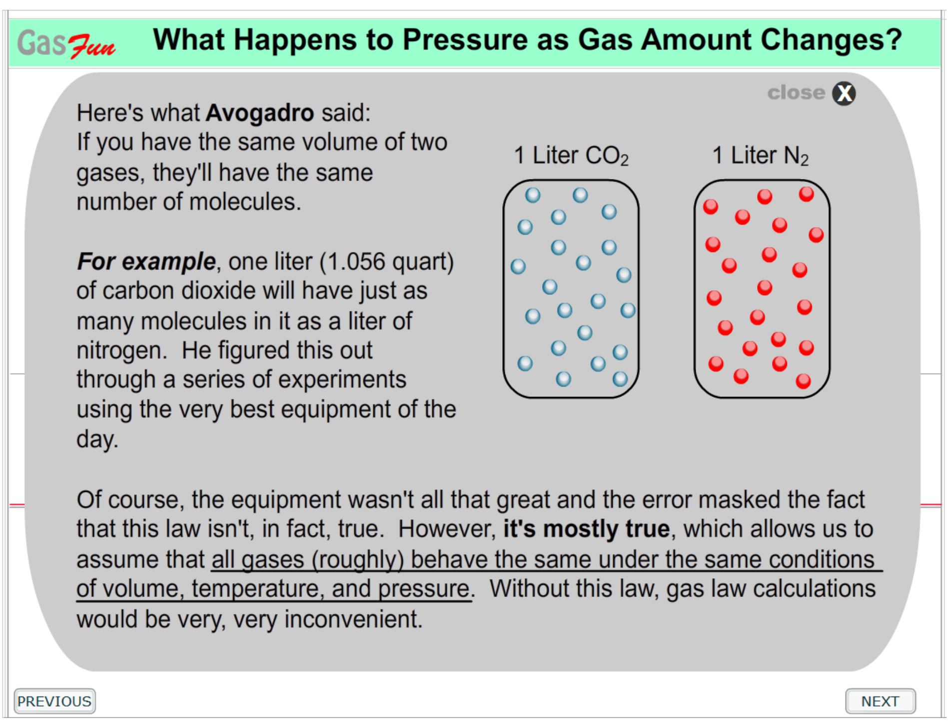


Figure 3.6 The screen shot illustrating the elaborative feedback response

On the next screen, the students were prompted to further elaborate on the concept of gas behavior (see Figure 3.7).

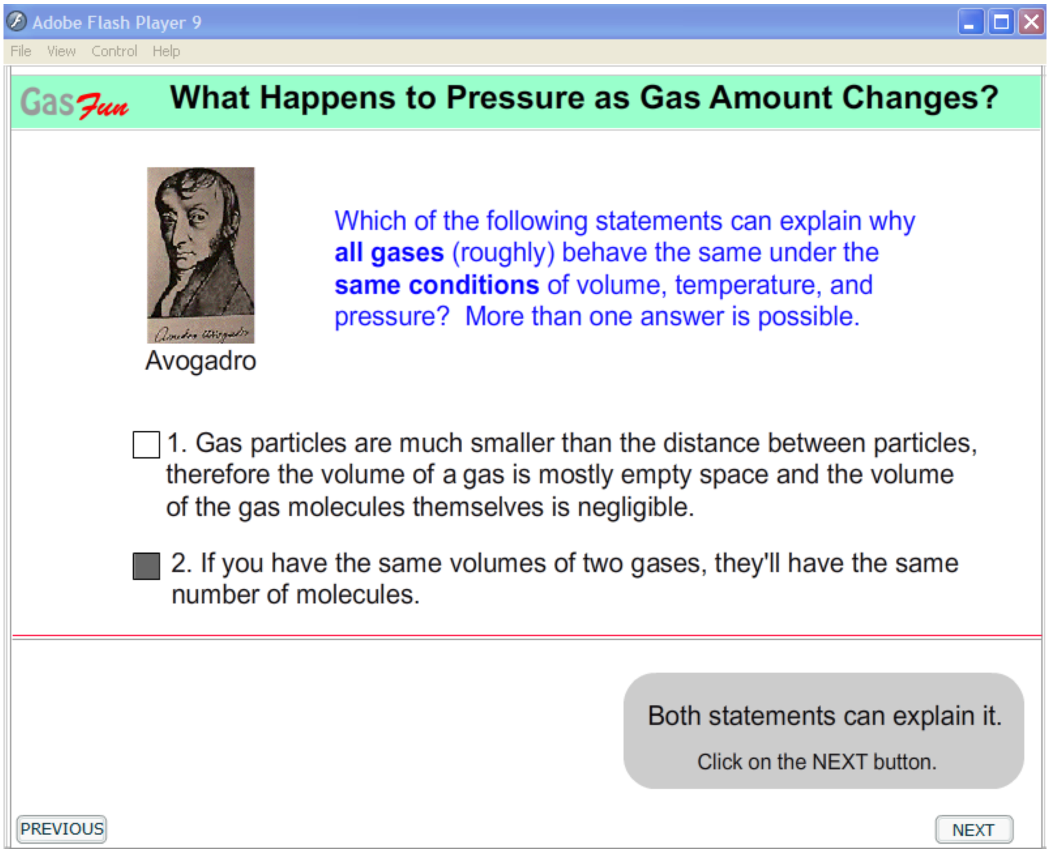


Figure 3.7 The screen shot illustrating the strategy of making students to further elaborate on gas behavior

As a further step, students were asked to describe in their own words their findings about gas behavior and summarize the newly acquired information by selecting responses from dropdown menus (see Figure 3.8).

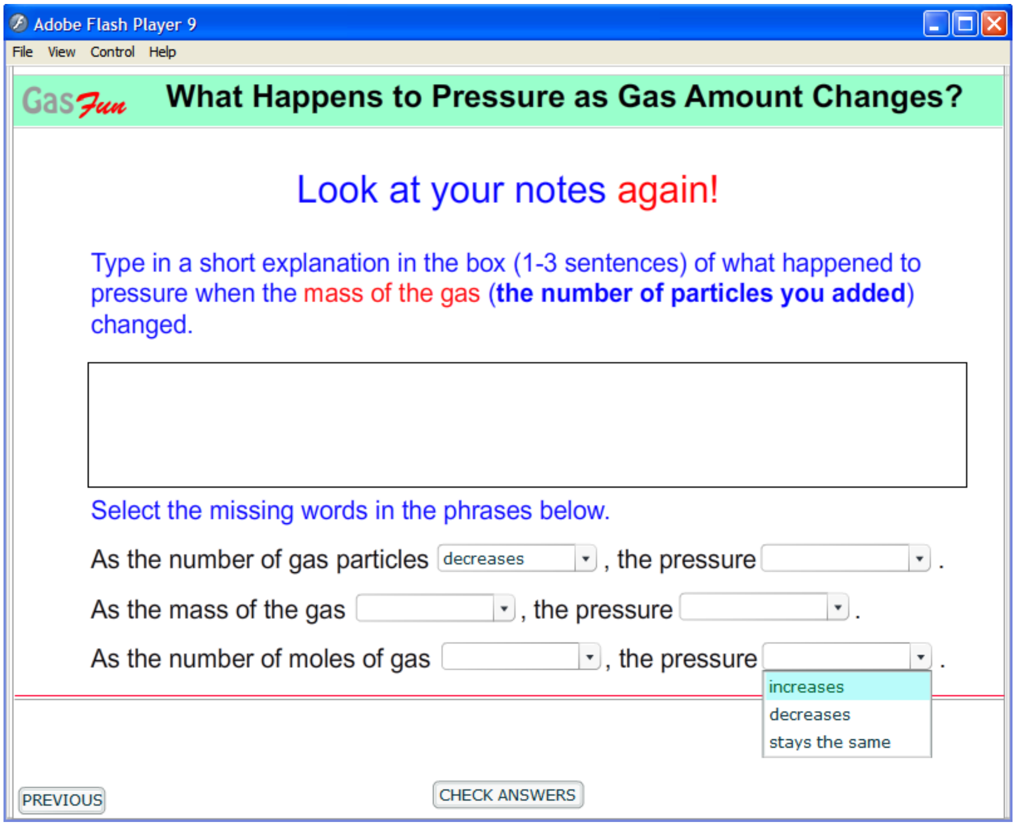


Figure 3.8 The screen shot illustrating the strategy of making students summarize the information acquired in Experiments 1 & 2

Finally, the elaborative feedback provided by the Program (see Figure 3.9) allowed learners to integrate proportional reasoning within the topic of gas laws.

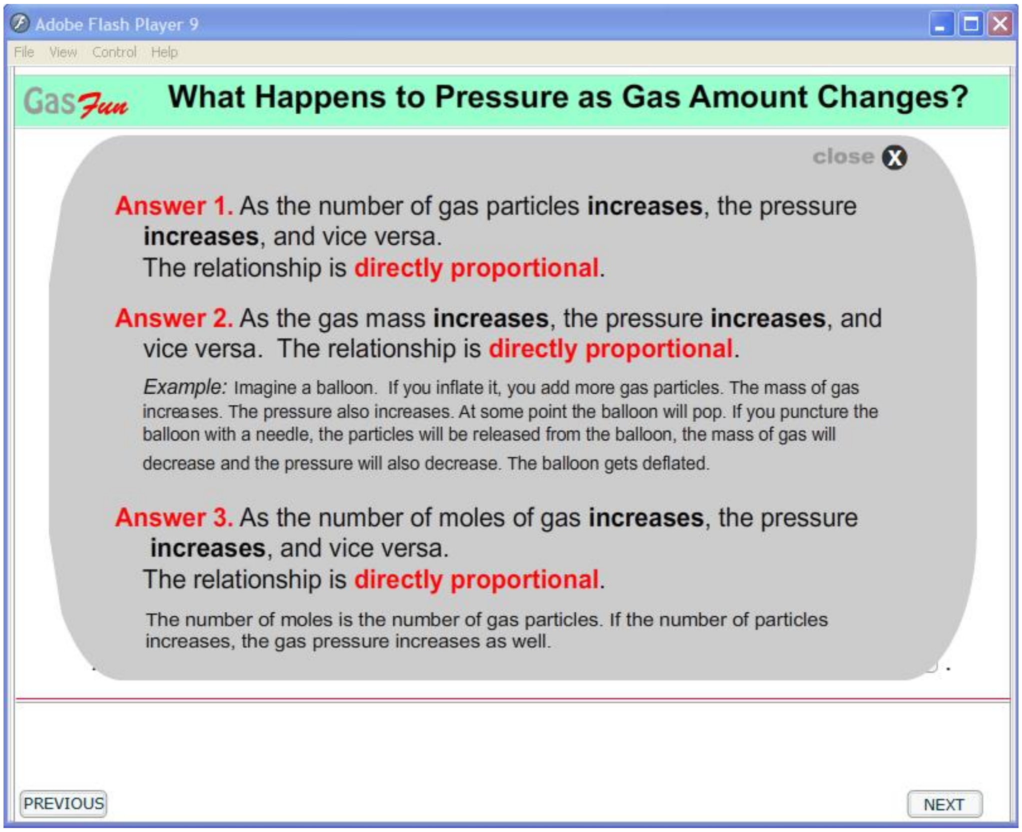


Figure 3.9 The screen shot illustrating the strategy of making students integrate proportional reasoning with the topic of gas laws

The next two experiments in the PhET simulation were also provided with detailed step by step instructions in the handouts (See Appendix B) to allow students work with the simulations efficiently. Students were asked to change temperature while keeping the volume constant in Experiment 3 and to change volume while keeping the temperature constant in Experiment 4 to observe what happened to gas pressure. The strategy of enabling translation between multiple representations of gas laws and proportional reasoning is presented on Figure 3.10.

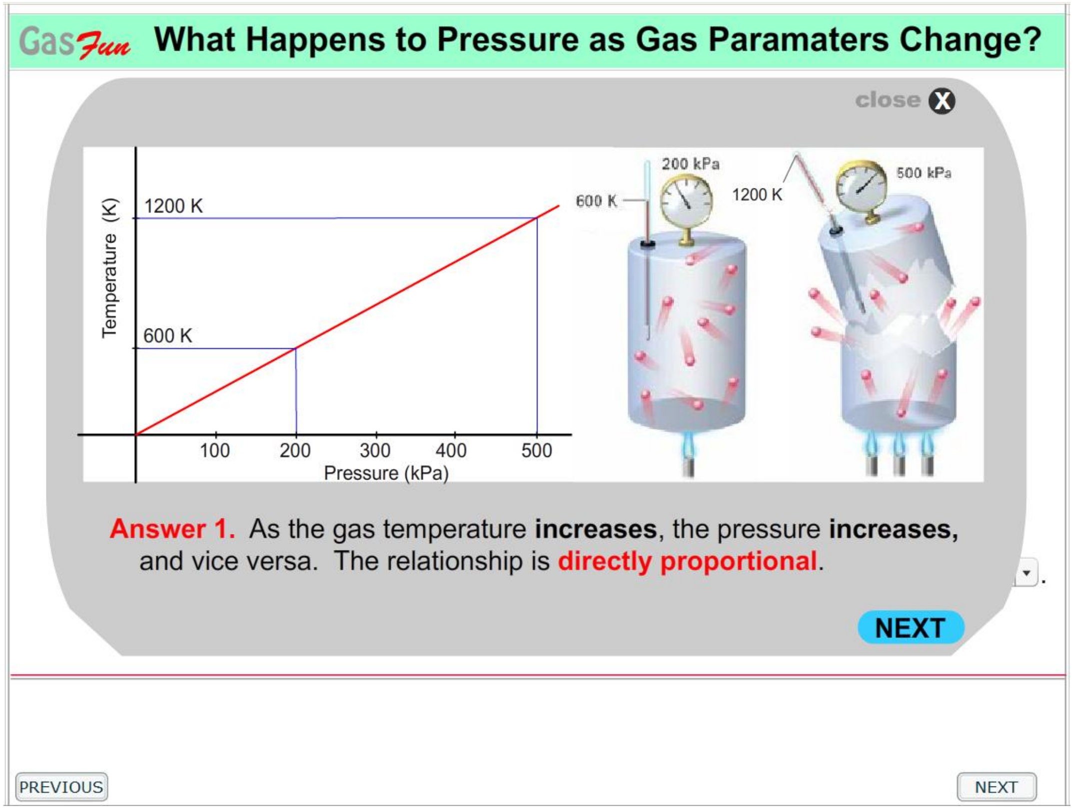


Figure 3.10 The screen shot illustrating the strategy of enabling translation between multiple representations of gas laws and proportional reasoning

Similar to Experiments 1 and 2, students were asked to describe in their own words what happened to pressure if volume or temperature changed, and summarize the newly acquired information by selecting responses from dropdown menus (See Figures 3.11 and 3.12). Making students select answers from dropdown menus served as a means of highlighting important information and helping students summarize the crucial concepts.

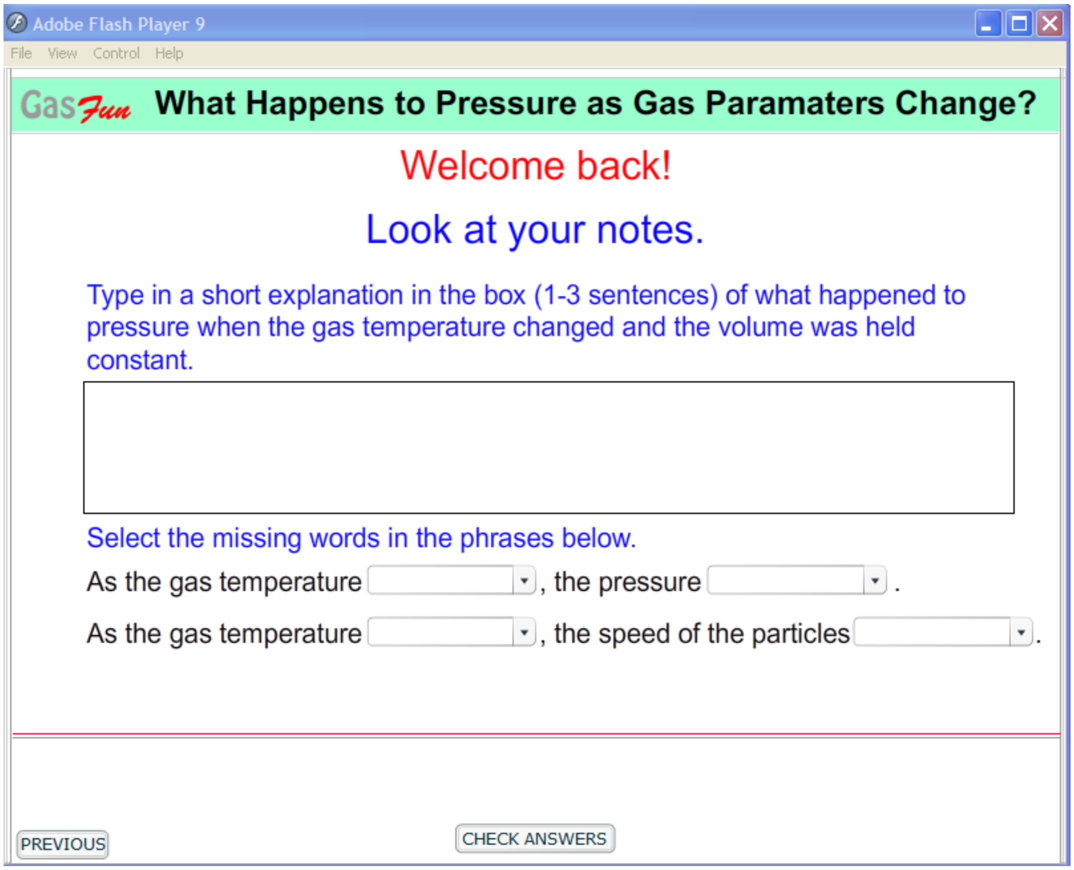


Figure 3.11 The screen shot illustrating the strategy of making students summarize the information acquired in Experiment 3

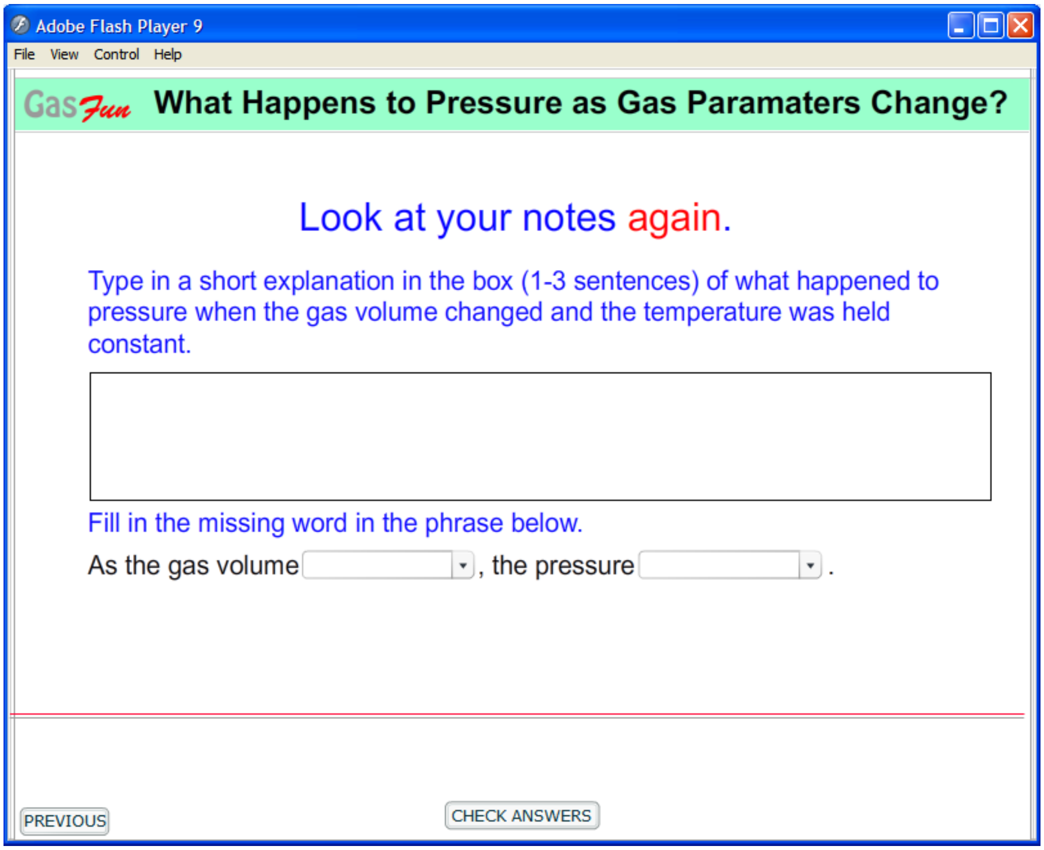


Figure 3.12 The screen shot illustrating the strategy of making students summarize the information acquired in Experiment 4

After that, the students were asked the same question about what might affect gas pressure: type of gas, temperature of gas, mass of gas, volume of gas, speed of gas molecules, and the number of gas moles (see Figure 3.3). This strategy was used to track the changes in students' understanding of gas behavior that could be attributed to their 40 minute experience with simulations and Flash-based formative assessment (the Program). Unlike the first time when students were asked to check their predictions in the simulations lab, this time, the elaborative feedback was presented to clarify the remaining students' misunderstandings of gas behavior, if any (see Figure 3.13).

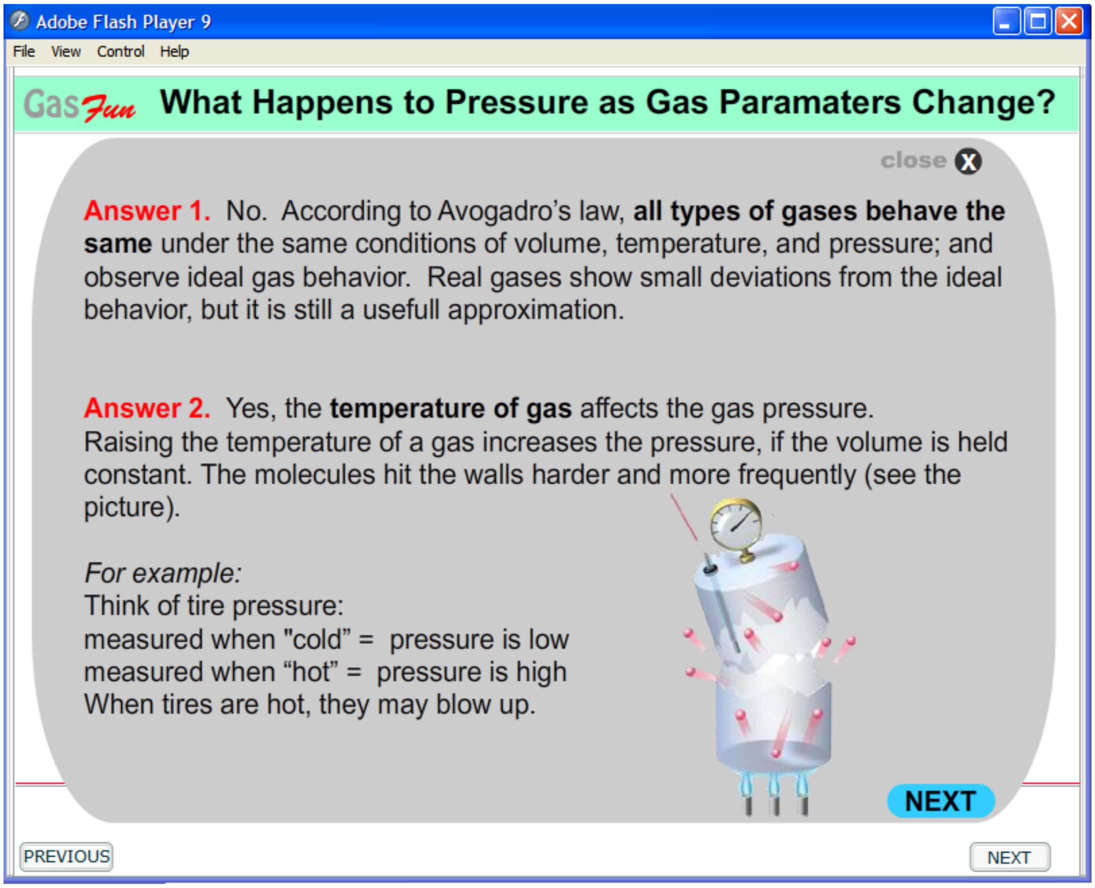


Figure 3.13 The screen shot illustrating the elaborative feedback clarifying possible remaining students' misunderstandings of gas behavior

On the last screen in the Program, students were asked to retrieve the target concepts from long-term memory by using a different context (see Figure 3.14). This strategy was used to enhance retention of the target information. As mentioned by Roedriger and Karpicke (2006) in their review of testing effects, "just as measuring the position of an electron changes that position, so the act of retrieving information from memory changes the mnemonic representation underlying retrieval—and enhances later retention of the tested information" (p.182).

Finally, as illustrated in several recent studies by Kornell, Hays, &Bjork (2009), using materials that ensured unsuccessful initial retrieval attempts, even failed tests can potentiate the effectiveness of subsequent study opportunities.

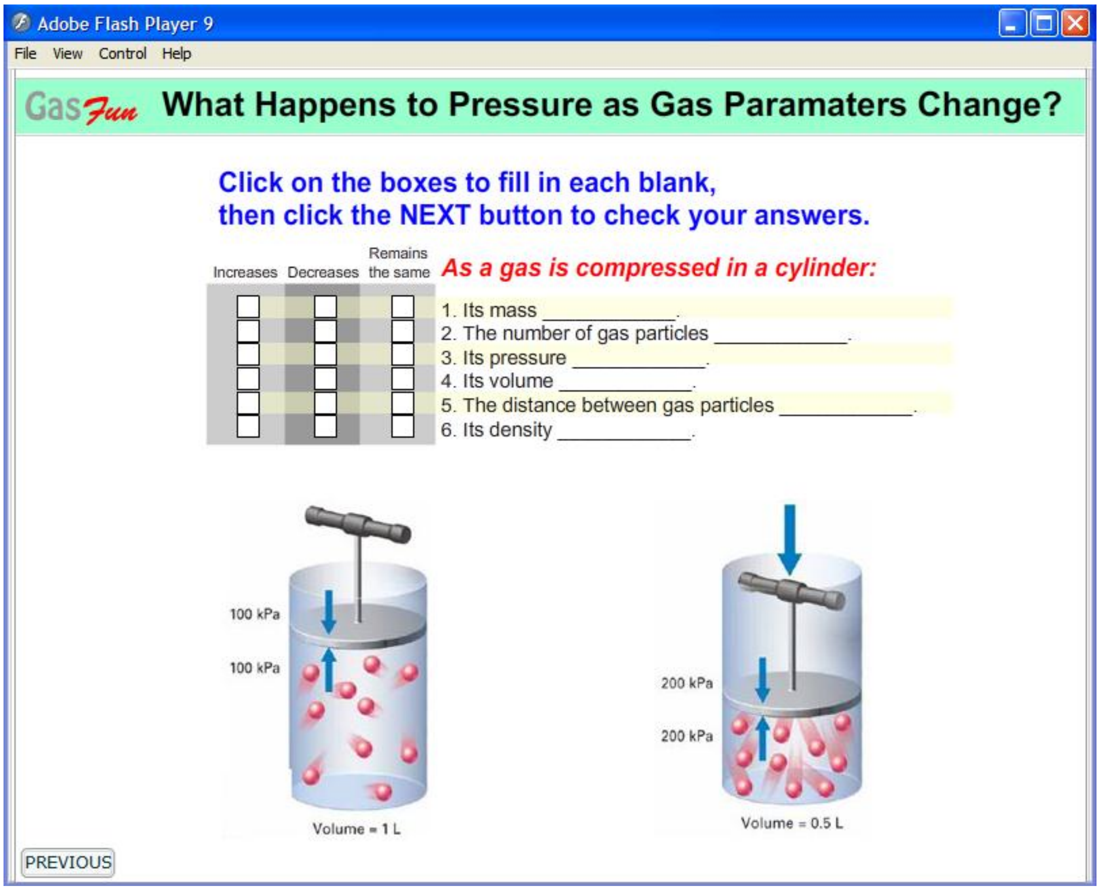


Figure 3.14 The screen shot illustrating the strategy of making students retrieve the target concepts from long-term memory by using a different context

In summary, all the scaffolding steps before and after the students' experiments in the simulation lab were designed to enable translation between multiple representations of gas laws concepts. Also, making students retrieve the target concepts all the way through the Program and at the end of the lesson was expected to ensure the effectiveness of subsequent study of gas laws on the macroscopic level.

Methods

Data Collection

The Program was tested in five 45 minute high school science classes as a part of their science curriculum. The 45 minute lessons using the Program and the Gas Laws PhET simulation were delivered in each of the five classes over a two week period. The data from the three sources: students' responses to three open-ended questions, students' survey ratings of the Program features, and a semi-structured interview with the instructor, were analyzed for recurring themes related to the helpfulness of the Program features and instructional strategies. The data were triangulated to ensure the reliability and validity of the instruments.

As a part of the classroom period, five out of the 45 minutes of the classroom time was used by students to answer four open-ended questions and rate the Program features in paper-based surveys. Students were asked to reflect on their experiences with the Program features and their overall educational experience with the Gas Laws PhET simulation augmented with the Program. The results are presented by research question.

Research question 1: What are the students' perceptions of how the Program supports their cognitive processing of the target concepts?

The results of the students' answers to three open-ended questions and students' ratings of the Program features are presented by question. The themes emerged from students' comments are presented in priority order, the most frequently highlighted themes being at the top of the list.

The list of open-ended questions:

1. What did you like about the program? How did it help you understand gas behavior?

2. What can be done to the program to make it better?
3. Would you recommend this program to other students who want to understand the gas laws on the microscopic level?

Research question 2: Which program features and instructional strategies implemented in the Program were the most and least helpful?

Students' average ratings of the features of their experience (5-strongly agree, 1-strongly disagree) were calculated. The following features were rated: pop-up theory explanations, images, simulations, guiding questions, and elaborated feedback.

Research question 3: What are the teacher's perceptions of how the objectives of the lesson were met?

The results of the semi-structured interview with the collaborating teacher were analyzed for recurring themes.

The list of questions to the teacher:

1. How easy was it for students to use the program?
2. Was the 45minute period enough for the use of the program?
3. Were there any unexpected outcomes?
4. How did the program help students learn?
5. Are there any comments you would like to share with the instructional designer?
6. How engaged were the students?
7. What do you think about the overall instructional design of the program?

Data Analysis

Qualitative methods were used to analyze the data obtained from students' answers to the three open-ended questions. Qualitative Data Analysis (QDA) is the range of processes and procedures whereby we move from the qualitative data that have been collected into some form of explanation, understanding or interpretation of the people and situations we are investigating (Lewins, A., Taylor, C. & Gibbs, 2005).

Content analysis of the students' responses to the first open-ended question started with simple coding for the identification of emerging patterns in the students' responses. After that, simple encoding was used to break the data into categories. In this way, the data were reduced to emergent categories. "Composite" responses that reflect the content of all the responses in each category were created and tabulated along with the count of students whose responses fell into certain categories. Some students' comments included more than one theme and fell into more than one category. The emerging themes from students' comments were compared to the instructional designer's assumptions underlying the instructional design decisions. This approach allowed the designers to test their instructional design decisions. The themes emerging from students' comments are presented in priority order, the most frequently highlighted themes being at the top of the list. One hundred and nine out of 114 students provided their comments. The results are presented by research questions in Tables 3.2, 3.3, 3.4 & 3.5.

In addition, pattern coding that identified the relationship between categories was applied. The researchers looked for patterns for high prior knowledge students (those who came to this class from advanced placement physics classes) and low prior knowledge students (those who come to this class from biology classes and did not have a background knowledge of physical chemistry concepts).

Results

Research question 1: What are the students' perceptions of how the Program supports their cognitive processing of the target concepts?

Open-ended question 1: What did you like about the program? How did it help you understand gas behavior?

Table 3.2 The themes from students' comments about how the Program helped them learn the target concepts

#	The themes from students' comments (composite responses)	Number of users commenting on the theme
1.	Changing variables in the simulation lab helped me understand gas behavior	71
2.	The program images and simulations helped me visualize gas behavior	36
3.	The design of the program made it easy to understand gas behavior	34
4.	It was fun and educational at the same time, a different way of learning	15
5.	The program provided interactive way of learning	10
6.	The program was easy to use	8
7.	I could learn at my own pace	7
8.	The guiding questions were informative	6
9.	The program provided important information and allowed me to test it by myself	5
10.	The diagrams and images in the program helped me understand the concepts	5
11.	The program design helped me retain information	3
12.	The explanations were simple and easy to understand	3

Open-ended question 2: What can be done to the program to make it better?

Students' responses were organized and sorted into emergent categories. The "composite" responses that reflected the content of all the responses in each category were tabulated along with the count of the students commenting on each category.

Table 3.3 The themes from students' comments about how the Program design can be improved

	The themes from students' comments	Number of users commenting on the theme
1.	The program is good the way it is now	40
2.	Making guiding questions easier and less confusing	9
3.	More explanations on the gas laws at the beginning	7
4.	More information, go into more depth	7
5.	More simulations	6
6.	Adding more guiding questions in the program	6
7.	Having fewer slides in the program	5
8.	Having clearer and shorter explanations	5
9.	Better instructions of how to use simulations	5
10.	Adding main review points at the end	5
11.	Using more simple language	5
12.	Making guiding questions harder	4
13.	Adding sound	4
14.	Making the program a little more entertaining	2
15.	Adding more visuals	1
16.	Adding more colors	1

Open-ended question 3: Would you recommend this program to other students who want to understand the gas laws on the microscopic level? Why?

Students' responses about why or why not they would recommend the Program are presented by the rating of 5, 4, 3, 2 or 1 in Table 3.4. The themes from these responses were categorized into the emergent categories and presented as "composite" responses. The counts of students' responses per category were tabulated as well as for providing a clear picture of what worked and what did not work for different students. The count of users commenting on different categories provided insights on which instructional strategies worked for the majority.

Table 3.4 The themes from students' comments about why they would recommend this program

Rating (5 strongly recommend, 1- not recommend at all)	Number of students who gave this rating	The themes from students comments (some students' comments include more than one theme)	Number of users who gave this comment
5	48	1. The program made is very clear and easy to understand the concept of gas behavior	15
		2. It is an easy hands-on simulation	11
		3. It helped me because of good visuals	11
		4. The explanations make it understand easily	8
		5. It is fun and helpful (engaging approach to learning)	4
		6. Very informative and a good way of communicating information	3
		7. Puts the ideas of gas into perspective, helps make connections to the previous topics	2
		8. Good idea of asking questions to make sure that the students are getting what they are supposed to get	2
		9. It is super-interactive	1
4	49	1. Pretty useful and effective learning tool	10
		2. It explains gas behavior well, helps understand gas behavior	8
		3. Simulations show gas behavior well	7
		4. It is good for its visual effects	4
		5. It is good, but a little confusing at times	4
		6. It is pretty good but could be a bit more engaging to involve students more	2
		7. It is much better than reading the concepts out of the book or listening to the teacher	2
		8. It makes learning easy	2
		9. It is good, but you would most likely want experience outside the program to learn	2
		10. It helps a lot, but there is room for improvement	1
		11. I am visual/verbal learner, I would like the teacher explain gas behavior, but I did learn	1
		12. The questions were a really good add in	1
		13. Hands on experience is good	1
		14. It is easy to use	1
		15. It needs some outside information before this lesson	1
		16. It has some good information and can help students develop accurate concept of gas behavior, but I am concerned that some of it could be easily misunderstood	1

Rating (5 strongly recommend, 1-not recommend at all)	Number of students who gave this rating	The themes from students comments (some students' comments include more than one theme)	Number of users who gave this comment
3	10	1. It helped me understand, but still a little bit confusing	3
		2. It is OK but anything really cool and special	2
		3. It conveyed the message semi-well. I myself learn better with just note-taking, reading lectures, but others probably like interactive programs	1
		4. I would rather have a teacher explain it to me	1
		5. If you need extra help, this is a good program, but may not be necessary for all students	1
		6. It was pretty simple, not too difficult	1
		7. It is a good program if you know it already, but otherwise it is OK	1
2	1	It did not help me that much	1
1	1	It did not help me at all.	1
Ave:4.30	Total:109		

Research question #2: Which program features and instructional strategies implemented in the Program were the most and least helpful?

The average students' ratings of the Program features and standard deviations of their ratings were reported. Based on the ratings, judgments about the necessity and helpfulness of the Program features were made. One hundred and ten students rated the Program features. Students' average ratings of the Program features (5-strongly agree, 1-strongly disagree) are presented below in Table 3.5.

Table 3.5 Students' average ratings of program features

Pop-up theory explanations	Images	Simulations	Guiding questions	Elaborated feedback
3.9 (0.8)	4.2 (0.8)	4.7 (0.6)	3.7(0.9)	3.8 (0.9)

Research question #3: What are the teacher's perceptions of how the objectives of the lesson were met?

The teacher found her students' learning experience with the Program effective and efficient. She mentioned that the level of students' cognitive engagement was high. According to the teacher's comment, the Program was easy to use, but it was a little long. The teacher's suggestion was to shorten the Program by reducing the number of experiments or having the students do Experiments 3 & 4 for one type of species, heavy or light, instead of both types. The most helpful program features identified by the teacher were: scaffolding questions, simulations, organization of the target information, self-paced learning, and explanatory feedback.

The teacher found the PRINT option at the end of the Program very helpful because giving students an option of printing the results of the training was valuable for students who would not be able to take the training in class. They can print out the results of the training and show them to the teacher. The results of the semi-structured interview with the teacher, the interviewer's questions, and teacher's supportive comments are presented in Appendix C. The nature of the researcher's claims is interpretative.

Discussion

Both the teacher and the students (except for two students) found the Program an effective and efficient way of learning the gas laws concepts on the microscopic level. The vast majority of the students (42.5% of the participants strongly recommended and 42.5% recommended the Program to other students) reported a positive experience and described it as worth spending their time (see the one of the most common comment below).

I thought that this program was set up very well. It helped me understand gas behavior because it provided good information, along with having me get to test it out myself in the simulation lab.

The participants found the hands on experience of manipulating gas parameters supported by the Flash-based Interactive Program helpful and enjoyable. The combination of the PhET simulation lab and the Program made their understanding of difficult concepts easy and helped them generalize the concepts. The students enjoyed the guidance provided by the Program and their self-paced learning. They considered their experience a creative and motivational way of learning. Overall, they valued the combination of different features and instructional methods supporting their learning. The majority found the concept explanations easy to understand. Many students mentioned the efficiency of their learning using the Program.

Even though the majority of students thought the combination of the Program and simulations helped them learn about gas behavior and their average rating of recommendation to other students was 4.24 (5-strongly recommend, 1-not recommend at all), there were two persons students who did not share these views. One of them commented:

Honestly, the program was more trouble than its worth, it would of been simpler, and more effective to have just straight out taught us the concepts.

The rest, 13%, of the students were partially satisfied with their experiences. The possible explanation of the difference in opinions might be the difference in students' prior knowledge of the target concepts. As specified by the teacher during the interview, several students out of 114 came from Advanced Placement classes and had the background knowledge that drastically exceeded the prior knowledge of the rest of the class. They would rather spend more time with designing their own experiments in the simulation lab. In contrast, there were some low prior knowledge students who needed direct instruction of the target concepts before they could work with the Program. The Program features found effective by students and supportive students' comments are presented in Table 3.6.

Table 3.6 Effective instructional design features and representative student's comments

Instructional design features	**Representative student's comments**
1. The hands-on experience with manipulating gas parameters	"The simulations helped me understand the relationship between volume, gas, and pressure. The questions asked also were informative."
2. The combination of program and simulation lab environments makes understanding of difficult concepts easy and helps generalize the concepts	"I thought that this program was set up very well. It helped me understand the gas behavior because it provided good information, along with having me get to test it out myself."
3. Program and simulation features supporting the visualization (simulations, images, diagrams)	"I like the idea of different pictures in the explanation pop-ups, the same application, but the pictures are different from the gas laws simulations, it helps generalize the concept, to see it from a different angle."
4. Guidance provided by the program	"Structure is good, asking questions and making me fill in their responses is good, structure helps me focus on important information, the combination of structure and freedom to make mistakes is good. If your prediction is wrong, the correct answer sticks better."
5. Creative and motivational way of learning	"It's a fun way of learning about gas behavior and it actually works."
6. The ease of use	"I liked that it was simple and interactive."
7. Self-paced learning	"You can learn at your own pace and the program is very visual."
8. Combination of different features and instructional methods	"Combination is really nice, that seems to pull everything together."
9. Simple explanations of difficult concepts	"The program addresses the main point, but does not go to detail too far so that you get confused overall."

Instructional design features	Representative student's comments
10. Efficiency of the program	"Simulations do not take as much time as real labs, but they help you grasp the concept quickly and better. Given the time it took, I am impressed with how much I learned."

Conclusions

The study provided preliminary evidence that the Program was an important scaffolding and formative assessment activity that made students' learning effective and enjoyable. Overall, the Program helped 50% of students to change their misconception and grasp the challenging concept that most gases behave almost the same way under the same conditions of volume, temperature, and pressure.

Interestingly, the teacher indicated during the interview held two weeks after the Program use that it helped students transfer the gained knowledge to other contexts. While solving problems on the macroscopic level, the students were referring to gas behavior on the microscopic level.

As to the opinions on what helped the most during the lesson, the teacher mentioned the formative assessment, the guiding questions and the directions for conducting experiments that helped the students learn better. The students valued the simulations the most. The difference in opinions might be explained by the fact that the teacher used the Gas Laws PhET simulations before, without having the interactive formative assessment program. When asked if the students' experiences last year were different from this lesson, the teacher admitted that they were different.

This is definitely different, because they had more guidance when they were working through the simulations. And they also had directions in the handouts. I would say overall more students got more out of this experience.

One of the problems that needs to be addressed is to make the Program design better address the needs of learners with different levels of prior knowledge of the domain. Also, students' performance measures need to be collected and statistically analyzed in the next study so that causal relationships can be drawn.

Suggestions for Program Modification

1. Design an alternative program for High Prior Knowledge (HPK) students (those who come from advanced placement physics classes).
   a) Provide more information that explains gas behavior on a higher level of complexity.
   b) Add more assignments with simulations that let students come up with the strategies to test their hypotheses.
   c) Make guiding questions more challenging.
2. Make changes to the existing program so that it better addresses the needs of Low Prior Knowledge (LPK) students (those who came from biology classes and do not have background knowledge of chemistry and physics).
   a) Add more explanations of gas behavior as pop-up windows at the beginning of the Program.
   b) Provide clear rationale for students about why they are asked certain questions to guide their learning.

c) Provide more detailed instructions supplemented with images in the handouts about how to use simulations.

d) Add main review points at the end of the Program.

Implications for Education

This study addressed a need to more effectively assess progress in knowledge development for the domain of physical chemistry. "Science teachers seek engaging, effective, and inquiry-based activities that are standards-aligned and convenient to implement in their classrooms" (Limson, Witzlib, & Desharnais, 2007).

References

Adams, W. K., Reid, S., LeMaster, R., McKagan, S. B., Perkins, K. K., Dubson, M., and Wieman, C. E. (2008). A Study of Educational Simulations Part I - Engagement and Learning. *Journal of Interactive Learning Research*, *19*(3), 397.

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.

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

Eylon, B., Ronen, M., & Ganiel, U. (1996). Computer simulations as tools for teaching and learning: Using a simulation environment in optics. *Journal of Science Education and Technology*, *5,* 93–110.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 989-998.

Lawson, A. E., Baker, W. P., DiDonato, L., Verdi, M., & Johnson, M. (1993). The role of hypothetico-deductive reasoning and physical analogies of molecular interaction in conceptual change. *Journal of Research in Science Teaching, 30*(7), 482-495.

Lewins, A., Taylor, C. & Gibbs, G. (2005). *What is qualitative data analysis?* School of Human & Health Sciences, University of Huddersfield. United Kingdom.

Limson, M., Witzlib, C. & Desharnais, R. (2007). Using web-based simulations to promote inquiry. *Science Scope*, *30*(6), 36-42.

Marzano, R. J. (2007). *The Art and Science of Teaching: A Comprehensive Framework for Effective Instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.

The PhET Team, (2006). PhET Interactive Computer Simulations [Computer software] Available at: http://phet.colorado.edu Colorado: Physics Education Technology Project.

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210.

Shavelson, R.J., Phillips, D.C., Towne, L., Feuer, M.J. (2003). On the science of education design studies. *Educational Researcher, 32*(1), 25-28.

Spector, J. M., Merrill,M. D., van Merrienboer, J., Driscoll, M. P.(Eds.) (2008). *Handbook of research on educational communications and technology*. New York, NY: Lawrence Erlbaum Associates.

Sweller, J. (1999). *Instructional design in technical areas*. Melbourne: ACER.

Sweller, J., & Chandler, P.(1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185–233.

APPENDIX A: EXPERIMENTS 1 & 2

What happens to pressure as gas amount (gas mass) changes?

What?

In this experiment, you will change the gas mass (add more particles) and observe the pressure while keeping the volume constant. This simulation allows the use of a gas with heavy particles (heavy species) and light particles (light species). In Experiment 1, you will use *Heavy Species,* in Experiment 2, you will use *Light Species*.

Why?

You will observe the pressure of two gases (with heavy and light particles) and how the pressure is affected by the number of gas particles (gas mass). Is it the same or different for both gases? Is the difference large or small?

Experiment 1

Follow the instructions:

1. Click on RESET to make sure that pressure and temperature are at the lowest level.
2. Keep *volume* constant: click on Volume.
3. Change the mass of the gas by adding 50 Heavy Species.
4. There are two ways to add species: by using the pump handle (to add the rough number of gas particles and by using the up arrow (to adjust the number of gas particles).
5. Keep *Light Species* at "0".
6. Notice that the temperature stays at 300K
7. Observe the movement of particles. Observe the pressure.

Record the pressure readings in the table.

For Heavy Species

Temperature (K)	Amount	Range of Pressure/Average Pressure (atm)
300	50	
300	100	
300	200	
300	300	

8. Add more Heavy Species until the number reaches 100. Keep the Light Species at “0”.
9. Observe the movement of particles, observe the pressure.

Record the pressure readings into the table above.

10. Add more Heavy species till the number reaches “200” and then “300”
11. Observe the movement of particles. Observe the pressure.

Record your responses in the table above.

Experiment 2

Follow the instructions:

Follow the same steps as you did for *Heavy Species*, but this time, keep Heavy species at “0” and manipulate with the number of *Light Species*.

Record your responses in the following table

For Light Species

Temperature (K)	**Amount**	**Range of Pressure/Average Pressure (atm)**
300	50	
300	100	
300	200	
300	300	

APPENDIX B: EXPERIMENTS 3 &4

Experiment 3

What happens to pressure as temperature changes?

Follow the instructions:

In this experiment, you will change the temperature and observe the pressure while keeping the volume constant.

1. Click on RESET to make sure that pressure and temperature are at the lowest level.
2. Set the *volume* as *constant* parameter.
3. Use the pump handle (another way to add species) to pump 400 *Heavy Species* into the container (select the heavy species option below the pump). To pump the exact amount of particles, start with the pump handle and then use the up arrow.
5. Notice that the temperature stays at 300K
6. Observe the movement of particles. Observe the pressure.
7. Change the temperature to 800K by moving the pointer of the *heat control* up (move and hold), observe the movement of particles, observe the pressure.
8. Change the temperature to the point (keep the temperature pointer in the up position) when pressure reaches its possible maximum level.

Record your responses in the following table

Temperature (K)	**Amount**	**Range of Pressure/Average Pressure (atm)**
300	400	
800	400	

9. Observe the movement of particles, changes in pressure and temperature. What happens to the lid? What happens to the pressure and temperature after the lid pops off?

Experiment 4

What happens to pressure as gas volume changes?

In this experiment, you will change the volume and observe the pressure while keeping the temperature constant.

Follow the instructions:

1. Click on RESET to make sure that pressure and temperature are at the lowest level.
2. Set the *temperature* as *constant* parameter.
3. Pump the handle to add some *Heavy Species* (a faster way to add species)
4. Notice that the temperature stays at 300K
5. Observe the movement of particles. Observe the pressure.
6. Decrease the gas volume by moving the handle to the right (the little man moves to the right).
7. Observe the speed of the particles, the temperature, and the pressure. Describe what happened.

____
____
____
____

8. Notice that after "squeezing the gas" in the container, the temperature goes up and then moves back to 300K because it is set to constant (300K)
9. Increase the gas volume by moving the handle to the left (the little man moves to the left)
10. Observe the speed of the particles, the temperature, and the pressure. Describe what happened.

____
____
____
____

11. Change the volume to the point when pressure reaches its possible maximum level.
12. Observe the movement of particles, changes in pressure. What happens to the lid? What happens to the pressure and temperature after the lid pops off? Describe.

____
____
____
____

APPENDIX C: TEACHER'S INTERVIEW QUESTIONS AND ANSWERS

Researcher's questions and teacher's comments	Researcher's explanatory comments	Researcher's claims
Q #1. How easy was it for students to use the program? "It went really well. I had 114 students working with the simulations in the lab. And there were only three students who closed too many windows. I told them ahead of time about it. And they were careful with not closing too many windows".	Since the PhET simulation was not imbedded in the program design, students had to go to the PhET site twice and return back to the program by closing the windows from the PhET sim lab.	It was easy for students to use the program
Q #2. Was the 45minutes period enough for the use of the program? Timewise the thing that I would probably change in the future is that 4 experiments probably would be too many. OR We may not need to do Experiment 3 and 4 with both types of species, light and heavy. Just one type will be enough.	During the pilot testing with three participants, timing was not a problem. It took the participants 40, 35, and 30 minutes correspondingly	Experiments 3 and 4 need to be a little shorter
Q #3 Any unexpected outcomes? a) "The only other thing that happened that the printer in the media center did not work very well. Some students were not able to print the results of the final multiple choice question for their reference." b) "There was one kid who finished in three minutes. In five minutes he was going to the printer to print the stuff. There was no way he could not even have read the questions."	According to teacher's comment, giving students an option of printing the results in the program was good. "For students who were absent I wanted it to be printed out to show that they have done it."	Logistics problems and student attitude problems are possible

Researcher's questions and teacher's comments	Researcher's explanatory comments	Researcher's claims
Q #4.How did the program help students learn? *Questions-answers format* "I think that it helped students learn because they had to look at the questions that they had to put an answer to. And the ones that I was watching closely " were not just choosing the answers from dropdowns, they were typing in the responses the text box and then choosing the dropdowns to make sure that they will match it." *Simulations* "They also think that the simulations side when they could put a different number of molecules in and see what happens. That was fun! It was nice to pump molecules in and see the top popped up.The interactive part of it was really good." *Information presentation* "And in the way that experiments were designed: let them control some things and vary some things and look more systematically on how it was going on. It was organized well to help them learn." *Independent self-paced learning* "It is the other thing that is really valuable for me is that I have students coming straight from biology and had not had any experience with gas laws and have students coming from AP physics who have a lot of gas laws experience. And I tell them: this is where we start it. And let them work at their own pace based on what they know, what their prior knowledge is. This is instead of me telling them that they have to wait and get my timeline. It gives them a stance to pick up really quickly and let them work on their own." *Feedback* "The feedback was sufficient. And I noticed a lot of them went back to the previous pages so that they could solve the problems. And the thing that when they were solving the question about Avogadro's law, they learned that both of them were right. So it was good" Making predictions and checking them in the lab It was absolutely good idea		Scaffolding questions, simulations, organization of the target information, self-paced learning, and explanatory feedback were effective learning tools

Researcher's questions and teacher's comments	Researcher's explanatory comments	Researcher's claims
Q #5 Any comments you would like to share with the instructional designer. a) While working with the program students did not ask many questions b) "When we came back one of the things that I asked them in class was if we have 4 particles and change it to 2 particles and the volume in the box is the same, the pressure is the same, what happens to the temperature? And 30% who would say that the temperature would decrease. And the rest would say that the temperature would increase. But it was not awful. I am sure that would happen if I taught it to them too."	Almost one third of the students gave the wrong answer to the teacher's question.	The program needs to provide more scaffolding on making students understand the relationship between gas pressure, temperature, and the number of gas particles
Q #6 Was the students' experience with gas laws in this semester different compared to the experience of other students (in the previous semester)? 'This is definitely different, because they had more guidance when they were working through simulations. They also had directions for the experiments that they needed to conduct. They had a measurable outcome when working through the slides answering questions. This was really nice. I would say more students got more out of this experience. I would say it was a fun and easy way for them to be accountable."	The same PhET simulation, but without interactive formative assessment (the Program) was used by the teacher before.	Program made a difference in students' understanding of gas laws.
Q #7 How engaged were the students? "The students were very much engaged. They worked hard for 45 minutes. They were thinking and writing things down"		The level of cognitive engagement was high
Q #8 Overall instructional design of the program a) **"**They had a good experience with the program and that brought everybody up to speed so that I do not need to spend more time on explaining gas laws." b) "It did exactly what I wanted it to do. And it helped them look at gas laws at their own pace with their individual backgrounds and come back and be able to understand that what 's unique about gases is that they are not unique, they all behave almost the same way." c) "It was organized well to help them learn. It helped them learn more systematically." d) "The general feeling at the end of the class was that students found it interesting and worth the time they spent on it."		It was an effective and efficient way of learning

CHAPTER 4: A DESIGN STUDY OF A MULTIMEDIA INSTRUCTIONAL GRAMMAR PROGRAM WITH EMBEDDED TRACKING

Abstract

This is a design study meant to demonstrate the feasibility of integrating three rather different theoretical perspectives for future efforts in multimedia instructional design. A multimedia instructional grammar program contextualized within the teaching of English as a Second Language (ESL) was developed and evaluated. The program design was grounded in Mayer's multimedia learning theory (2001), Sweller's cognitive load theory (CLT, 2005), and cognitive training theory using an inductive reasoning paradigm (Klauer & Phye, 2008). A successful integration of cognitive training theory into program design is expected to facilitate the transition of student's declarative knowledge of a grammar concept of passive voice to procedural knowledge (Phye, 1991, Phye et al., 2005). Two studies involving ten and four adult ESL learners were conducted in a Midwest community college. Grammar teaching occurred within the context of history and geography of the USA. Students with low prior knowledge of passive voice grammar concepts, intermediate level of general vocabulary, and adequate basic knowledge of content (basic geography and history) benefited most from the program. Preliminary results are encouraging for the aforementioned integrative efforts.

Introduction

Historically, the teaching of English grammar can be identified with two approaches. The deductive approach focuses on mastering grammar form (Biber et al., 1999; Bresnan, 2001; Chomsky, 1986; Halliday, 1994; Huddleston & Pullum, 2002; Langacker, 1987, 1991; and Quirk, 1985). The inductive approach has been identified with Krashen, 1988. Rosetta

Stone, language-learning software, produced by Rosetta Stone, Ltd. uses inductive approach to teach grammatical functions without drills or translation by applying ‘‘Dynamic Immersion method.’’ Learning occurs naturally through direct association of words, structures and images (Marcy, 2007).

Practitioners around the world often use modified deductive or modified inductive approaches. The techniques used in these approaches vary according to whether explanation of the rules takes place before (modified deductive) or after practice (modified inductive) and according to the degree of guidance the students are given in working out the rules. Combined approaches take several forms and are frequently seen in the design of computer-assisted language learning (CALL) approaches. Common examples would include Azar Grammar Series (Azar, 2007) and the Grammar ROM (Freebairn &Parnall, 1995).

As computer technology has advanced, more sophisticated designs such as natural language processing techniques (NLP) have evolved (Nagata, 2002). Assessment-Based Learning Environments (ABLE) represent the most advanced intelligent tutoring programs that use the assessment information coming from formative and summative sources to help English language learners (ELLs) learn about English grammar. ABLE are adaptive scaffolded learning environments. The main functional features of English ABLE include item/task reuse items, a Bayesian psychometric student model that makes use of item statistics, adaptive feedback, adaptive sequencing of tasks, pedagogical agents, and indirectly visible student model (Zapata-Rivera, 2007).

Despite all of the advantages of NLP techniques (Nagata, 2002) and ABLE (Zapata-Rivera, 2007), the former does not focus on helping students develop conceptual understanding of grammar concepts, and the teaching of grammar does not occur within a cultural or academic learning context such as geography or history. Further, ABLE is designed for advanced ELLs.

The web-based multimedia instruction for teaching English grammar program described in this paper was designed to address the aforementioned problems. Also, it is common knowledge that passives have always presented a challenge to ESL/EFL teachers. Consequently, the grammar concept of passive voice was chosen as the procedural knowledge skill for testing the robustness of the multimedia instructional design elements.

Because effective instructional design is typically based on a design-engineering-develop approach to innovation (Bryk & Gomez,2008), the first step was the identification of cognitive theories that promote learning, the next step was the implementation of these theories in the design, and the last step was evaluation of multimedia design decisions grounded in these models.

Literature Review

The cognitive theory of multimedia learning by Mayer (2001) has shown that learners are better able to transfer their learning when given multimodal instruction. According to the cognitive theory of multimedia learning, ‘‘multimedia presentations have the potential to result in deeper learning and understanding than do presentations that are presented solely in one format’’ (Mayer, 2001, p.68). Multimedia presentations allow learners to hold verbal and pictorial representations at the same time, thus increasing the chances that learners will be able to build mental connections between them. Building connections between verbal and pictorial mental models is an important step in conceptual understanding. Therefore, students who receive well-constructed multimedia messages should perform better on transfer tests.

In addition to providing multiple retrieval cues as proposed by Mayer, schema development is a major part of developing procedural knowledge. According to Klauer and Phye (2008), cognitive training activities stimulating students' cognitive abilities of finding similarities and/or differences among attributes and relations of objects can significantly improve learning by facilitating the construction of inductive reasoning schemas. Moreover, the development of an inductive reasoning schema is knowledge from a well understood domain to one that is unfamiliar (from the context of geography of the USA to any other contexts as a measure of between-domain transfer).

In summary, there is a place for both explicit and implicit learning in teaching ESL/EFL as well as both declarative and procedural knowledge because the major role of teaching ESL/EFL is to help learners develop knowledge that is both easily accessible in real-time interaction and can be generalized and used in different contexts. In this research, an instructional design strategy that combines the advantages of both implicit and explicit learning and speeds up the process of "proceduralization" of declarative knowledge is implemented in the design of a web-based multimedia program for contextualized teaching English grammar.

The third model that served as a foundation for the program was Sweller's cognitive load theory (CLT). ''Cognitive load theory (Paas et al., 2003, 2004; Sweller, 1999, 2003; Sweller et al., 1998) and the instructional principles it has generated are all based on the assumptions concerning human cognitive architecture'' (Sweller, 2005, p.26). CLT is a universal set of learning principles that are proven to result in efficient instructional environments as a consequence of leveraging human cognitive learning processes (Clark et al. 2006). Sweller's theory emphasizes the limitations of human's working memory load on learning during instruction.

‘‘Novices need support from the instructional environment to substitute for their lack of schemas’’ (Clark et al. 2006). At intermediate levels, an optimal executive function should be knowledge-based when dealing with familiar elements of information and externally based when dealing with new elements of information (Kalyuga, 2007, p.390). In both cases, students need to be provided with different levels of scaffolding because working memory is capacity-limited.

A multimedia instructional program grounded in engineering activities and cognitive theories was developed. It was expected to promote conceptual understanding and knowledge transfer of target grammar concepts and assist users in several ways:

- Help learners build connections between verbal and pictorial mental models (grounded in Mayer's multimedia theory).
- Facilitate the transition of student's declarative knowledge of a grammar concept of passive voice to procedural knowledge as well as analogical reasoning and transfer by providing cognitive training: matching situations that share both attributes and relations between attributes (grounded in Klauer & Phye's cognitive training theory).
- Provide scaffolding activities and allow students with different levels of expertise adapt to instruction by switching from one instructional method to the other (from modified inductive to modified deductive and vice versa) when they need to maximize their intellectual performance (grounded in Sweller's CLT).

The focus of this project is to describe the design and evaluation of a low-cost multimedia instructional program that provides an interesting and meaningful environment to learn ESL grammar consciously, and geography and history of the United States incidentally. The goal was to design a simple program that has linear navigation and predetermined, explanatory feedback and effectively helps users master grammar concepts by combining the advantages of both modified deductive and modified inductive approaches for teaching ESL grammar.

The modified deductive teaching occurs when students read the explanations of grammar concepts from grammar help pop-ups activated by students' clicking on the Grammar Help button. After that, students reinforce their understanding of grammar concepts by going through a set of compare and contrast cases and exercises that help them build cognitive schemas. The modified inductive teaching occurs when students go through the compare and contrast cases and exercises (cognitive training) first, and at the end of each unit, read the summary of grammar concepts taught within the unit.

The program tracks learners' use of performance aids, which allows the designers to see when, how often, at what particular moment during the training, and how long students use performance aids (the number of milliseconds spent on a screen). These data can help the designer justify the implementation of those aids. Also, the data are analyzed for patterns to determine how the performance aids are used by students with different level of expertise. Pop-up validation error messages on each screen are counted. The screens with many validation error messages are considered for redesign.

In order to test how the implemented ideas worked for ESL learners, data were collected to see if the program delivered its promise. The engineering questions in the study were focused on the following issues:

1. What is the impact of the intervention on students' ability to use present and past simple passive voice in affirmative sentences for describing the USA?
2. Does the design of the program facilitate the grammar concept knowledge transfer to contexts other than geography and history of the country?
3. How do students use the program?
4. What is the relationship between students' learning outcomes and their computer skills, between their learning outcomes and prior knowledge?

The engineering section located in Appendix A is a survey of the instructional design literature that served as the empirical basis for design elements incorporated into the current program.

Program Description

Students learn how to apply passive voice for the description of the geographical location, terrain, and some cities in the USA. The idea is to engage the target audience in meaningful project-based learning. The main character, an ESL learner, wants to write a letter to her brother, who is learning English in Mexico, and attach some pictures along with the description of the country. She wants to learn passive voice grammar structures to be able to describe the geography of the USA in her letter. The users of this program are asked to assist her every step of the way through the project. As the final assignment, the users help her describe the pictures of the USA that she wants to send with her letter. The program units are shown in Appendix B. The program can be accessed at: http://training.perl.hs.iastate.edu/esl1.html.

Features in the Program

Students can choose between modified deductive and modified inductive teaching approach (Gollin, 1998) or use a combination of both. The pre-determined interpretive feedback is provided to learners every step of the way through the whole program, so that users can check their answers, see the flaws in their logic, if any, and compare their responses to the correct answers. The assignments lead users through the process of constructing new cognitive schemas and assist students in facilitating analogical transfer.

The implementation of the story-telling format in the program allows the use of informal language in the dialogs between the main characters of the story (Mayer et al., 2004). In contrast, formal language is used in the Grammar Help aids.

Students can use a variety of tools to adjust the instruction to their needs:

- To adjust the pace of the instruction in the system-paced parts (Pause, Play, Replay the last sentence, Replay the whole dialogue).
- To enhance comprehension (word annotations to read and listen to word definitions, image pop-ups, map pop-ups, simple animation) and listening comprehension (captions on/off).
- To learn grammar (Grammar pop-ups and Irregular Verb pop-ups).

Teaching is contextualized so that students can learn about the country and how to apply the target concept for the description of the country. Ninety-five percent of the training appears in the context of USA geography and history, and 5% of the training constitutes Extra Practice exercises at the end of each unit. Contexts other than USA geography and history are used in these exercises. The embedded user-behavior tracking technology collects information about students' performance and navigation patterns. The collected data can be used for further research and provide data to help develop ideas about how the program can be improved (see Appendix C).

The contextualized teaching of grammar is complemented by the assessment that checks students' ability to apply the acquired knowledge. At the end of the program users are asked to describe 12 images. The embedded user-behavior tracking technology collects students' descriptions of the images (see Appendix D). The demographics survey, pre-test, post-test, transfer-test, and Likert survey are embedded in the program and the responses are captured in the database.

Program Evaluation

Method

According to Cronbach (1982), if the goal of the evaluation is to develop a clear understanding of how the program works and how to improve it, one can use multiple small evaluations rather than one large study. The design study consisting of two studies was conducted within 1 year: an initial pre-post test evaluation with ten participants (Study Group A) and user interface analysis with four participants (Study Group B). The research questions in the second study (Study Group B) were driven by the findings from the first study (Study Group A). The research method used the combination of the theory-based approach to interpret evaluative information within a theoretical framework of three cognitive theories underlying the design of the program and evidence based approach to determine students' learning outcomes, perceptions about their use of the program, and frequencies of performance aids use. A non-experimental, descriptive approach was considered appropriate for both experiments. Because of the small scale of the design study, the findings can be interpreted as preliminary evidence before conducting a large scale study.

In order to insure the construct validity of the design, the quantitative data about students' performance and the data from the database about students' use of performance aids were collected to validate the qualitative data about students' perceptions of the program from surveys and debriefing. Numerical and visual descriptive statistics to describe the central tendency (means, standard deviations) were used in the first experiment for the description of students' scores at the pre-test, delayed post-test, and transfer test.

Study Group A

Participants.

Data were collected at a Midwest Community College with 15 adult ESL learners in December 2008. The participants ranged from 22 to 60 years old and were selected because they were easily accessible and available. The level of their English language proficiency: intermediate and advanced. The level of familiarity with passive voice concept: low, intermediate, and high. Participants were from a wide range of native languages, educational backgrounds, and computer experience. They had Chinese, Spanish, Slovak, Korean, Danka, Taiwanese, and Russian as their first languages. However, due to the lack of consistent attendance of five students only ten received the whole treatment. Among the ten students, four have lived in the USA for less than a year, three for less than 4 years, and three for 5 years or longer. The typing and computer skills of the participants ranged from low to average. Research questions and data collection instruments are shown in Table 4.1.

Table 4.1 Research questions and data collection instruments

Research Questions	Data Collection Instruments
1. What is the impact of the intervention on students' ability to use present and past simple passive voice in affirmative sentences for describing the USA?	1. Students' pre-test and delayed post-test scores stored in the database. 2. Students' writing assignments administered within the training episode and saved in the database.
2. Does the design of the program facilitate the grammar concept knowledge transfer to the contexts other than geography and history of the country?	1. Students' pre-test. 2. Students' knowledge transfer-test scores.
3. How do students use the program?	1. Likert-scale survey. 2. Debriefing. 3. Navigation patterns from a database report. 4. Informal observations of program use.
4. What is the relationship between students' learning outcomes and their computer skills, students' learning outcomes and their prior knowledge of passive voice?	1. Demographics survey (computer skills). 2. Pre-test (prior knowledge). 3. Delayed post-test. 4. Transfer-test.

Note Each of the 3 tests consisted of 20 items, 10 of them being selected response (multiple choice) and 10 of them being constructed response (fill in the correct response).

Data collection.

The participants received instruction through the multimedia instructional program in 2 sessions for a total of 2-3 hours without any help from the teachers. All three tests and the training episode took place during the ESL class periods: Wednesday (Day1), Friday (Day3), Monday (Day 6), Friday (Day10), Wednesday (Day 16). Because the instruction could be done at the students' own pace, it took some students longer than others. During the first session, the concepts of passiveness and the use of present tense passive voice were covered. During the second session, the participants learned how to use past tense passive voice, transformation of passive sentences to active, alternatives to passive voice, and suggestions on the choice and use of active and passive voice. One student, who had poor prior knowledge of general geography (observations, debriefing) and had low prior knowledge of passive voice, used the third session to finish the training on the next day at her request.

The data collection process is presented in Appendix E.

Data analysis.

Students' pre-test scores served as indicators of their prior knowledge of passive voice. The difference between delayed post-test scores and pre-test scores served as an indicator of the students' retention of the grammar concept of passive voice. The difference between transfer-test scores and pre-test scores was used to measure students' ability to transfer information to contexts other than geography and history of the country.

All the tests and surveys were embedded in the program and the students' responses were captured in the database. Each of the 3 tests consisted of 20 items total, 10-selected response items and 10 constructed response items. The following concepts were measured through 10 "multiple choice" questions and 10 "fill in the blanks questions" (2 items "multiple choice" and 2 items "fill in the blank" per each of the 5 concepts for a total of 20 questions):

All the items in the tests were designed to check the students' ability to differentiate between:

1. Present active and passive voice
2. 3rd person singular and 3rd person plural present passive voice
3. Present and past passive voice
4. 3rd person singular and 3rd person plural past passive voice
5. Passive voice with regular and irregular nouns

The 10 minute debriefing sessions were conducted individually with each student upon the completion of the survey. The list of the debriefing questions is shown in Appendix E. The participants were encouraged to engage in a learning-focused dialogue about the use of the program. The feedback from the participants was analyzed for recurring themes about the navigation system, visual design, and the language used in the grammar help pop-ups, instructions, and word annotations. In addition, the feedback on the appropriateness of various aids and sufficiency of the implemented instructional decisions was collected from the students. The database was the source of data for the frequency of the use of different performance aids in the program. Informal observations were conducted with four participants with different levels of expertise. The range of expertise of the observed participants contributed to understanding a broader picture of how the program suited the needs of different target audiences. The participants were encouraged to engage in a learning-focused dialogue about the use of the program. The observation sheet is presented in Appendix F. The feedback from the participants was analyzed for recurring themes about the navigation system, visual design, and the language used in the grammar help pop-ups, instructions, and word annotations.

The students' background knowledge of passive voice was determined from the pre-test scores. Students with pre-test scores 10-13 out of 20 and were considered as low prior knowledge students, and those with scores of 17-18 out of 20 were considered high prior knowledge students. The relationship between the users' background knowledge of passive voice and performance on the post-test and knowledge transfer-test was examined.

The writing assignments of low and high prior knowledge students were compared to determine the similarities and differences in the use of passive voice structures, appropriate vocabulary for describing maps and pictures, overall clarity, the use of correct passive voice grammar, the correctness of the factual information. In addition, close attention was paid to students' use of appropriate prepositions for describing locations and articles with proper nouns. There were no exercises in the program to practice prepositions and articles except for the preposition "by" because it was an integral part of the passive voice instruction. The idea of pure exposure to prepositions and articles (students were exposed to many sentences using those prepositions and articles) was checked to determine whether pure exposure was enough for the students to pay attention to articles and prepositions and use them correctly in their writings.

The data from the tests, observations, surveys, and database were triangulated to cross check the accuracy and reliability of the instruments and get a more complete picture.

Results and discussion.

The findings are organized by the four research questions.

Research question 1: What is the impact of the intervention on students' ability to use present and past simple passive voice in affirmative sentences for describing the USA?

An examination of the pre-test and delayed post-test scores revealed that the students' ability to use present and past simple passive voice in the context of geography of the USA on average increased by 3.6 points out of 20 (see Table 4.2).

All the users showed knowledge gain at the retention test (delayed post-test) administered on the fifth day after the training. Interestingly, students who showed the lowest scores at the pre-test gained the most from the program. One can notice that students' performance on the constructed response items shows a larger increase than on the selected response items (see Appendix H).

The examination of students' writing assignments, the descriptions of 12 pictures, revealed that the low prior knowledge students made correct sentences with the verbs which had been practiced the most through the program. Such verbs as ‘‘is/are located, is/are occupied, is/are known, and is/are composed’’ had been used by students in various activities during the training.

Table 4.2 Means and standard deviations of students' pre and post-test raw scores

	Pre-test			Post-test	
Mean	SD	N	Mean	SD	N
14.4	2.675	10	18	1.155	10

Advanced students with high prior knowledge of passive voice were able to correctly use even those verbs which were not used in the program as often as the verbs mentioned above. Three low prior knowledge students made mistakes in the use of the verb ‘‘honor’’ (database report). For example, instead of making up a sentence ‘‘The Jefferson Memorial was built to honor the third American president’’ one of the students made up the sentence ‘‘Jefferson Memorial was honored the third American president’’ (database report). A possible explanation for this mistake is that the intermediate student was not familiar with complex infinitive grammar structures such as ‘‘was built to honor.’’

All the students used at least one of the alternative active verbs instead of passive verbs (for example: lies instead of is located) in their writings and did not make any mistakes in the use of factual information. Interestingly, seven out of ten participants made mistakes in the use of prepositions and articles, which showed the need in additional assignments to practice those skills in the program. For example, a low intermediate student who was neither familiar with the concept of passive voice (the data from pre-test analysis and observations) nor with the geography of the country, had problems remembering prepositions needed for the description of the country and excessively used the preposition ‘‘by’’ instead of other appropriate prepositions in their writings. Possible explanation for this fact is that the preposition ‘‘by’’ was the only one that was taught in the program. Students had an opportunity to use this preposition in different activities during the training. The findings are consistent with the CLT (Sweller, 2005). Possibly, because the cognitive loads were high for her, she could not process all the information on the screen on the level of ‘‘recall’’ (not enough practice), but just on the level of recognition. She was able to understand the meaning of prepositions in the sentences, but was not able to recall them for the production of the language output. Her sentence: ‘‘Hawaii is known by its beautiful beaches and tropical climate.’’

The findings from the analysis of the writings are in tune with the findings from the debriefing and informal observations. All low prior knowledge students emphasized the fact that they needed more practice with prepositions and articles and they needed special exercises to practice the use of prepositions and articles to describe locations and places.

Students with the lowest pre-test scores (10–11 out of 20) pointed out that they needed more practice with the target grammar concept of passive voice. Although some students found the description of the pictures challenging, all of them emphasized the importance of this activity (the findings from observations, debriefing) because it helped them retrieve and recall the knowledge that they had gained. Student’s comment during the debriefing:

> It is not easy, but very important because when you need to write something you have to recall what you have learned.

Research question 2: Does the design of the program facilitate the grammar concept knowledge transfer to the contexts other than geography and history of the country?

An examination of the pre-test and transfer-test scores revealed that the students' ability to use present and past simple passive voice in the context other than geography of the USA on average increased by 3.3 points out of 20 (see Table 4.3).

Table 4.3 Means and standard deviations of students' pre and transfer-test raw scores

	Pre-test			**Transfer-test**	
Mean	**SD**	**N**	**Mean**	**SD**	**N**
14.4	2.675	10	17.7	1.703	10

Even though 95% of the training occurred in the context of geography and history of the USA, participants showed knowledge gain on the transfer-test administered on the sixth day after the post-test.

Similar to the results of the retention test, the students who showed the lowest scores at the pre-test had the highest knowledge gain at the transfer-test. Consistent with data in Appendix H, students' performance on the constructed response items shows a larger increase than on the selected response items (see Appendix I).

The data from the survey support these findings (see Table 4.4). Seven students strongly agreed and three students agreed that the Extra Practice helped them understand the passive voice concept (database report, survey). The data from debriefing also indicate that all ten students liked the combination of teaching the grammar concept of passive voice in a context of geography of the USA (''I learned a lot about the country'') and do ''Extra Practice'' exercises in other contexts (''It helped me generalize the information'').

Research question 3: How do students use the program?

The data from the Likert-scale survey, debriefing, informal observations, and database report were used to determine how students used the program. In the survey, the participants ranked the effectiveness of different features in the program on a 5-point Likert scale that ranged from strongly disagree to strongly agree. Students mean scores for all the features were favorable and ranged from 4.9 to 4.1 as listed in Table 4.4

Table 4.4 Results of student survey (5-strongly agree, 4-agree, 3-somewhat agree, 2-disagree, 1-strongly disagree)

#	Program Component	Rating	#	Program Component	Rating
1	This program helped me understand how to use passive voice.	4.9	8	The material is presented in an interesting way.	4.7
2	I like the sound in the program.	4.2	9	The Grammar Help pop-ups are helpful.	4.4
3	The pictures in the program are helpful.	4.6	10	The Review Pages summarizing each unit are helpful.	4.5
4	The clickable words highlighted yellow are helpful.	4.5	11	The Extra Practice helped me understand passive voice.	4.7
5	I can understand all sentences.	4.3	12	The captions that can be turned on and off are helpful.	4.4
6	The maps are helpful.	4.7	13	The Stop and Replay Sentence tools are helpful.	4.3
7	It is easy to navigate through the program.	4.1	14	The feedback with explanation is helpful.	4.7

According to the Likert scale ratings, the participants found all the features in the program helpful. The most highly ranked features were feedback, images, the highlighted words that can be clicked to hear and see their definition (word annotations), and extra practice. Also, most students strongly agreed that the material was presented in an interesting way. Performance aids were used by all 10 students for a total of 996 times (database report).

Grammar Help pop-up (database, observations, debriefing).

The frequency of use of the Grammar Help pop-up during the training episode ranged from 1 to 21 times. On average low-prior knowledge students used it more often (15 times) than high prior knowledge students (4 times). The findings suggest that these students were more inclined to use the direct instruction (Grammar Help pop-up) when they encountered a completely new concept (observations). High prior knowledge students encountered new concepts less frequently than low prior knowledge students. This might explain the difference in the frequency of the grammar help use. The common comments from high and low prior knowledge students about different program aspects and features are presented in Appendix J.

Word annotations (database report).

Nine out of ten students used the clickable highlighted words to hear and see their definitions. The frequency of use of the clickable words ranged from 10 to 29 times.

Maps, images.

All the participants liked the combination of text, maps, images, and audios (debriefing, informal observations, survey). All of them perceived it as an advantage of this program compared to textbooks and regular classroom instructions. The participants liked the idea that they could use images, maps, and play the audios of the sentences at their convenience (see Appendix J).

Most students extensively used maps and images in the program (up to 34 times). The students with poor prior knowledge of general geography (observations) used maps less frequently (5-8 times) than the ones with better knowledge of geography (20-30) times (database report). A possible explanation for this fact is that they had problems reading maps.

Listen to a sentence (database report, debriefing).

Many students favored the option of listening to sentences during the training session (19-33 times). Some of them listened to the same sentence 2-3 times in a row (see Appendix J).

Tools to adjust instruction to students' needs (database report, debriefing).

Five students had captions turned off all the time. The other five students turned captions on and off from 3 to 11 times. Six students used the “Repeat the Scene”, “Play” and “Pause” tools. On average, they replayed scenes 4 times and used the button to stop and replay the last sentence 10 times. Students who have lived in the country for less than 1 year (demographic survey) used the above tools much more often compared to students who have lived in the country for a longer time. They might have had problems with listening and comprehension and wanted to practice their listening skills.

As to the findings from the debriefing, there was one person who was not happy with the quality of the sound and complained about the background noise. This person was the only one who used a headset with the sound coming in only one ear, which might have been a problem.

Irregular verbs table pop-up (database report, debriefing).

The range of use was from 8 to 30 times (Appendix J).

Story-telling format (survey, debriefing).

Nine of ten students liked the story telling format for different reasons (debriefing). Some of them liked it because they identified themselves with Maria, an ESL student. Others considered it as a good practice for listening skills. Two students mentioned that the use of informal language in a conversation between the main characters helped them understand the material better (see Appendix J).

Cognitive training theory at work (observations, debriefing, tests scores from the database).

Nine of ten students strongly agreed and one student agreed that the program helped them understand how to use passive voice (survey) (Appendix J).

Extra practice in the contexts other than country studying (debriefing, survey).

Seven students strongly agreed and 3 agreed that the extra practice was helpful. The findings from the debriefing session validate the data from the survey (see Appendix J). The extra practice helped the students make connections between the concepts of passive voice and other contexts, not just the context of geography and history of the USA. The fact of building connections is an important step in conceptual understanding. This could be one of the reasons for students' high performance scores at the transfer tests. Further investigation of the possible effect of this instructional design strategy on students transfer test scores needs to be conducted with a larger group of students and experimental design of the evaluation study.

Students' suggestions for modification:

1. In several cases the predetermined explanatory feedback was not enough for advance students
2. A few typos
3. Minor changes in visual design (some screens were cluttered, on some screens the instructions were too long)
4. Redesign of the unit with alternatives (it was confusing for some students, especially for low prior knowledge students
5. Need a *previous* button (some students wanted to go back to the previous screen)
6. More practice with prepositions and articles (all the students)

Research question 4: What is the relationship between students' learning outcomes and computer skills, learning outcomes and prior knowledge?

The results of the study made it clear that no relationship between computer skills (demographics survey) and performance on the tests was found. All the participants were able to use a keyboard and a mouse (demographics survey, observations) and did not have problems navigating through the program (debriefing, observations). The knowledge gain of the participants ranged from 1 to 8 correct answers out of 20 on the post-test and from 2 to 6 correct answers out of 20 on the transfer test. The conclusion was made that students' performance on the tests was not tied to their computer skills. The goal of the designers was to make the program simple enough so that people did not need special computer skills to navigate through the program. However, low typing skills slowed down some students (observations, database report).

Students' prior knowledge had an effect on students' knowledge gain. The gain was higher among the intermediate students whose familiarity with passive voice was limited. On average, low prior knowledge students' knowledge gain on the retention test was 6.3 out of 20 and on the transfer test 6.2 out of 20. As to high prior knowledge students, their gain on the retention test was 2.6 out of 20 and on the transfer test 2.4 out of 20.This might have happened because of the ceiling effect that occurred with high prior knowledge students. On average, low prior knowledge students used grammar help, irregular verbs tables, word annotations, and tools to adjust the pace of the instruction to their needs more often compared to high prior knowledge students (database report).

Summary

Despite the fact that overall the designed program showed that it delivered its promise, some questions arose during the initial evaluation (Study Group A) and required further investigation. When asked about what new things they learned through the program, all participants mentioned that they learned a lot about geography and history of the country. But when it came to grammar, their answers were different. The student with the highest pre-test score (18 out of 20) learned more about when it is appropriate to use passive voice, alternatives to passive voice, prepositions and articles, spelling of geographical names (analysis of writing assignments, debriefing).

The student with the lowest pre-test score (ten out of 20) learned about the concept of passive voice versus active voice and formation of different forms of passive voice. This student made mistakes in the use of prepositions and articles and some grammar structures other than passive voice and had a hard time understanding alternatives to passive voice. The participant even requested one more session. Possibly, the cognitive loads were high and she was too tired to proceed. Seven out of ten participants also had problems with the unit about alternatives to passive voice. Overall, it appeared that the participants with different level of prior knowledge learned different things through the program. This can be explained by low prior knowledge students' working memory limitations in accord with CLT (Sweller, 2005). Low prior knowledge students' were able to learn only the major grammar concepts taught in the program.

Given these findings, a question arose as to whether it was worth designing a program with linear navigation and pre-determined feedback to address both the needs of low and high prior knowledge students in the same program. Linear navigation assumes that different learners have to go through the program unit by unit regardless of their level of expertise. However, some exercises may be too easy for high prior knowledge students and others are too difficult for low prior knowledge students. Such experiences may negatively affect learning of both high and low prior knowledge learners. Redundant exercises may interfere with the existing cognitive schemas of high prior knowledge students (Kalyuga, 2007). In case of low prior knowledge students, the process of building cognitive schemas may have been slowed down because new information interfered with the previous information that had not been organized and automated well enough. The linkage of the new information might have been problematic in this situation. (Phye & Sanders, 1994).

In order to make sure that it was not a flawed instructional design that caused the above problems, the following modifications to the program were made:

1. The unit about alternatives to passive voice was redesigned; more practice exercises were added to the unit. The changes were made to make sure that it was not the flawed design that led to low prior knowledge students' confusion.
2. Four more items were added to each of the tests (pre-test, post-test, and transfer-test): conversion from active voice to passive and vice versa, 2 selected response (multiple choice) and 2 constructed response (convert the sentence into passive voice or vice versa) items. Conversion is considered in TESL as a high element interactivity assignment. Adding those items helps better identify high prior knowledge students.
3. Additional unit for reviewing articles and prepositions related to the description of geography and history of the USA, the ones used in the program, was added. This change was made to help low prior knowledge students to master the above prepositions and articles.
4. Several cluttered screens were redesigned.
5. Previous button was added on several screens to make the navigation easier.

After making changes to the program, one more experiment, a user interface analysis was attempted.

Study Group B

The investigation focus of the study.

In contrast to the focus of the first study, different features in the program and their use by students, the focus of the second experiment was to find out how the program addressed individual learning needs of students. The interest in learning differences was caused by the findings from the previous experiment. Students with different levels of prior knowledge learned different things through the program, and some of students' preferences were similar while others were different. For this reason, further investigation was attempted to examine how the program addresses the needs of students with individual learning differences and how prior knowledge of students accounts for those differences.

Participants.

The second study was conducted at a Midwest Community College with a different set of students, 4 adult learners, in June 2009. The participants were females and ranged from 22 to 45 years old. Their English language proficiency ranged from intermediate to advanced level. The level of familiarity with passive voice concept was beginners and intermediate. Participants had Portuguese, Spanish, Vietnamese, and Slovak, as their first languages. Because the focus of the experiment was to investigate how the program addressed the needs of students with individual differences including the difference in prior knowledge, the purposive sample was selected. Recommendations from English teachers were used to identify participants with different levels of prior knowledge of the grammar concept of passive voice, general vocabulary, and general knowledge of geography.

Research questions and data collection methods.

The research questions in the second study were driven by the findings from the first study.

1. How does the program address the needs of students with individual learning preferences?
2. How does prior knowledge affect students' learning with the program?

The qualitative and quantitative data from 3 sources, students' perceptions of their experiences with the program, their learning outcomes, and the data from a database report about their use of performance aids were collected. In Table 4.5, the categories under analysis are presented.

Table 4.5 Research questions and categories under analysis

Research Questions	Categories under Analysis
1. How does the program address the needs of students with individual learning preferences?	Individual learning preferences: 1. Need for visualization and audio (database report, survey, and debriefing) 2. Context for teaching grammar (survey, debriefing) 3. Preferred type of instructional setting for the use of the program (debriefing)
2. How does prior knowledge affect students' learning with the program?	Students prior knowledge: 1. Prior knowledge of passive voice concept (pre-test scores) 2. Prior knowledge of general vocabulary (database report) 3. Prior knowledge of basic geography (observations)
	Students' navigation patterns: 1. The count of use of different performance aids (database report) 2. Preference for teaching approach: modified deductive or modified inductive: 3. Database report: count of the use of grammar help pop-ups 4. Data from the survey and debriefing about students'

Research Questions	Categories under Analysis
	preferences of teaching approach Preference for different kinds of assignments in the program: 1. Survey 2. Knowledge gain in the use of passive voice within the context of geography of the USA 3. Database report: the difference between pre- test and post-test scores 4. Knowledge gain in the use of passive voice within contexts other than geography 5. Database report: the difference between pre-test and knowledge transfer-test scores

Data collection process.

The participants received the instruction during two sessions for a total of 4 h without any help from the teachers at different times during 1 month, so that the principal investigator could conduct the informal observations individually with each student who was going through the training. Each participant went through the set of procedures described in Appendix K.

The participants did not receive any formal instruction on the concept of passive voice between the pre-test and knowledge transfer-test. One of the participants received 5 min of pre-training at her request because she was not familiar with basic grammar terminology. The data collection process is presented in Appendix K.

Data collection instruments.

The following instruments were applied for data measurement (see Appendix K):

1. Paper-based survey administered by the principal investigator upon students'
2. completion of each unit in the program.
3. Database report.

4. Informal observations.

5. Debriefing after the training.

6. Demographics survey imbedded in the program taken by each student before the pretest.

Data analysis.

The analysis was conducted in a fashion similar to the first experiment. The differences between the experiments are presented below and in Appendix L. Each of the three tests consisted of 24 items total, 12 selected response and 12 constructed response items. Students' scores were recorded to validate students' perceptions, but because of the small purposive sample (four participants) the descriptive statistics was not used to describe the central tendency data. This method was excluded as inappropriate because the study was intentionally designed as the study of extremes.

The five point Likert survey was administered after the students' completion of each program unit, not at the end of the training. In contrast to the first study (Study Group A), the focus was not on the use of features in the program, but on the implemented instructional decisions and how they helped students learn through the program. Moreover, the survey items were designed to check if students' interest was contributing to their learning.

Example of a survey item: (rate on a scale of 5, 5—strongly agree, 1—strongly disagree).

How do you like the use of maps in the program overall? Why?

1. They help me learn geography of the country.

2. They help me learn the grammar concept.

3. They help me integrate the information about geography with the grammar concept.
4. They help me memorize new words.
5. They make learning more interesting.

The 15 min debriefing session was conducted individually with each student upon completion of the survey. The principal investigator asked the participants to rate the choices and typed their ratings into a Word document.

Results and discussion.

The results of the second experiment are presented in Appendix L and Appendix M. The participants shared several learning preferences. All of them appreciated the narration and animation presented at the same time (observation, debriefing), which is in tune with the temporal contiguity principle (Mayer, 2008). Also, they would rather periodically switch modalities, alternate text for describing visuals with narrated animation. ‘‘It helps avoid monotony, change is good’’ (all students). Seemingly, ESL students may have a different perspective on modality effect (Mayer, 2008).

All of them gave a rating of five to the presentation of the new information through two channels: verbal and visual. The combination of text, audio, maps, and images in the program helped them memorize and organize the information (survey, observations, debriefing, and database report). These findings are consistent with the previous research (Deno, 1968; Paivio et al., 1988; Paivio & Desrochers, 1979). ‘‘It helps you memorize new information much better compared to learning with textbooks’’ (student’s comment). All the participants valued the option of clicking on maps, audios, and images when they choose to do it. ‘‘It gives you control over your learning’’.

All the users gave a rating of five to the contextualized teaching of grammar because it provided an interesting environment to learn grammar and vocabulary for describing the country. All of them gave a rating of five to extra practice in different contexts at the end of each unit. It also appeared that the interesting content did not get in the way of their learning. The users found most of the assignments in the program both interesting (rating of five) and helping them organize and integrate the information (rating of five). Seemingly, the match of students' cognitive and emotional motivation was achieved.

All the participants liked the characters in the program because the dialogues between characters allowed the use of informal language and provided concise presentation of the important information (survey, observations, and debriefing). The findings are consistent with the personalization principle (Mayer, 2008). All the participants appreciated the fact that essential material was highlighted in the program (debriefing, observations), which is in tune with the signaling principle (Mayer, 2008).

Production of the language output was favored by all the students; low prior knowledge students found it challenging, but interesting. All four participants enjoyed the tools to adjust instruction to their own pace in the system-paced parts of the program (survey, debriefing). The findings are in tune with the interactivity principle (Mayer, 2008).

One out of four students, a high prior knowledge student, would rather have had more extraneous material about the country to maintain her interest. In the program, all the material was used to teach grammar concepts, in tune with the coherence principle (Mayer, 2008). Nevertheless, it remains unclear how the inclusion of extraneous materials would affect students' learning outcomes.

All the participants appreciated the text describing images included in the image popups, so that the users did not need to go back and forth between the text on the screen and images on the image pops. The findings are in accord with the spatial contiguity principle (Mayer, 2008). Interestingly, all the participants would rather have image pop-ups movable, so that they could see the image and all the text on the screen (not just the text that describes this particular image). In addition, all the participants wanted to have audio included with the text and images in the pop-ups.

There was only one out of four students who was able to process narrated animation (visuals and audio) along with reading captions (her comment at the debriefing). Some alternated captions on/off to focus on either listening or reading the captions, others had captions off all the time.

Interestingly, two students with higher pre-test scores sometimes had attention problems; they did not read the instructions closely during the training and had problems focusing on the worked examples. This might have happened because some of the exercises were redundant for them given their level of expertise.

The student with the lowest pre-test score (seven out of 24) mentioned that she needed more practice to master the grammar concepts. She also mentioned that it was very important for her to understand the grammar concepts before practicing. She even shared her experience with Rosetta Stone Software. She pointed out that this program and Rosetta Stone complement each other. ‘‘Rosetta Stone has more practice. Rosetta Stone helps memorize because it is more visual. In this program, I like the way grammar is taught. It will help me in the future. I like summary pages, compare and contrast exercises. The same images and maps are used through the whole program. I want to understand grammar concepts before I practice. This program helps me with that. But there is more practice in Rosetta Stone, there is not enough practice for me in this program. Understanding helps me memorize and it will also help me in the future. I could also learn about geography of the country and its culture. It is very important''.

According to the findings (see Appendix N), it is not only prior knowledge of passive voice, but also prior general knowledge of vocabulary and geography that affected students' progress. The participants with low prior knowledge of passive voice grammar concepts, intermediate level of general vocabulary, and adequate basic knowledge of geography benefited the most from the program.

As to preference of a teaching approach; modified deductive or modified inductive, it was not consistent for all students and it was changing depending on their level of familiarity with a grammar concept. Low prior knowledge students used grammar help less extensively at the end of the program. High prior knowledge students used grammar help when dealing with unfamiliar grammar concepts. Because the pre-test scores were different for different students, the number of unfamiliar concepts was also different, and so was the count of grammar help use at different times during the training episode (see Appendix M and Appendix N).

Even though the students were happy with the program and their learning outcomes proved its benefit, the unit with alternatives to passive voice was still a problem for students with low prior knowledge. Moreover, the low prior knowledge students were still making mistakes in the use of prepositions and articles even though they appreciated the exercises and grammar pop-ups related to prepositions and articles. Apparently, the design changes still did not help low prior knowledge students master alternatives to passive voice, prepositions, and articles for describing places and locations. The program did not provide enough training for them (not enough practice) to be able to cope with those concepts. It seemed like the idea of creating a program with linear navigation and interpretative predetermined feedback for use in mixed group classes did not work. In the future, alternative ways to address the needs of mixed group classes needs further investigation. In the final version, the unit teaching alternatives to passive voice was deleted. In the current version, the program could be used more effectively by low prior knowledge students who demonstrated the best progress during the two experiments.

Overall, the findings indicated that some of the differences in learning preferences can be attributed to the differences in prior knowledge of the target grammar concept, vocabulary, and basic knowledge of geography among the participants (see Appendix M) while others presented in Appendix L cannot. These findings should be treated with caution because of the small size of the group of participants.

General Discussion

The findings provided preliminary evidence of the ability of the program to facilitate retention and transition from declarative to procedural knowledge of the grammar concept of passive voice as well as knowledge transfer of the grammar concept to contexts other than geography and history of the USA. Overall, low prior knowledge students showed higher knowledge gain on both retention and transfer tests compared to high prior knowledge students. On average, low prior knowledge students' knowledge gain on the retention test was 6.3 out of 20 in the Study Group A and 8 out of 24 in the Study Group B. On average, high prior knowledge students' knowledge gain on the retention test was 2.6 out of 20 in the Study Group A and 3 out of 24 in the Study Group B. As for the transfer test, it was 6.2 out of 20 in the Study Group A and 7 out of 24 in the Study Group B for low prior knowledge students and 2.4 out of 20 in the first experiment and 5 out of 24 in the Study Group B for high prior knowledge students. The scores of one low prior knowledge student were considered as outliers and were excluded from the calculations. This student had the highest knowledge gain on the retention test (12 out of 24). In contrast to other students, she took a 5 minutes pre-training session before the training episode, had the lowest pre-test score (7 out of 24), had very high spatial abilities, and took the transfer test two weeks after the post-test due to personal circumstances.

The instructional design decisions based on Sweller's cognitive load theory and Mayer's multimedia learning theory worked. Even though the invested mental effort was not directly measured in this study, it was assumed that on some occasions the working memory of low prior knowledge participants was overloaded (informal observations, students' responses from the database reports). One of the reasons why it may have happened was the attempt to design a program with linear navigation and predetermined feedback that could address the needs of students with different levels of prior knowledge.

The assumption underlying such a design was that students with different levels of prior knowledge will learn different things through the program. Even though it was actually the case, low prior knowledge students were frustrated that they could not master the concept of alternatives to passive voice, because the program did not provide enough practice for them to automate their skills of a new grammar concept before another new concept was introduced. For this reason, in the final version, the unit with alternatives to passive voice was deleted. All the features implemented in the program were used by all the students to a different extent in accord with their needs.

As to instructional design decisions based on cognitive training theory, they worked for all the participants in both experiments. All the participants strongly agreed that that understanding of grammar concepts was important for mastering ESL, and the practice of generalization, discrimination, and integrations of the target grammar concepts helped them to memorize, organize, and transfer the target grammar concepts.

Overall, the design and formative evaluation of the multimedia program for teaching ESL grammar was an insightful experience that contributed not only to program modification but also to a collection of preliminary evidence of what instructional design decisions can be beneficial for teaching ESL grammar through multimedia instructional programs. In other words, this research identified themes for further research.

According to the preliminary evidence received both from the participants and database reports, the multimedia environment was instrumental for the learners in building cognitive schemas in several different ways:

- By allocating different performance aids that can be enabled when needed, such as images, maps, graphs, tables, and audios.
- By providing an environment for both implicit and explicit learning and allowing students to adjust the level of instructional support according to their needs at any moment during the training.
- By delivering explanatory feedback on each students' action.

Overall, the design and formative evaluation of the multimedia program for teaching ESL grammar was an insightful experience that contributed not only to program modification but also to a collection of preliminary evidence of what instructional design decisions can be

beneficial for teaching ESL grammar through multimedia instructional programs. In other words, this small scale research identifies themes for further research.

Future Directions and Limitations of the Research

1. The next phase of the program evaluation on a larger sample of participants needs to be conducted to measure the effectiveness of the program by using statistical analysis for generalization purposes.
2. Since language production is a task with high element interactivity, there is need for research on how to better blend newly learned vocabulary with the newly learned grammar concepts so that the users have an opportunity to practice all the newly learned items as a whole including articles, prepositions, and anticipated additional grammar structures that they will need for producing correct comprehensible language output. The research providing evidence of the effectiveness of blended teaching is important for instructional designers, practitioners, and scholars.
3. Taking into account the cost of building intelligent tutoring systems, the decision needs to be made as to which path to take: to build several short programs with linear navigation and pre-defined feedback focused on a particular level of students' expertise or lack of expertise (novices) or to build sophisticated intelligent tutors that can recognize the prior knowledge of learners and take them to the appropriate level. There is a need for further research to compare the learning outcomes of students exposed to both methods.

References

Al-Jarf, R. S. (2005). The effects of online grammar instruction on low proficiency EFL college students'achievement. *Asian EFL Journal, 7*(4), 166–190. Retrieved from http://www.asian-efl-journal.com/December_05_rsaj.php.

Azar, B. (2007). Grammar-based teaching: A practitioner's perspective. *TESL-EJ, 11*(2), 1–12. Retrieved from http://tesl-ej.org/ej42/a1.pdf.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Essex: Addison Wesley Longman, Inc.

Bresnan, J. (2001). *Lexical-functional syntax*. Oxford: Blackwell Publishing.

Bryk, A. S., & Gomez, L. M. (2008). Ruminations on reinventing an R&D capacity for educational improvement. In F. M. Hess (Ed.), *The future of educational entrepreneurship: Possibilities for school reform.* Cambridge: Harvard Education Press.

Chomsky, N. (1986). *Knowledge of language*. New York: Praeger.

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

Collentine, J. (2000). Insights into the construction of grammatical knowledge provided by user-behavior tracking technologies. *Language Learning &Technology, 3*(2), 44–57. Retrieved from http://llt.msu.edu/vol3num2/collentine/.

Conrad, D. (1999). The student view on effective practices in the college elementary and intermediate foreign language classroom. *Foreign Language Annals, 32*(4), 494–512. doi:10.1111/j.1944-9720.1999.tb00878.x.

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey Bass.

Deno, S. (1968). Effects of words and pictures as stimuli in learning language equivalents. *Journal of Educational Psychology*, 59(3), 202–206. doi:10.1037/h0025772.

Freebairn, I., & Parnall, H. R. (1995). The Grammar Rom [Computer software]. London: Longman Group.

Gollin, J. (1998). Deductive vs. inductive language learning. *ELT Journal, 52*(1), 87–88. doi:10.1093/elt/52.1.88.

Grgurovic´, M., & Hegelheimer, V. (2007). Help options and multimedia listening: Students' use of subtitles and the transcript. *Language Learning & Technology, 11*(1), 45–66. Retrieved from http://llt.msu.edu/vol11num1/grgurovic/default.html.

Halliday, M. A. K. (1994). *An introduction to functional grammar*. London: Edward Arnold.

Huang, H., & Eskey, D. E. (1999). The effects of closed-captioned television on the listening comprehension of intermediate English as a second language (ESL) students. *Journal of Educational Technology Systems, 28*, 75–96. Retrieved from http://www.dcmp.org/caai/nadh163.pdf.

Hubbard, P., & Siskin, C. (2004). Another look at tutorial CALL. *ReCALL, 16*(2), 448–461. Retrieved from http://www.edvista.com/claire/pres/tutorial2003/ppt/index.html.

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge:Cambridge University Press.

Kalyuga, S. (2007). Enhancing instructional efficiency of interactive e-learning environments: A cognitive load perspective. *Educational Psychology Review, 19*, 387–399. doi:10.1080/01443410124681.

Klauer, K., & Phye, G. (2008). Inductive reasoning: A training approach. *Review of Educational Research, 78*(1), 85–123. doi:10.3102/0034654307313402.

Krashen, S. D. (1988). *Second language acquisition and second language learning*. Upper Saddle River, NJ: Prentice-Hall International.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Vol. I: Theoretical prerequisites.* Stanford: Stanford University Press.

Marcy, P.(2007). Rosetta Stone. *TESL-EJ, 11*(3). Retrieved June 17, 2010, from http://tesl-ej.org/ej43/m1.html.

Marsden, E. (2007). Can educational experiments both test a theory and inform practice? *British Educational Research Journal*, *33*(4), 565–588. doi:10.1080/01411920701434094.

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

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., Fennell, S., Farmer, L., & Campbell, J. (2004). A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology, 9*(2), 389–395.

Nagata, N. (2002). An application of natural language processing to web-based language learning, *CALICO Journal, 19*(3), 583–600. Retrieved from https://www.calico.org/a425BANZAI%20An%20Application%20of%20Natural%20Language%20Processing%20to%20Webbased%20Language%20Learning.html.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design. *Educational Psychologist, 38*, 1–4. doi:10.1207/S15326985EP3801.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interactions between information structures and cognitive architecture. *Instructional Science, 32*, 1–8. doi:10.1023/B:TRUC.0000021806.17516.d0.

Paivio, A., Clark, J., & Lambert, W. (1988). Bilingual dual-coding theory and semantic repetition effects on recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 163–172. doi: 10.1037/0278-7393.14.1.163.

Paivio, A., & Desrochers, A. (1979). Effects of an imagery mnemonic on second language recall and comprehension. *Canadian Journal of Psychology, 33*, 17–28. doi:10.1037/h0081699.

Phye, G. D. (1991). Advice and feedback during cognitive training: Effects on acquisition and delayed transfer. *Contemporary Educational Psychology*, *16*, 87–94. doi:10.1016/0361-476X(91)90008-9.

Phye, G. D., Robinson, D. H., & Levin, J. R. (Eds.). (2005). *Empirical methods for evaluating educational interventions*. San Diego: Academic Press.

Phye, G., & Sanders, C. (1994). Practice and feedback: Elements of training for transfer. *Contemporary Educational Psychology*, *19*, 286–301. doi:10.1006/ceps.1994.1022.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Essex: Longman Group Limited.

Stepp-Greany, J. (2002). Student perception on language learning in a technological environment: Implications for the new millennium. *Language Learning & Technology*, *6*(1), 165–180. Retrieved from http://llt.msu.edu/vol6num1/steppgreany/default.html.

Sweller, J. (1999). Instructional design in technical areas. Melbourne: ACER Press.

Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp.215–266). San Diego, CA: Academic Press.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp.19–30). Cambridge, UK: Cambridge University Press.

Sweller, J., Van Merrie¨nboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. doi:10.1023/A:1022193728205.

Widodo, H. (2006). Approaches and procedures for teaching grammar. *English Teaching: Practice and Critique*, *5*(1), 122–141. Retrieved from http://www.scribd.com/doc/24456708/English-Grammar.

Zapata-Rivera, D., Vanwinkle, W., Shute, V., Underwood, J. S., & Bauer, M. (2007). English ABLE. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Proceeding of the 2007 conference on artificial intelligence in education: Building technology rich learning contexts that work. Frontiers in Artificial Intelligence and Applications, 15,* 323–330. Amsterdam, The Netherlands: IOS Press.

APPENDIX A: MULTIMEDIA DESIGN DECISIONS GROUNDED IN THE FINDINGS FROM EVALUATION STUDIES AND COGNITIVE THEORIES

Instructional design decision grounded in a cognitive theory	Part of the program where it will be used	Supportive evaluative and descriptive studies	Supportive findings from evaluative studies and cognitive theories
Contextualized teaching of grammar.	The content of USA geography and history is used in the whole program.	1) Conrad, 1999 2) Marsden, 2007 3) Stepp-Greany, 2002	
The program makes the users compare and contrast different grammar concepts to help them organize information by making implicit thinking explicit. (cognitive training theory)	Tutorial parts.	1) Gollin, 1998 2) Widodo, 2006 3) Klauer & Phye, 2008)	
Sequential presentation of the material with drill, practice, and predetermined interpretive feedback. (cognitive load theory) (multimedia learning theory)	Through the whole program.	1) Conrad, 1999 2) Marsden, 2007 3) Stepp-Greany, 2002 4) Clark et al., 2006 5) Mayer, 2008	
The possibility for students to choose modified inductive or modified deductive approach for teaching/learning grammar in the program. (cognitive load theory)	Through the whole program.	1) Conrad, 1999 2) Marsden, 2007 3) Stepp-Greany, 2002 4) Clark et al., 2006 5) Kalyuga, 2007 6) Sweller, 1998	
The use of a story-telling format. (multimedia learning theory)	It is used to personalize instruction, use informal language, and stitch together the tutorial parts of the instruction.	1) Mayer et al., 2004	"Personalized instruction leads to deeper processing of information and better students' performance on transfer-tests."

Instructional design decision grounded in a cognitive theory	Part of the program where it will be used	Supportive evaluative and descriptive studies	Supportive findings from evaluative studies and cognitive theories
Subtitles and control over pacing of the presentation. (multimedia learning theory -interactivity principle)	In the system-paced parts of the program.	1) Grgurovic & Hegelheimer, 2007 2)Clark et al., 2006 3) Huang & Eskey, 1999 4) Mayer, 2008	It is a must to give learners control over pacing when there is mismatch in delivery and native language (Clark, 2006). "ESL students score significantly better on listening comprehension if subtitles or scripts are provided" (Grgurovic, et al., 2007). Subtitles are preferable compared to scripts (Huang & Eskey, 1999).
The program is web-based.	Through the whole program.	Al-Jarf, 2005	Web-based grammar instruction integrated into in-class grammar instruction significantly improved EFL freshmen college students' achievement and attitudes.
Students' performance and behavior tracking embedded in the program.	Through the whole program.	1) Collentine, 2000 2) Hubbard and Siskin, 2004	CALL containing user-behavior tracking technologies can provide important insights into the construction of grammatical knowledge and access if the program promotes grammatical development.

APPENDIX B: TRAINING EPISODE UNITS

Pre-unit: What is wrong with Maria's sentence?

Unit 1: David's intelligent tutor/ Passive voice or active voice?

Unit 2: American states/ Formation of present tense passive voice

Unit 3:Landforms of the USA/Present tense passive voice with singular and plural nouns

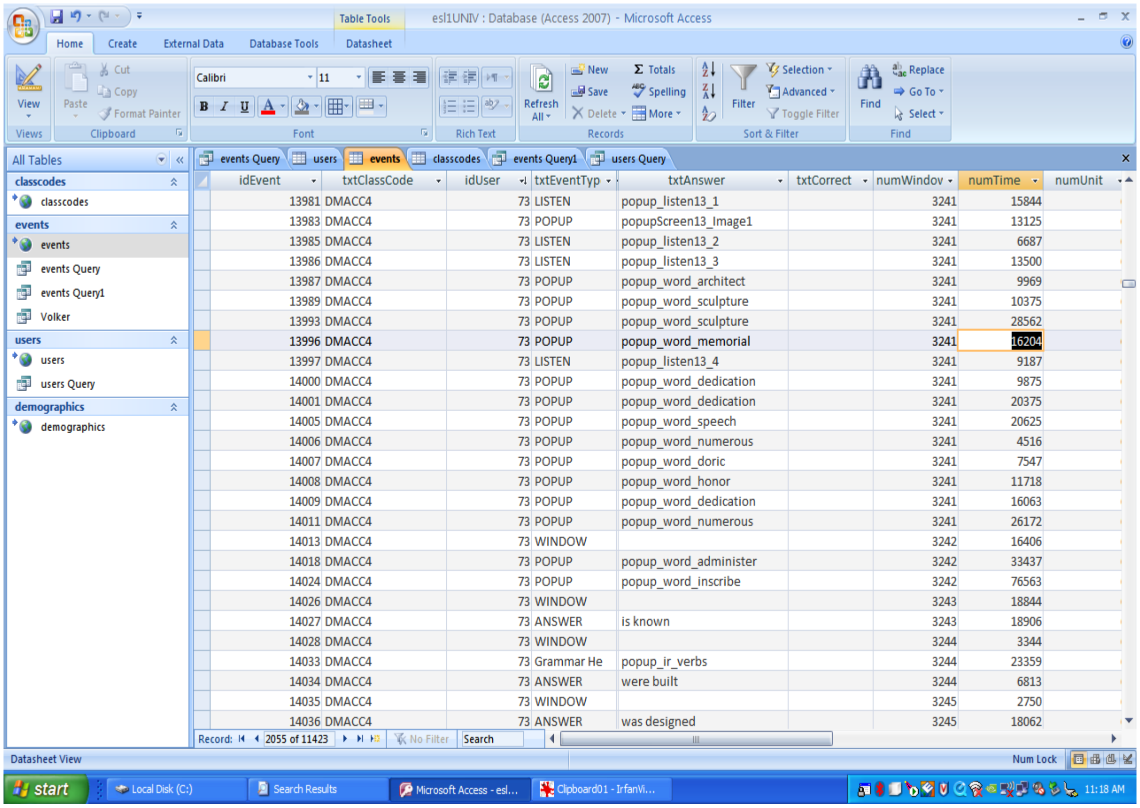
Unit 4: Geography of the USA/ Revision of present tense passive voice

Unit 5: Washington D. C. /Using "by" in passive voice sentences

Unit 6: Places of interest in Washington D. C./ Past tense passive voice vs. present tense passive voice

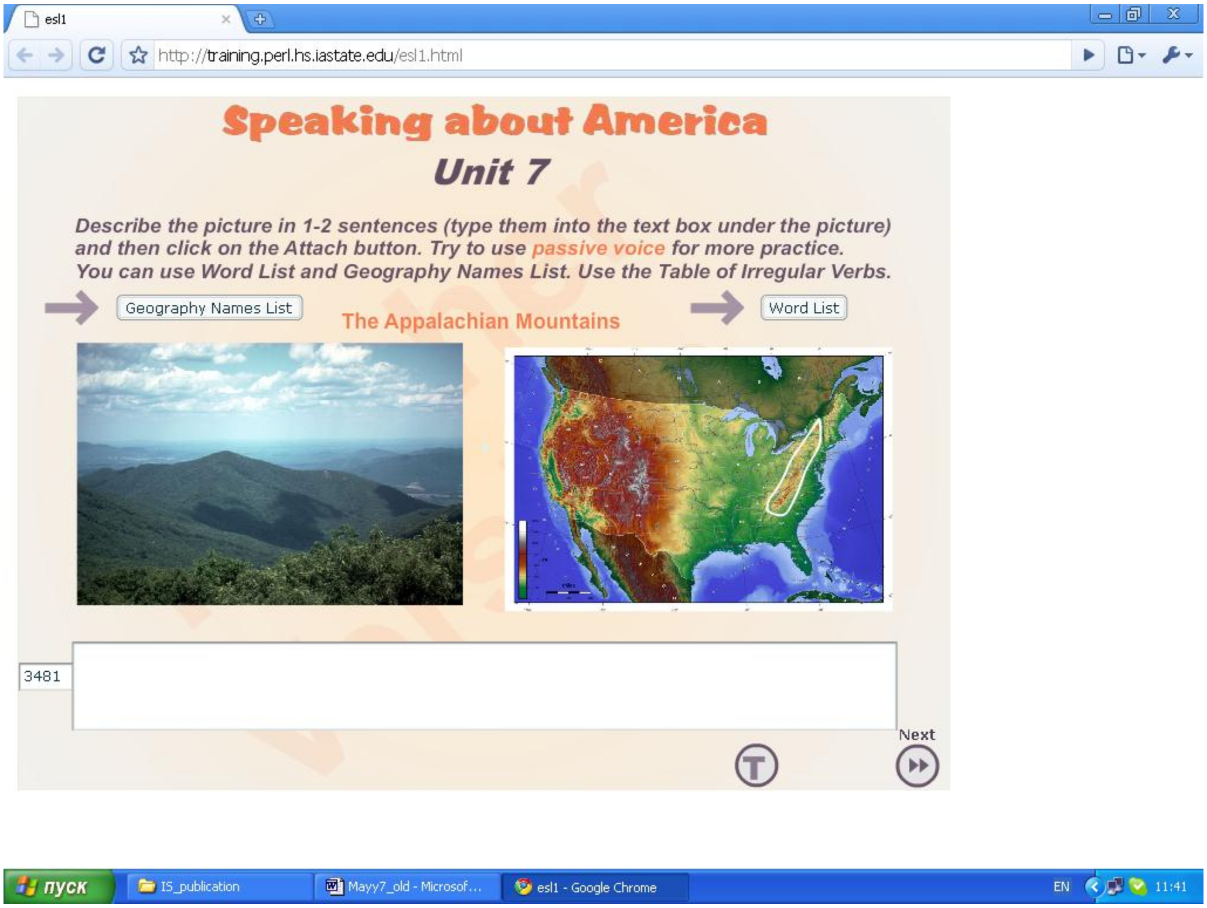
Unit 7: Test what you have learned about the USA

APPENDIX C: EXAMPLE OF DATABASE RECORD OF THE USE OF PERFORMANCE AIDS



idEvent	txtClassCode	idUser	txtEventTyp	txtAnswer	txtCorrect	numWindov	numTime	numUnit
13981	DMACC4	73	LISTEN	popup_listen13_1		3241	15844	
13983	DMACC4	73	POPUP	popupScreen13_Image1		3241	13125	
13985	DMACC4	73	LISTEN	popup_listen13_2		3241	6687	
13986	DMACC4	73	LISTEN	popup_listen13_3		3241	13500	
13987	DMACC4	73	POPUP	popup_word_architect		3241	9969	
13989	DMACC4	73	POPUP	popup_word_sculpture		3241	10375	
13993	DMACC4	73	POPUP	popup_word_sculpture		3241	28562	
13996	DMACC4	73	POPUP	popup_word_memorial		3241	16204	
13997	DMACC4	73	LISTEN	popup_listen13_4		3241	9187	
14000	DMACC4	73	POPUP	popup_word_dedication		3241	9875	
14001	DMACC4	73	POPUP	popup_word_dedication		3241	20375	
14005	DMACC4	73	POPUP	popup_word_speech		3241	20625	
14006	DMACC4	73	POPUP	popup_word_numerous		3241	4516	
14007	DMACC4	73	POPUP	popup_word_doric		3241	7547	
14008	DMACC4	73	POPUP	popup_word_honor		3241	11718	
14009	DMACC4	73	POPUP	popup_word_dedication		3241	16063	
14011	DMACC4	73	POPUP	popup_word_numerous		3241	26172	
14013	DMACC4	73	WINDOW			3242	16406	
14018	DMACC4	73	POPUP	popup_word_administer		3242	33437	
14024	DMACC4	73	POPUP	popup_word_inscribe		3242	76563	
14026	DMACC4	73	WINDOW			3243	18844	
14027	DMACC4	73	ANSWER	is known		3243	18906	
14028	DMACC4	73	WINDOW			3244	3344	
14033	DMACC4	73	Grammar He	popup_ir_verbs		3244	23359	
14034	DMACC4	73	ANSWER	were built		3244	6813	
14035	DMACC4	73	WINDOW			3245	2750	
14036	DMACC4	73	ANSWER	was designed		3245	18062	

APPENDIX D: EXAMPLE OF A PICTURE DESCRIPTION ASSIGNMENT AT THE END OF THE PROGRAM



APPENDIX E: DATA COLLECTION PROCESS IN STUDY GROUP A

Time Schedule	Procedures
Day1	Collecting demographics (Demographics survey). Pre-test (USA geography and history context).
Day3	Training episode: Unit1-4. Informal observations.
Day6	Training episode: Unit5-7 (picture description included). Likert survey. Debriefing. Informal observations.
Day10	Delayed post-test (USA geography context).
Day 16	Knowledge transfer-test (context other than USA geography and history).

APPENDIX F: THE LIST OF QUESTIONS DURING THE DEBRIEFING SESSION (STUDY GROUP A)

1. Did you have any problems finding something on the program pages? If yes, what distracted your attention?
2. Were there any words or sentences you did not like or found difficult?
3. Were there any assignments you did not like or found difficult? If yes, which assignments?
4. Did you have any problems navigating through the program? If yes, what problems?
5. Did you have any problems understanding the instructions? If yes, what problems?
6. How did you like the colors?
7. How did you like the fonts (size and color)?
8. Were there any screens on which it was difficult for you to find something? What was it?
9. Did you learn anything new today? Please specify.
10. How helpful were extra practice assignments? Why?
11. How did you like describing pictures at the end of the program? Why?
12. Does the strategy of comparing and contrasting sentences help you understand concepts? If yes, why? If, no, why?
13. What would be the best place to use the program?
14. What would you change in the program to help you learn better?

APPENDIX G: INFORMAL OBSERVATIONS SHEET (STUDY GROUP A)

(Because of a confidentiality issue the real name of the participant is not disclosed and the pseudonym Amy is used).

Dates: December 10-15, 2008

Typing speed: low, uses two fingers.

Login-1min

Demographics survey-3min

Pretest: 10 min

Training Session 1.

Unit 1

Amy: I like these dialogues with captions and clickable words. They are very helpful. I like to listen. It helps me learn better.

Frame 997

She cannot understand the idea of passiveness (the difference between " acts" or "receives the action").

Frame 998

She reads the explanation and looks like she is starting to understand.

Frame 999 (Summary page)

Amy: (talking to the program) OK. Now I understand what you mean.

Investigator: How do you like the idea of Review Lesson screen? Grammar Help pop up?

Amy: I do like them. It helps. I have a suggestion. I want to have a previous button to go to the previous screen and review the page that I had problem understanding.

Investigator: How about feedback?

Amy: Feedback is important because it explains. It is not that it says that your answer is right or wrong, it explains to you why it is wrong or right.

Unit 2

Frame 1038.

She has problem understanding that she needs to click on the Map button to see the map.

She prefers to use the Grammar Help pop up first and then do the exercises.

Amy: This table of irregular verbs pop up is great, very convenient.

Investigator: What would you prefer: to use the textbook to do the exercises or the program?

Amy: The program is better because I can click and type and there is audio to listen to. I also like to practice with the feedback, it helps. It is more interesting than the textbook. Emotionally more interesting.

Unit3

Frames 1879-1880

She likes the idea of putting the words in sentences in the correct order. She also likes the idea of maps and that on the pop-up maps she can see her answer, the jumbled sentence, and the correct sentence, all of them in one place.

Amy: It takes long to type in the sentences. It would be nice just to click on a word and it jumps to the right place.

She likes the images. She thinks that they help learn new information better.

Investigator: How easy is it for you to work with the program?

Amy: Easy. It is because I can see an unfamiliar word in a sentence, read it, see the image, and listen to it. (very important to understand which preposition to use to describe the position of places)

Frames 2096 (Review Lesson 3 screen)

Amy: I like the review screen, it helps me review what I have learned.

Unit 4

She likes the animation about exceptions.

Frame 2595

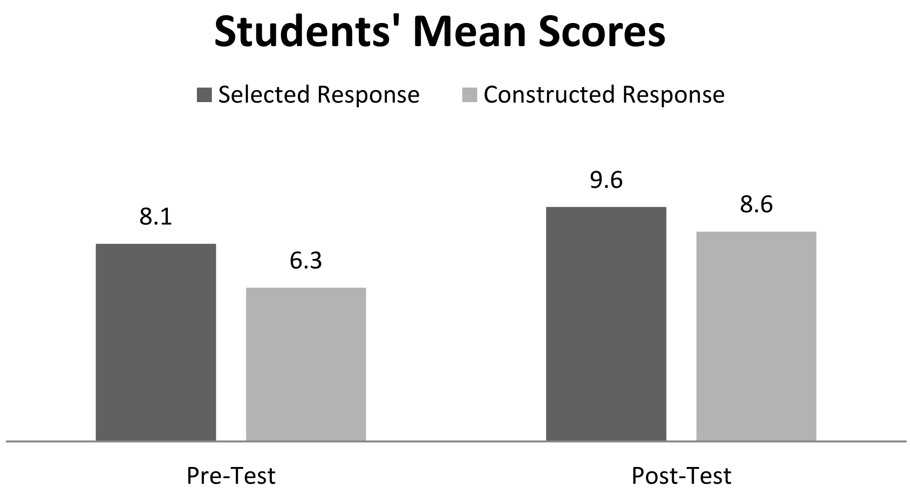
Amy: There is a typo on this frame.

I like the assignments with exceptions and alternative phrases. This is an important thing to know for ESL students. I have a suggestion. It would be nice to learn about which articles to use with geographical names.

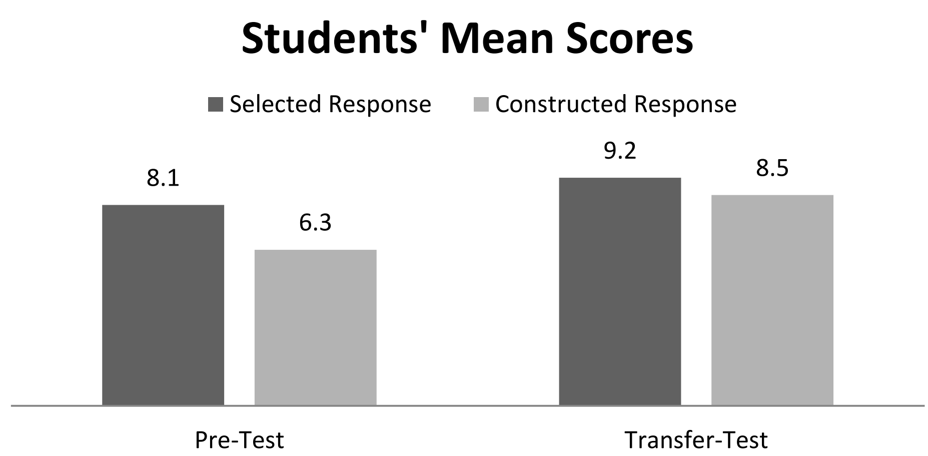
Investigator: What do you think about the screens divided into two so that you can compare and contrast?

Amy: It is very helpful. I would like more programs with other tenses used in passive voice. I would like to continue learning in this way.

APPENDIX H: STUDENTS' MEAN SCORES AT THE PRE-TEST AND POST-TEST (STUDY GROUP A)



APPENDIX I: STUDENTS' MEAN SCORES AT THE PRE-TEST AND TRANSFER-TEST (STUDY GROUP A)



APPENDIX J: STUDENTS'COMMENTS ABOUT PROGRAM FEATURES AND ASPECTS (STUDY GROUP A)

Program features and aspects	Researchers' Claims	Students' comments	Source
Grammar help pop-up	Low prior knowledge used it more often (15 times) than high prior knowledge students (4 times)	1. The way the program teaches grammar is enough and there is no need to use Grammar Help because the exercises make you think and you can read them as many times as you need. I cannot do it in class, when the teacher teaches grammar. I have to catch up with the class (high prior knowledge student). 2. I click on the Grammar Help and read the explanations to make sure that my guesses are correct and I do not miss anything (low prior knowledge student). 3. I click on the grammar help when I cannot guess the correct answer (low prior knowledge student).	Database report, observations, debriefing
Word annotation pop-up	1. The students appreciated holding verbal and audio representations at the same time. 2.The learners suggested that pictorial representations for the verbs describing locations would be a valuable addition to the pop-ups)	1. Very convenient and helpful, better than a dictionary because it is at hand. 2. Audio that goes with the word definition is very helpful. I can learn the pronunciation of unfamiliar words. 3. The verbs to describe locations such as *bound*, *border* are difficult to memorize. Images that will help understand and memorize the meaning of those verbs would be instrumental.	Database report, observations, debriefing
Maps, images	1. The learners appreciated holding verbal and pictorial representations at the same time. It helped them build mental connections. 2. The students suggested that audio of the text in the image/map pop-ups would be a valuable addition.	1. It helps memorize new words. 2. It has been 10 days since I used the program, but I still remember. I see the images of the places and maps with their location in my head. It's amazing. 3. Adding audio to a pop-up would be better than having text and audio on the main screen, and text and picture on the pop-up screen. It is better to have text, audio, and image/ map together in a pop-up	Database report, observations, debriefing
Listen to sentence	Listening to sentences is an important part of language learning process	1. Geographical names are difficult to pronounce. Listening helps. 2. Listening helps memorize the pronunciation of words.	Database report, observations, debriefing
Tools to adjust instruction to students' needs	The tools were used to meet the needs of individual students.	1. I can listen to the dialogues and read captions at the same time; it helps. 2. I cannot pay attention to both captions and listening at the same time. 3. I can train my listening skills. I listen to the scene without captions and after that, with captions.	Database report, observations, debriefing
Irregular verbs table pop-up	The tool was appreciated by all the students.	1. Very convenient, you just click on the button and check whether the verb is regular or irregular.	Database report, observations, debriefing

Program features and aspects	Researchers' Claims	Students' comments	Source
Story-telling format	The findings support the idea of personalization effect by Mayer (2005)	1. I like the story because it is about an ESL learner and her family. If she can do it, I can also do it. 2. It is a good way to introduce what will be taught in the unit. 3. Good for practicing listening skills. 4. It is a concise and simple presentation, easier than in the grammar help pop up (a comment from a novice student).	Survey, debriefing
Cognitive training theory at work	The implementation of the cognitive training theory allowed the researchers to improve academic learning of the subject matter, ESL grammar and vocabulary (Kalyer & Phye, 2008).	1. It is a good way to learn grammar, makes you think. The explanation in the feedback is sufficient. 2. It helps you organize the information and see patterns. 3. It helps you understand how grammar works and memorize new information. 4. This program will work for lazy people; it is easy to memorize new words and understand grammar.	observations, debriefing, tests scores
Contextualized learning	Contextualized learning was motivational for all the students, but difficult for those who had poor background of basic geography. Despite the challenge, those students did well at the tests and learned some facts about the geography of the country.	1. Helps learn new grammar concepts faster because it is interesting 2. I like it, but it was not easy for me. I learned a lot about geography of the US. (This student did not know that Alaska belongs to the US) 3. I can say something about the geography of the country. I have lived in this country for 25 years, but I could not do it before 4. Grammar is used in real life situations, meaningful practice. 5. I want to use this program at home again to learn more about the geography of the US.	Students' test scores from the data base report, debriefing
Comparison of the program to regular classroom teaching	This method provides better chances for learners to build mental connections between: 1. Verbal, pictorial, and auditory representations (multimedia effect, Mayer, 2001) 2. Different passive voice grammar concepts: active vs. passive, singular vs. plural, present vs. past passive voice. 3. It also better addresses individual learning needs.	1. Images, maps, audios in addition to text help me memorize new information. 2. It makes you think and understand. Not many teachers can help you with that. 3. I can learn at my own pace, I cannot do it in the classroom; I have to catch up with the class. 4. The program is better than a textbook because it gives you correct answer and feedback right away.	Informal observations, debriefing

APPENDIX K: DATA COLLECTION PROCESS (STUDY GROUP B)

Time Schedule	Procedures
Day 1	Collecting demographics (background questionnaire) Pre-test (USA geography context)
Day 3	Training episode: Unit1-4 Informal observations Paper-based Likert survey administered at the end of each unit
Day 6	Training episode: Unit5-7 (picture description included) Paper-based Likert survey Debriefing Informal observations
Day 11	Delayed post-test (USA geography context)
Day 16	Knowledge transfer-test (context other than USA geography)

Note Each of the 3 tests consisted of 24 items, 12 selected response and 12 constructed response items.

APPENDIX L: DATA COLLECTION INSTRUMENTS AND ANALYSIS OF THE DATA (STUDY GROUP B)

Data Collection Instruments	How the Data Were Collected	How the Data Were Analyzed
1. Paper-based survey administered after each unit in the program by the principal investigator.	Students rated their preferences for different kinds of assignments in the program on a 5 point rating scale (5- strongly agree, 1- strongly disagree).	Students' ratings were compared with their prior knowledge of the grammar concept, general vocabulary, and basic geography.
	The users ranked each assignment on: 1. How interesting it was 2. How it helped them select, organize, and integrate the information	Was the students' interest in tune with their basic cognitive processing of the information?
2. Database report.	Students' knowledge gain between: 1. Pre-test and post-test 2. Pre-test and knowledge transfer-test	Students' knowledge gain was compared with their prior knowledge of the grammar concept, general vocabulary, basic geography, and compared to their ratings in the survey.
	The count of performance aids use.	Served as an indicator of students' preferences for different modes of instruction, verbal or visual.
	The count of the f Grammar Help pop-up and Irregular Verbs pop-up use.	Served as an indicator of students 'prior knowledge of grammar.
	The count of word annotations pop-up use.	Served as an indicator of prior knowledge of vocabulary.
3. Observations conducted by the principal investigator with each participant during the 2 training sessions.	The principal investigator was sitting beside each participant and typing the data from her observations in a Word document. A close look was given to students' attention during the training and their ability to read maps (locating oceans, continents).	How was users' attention related to their prior knowledge? How was students' ability to read maps related to their learning outcomes?
4. Debriefing after the training.	The principal investigator asked the participants to rate the choices below and typed their ratings in a Word document. 1.Need for visualization and audio. 2.Need for context for teaching grammar. 3.Preference for different assignments. 4.Preferred instructional setting for the use of the program. 5.Need for characters in the program.	How were students' individual preferences related to their prior knowledge?
5. One page demographics survey administered before the pre-test.	Students' demographics.	The data were used to determine students' basic demographics and their level of computer skills.

Note The program used behavior tracking technology to collect data about students' performance, navigation patterns, and the use of performance aids.

Item	Student 1	Student 2	Student 3	Student 4
Preference of instructional setting (debriefing, survey ratings).	She would rather use the program in the classroom with a teacher, but would also like to use the program at home.	At the beginning (Session 1) she would rather use it in the classroom so that she can ask questions. At the end (Session 2) she would rather use it at home and work at her own pace.	She does not need a teacher. She would rather use the program at home.	In class with a teacher.
Preference for narrated animation versus text for description of visuals (debriefing, survey ratings).	She likes the characters and the animation a lot. She likes the dialog between the characters because it explains information in a simple way.	It is good, but she would rather have text for describing images.	Needs more narrated animation. It makes learning more interesting and helps maintain my attention.	Excellent. It helps her focus and it is interesting for her.
Contextualized teaching of grammar (debriefing, survey ratings).	She thinks that it is interesting and motivational. A little difficult at the beginning. She likes that the primary focus of the program is grammar and the secondary focus is USA geography and history.	Very interesting and helps maintain attention.	Excellent. It was interesting and helped in maintaining attention.	Excellent. She liked it a lot. She would fall asleep without the contextual presentation.

APPENDIX M: RESULTS: STUDENTS' DIFFERENCES NOT RELATED TO PRIOR KNOWLEDGE

Item	Student 1	Student 2	Student 3	Student 4
Comparing and contrasting grammar concepts (debriefing, survey ratings).	She likes it a lot. She needs to understand before she can practice. Understanding goes first and helps memorize.	It is difficult, but useful. It helps her organize and integrate the information.	It is challenging but useful. It helps her organize and integrate the information.	She likes it.
Using a combination of text, audio, and images (debriefing, survey ratings).	Excellent. It helps memorize new information.	It helps her memorize the information better compared to working with a textbook.	It helps her memorize the information better compared to working with a textbook, helps maintain attention.	It helps her memorize the information better compared to working with a textbook and maintain attention.
Preference of instructional setting (debriefing, survey ratings).	In class with the teacher during the class period or sometimes at home "for more practice at my own pace".	In class with the teacher during the class period or without a teacher during the class period "At the beginning I needed help, during the second session I did not need a teacher".	In class with the teacher during the class period or without a teacher during the class period.	In class with the teacher during the class period.

APPENDIX N: RESULTS: STUDENTS' DIFFERENCES RELATED TO PRIOR KNOWLEDGE

Student	Test Scores		Prior Knowledge: Grammar Concept	Prior Knowledge: General Vocabulary	Prior Knowledge: Geography	Grammar Teaching Approach			Attention	Knowledge Gain	
	Post-Test (out of 24)	Transfer-test (out of 24)	Pre-Test Score (out of 24)	Word Annotation pop-up usage (times)	Observations	Use of grammar pop-up (times)	Overall preference grammar teaching approach: Inductive	Overall preference grammar teaching approach: Deductive	Observation	Retention (Post-Pre difference)	Transfer (Trans-Pre difference)
1	19	11*	7	22	Adequate	23		Y	Good	12	4*
2	20	19	12	35	Adequate	8	Y		Good	8	7
3	22	23	16	11	Adequate	6	Y		Sometimes did not read instructions	6	7
4	19	22	19	41	Limited – could not locate cardinal directions on a map	6	Y		Sometimes did not read instructions	0	3

CHAPTER 5: GENERAL CONCLUSION

Cognitive science has taught us a lot about how humans learn. Now computer-based learning programs are putting those principles into action and improving student knowledge gains (Graesser, 2011). In his research, Mayer (2008) suggests that the design of multimedia instruction can be informed by the science of learning and the science of instruction because the relationship between the science of learning and the science of instruction is reciprocal.

The first contribution of this dissertation has been to demonstrate this reciprocity. The instructional design of three interactive multimedia instructional programs serving as interventions that provide solutions to instructional problems was grounded in cognitive theories. The instructional principles based on research findings from cognitive psychology were evaluated in authentic instructional situations to make sure that the implemented instructional principles were consistent with what is known about how people learn. This was done in response to the need of deriving instructional principles that are both grounded in theory and supported by evidence from authentic tasks (Mayer, 2008).

The second contribution of this dissertation is an attempt of conducting holistic formative evaluations of the designed interactive multimedia products. The research on formative evaluation of interactive multimedia instructional programs has been composed of usability studies or experimental research on the effectiveness of multimedia program features. In contrast, this research provides an in-depth analysis of both the overall effectiveness of the web-based multimedia instructional programs with embedded tracking and program features at the same time. Behavior tracking and embedding most of the data collection instruments in the programs allow the researchers efficiently collect data from three perspectives:

- students'(their perceptions of the programs collected through interviews, informal observations, and imbedded Likert-scale surveys)
- instructional designers'(students' test scores, think-aloud protocols)
- multimedia instructional designers (navigation patterns, the count of program features use from a database report, interviews, observations).

Key Findings

Chapter 2 presents a design-engineering-development approach to innovation (Bryk & Gomez, 2008) based on the use of the instructional program as a research platform. This approach offers more possibilities for testing early in the design process the potential program features identified as effective in the previous research. The potential program features can be compared during the design study experiments since most of the data collection instruments are embedded in the program. By using this approach, the features that show the most effectiveness are implemented in the final product. The insights on how this approach can contribute to the design process are provided.

In chapter 3, the researchers argue that model explorations vs. model building can be effectively used by for introducing unfamiliar chemistry concept to students. While model building could be effective at later stages of the instructional process, model exploration could be more appropriate for earlier stages such introductory lessons.

The interactive Flash-based program was designed as a formative assessment around the PhET simulation lab. According to the teacher's and students' comments, the program helped students effectively learn about gas behavior by allowing them to test out their hypotheses and correct errors and misconceptions in their thinking (Rieber, Tzeng, & Tribble 2004). Students' experience was designed as model exploration and worked well for low prior knowledge students, the vast majority in the classrooms in which it was used. According to the comments from high prior knowledge students, they could have benefited from a more challenging activity such as model building.

In this study, an innovative instructional approach based on the integration of three cognitive theory models was implemented and tested:

- Mayer's multimedia learning theory (2001)
- Sweller's cognitive load theory (CLT, 2005)
- cognitive training theory using an inductive reasoning paradigm (Klauer & Phye, 2008)

The preliminary results were encouraging for this integration.

Recommendations for Future Research

This dissertation is focused on design studies of interactive multimedia programs. The findings from the research in Chapter 2 suggest that the next stage of the design study needs to focus on experimental research comparing three conditions (problem scenario augmented with animation, problem scenario augmented with still images, and text only scenarios) to make a choice as to which type of graphics should be implemented in the final version of the multimedia instructional program.

The specific suggestions for future research for the paper described in Chapter 3 include gathering data of students' performance on pre-test, delayed post-test, and knowledge transfer test in order to collect experimental data supporting the effectiveness of the program.

As to the paper described in Chapter 4, there are three specific suggestions for future research that may be important to understanding of how multimedia instructional programs focused on a particular level of students' expertise with linear navigation and predetermined feedback are comparable with the multimedia programs with intelligent feedback. The experimental research is necessary to compare the learning outcomes of students exposed to both methods.

Another important aspect of future research is conducting empirical studies of the effectiveness of the software program described in this dissertation on a larger number of participants in various classroom settings. This empirical research is meant to yield generalizable findings as a basis for further investigation on a larger scale. The use of the software applications on a larger scale will need evaluation of possible teacher and school factors that promote the quality of software integration into the curriculum.

References

Bryk, A. S., & Gomez, L. M. (2008). Ruminations on reinventing an R&D capacity for educational improvement. In F. M. Hess (Ed.), *The future of educational entrepreneurship: Possibilities for school reform*. Cambridge: Harvard Education Press.

Graesser, A.C. (2011). Learning, thinking, and emoting with discourse technologies. *American Psychologist, 66,* 743-757.

Klauer, K., & Phye, G. (2008). Inductive reasoning: A training approach. *Review of Educational Research, 78*(1), 85–123. doi:10.3102/0034654307313402.

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

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.

Rieber, L. P., Tzeng, S., & Tribble, K. (2004). Discovery learning, representation, and explanation within a computer-based simulation: Finding the right mix. *Learning and Instruction, 14*, 307–323

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp.19–30). Cambridge, UK: Cambridge University Press.