The impact of web-based worked examples and self-explanation on performance, problem solving, and self-efficacy

Kent J. Crippen a,*, Boyd L. Earl b

a Department of Curriculum and Instruction, University of Nevada Las Vegas, P.O. Box 453005, 4505 Maryland Parkway, Las Vegas, NV 89154-3005, United States
b Department of Chemistry, University of Nevada Las Vegas, 4505 Maryland Parkway, Las Vegas, NV 89154-4003, United States

Received 3 November 2005; accepted 13 November 2005

Abstract

Studying worked examples and engaging in self-explanation are well-supported strategies for developing self-regulated learning and improving student performance. Our efforts involve a design theory approach to creating a Web-based learning tool that uses the notion of a weekly quiz to draw students into an environment that supports the development of fundamental strategies for improving performance and well-structured problem solving. This manuscript describes a quasi-experimental study to isolate the specific impact of our tool on student learning and motivation. Results are inconclusive about an impact for students only provided with worked examples. The combination of a worked example with a self-explanation prompt produces improvement in performance, problem solving skill, and self-efficacy.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Applications in subject areas; Architectures for educational technology system; Distance education and telelearning; Interactive learning environments; Multimedia/hypermedia systems
1. Introduction

Our work focuses on using a Web-based software environment to develop well-structured problem solving in the domain of chemistry. Our context is large undergraduate, introductory sections of students. Our software is proprietary, but not commercial. Our efforts involve a design theory approach to creating a Web-based learning tool that uses the notion of a weekly quiz to draw students into an environment that supports the development of fundamental strategies for improving well-structured problem solving using empirically supported techniques.

This manuscript describes a quasi-experimental study to isolate the specific impact of our tool on student learning and motivation. The project is grounded in the literature supporting the use of worked examples and self-explanation strategies for developing student problem solving and leverages the hypermedia characteristics of a Web-based repeatable testing environment for delivery. Our previous research (Crippen & Earl, 2004) has shown that: (a) students make extensive use of both worked examples and self-explanation prompts (b) students self-report that both strategies are helpful for improving both their learning and performance (c) use of the learning system correlates significantly with performance. The current effort aimed to establish a causal relationship between use of the Web-based learning tool and increases in student performance and self-efficacy.

This paper will: (1) provide a theoretical rationale for our project, (2) describe our Web-based learning system and the dimensions of the current study, (3) report the results of the intervention, and (4) discuss conclusions and future research. Further information from this project, including copies of instruments, is available via the Web (http://crippen.nevada.edu/chemistry/WE_Study/).

2. Theoretical framework

Problem solving has long been held as an important part of mathematical and scientific literacy. This value is reflected in the national K-12 standards for both mathematics and science instruction (National science standards, 2003, Principles and standards, 2003). We view problem solving as falling into one of two categories. Depending upon the nature of the problem-solving situation, the task may be classified as well-structured or ill-structured. Ill-structured tasks tend to be open-ended and are characterized by the lack of a socially accepted answer (e.g., resource allocation issues, global warming, water quality). Well-structured problems have an accepted answer and require students to use a mental heuristic to determine an appropriate algorithm for generating a problem solution.

Well-structured problems have been used traditionally as the basis for mathematics learning, and to a lesser, but significant extent, science learning. As curriculum has evolved, focus from the science-technology-society (STS) movement and the emphasis on a more general notion of scientific literacy have pushed the use of ill-structured tasks. Though the degree of emphasis and chronology of introduction can be hotly debated, most scientists and science educators agree that both types of problems are appropriate for introductory students. Regardless of the problem type, or degree of difficulty, developing problem solving skill is complex.

Our project suggests that providing worked examples and self-explanation prompts within test items in a Web-based, repeatable testing environment has the potential to improve problem solving skill and conceptual understanding. This is accomplished by giving students the opportunity...
to develop their own problem solving strategies as a result of focusing their attention on problem states and problem solving operators.

2.1. Worked examples

Worked examples are detailed problem solutions that contain identifiable qualities and characteristics (Ward & Sweller, 1990). These representations are constructed in such a way as to provide the learner with some structure for understanding how the solution was established without providing a script or algorithm (Atkinson, Derry, Renkl, & Wortham, 2000). Depending upon the topic area, these solutions can take many forms. Snippets of computer program code can be effective worked examples for computer programming while structural diagrams work well for architecture and engineering. Worked examples have been shown successful in computer programming, algebra, and geometry instruction (Carroll, 1994; Paas & van Merrienboer, 1994). Worked examples instruction has shown to be more appropriate for inexperienced learners while problem solving practice is more appropriate for experienced students (Kalyuga, Chandler, Tuovinen, & Sweller, 2001).

From a cognitive perspective, studying worked examples lessens the load on working memory while focusing student attention on problem states and problem solving operators (moves) (Sweller, 1988, 1994). Students who have been engaged in studying worked examples as an instructional strategy adopt problem solving techniques more quickly and have improved problem solving performance (Chandler & Sweller, 1991; Cooper & Sweller, 1987; Sweller & Cooper, 1985; Tarmizi & Sweller, 1988). This typically manifests itself as instruction that is more efficient and improves student performance on problems requiring transferred knowledge and skills.

Using worked examples to improve problem solving in science instruction is promising (Taconis, Ferguson-Hessler, & Broekkamp, 2001), but validation of these strategies in large scale, practical implementations is needed and noticeably missing from the literature. Further, these techniques are not easily replicated with the Web and simply providing Web-based worked examples is not likely to be successful (Smith & Jacobs, 2002).

2.2. Self explanation

Self-explanation is a form of self-talk where a learner engages in an iterative personal dialog while engaged in problem solving. This dialog aids the learner in identifying problem states and potential solution moves. Good problem solvers exhibit a larger volume of focused self-explanations than do novices and poor problem solvers (Chi, Bassok, Lewis, Reimann, & Glaser, 1989). This form of self-regulation is intricately tied to successful problem solving (Bielaczyc, Pirolli, & Brown, 1995a).

Engaging in self-explanation complements studying worked examples. Together, both strategies are useful for improving student problem solving. In fact, students who engage in self explanation of worked examples have access to those examples for creative problem solving strategies (Didierjean & Cauzinille, 1997).

2.3. Self-efficacy

Self-efficacy, a motivational belief, plays a crucial role in science learning (McRobbie & Thomas, 2000; Roth & Tobin, 2001). Self-efficacy refers to the degree to which an individual is confident that
they can perform a specific task or accomplish a specific goal (Bandura, 1997). Self-efficacy is extremely important for self-regulated learning because it affects the extent to which learners engage and persist at challenging tasks. Students with higher self-efficacy are more likely to engage in a difficult task and more likely to persist at a task even in the face of initial failures compared to low-efficacy students (Pajares, 1996). Considering that students generally lose confidence and interest in science with age, improving self-efficacy is an important goal (Pell & Jarvis, 2001). A number of studies indicate that teacher and student self-efficacy plays an important role in science education (Cannon & Scharmann, 1996; Schoon & Boone, 1998).

Expert modeling can improve self-efficacy and cognitive strategies. Modeling occurs when students learn intentionally from other individuals. Students are most likely to increase their self-efficacy when observing a model of similar ability level performing the skill (Schunk, 1996). Worked examples serve as expert models.

2.4. Web-based repeatable testing

As an instructional strategy, Web-based repeatable testing has shown to improve student performance for well-structured problems (Dufresne, Mestre, Hart, & Rath, 2002; Penn & Nedeff, 2000). The design of nearly every delivery system employs a simple feedback model where a learner’s response generates performance related feedback. This feedback is typically an acknowledgment of the correctness of the response and suggestions for improvement. This is an electronic application of the classic model where instruction is followed by large amounts of individualized practice. The basic model operates on the assumption that the learner is engaged with the system and is modifying their understanding accordingly.

While use of Web-based repeatable testing produces performance gains, learning rates are shallow (Crippen & Brooks, 2002). Close examination of the reported results reveals that in nearly every study, the number of items required for students to achieve mastery can be staggering. Experience suggests that simple repeatable testing systems, while successful on some level, are largely inefficient and may not improve problem solving skill. Further, issues of design play a critical role (Brooks & Crippen, 2001).

Our primary goal is to develop students who are better solvers of well-structured problems. Our project involves a contextualized research and design study focused on a more robust Web-based repeatable testing system where project evaluation focuses on the outcomes of student performance and motivation. This research emphasizes formative assessment where students are provided with research supported techniques embedded in traditional assessment items. The goal, consistent with national efforts, is to improve instructional efficiency, student performance, and problem solving skill using Web-based efforts (Donovan, Bransford, & Pellegrino, 2001).

3. The study

The centerpiece of our project is a Web-based learning tool that uses the notion of a weekly quiz to draw students into an environment that supports the development of fundamental strategies for improving well-structured problem solving. Our work involves 3 years of research and development using a design theory methodology that emphasizes an ongoing repetitive cycle of
design, testing, data collection, and analysis (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003). Consistent with design theory, our analysis and development cycles rely heavily on retrospective analysis.

This study investigated the differential impact of two Web-based instructional interventions on student performance and self-efficacy. Using a quasi-experimental design during one semester of instruction, the conditions of worked examples instruction and worked examples with self-explanation prompts were compared to a control situation. It was anticipated during the design of this study that the small sample size for the intervention groups would be problematic for significance testing of any single factor. Our hope was to gain a better understanding about the larger interplay of worked examples by examining the summative descriptive impact on a series of measures.

The study was designed to use a control group comparison to answer the following research questions:

- Does the worked example condition improve student performance?
- Is the worked example condition improving well-structured problem solving and student self-efficacy?
- Does combining self-explanation prompts with worked examples increase the impact of the worked examples condition on the dependent variables of student performance and self-efficacy?

Chemistry 115 – General Chemistry I for all science majors represents the context for this study. This is a large introductory lecture course with a laboratory component where enrollment typically totals about 120 students. The Web-based learning system used to deliver the interventions in this study was open to all students. While use of the system may have implied a large amount of a student’s time, their scores only accounted for 5% of the overall grade. This study represents one semester of data in an ongoing project that has been in effect for 3 years (Crippen & Earl, 2004).

Ours is a sample of convenience and our research method is quasi-experimental. The students in our experimental section were assigned by the campus course information system and participants provided informed consent. Regardless of inclusion in our study or experimental condition, all students had access to the Web-based learning system.

4. Methods

Using our combined expertise, the authors have personally been building the item and example databases since the inception of our project. Quiz items are written and assigned based upon content, level of difficulty, and problem characteristics. Quiz content is matched with the weekly lecture topics and worked examples are built in support of techniques taught in lecture and described in the textbook.

In our system, students have access to a Web-accessible content quiz for 1 week. During this period, students can modify their responses at any time as their skills and understanding of the material change. The quizzes are graded, with correct/incorrect results and the correct answers provided, at the end of the week. Principally, these assessments are designed as a learning opportunity. Students are evaluated, but the instructional objective is to use the assessments as a mechanism for engaging
students in strategies for improving their well-structured problem solving skill. Due to the nature of the content, not all of the items for every quiz focus on well-structured problem solving. Typically, three of the five items focused on well-structured problem solving.

Students were randomly assigned to one of three conditions; control ($n = 18$), worked example ($n = 24$) or worked example/self-explanation ($n = 24$). All three groups had access to the Web-based learning system and operated under the same quizzing structure. The experimental conditions of worked example (WE) and worked example/self-explanation (WE and SE) included additional content delivered with the assessment items of the weekly quizzes (Fig. 1).

Both experimental conditions offered the possibility of up to three examples (appearing on the Web page as buttons) for each of five items on every quiz. Worked examples were provided for every item that explicitly required well-structured problem solving. Each example was unique, but the collection of examples for any given item were chosen to be consistent in the problem solving strategy needed to successfully answer the item. The notion of providing three unique examples and the design of each example were consistent with guidelines found in the literature (Ainsworth & Loizou, 2003; Atkinson et al., 2000; Bielaczyc, Pirolli, & Brown, 1995b; Catrambone, 1996).

When a student in either experimental condition clicked an example button, a new window opened presenting the worked example. The two experimental conditions differed in the nature and amount of information provided when the student clicked any one of the example buttons. Both experimental groups saw the worked examples. The WE and SE group had additional information in the form of a self-explanation prompt (labeled a suggestion) with every example (Fig. 2). Like the content of the worked examples, the prompt for each example was unique. The impetus for the WE and SE condition is the notion that combining problems with examples

3. For the reaction shown below, which statement is true?

$$2\text{Fe} + 3\text{CdCl}_2 \rightarrow 2\text{FeCl}_3 + 3\text{Cd}$$

- Fe is the oxidizing agent
- Cd undergoes oxidation
- Fe undergoes oxidation

3. For the reaction shown below, which statement is true?

$$2\text{Fe} + 3\text{CdCl}_2 \rightarrow 2\text{FeCl}_3 + 3\text{Cd}$$

<table>
<thead>
<tr>
<th>Example 1</th>
<th>Example 2</th>
<th>Example 3</th>
</tr>
</thead>
</table>
| Fe is the oxidizing agent
| Cd undergoes oxidation
| Fe undergoes oxidation

Fig. 1. A representative quiz item from the control condition (a) and from either of the experimental conditions (b).
is better than examples alone (Stark, Gruber, Renkl, & Mandl, 2000). The two experimental conditions allowed for isolation of the learning enhancements for worked examples and possible combined effects of worked examples with self-explanation prompts.

Data collection involved harvesting the four in-class mid-term examinations, the final examination, and a 34-item self-efficacy measure delivered post instruction. Exam scores were used as a general performance indicator. Exam items were further sorted and classified to develop a measure for well-structured problem solving. Homogeneity of the groups for ability was established using one-way between-groups analysis of variance (ANOVA) with college entrance exam scores (Scholastic Aptitude Test, SAT), $F(2, 45) = 0.335, p = 0.717$ and undergraduate grade point average (GPA) $F(2, 69) = 2.313, p = 0.107$ scores prior to instruction.

The mid-term exams were prepared by the second author (BE) and were not specifically designed to address the worked examples content; rather to address all of the learning for the course. The first author (KC) matched exam items with worked example concepts ex post facto. Only a portion of the hour exam items measured worked examples content. The identified items were classified as either knowledge/conceptual or well-structured based upon the type of question.

A 34-item measure of self-efficacy was constructed and delivered electronically post instruction. The concepts and strategies represented in the worked examples were used to construct task statements (e.g., “Identify the oxidizing/reducing species in a re-do x reaction.”). Students were asked to rate their confidence in completing each task and responded using a 6-category Likert-type scale ranging from no confidence to complete confidence. The sum of the item responses was used as a continuous scale.
Statistically, a one-way between-groups ANOVA was applied to all metrics, but significant results were neither anticipated, nor found, due to the small sample size. We acknowledged in the design phase that our study would not provide adequate statistical power for meaningful significance testing. Our analysis focused on a holistic look, taking into account the descriptive results of each measure.

5. Results

5.1. Student performance

Table 1 presents a comparison of the exam scores for each condition. The performance of both experimental groups is qualitatively more consistent than the control group. The control group has a more downward trend for exams two and four than either of the experimental groups. With the exception of exam four, the standard deviations suggest a consistency in the distributions of scores among groups.

The WE group performed below the standards of the control group. With the exception of the fourth hour exam, the mean score for the WE condition are below the control, suggesting no effect for this condition. In comparison to the control group, the first and final exams were disastrous for the WE condition.

The WE and SE group outperformed both the control and WE groups on every exam. Though not statistically significant, the mean scores of the WE and SE group were above the control group in every case and show a degree of consistency. This result suggests an effect for this condition.

5.2. Well-structured problem solving skill

Performance on the multiple choice and well-structured items from the hour exams were disaggregated for comparison. In general, the multiple choice items focused on factual and conceptual knowledge. The well-structured items were short response problems requiring a multi-step solution. Multiple choice items were graded correct/incorrect. Well-structured items were allotted a given number of points and students earned partial credit by accurately completing some of the required steps. Most, but not all, of the well-structured problems contained content parallel

<table>
<thead>
<tr>
<th>Exam</th>
<th>Control</th>
<th>WE</th>
<th>WE and SE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>75.30</td>
<td>69.42</td>
<td>76.92</td>
</tr>
<tr>
<td></td>
<td>17.36</td>
<td>16.74</td>
<td>14.45</td>
</tr>
<tr>
<td>1</td>
<td>64.08</td>
<td>63.08</td>
<td>67.21</td>
</tr>
<tr>
<td>2</td>
<td>18.82</td>
<td>16.71</td>
<td>18.83</td>
</tr>
<tr>
<td>3</td>
<td>71.20</td>
<td>70.25</td>
<td>72.08</td>
</tr>
<tr>
<td>4</td>
<td>23.84</td>
<td>23.30</td>
<td>16.29</td>
</tr>
<tr>
<td></td>
<td>61.76</td>
<td>62.92</td>
<td>66.71</td>
</tr>
<tr>
<td></td>
<td>63.32</td>
<td>60.19</td>
<td>65.28</td>
</tr>
<tr>
<td></td>
<td>18.20</td>
<td>17.04</td>
<td>19.87</td>
</tr>
<tr>
<td>Final exam</td>
<td>63.32</td>
<td>60.19</td>
<td>65.28</td>
</tr>
</tbody>
</table>
to that found in the on-line worked examples. We felt that the well-structured items were a good measure of the transfer effects for the worked example content to the classroom and any difference in the experimental group scores would illustrate a differential development in problem solving skill for this context.

With the exception of exam four, the mean score of the WE condition was never above that of the control group (Fig. 3). In addition, the mean score for multiple choice items for exam four in the WE condition was also higher than the WE and SE condition. The mean score on the multiple choice items for the WE and SE condition was always above the control group. The degree of difference in mean scores between the groups was inconsistent.

When comparing mean scores on well-structured items, exam three was the only case where the control group was higher than either experimental condition (Fig. 4). The mean scores for the experimental conditions are well above those of the control group for three of the five exams. As the semester progressed, a qualitative positive difference appears in mean scores for well-structured items for each experimental group.

5.3. Self-efficacy

A comparison of raw scores on the self-efficacy measure illustrates a marked positive difference for the WE and SE group (Table 2). In fact, regardless of the small group sizes, the difference
among the groups was nearly statistically significant \(F(2, 64) = 2.830, p = 0.067\). Consistent with the exam results, the mean self-efficacy score for the WE group is below the control group as well as the WE and SE group. Worked examples alone are not producing differences in self-efficacy, but the combination of a worked example with a self-explanation prompt is having a positive impact on student self-efficacy.

6. Conclusions

We believe that the collection of descriptive measures from this study have some very powerful things to say about our project. Results are inconclusive about an impact for students provided only with worked examples. Considering the results of performance on the multiple choice items, worked examples alone did not improve student performance. Student performance in the WE condition improved on average for well-structured problems as the semester progressed. Without guidance, it may take time for students to figure out how best to use the worked examples they are provided. Worked examples alone may have a very refined impact on student performance. By themselves, they may work best for only specific types of well-structured problems and not for others. Clearly, the WE condition had no impact on student self-efficacy for worked example concepts. Simply providing worked examples may be impacting student performance for very specific types of problems, but neither a general problem solving improvement nor an improvement in self-efficacy are supported by the results of this study.

The combination of a worked example with a self-explanation prompt seems to produce a difference in performance, problem solving skill, and self-efficacy. The results of the WE and SE condition are very encouraging. The use of Web-based worked examples that include a tailored self-explanation prompt consistently improves student performance on well-structured problems when measured with a traditional exam. This result manifests as improved overall exam performance and final course grades. The results of the self-efficacy measure suggests that the combination of these two interventions in the WE and SE condition has a big impact on student motivation.

7. Discussion

The learning aid of a worked example with a tailored self-explanation prompt improves student performance, well-structured problem solving skill, and motivation within a Web-based learning system. This study suggests that these improvements require both interventions; worked examples alone are not enough. We believe that the added benefit of having the self-explanation prompt

<table>
<thead>
<tr>
<th>Control</th>
<th>WE</th>
<th>WE and SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>148.50</td>
<td>143.13</td>
</tr>
<tr>
<td>SD</td>
<td>39.31</td>
<td>30.84</td>
</tr>
</tbody>
</table>
provides students with a contextual guide for interpreting the examples and a queuing mechanism for activating learning strategies. Both of these outcomes would account for the increase in self-efficacy (motivation) and performance. Without the self-explanation prompts, the degree of detail provided in the worked example may be overwhelming. If students are overwhelmed and do not know what to do with the worked examples, they may have an intuitive sense that they are helpful, but they fail to engage with them. Failure to engage, or an inconsistent engagement, would drastically hinder any positive effects on performance or motivation.

The positive impact of the combination of worked example with a self-explanation prompt on student self-efficacy is especially intriguing. This outcome has important implications considering the known connection between self-efficacy and persistence with difficult content. Coupled with our previous work showing that students use these interventions heavily and have positive feelings about them, the results of the current study make a strong case for including embedded worked examples with self-explanation prompts in the design of on-line testing systems.

7.1. Implications for future research

Small sample sizes are a very real problem in evaluating the effectiveness of our Web-based learning system. Instructor differences create seriously confounding contextual variables for identifying a suitable control group. Further, we find it difficult to justify not providing interventions to groups of students in a vain effort to create a control group for statistical significance testing. We believe the volume of data collected to date supports our method of embedding worked examples with a self-explanation prompt within a Web-based learning system for improving performance, well-structured problem solving, and self-efficacy. In the future, all students will have access to this condition and it will be used as a comparison for other treatments. Our future research focuses on the content and design of the examples and prompts to maximize the positive outcomes illustrated in the current study. For instance we are developing a mechanism for providing the self-explanation prompt in an audio format and experimenting with a modified multiple choice item that forces the student to justify their response by selecting an appropriate worked example.

References


