CrossMark

ARTICLE

Delayed Learning Effects with Erroneous Examples: a Study of Learning Decimals with a Web-Based Tutor

Bruce M. McLaren[1] · Deanne M. Adams[2] ·
Richard E. Mayer[3]

Abstract Erroneous examples – step-by-step problem solutions with one or more errors for students to find and fix – hold great potential to help students learn. In this study, which is a replication of a prior study (Adams et al. 2014), but with a much larger population (390 vs. 208), middle school students learned about decimals either by working with interactive, web-based erroneous examples or with more traditional supported problems to solve. The erroneous examples group was interactively prompted to find, explain, and fix errors in decimal problems, while the problem-solving group was prompted to solve the same decimal problems and explain their solutions. Both groups were given correctness feedback on their work by the web-based program. Although the two groups did not differ on an immediate post-test, the erroneous examples group performed significantly better on a delayed test, given a week after the initial post-test ($d=.33$, for gain scores), replicating the pattern of the prior study. Interestingly, the problem solving group reported liking the intervention more than the erroneous examples group ($d=.21$ for liking rating in a questionnaire) and found the user interface easier to interact with ($d=.37$), suggesting that what students like does not always lead to the best learning outcomes. This result is consistent with that of *desirable difficulty* studies, in which a more cognitively challenging learning task results in deeper and longer-lasting learning.

✉ Bruce M. McLaren
   bmclaren@cs.cmu.edu

   Deanne M. Adams
   deanne.adams@gmail.com

   Richard E. Mayer
   mayer@psych.ucsb.edu

[1]   Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA

[2]   Department of Psychology, University of Notre Dame, South Bend, IN 46556, USA

[3]   Department of Psychological and Brain Sciences, University of California, Santa Barbara,
   CA 93106-9660, USA

Introduction

A somewhat unusual but potentially productive instructional technique is learning from *erroneous examples*, problem examples with step-by-step solutions that have one or more errors, and for which students are prompted to find and fix the error(s). Interestingly, such examples have been controversial in education (Tsamir and Tirosh 2003). This is likely due to behaviorist theory (Skinner 1938), and more specifically stimulus–response theory (Guthrie 1952; Hull 1952), that proposes that exposing students to errors will make them more prone to make the errors themselves. Yet, some theorists propose that erroneous examples provide unique learning opportunities, particularly in mathematics, where students might improve their understanding and problem solving skills, as well as develop reflection and critical thinking skills, by grappling with errors in example solutions (Borasi 1996). According to this theory, directly confronting students with errors and prompting reflection may lead to the eradication of the errors, similar to what has been shown in learning research on misconceptions (Bransford et al. 1999). Yet, the argument for the potential instructional value of erroneous examples appears to have swayed few educational practitioners, with medical training being one of the few areas that has embraced learning from errors (e.g., Gunderman and Burdick 2007). Surgeons routinely use "Morbidity and Mortality" (M&M) rounds, discussions of what went wrong in actual surgical procedures, as an instructional opportunity for other surgeons and residents and to avoid these errors in the future (Dr. Janet Durick, personal correspondence). Also, a variety of medical websites use erroneous examples as a key instructional technique (WHO 2014; The Doctor's Company 2013; National Health Care 2013). There are other examples of students learning from errors, such as students being asked to debug buggy computer code (Swigger and Wallace 1988) or find and correct errors in writing (Shoebottom 2015; CollegeBoard 2015). Nevertheless, learning from erroneous examples is far from a routine method of learning in most educational contexts.

Our goal in this study was to explore whether middle-school math students could learn better from erroneous examples than from the more traditional instructional approach of problem solving. Furthermore, our goal was to conduct the study with the support of educational technology, providing students with web-based, interactive erroneous examples in which they received feedback on the correctness of their work and were interactively prompted to find, explain,[1] and fix the errors. In comparison, students who did more traditional problem solving also worked with web-based instructional materials and were also supported with correctness feedback on their work.

---

[1] The self-explanations in this study were selected from a menu rather than generated by the learner. Since the literature reports studies with both approaches, it is important to make clear the type used in this study. There is some evidence that selecting from a menu is more effective than generating explanations when students work in a fast-paced, computer-based learning environment (Johnson and Mayer 2010; Mayer and Johnson 2010).

Our hypothesis, which we refer to as the *erroneous examples hypothesis*, is that students learn and understand mathematics at a deeper level when they are prompted to engage in the active cognitive processes of identifying, explaining, and fixing errors in the erroneous solutions of others. Further, we propose that students might find erroneous examples less desirable and more challenging to work with, even if such materials could help them learn and understand mathematics at a deeper level. Erroneous examples include an element of problem solving, through prompting students to find and fix the errors, and this is likely to tax working memory and increase cognitive load, as has been seen with conventional problem solving (Sweller et al. 1998). In addition to the problem solving aspect of erroneous examples, students are confronted with a deceptive and incorrect solution, which is something they are expected to find particularly challenging, due to their unfamiliarity with this type of example. For these reasons, we conjecture that students will like learning from erroneous examples less than conventional problem solving. Finally, we propose that exposing students to erroneous examples of decimals might make them more aware of their own decimal misconceptions, an important step toward addressing and ameliorating the misconceptions.

Prior Research on Learning from Erroneous Examples

A plethora of research has shown the advantages of learning from correct worked examples (Catrambone 1998; Kalyuga et al. 2001; McLaren et al. 2008; Paas and van Merriënboer 1994; Renkl 2014; Renkl and Atkinson 2010; Schwonke et al. 2009; Sweller and Cooper 1985; Zhu and Simon 1987). The theory behind the worked examples effect is that human working memory, which has a limited capacity, is taxed by strictly solving problems, which requires focused thinking, such as setting subgoals (Catrambone 1998). As mentioned above, problem solving has been shown to consume cognitive resources that could be better used for learning. Worked examples free cognitive resources for learning, in particular, for the induction of new knowledge by generative processing (Sweller et al. 2011).

In contrast, the case for erroneous examples is that they may stimulate generative processing and active learning through the prompting of students to determine what is wrong with a given problem solution and how to fix the error(s). It also appears that erroneous examples may help students become better at evaluating and justifying problem solutions, which, in turn, may help them learn material at a deeper level, with more lasting effects.

Surprisingly, there has not been much empirical research on the learning benefits of erroneous examples, particularly in the context of learning with educational technology. One of the first researchers to experiment with erroneous examples as a possible instructional technique was Siegler (2002). He investigated whether presenting third and fourth grade students with both correct and erroneous examples of mathematical equality, and asking them to self-explain those examples, was more beneficial than asking them to self-explain correct examples only or to self-explain their own solutions. He found that students who studied and self-explained both correct and erroneous examples led to the best learning outcomes of the three groups. Groβe and Renkl (2007) studied whether explaining both correct and incorrect examples made a difference to university

students as they learned mathematical probability. Their studies showed learning benefits for erroneous examples for learners with higher prior knowledge on far transfer learning. When errors were highlighted, low prior knowledge individuals did significantly better, while high prior knowledge students did not show any benefit, presumably because they were already able to identify errors on their own. Durkin and Rittle-Johnson (2012) tested whether comparing incorrect and correct decimals worked examples (the "incorrect" condition) promotes greater learning than comparing two correct decimals examples (the "correct" condition). They found that the "incorrect" condition helped students learn more procedural knowledge and key concepts, and also lessened their misconceptions. Unlike Groβe and Renkl, they did not find this effect to be exclusive to higher prior knowledge students.

A recurrent theme of empirical research on both correct worked examples and erroneous examples is the prompting of self-explanation to encourage students to process examples at a deeper level as they study them. Both the Siegler (2002) and Groβe and Renkl (2007) studies led to an erroneous example effect when students were not only prompted to study the erroneous examples but also to self explain those examples. It is thought that self-explanation triggers generative processing, which, in turn, supports learning. Chi et al. (1989) were the first to explore this phenomenon, the now well known and instructionally robust *self-explanation effect* (Chi 2000; Renkl 2002), finding that good problem solvers are more likely to self-explain when studying worked examples of physics problems. Explicitly prompting for self-explanation has also been found to be valuable for learning (Chi et al. 1994; Hausmann and Chi 2002; King 1994) and for better performance on transfer items (Atkinson et al. 2003; Hausmann and Chi 2002; Wylie and Chi 2014). Given the robustness of these findings and this line of research, our use of erroneous examples also involves prompting for self-explanation.

While the earlier described studies on erroneous examples were paper based, there have been a few studies in which students learned by interacting with erroneous examples supported by educational technology. For instance, Tsovaltzi et al. (2012) presented erroneous examples of fractions to students using an interactive intelligent tutoring system with feedback. They found that 6th grade students improved their metacognitive skills when presented with erroneous examples with interactive help, as compared to a problem solving condition and an erroneous examples condition with no help. Older students – 9th and 10th graders – did not benefit metacognitively but did improve their problem solving skills and conceptual understanding by using erroneous examples with help.

A study by Booth et al. (2013) with a computer-based algebra cognitive tutor found that prompting students to explain both correct and erroneous examples significantly increased posttest performance compared to students who only explained correct solutions. In addition, students who received only erroneous examples showed higher encoding of conceptual features compared to students who received only correct examples. The authors concluded that combining incorrect examples with correct examples can increase conceptual understanding of algebra. Huang et al (2008), experimenting with a software tutor focused on decimals and fractions, found that having students address cognitive conflicts associated with their own errors significant-ly increased learning compared to students who studied by working with review sheets

only. After committing an error, students in the tutor group were not confronted with their mistake directly but were presented with a cognitive conflict screen related to the misconception. The cognitive conflict screen was designed to help students recognize the error in their thinking and was followed by an instruction screen to clarify misconceptions. Students in the tutor group scored significantly higher on an immediate and a delayed posttest than the review sheets group. The results also showed that the tutor was significantly more effective for students with the lowest scores on the pretests.

Adams et al. (2014) compared an interactive erroneous examples condition to a supported (i.e., correctness feedback) problem solving condition. In this study, sixth-grade students learned about decimals using the web-based instructional technology described in the current paper. With 100+ students per condition, a delayed erroneous example effect was found. Although there were no significant differences on an immediate posttest, students who worked with the erroneous examples did significantly better on a delayed posttest than the problem solving students. There was no interaction between prior knowledge and condition, showing that erroneous examples were beneficial to both high and low prior knowledge students, contrary to the findings of the Große and Renkl (2007) study, in which only high prior knowledge students benefited from erroneous examples, or the Huang et al (2008) study, in which low prior knowledge students benefitted more from erroneous examples than high prior knowledge students.[2] The current study is a replication of the Adams et al. (2014) study, with a larger population of students. Given the previous pattern of results in which the erroneous examples treatment resulted in improved performance on a delayed test but not on an immediate test, our goal was to determine whether the pattern from the earlier study would be replicated in a larger-scale study.

A key distinction between the present study and past studies of erroneous examples is the exploration into the relationship between *liking* and *learning*. An implicit assumption of many educators, and even learning scientists, is the notion that students should *like* what and how they are learning. This is certainly a key reason behind the recent surge to investigate educational games (cf. Gee 2003; Aleven et al 2010; Lomas et al 2013). The current study investigates this important issue of whether liking is necessary or important to learning.

Background on Decimal Learning and Common Decimal Misconceptions

It is well documented that students often have difficulty understanding decimals, a fundamental and gateway topic in mathematics (Glasgow et al. 2000; National Mathematics Advisory Panel 2008; Rittle-Johnson et al. 2001). Many of the decimal misconceptions young learners have can persist to adulthood (Putt 1995; Stacey et al. 2001; Widjaja et al. 2011). Isotani et al. (2010) conducted an extensive review of the math education literature, covering 32 published papers and extending as far back as 1928 (e.g., Brueckner 1928; Glasgow et al. 2000; Graeber and Tirosh 1988; Hiebert 1992; Irwin 2001; Resnick et al. 1989; Sackur-Grisvard and Léonard 1985; Stacey et al.

---

[2] From Huang et al (2008) it is unclear whether low prior knowledge students received more instruction than higher prior knowledge students due to producing more errors during instructions. In other words, low prior knowledge students may have had more opportunities to encounter cognitive conflict instruction and thus, for this reason, had more opportunity to benefit from it.

2001) and compiled and analyzed a taxonomy of 17 common and persistent decimal misconceptions.

For instance, a very common decimal misconception is a student thinking that longer decimals are larger (Stacey et al. 2001). This happens when students confuse decimal numbers with whole numbers, which they learn before decimals. With this misconception a student might order decimal numbers from smallest to largest as follows: 0.9, 0.65, 0.731, 0.2347. Another common misconception is "negative thinking" where students think that a decimal between 0 and 1, e.g., 0.2, is actually smaller than 0 (Irwin 2001; Widjaja et al. 2011). This misconception seems to arise from a misunderstanding of the role of the decimal point. Misconceptions such as these two are surprisingly resilient to remediation and cause problems for many adults (Putt 1995; Stacey et al. 2001).

Furthermore, these misconceptions interfere with a conceptual understanding of decimals that leads to difficulty in later tackling mathematical problems involving decimals (Hiebert and Wearne 1985). For example, when asked to add or subtract two decimals, students often do not know how to align the numbers properly, probably due to relying on learned procedures without a solid conceptual understanding of the role of the decimal point.

The study presented in this paper focuses on four of the misconceptions that prior research has shown are most common and contributory to other misconceptions (Stacey 2005; Sackur-Grisvard and Léonard 1985; Resnick et al. 1989). Isotani et al. (2010) gave these misconceptions short and memorable names, as follows: Megz ("longer decimals are larger", e.g., 0.59>0.8), Segz ("shorter decimals are larger", e.g., 0.1>0.68), Negz ("decimals between 0 and 1 are viewed as less than 0"), and Pegz ("the numbers on either side of a decimal are separate and independent numbers", e.g., 12.8+4.5=16.13). The instructional approach of the web-based materials, both erroneous examples and problem solving, is to have every item target at least one of these four misconceptions.

Relationship to AI in Education Research

All of the erroneous examples and problem solving materials used in this study were implemented and rendered interactive using the Cognitive Tutor Authoring Tools (CTAT: Aleven et al. 2009), a well-known intelligent tutoring authoring tool within the Artificial Intelligence in Education (AIED) community. While not all of the technical capabilities of CTAT were used in this project, the fundamental representational construct of CTAT, behavior graphs, was used to model how students can solve the erroneous examples and decimal problems. Behavior graphs are a graphical representation provided by CTAT that model all possible correct solution paths to given problems, as well as typical errors made by students along those solution paths. Decimal misconceptions were modeled and represented as errors within the CTAT behavior graphs.

Some of the more advanced features of CTAT, such as allowing student responses to be provided in varying orders (i.e., unordered behavior graphs) and using variables to reference various elements in the behavior graph, were not used due to the relative simplicity of the decimal problems. On the other hand, erroneous examples necessitated extensions to the

CTAT software, in particular, in developing components to guide the user interface through the specific steps of identifying, explaining, and fixing errors in the erroneous examples, as described in the "Intervention Design" section later in this paper.[3]

The research reported here is related to the search for the right combination of intelligent tutors, examples (correct and incorrect with interactive features), and problem solving for optimal learning. A thread of research within AIED has shown, in general, that alternating interactive examples and intelligently tutored problems can sometimes increase learning benefits and usually reduces learning time (Anthony 2008; McLaren et al. 2008; Salden et al. 2010; Schwonke et al. 2009). All of the examples in these earlier studies, like those of the present study, involved *interactive* examples, for instance providing feedback on the correctness of work, prompting students to self-explain their answer steps, and supporting students in finishing partially completed examples. The examples of older, pure educational psychology studies (e.g., Siegler 2002; Sweller and Cooper 1985; Zhu and Simon 1987) were paper based, static, and, therefore, without interactive features. Thus, another important strand of active AIED research, for which the present study is representative, is exploring the best way to optimize learning by imbuing both correct and erroneous examples with interactive, computer-based features.

Method

Participants and Design

The original set of participants included 463 sixth grade middle-school students from Pittsburgh-area schools. Seventy participants were removed due to having missed either the immediate or the delayed posttest.[4] Two additional participants were removed from the sample due to having negative gain scores 3 standard deviations from the mean between the pretest and immediate posttest. Finally, one student repeated the intervention; thus, their second data set was removed from the analysis. This left a total of 390 participants in the final sample (197 females, 193 males). The students' ages ranged from 10 to 13 ($M=11.57$, $SD=.61$). There was a significant difference between participants who dropped out and those who stayed in the study $F(1456)=23.33$, $p<.001$. However, there was no significant interaction between condition and participants who dropped out $F(1456)=.04$, $p=.85$, therefore, one group did not lose a larger number of higher or lower prior knowledge participants. The study took place at two Pittsburgh-area schools over two school years, with two test runs in the spring of 2012, one at each school, and two in the fall of 2012, again one at each school, but with a different population of students.

---

[3] While CTAT was a useful tool in the development of the study materials, a limited description of the software is provided here, since a deep understanding of CTAT is not essential to understanding the study design.

[4] Virtually all of the deleted students were removed due to illness or otherwise missing class time, conditions outside of the experimenters' control and not necessarily indicative of specific learner characteristics, e.g., weak learners. Furthermore, there was not a significant difference on the pretest between the deleted students assigned to each of the two conditions t (68)$=-.63$, $p=.53$, indicating that the missing students from each condition were not significantly different from one another in terms of prior knowledge.

Materials, Apparatus, and Procedure

The materials, apparatus, and procedure used in this study were identical to our previously published study (Adams et al. 2014). All of the materials, including the three decimal assessment tests, a demographic questionnaire, an evaluation questionnaire, and two different versions of an online lesson on decimals (erroneous examples and problem-solving), were implemented using the aforementioned CTAT authoring tool (Aleven et al. 2009).

Assessment Tests

For the pretest, immediate posttest, and delayed posttest three isomorphic versions of a 46-item decimal assessment test were created (called, henceforth, Tests A, B. and C). The three tests included matched test items (i.e., an equal number of questions, appearing in the same test item position in each test) although the cover stories and values of the test items varied across tests. Each test had a grand total of 50 possible points, due to some test items having multiple components. Every test item was designed to probe for a specific misconception. Test items included a variety of decimal problems:

- Adding decimal numbers together (e.g., 11.90+0.2=_____);
- Ordering decimals according to magnitude (e.g.,. ("Put the following list of decimals in order of size, smallest to largest: 0.899, 0.89, 0.8, 0.8997");
- Answering multiple-choice questions (i.e., "If a decimal number starts with a 0 before the decimal point, would it be less than 0? Yes, No, It Depends, Don't Know");
- Placing decimals on a number line (i.e., "Place 0.6 on a number line between −1 and 1");
- Providing the next decimal number in a sequence (".201, 0.401, 0.601, 0.801, ____); and
- Choosing the largest or smallest decimal from a list (e.g., "Choose the largest of the following three numbers: 0.22, 0.31, 0.9)

In addition to looking at overall accuracy, we were also interested in the students' meta-cognitive awareness of their decimal knowledge. If students become more aware of their misconceptions, they are theoretically better prepared to address and ameliorate those misconceptions. Thus, for 15 of the test items students were asked to rate their confidence on a 5-point Likert scale ranging from "Not at all sure" (1) to "Very sure" (5). The rationale for this data collection was that students with high awareness would be more likely to give high confidence ratings for correct answers and low confidence ratings for incorrect answers. These judgments were collected across the three testing sessions (pretest, posttest, delayed posttest) to examine whether erroneous examples or problem solving would increase the students' awareness of their own misconceptions.

Questionnaires

The demographic questionnaire solicited basic information about age, gender, and grade level. In addition students were asked a series of questions relating to their prior

experience with decimals, experience working with computers, and questions relating to math self-efficacy. Upon completion of the intervention students were given an evaluation questionnaire to rate how they felt about their lesson. The questionnaire included 10 items, which were later combined into 4 categories: "Lesson Enjoyment" (How well students liked the lesson - 2 items): "Ease of Interface Use" (How easy it was for the student to interact with the tutor and its interface - 4 items); "Feelings of Math Efficacy" (Whether the student had positive feelings about mathematics after using these materials - 2 items); and "Perceived Material Difficulty" (Whether the student perceived that the lesson was difficult - 2 items). Responses were given using a 5-point Likert scale ranging from "Strongly agree" (1) to "Strongly disagree" (5).

Intervention Design

The two versions of the lesson, erroneous examples and problem solving, each comprised 36 total items, as illustrated in Table 1.

The two interventions were arranged into 12 groups of three items, with each group targeting one of the four misconception types discussed previously (i.e., Megz, Segz, Pegz, Negz). Within each group of three items there were two intervention-specific items (i.e., two erroneous example items or two problem solving items) with the final item of each group being a supported problem to solve to allow practice of the just exercised problem type. For the first two items in each group – either erroneous example or problem solving items – students were prompted for self-explanation (i.e., they selected possible explanations from a menu) and received correctness feedback on all of their steps. The third item in each group – the problem to solve – prompted students to solve a problem targeted at the specific misconception with feedback provided, but without prompted self-explanation. Figure 1 contains a step-by-step comparison of the items in the two conditions.

Figure 2 illustrates what happens in the actual interfaces students used to tackle each of the steps for erroneous examples. In the sample erroneous example of Fig. 2, a fictional student is asked to order three decimal numbers from smallest to largest and commits the Segz misconception ("shorter decimals are larger") by putting the decimals in order from shortest to longest. To tackle erroneous example items, students first read and reviewed the error made by the fictional student (top left panel). After pressing a "Next" button – something the student does after tackling the subtask in each of the panels of Fig. 2 – students are asked to identify what the fictional student has done wrong from a list of 3 to 4 options, one of which is the misconception exhibited by that student (in this case, the final option "He thinks that a decimal is smaller if it has more digits"). In the left middle panel students are then asked to correct the mistake. This involves, for instance, correcting an incorrect sequence of decimals (as in this case), moving a decimal to the correct position on a number line, or correctly adding two decimals. In the right middle panel participants next explain why their new answer is correct or confirm the correct solution (i.e., the "Confirms Correct Solution" step of Fig. 1). Finally, in the bottom left panel the students are prompted to give advice on how to solve the problem correctly. This is the step where the student is effectively explaining the solution (i.e., "Explains Correct Solution" in Fig. 1). The prompted explanation here, and for most of the erroneous examples and problems to solve, is for an explanation of the procedure used to solve the problem. For every panel that requires

Table 1 This table shows the sequence of materials for the two versions of the lesson, erroneous examples and problem solving

	Erroneous examples (ErrEx)	Problem solving (PS)
Group 1: Longer decimals are larger (Megz)	1. ErrEx (Megz1) 2. ErrEx (Megz2) 3. Practice Problem (Megz1)	1. PS (Megz1) 2. PS (Megz2) 3. Practice Problem (Megz1)
Group 2: Shorter decimals are larger (Segz)	4. ErrEx (Segz1) 5. ErrEx (Segz2) 6. Practice Problem (Segz1)	4. PS (Segz1) 5. PS (Segz2) 6. Practice Problem (Segz1)
Group 3: independent #s left & right of decimal (Segz)	7. ErrEx (Pegz1) 8. ErrEx (Pegz2) 9. Practice Problem (Pegz1)	7. PS (Pegz1) 8. PS (Pegz2) 9. Practice Problem (Pegz1)
Group 4: decimals between 0 and 1 are < 0 (Negz)	10. ErrEx (Negz1) 11. ErrEx (Negz2) 12. Practice Problem (Negz1)	10. PS (Negz1) 11. PS (Negz2) 12. Practice Problem (Negz1)
Group 5: Longer decimals are larger (Megz)	13. ErrEx (Megz3) 14. ErrEx (Megz4) 15. Practice Problem (Megz2)	13. PS (Megz3) 14. PS (Megz4) 15. Practice Problem (Megz2)
Group 6: Shorter decimals are larger (Segz)	16. ErrEx (Segz3) 17. ErrEx (Segz4) 18. Practice Problem (Segz2)	16. PS (Segz3) 17. PS (Segz4) 18. Practice Problem (Segz2)
Group 7: independent #s left & right of decimal (Segz)	19. ErrEx (Pegz3) 20. ErrEx (Pegz4) 21. Practice Problem (Pegz2)	19. PS (Pegz3) 20. PS (Pegz4) 21. Practice Problem (Pegz2)
Group 8: decimals between 0 and 1 are < 0 (Negz)	22. ErrEx (Negz3) 23. ErrEx (Negz4) 24. Practice Problem (Negz2)	22. PS (Negz3) 23. PS (Negz4) 24. Practice Problem (Negz2)
Group 9: Longer decimals are larger (Megz)	25. ErrEx (Megz5) 26. ErrEx (Megz6) 27. Practice Problem (Megz3)	25. PS (Megz5) 26. PS (Megz6) 27. Practice Problem (Megz3)
Group 10: Shorter decimals are larger (Segz)	28. ErrEx (Segz5) 29. ErrEx (Segz6) 30. Practice Problem (Segz3)	28. PS (Segz5) 29. PS (Segz6) 30. Practice Problem (Segz3)
Group 11: independent #s left & right of decimal (Segz)	31. ErrEx (Pegz5) 32. ErrEx (Pegz6) 33. Practice Problem (Pegz3)	31. PS (Pegz5) 32. PS (Pegz6) 33. Practice Problem (Pegz3)
Group 12: decimals between 0 and 1 are < 0 (Negz)	34. ErrEx (Negz5) 35. ErrEx (Negz6) 36. Practice Problem (Negz3)	34. PS (Negz5) 35. PS (Negz6) 36. Practice Problem (Negz3)

Items are organized into 12 groups of three items each, with each group targeting one of the four misconception types. The ErrEx and PS items in the same row are the same problem, but presented differently according to condition. The third item in every group, the practice problem, is the same across conditions

Steps for Erroneous Example (ErrEx) Items
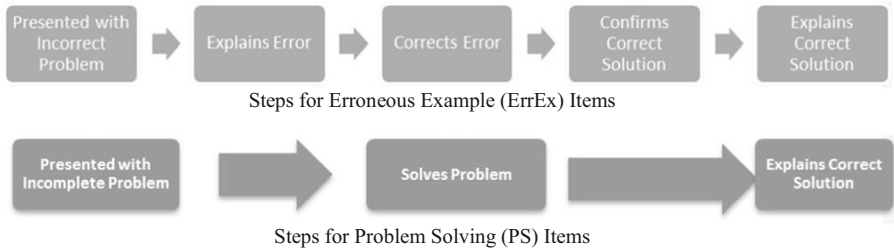
Steps for Problem Solving (PS) Items

Fig. 1 Comparison between the sequences of steps in the two experimental conditions

students to make a selection, feedback is provided, with the answer turning green for correct answers, or red for incorrect answers. Students also receive text feedback from a message window in the bottom right corner of the intervention screen. Messages include encouragement for students to try incorrect steps again (e.g., "Can you try that again? That answer is not correct") or "success" feedback to continue on to the next step or problem after correctly solving a step (e.g., "You've got it. Well done.", as in Fig. 2).

Figure 3 illustrates what happens in the actual interface students use to tackle a problem solving item. Figure 3 shows the isomorphic problem-solving item of Fig. 2. For the problem-solving condition of Fig. 3, the items contain the same numbers and problem requirements (e.g., order the three decimals 1.932, 1.9, 1.63 from smallest to largest) as the corresponding erroneous example items except students are prompted to solve the problem on their own, rather than review the erroneous solution of a fictitious student. The explanation prompts, which are multiple-choice questions, include one

Fig. 2 Example of an Erroneous Example item focused on the Segz misconception ("shorter decimals are larger")

Fig. 3 Example of a Problem Solving item focused on the Segz misconception ("shorter decimals are larger")

correct explanation and misconception distracters. Students in this group also receive feedback from a message window in the bottom right panel as well as green/red feedback on their solution and multiple-choice explanation questions.

Procedure

The study was conducted in each school's computer lab, and replaced the students' regular math class. The grades students received on the tests were used as part of the students' grades in their regular math class. Students worked on either Apple or PC computers, depending on what each school's computer room provided, with full Internet connectivity.

The students were randomly assigned to either the erroneous examples group (188) or the problem-solving group (202).[5] Within each group, students were also randomly assigned to receive one of the six possible pretest/posttest/delayed-posttest orderings (ABC, ACB, BAC, BCA, CAB, CBA). The study took place over five 43-min sessions (the first four sessions on consecutive days), in which students took the pretest and filled out the demographic questionnaire during the first session, received the intervention during the second and third sessions, completed the evaluation questionnaire during the third session, took the immediate posttest during the fourth session, and took the delayed-posttest during the fifth session which took place 1 week after the immediate posttest. The students did not work on decimal-related homework or

---

[5] An adaptive erroneous examples version of the intervention was also piloted during the two Fall 2012 runs. However, not enough data was collected from the adaptive erroneous examples group to draw clear comparisons with the erroneous examples and problem solving conditions.

assignments during the intervening time between the immediate and delayed posttest. In each session, if students finished early, which occurred somewhat frequently since more class time was reserved for the study than was needed by the average student, they received non-decimal math homework to work on. All of the 390 students analyzed and reported in the results completed the 36 items on the intervention.

Results

Are the Groups Equivalent on Prior Knowledge and Basic Demographic Characteristics?

The first row of Table 2 shows the mean (and standard deviation) of the erroneous example group (ErrEx) and problem-solving group (PS) on the pretest. An ANOVA showed there were no significant differences between the ErrEx and PS groups on the pretest, $F(1, 389)=.92$, $p=.34$. While there was a significant difference in pretest performance between the students tested in the spring versus the fall, $F(1388) =16.44$, $p<.001$, a chi-squared analysis looking at condition and testing time showed there were no significant differences between the two conditions in terms of percentage of data collected in the spring versus the fall between the two conditions, $X^2 (1, N=390)=.19$, $p=.66$. Therefore, neither condition was biased in terms of having more students from a particular testing time. In addition there was an equal distribution across the two conditions of participants from the two schools, $X^2 (1, N=390)=.43$, $p=.51$ as well as an equal distribution of male and female participants across the two conditions, $X^2(1, N=390)=.36$, $p=.55$[6]

   Looking at reported experience and self-efficacy with decimals, all of the scores from the demographic survey that dealt with decimals were added together and then averaged to determine familiarity with decimals. There were no significant differences between the groups in terms of self-perceived competence with decimals, $t(388)=.04$, $p=.98$. Due to participants being randomly assigned to a test order for the three different versions of the test (i.e., A, B, and C), ANOVAS were used to examine whether test version significantly affected performance. The analysis showed that there were no significant differences between the three versions of the pretest ($p=.85$), immediate posttest ($p=.50$), or delayed posttest ($p=.12$). Due to the lack of difference all subsequent analyses were collapsed across this factor.

Do the Groups Differ on Learning Outcomes?

Means and standard deviations for the immediate and delayed posttest can be found in the second row of Table 2. Gain scores were calculated by subtracting each student's

---

[6] There were 255 students at School A and 135 students at School B. Based on an ANOVA, students at School A ($M=30.56$, $SD=10.54$) scored significant higher on the pretest than did students from School B ($M=25.73$, $SD=11.88$), $F(1, 386)=17.52$, $MSE=2024.19$ $p<.001$; and there was no significant interaction between school and treatment group, $F(1, 386)=1.52$, $MSE=175.14$, $p=.22$. In addition, students from school A (M=10.60, SD=4.01) rated their competence with decimals significantly higher compared to students from school B (M= 11.71, SD=4.43), $F(1,386)=6.14$, $p=.01$; and there was no significant interaction between school and treatment group for decimal self-efficacy.

Table 2 Mean and Standard Deviation on Pretest, Immediate Test, and Delayed Test for the Two Groups

	Condition	
	Erroneous Examples $N=188$	Problem Solving $N=202$
Pretest	28.35 (10.64)	29.39 (11.31)
Immediate Posttest	33.61 (10.67)	33.20 (11.04)
Delayed Posttest	35.70 (10.13)	34.46 (10.78)
Pretest-Immediate Posttest Gain Score	5.26 (7.08)	3.81 (6.34)
Pretest-Delayed Posttest Gain Score	7.35 (7.07)	5.07 (6.56)

pretest total scores from the immediate and delayed posttest scores. Looking at gain scores between the pretest and immediate posttest, an ANCOVA with pretest score as a covariate, revealed that there was a marginally significant effect with ErrEx showing higher gains between the pretest and immediate posttest compared to the PS condition, $F(1387)=3.72$, $MSE=150.03$, $p=.055$, $d=.22$ For the gains scores between the pretest and delayed posttest, an ANCOVA with pretest score as a covariate showed that students in the ErrEx group had significantly higher gains than students in the PS condition, $F(1387)=10.15$, $MSE=402.09$, $p=.002$, $d=.33$. The superior performance of the ErrEx group on the delayed test is the major empirical finding of this study[7].

Are there Group Differences in Learning Outcome Greater for Students with Low or High Prior Knowledge?

An additional analysis was conducted to determine whether the intervention had differential effects for students with low versus high prior knowledge. First, we classified students based on a median split on pretest score, with 200 students classified as low prior knowledge (i.e., pretest score from 7 to 28 points) and 190 students classified as high prior knowledge (i.e., pretest score from 29 to 49 points). In general, low prior knowledge participants had significantly higher gains compared to the high prior knowledge students between the pretest and the immediate posttest, $F(1386)=33.59$, $MSE=1396.40$, $p<.001$, $d=.59$, and between the pretest and delayed posttest, $F(1386)=54.17$, $MSE=2211.29$, $p<.001$, $d=.74$. However, there was no significant interaction between condition and prior knowledge level for gains between either the pretest to the immediate posttest ($F(1386)=.36$, $MSE=145.69$, $p=.55$) or pretest to the delayed posttest $F(1386)=.67$, $MSE=27.44$, $p=.41$). This suggests that both of the

---

[7] Looking at the results from school A, there was no significant difference in gains scores between the pretest and immediate posttest between PS group ($M=3.79$, $SD=6.53$) and the ErrEx group ($M=5.67$, $SD=7.52$) $F(1252)=2.44$, $p=.12$, $d=.27$. Between the pretest and the delayed posttest, participants in the ErrEx group ($M=7.75$, $SD=7.35$) had significantly higher gain scores than the PS group ($M=5.09$, $SD=6.79$), $F(1252)=6.26$, $p=.01$, $d=.38$. When looking at the results for school B, there was no significant difference between the PS group ($M=3.84$, $SD=6.02$) and the ErrEx group ($M=4.42$, $SD=6.05$) for gains between the pretest and the immediate posttest, $F(1133)=.31$, $p=.58$, $d=.10$. In contrast to School A, there was no significant difference between the PS group (M=5.03, SD=6.16) and the ErrEx group ($M=6.53$, $SD=6.44$) of School B for gains between the pretest and the delayed posttest, $F(1, 133)=1.92$, $p=.17$, $d=.2$.

interventions were beneficial for low prior knowledge students, with no significant difference between the interventions.

High prior knowledge students had, of course, less room for growth due to having higher scores on the pretest. Separate analyses were conducted on both the low and high prior knowledge participants to determine whether the benefit for erroneous examples on the delayed posttest was significant for both groups. For low prior knowledge individuals an ANCOVA, with pretest as a covariate, was conducted looking at gains between the pretest and immediate posttest and pretest and delayed posttest. Low prior knowledge participants in the ErrEx and PS conditions did not show significant differences in gains between the pretest and immediate posttest, $F(1197)=2.47$, $MSE=150.84$, $p=.12$, $d=.23$.; however, the ErrEx condition had significantly higher gains between the pretest and the delayed posttest, $F(1197)=6.06$, $MSE=367.21$, $p=.02$, $d=.35$. High prior knowledge individuals showed the same pattern with no significant difference for gains between the pretest and the immediate posttest, $F(1187)=1.00$, $MSE=18.39.59$, $p=.32$, $d=.21$, and ErrEx participants having significantly higher gains compared to the PS student between the pretest and the delayed posttest, $F(1187)=4.28$, $MSE=70.60$, $p=.04$, $d=.37$. Therefore although high prior knowledge students had lower gains overall, the higher prior knowledge students in the ErrEx condition still had larger gains than the higher prior knowledge students in the PS condition between the pretest and delayed posttest.

Along with separating participants into high and low prior knowledge groups, performance on the pretest was also used as a continuous variable in a stepwise regression analysis to determine if there was any significant interaction between the intervention condition and the student's prior knowledge level on immediate and delayed posttest performance. Step 1 for both analyses examined the effects of the pretest as well as condition on test performance, while Step 2 examined whether the interaction between the two variables could account for any additional variance in test performance. For Step 1, prior knowledge and condition accounted for a 65.9 % of the variance for immediate posttest performance, $F(2387)=373.93$, $p<.001$. Performance on the pretest had a significant effect on the immediate posttest, as reveal by the standardized partial regression coefficients, $\beta=.81$, $t=27.34$, $p<.001$, however, condition had only a marginally significant effect on the immediate posttest, $\beta=.06$, $t=1.93.88$, $p=.055$. The coefficient for the interaction term entered at Step 2 showed no significant interaction between pretest performance and condition on immediate posttest performance, $\beta=-.04$, $t=-.65$, $p=.52$. On the delayed posttest, pretest performance and condition account for 64.1 % of the variance in test performance, $F(2387)=345.51$, $p<.001$. Both pretest, $\beta=.80$, $t=26.22$, $p<.001$, and condition, $\beta=.10$, $t=3.19$, $p=.002$, significantly affected performance on the delayed posttest performance, mirroring earlier analyses. There was no significant interaction between condition and pretest performance on delayed posttest performance as indicated by the interaction coefficient on Step 2, $\beta=-.04$, $t=-0.92$, $p=.36$.

Combined with the median split analysis, these analyses suggest that erroneous examples were not more or less effective for students with high or low prior knowledge.

Do the Groups Differ on Their Awareness of Misconceptions?

An additional goal of the erroneous example treatment was to improve students' metacognitive skills, particularly their awareness of their own decimal knowledge

and misconceptions. To explore this question, the strength of students' misconception awareness was calculated through self-assessed confidence in correctness of test responses. It should be noted that confidence ratings are only a rough metric that do not fully capture the students' awareness of misconceptions. For instance, a student being aware of having made a computational error is not the same as being aware of a misconception. On the other hand, awareness of many other errors would arguably be the same as awareness of misconceptions.

One of the items was dropped from the analysis across the 3 tests due to a data logging issue. This left a total of 17 test items per test that the students were asked to give a confidence rating on after answering the question. Due to an error in logging some of the confidence data, six participants were removed from the confidence calibration analysis. To examine how confident the students were of their answers on the pretest, immediate posttest, and delayed posttest the mean confidence level for each student was calculated using the data from the 5-point Likert confidence scales. A repeated measures ANOVA was conducted with testing session as a within subjects factor and condition as a between subjects factor. There was no significant main effect for condition, $F(1, 381)=.10$, $MSE=.14$, $p=.76$. There was a significant main effect of testing session, $F(2, 762)=75.04$, $MSE=7.89$, $p<.001$. Post-hoc Bonferroni pairwise comparison between the testing sessions showed the participants significantly increased in confidence across the three sessions with an overall average increase in confidence of .28 points ($SE=.03$) on a five point scale. There was no significant interaction between test and condition, $F(2, 762)=1.14$, $MSE=.12$, $p=.32$, therefore there was no significant difference in terms of increase in confidence across the three tests between the ErrEx and PS conditions.

Students' responses were then categorized by confidence level and accuracy, which led to four response categories: high confidence error, low confidence error, high confidence correct, and low confidence correct. Students' responses were categorized as being low confidence if they were a 1 or 2 on the 5-point scale and high confidence if they were a 3, 4, or 5 on the 5-point scale. There were no significant differences between conditions for any of the responses on the pretest. For each of the four response types categories an ANCOVA was conducted, with pretest rate of the respective response type as a covariate, to examine whether there were significant differences between the two conditions for any of the response types on the immediate or delayed posttest. There were no significant differences in response type percentage on the immediate posttest for any of the response types. For the delayed posttest, the only significant difference was for high confidence correct answers, $F(1, 380)=5.07$, $MSE=.15$, $p=.03$. Students in the ErrEx condition were more likely to make high confidence correct responses ($M=66.27$ %, $SD=24.20$ %) than students in the PS condition ($M=63.45$ %, $SD=26.33$ %). While it appears that erroneous examples did not raise students' awareness of their misconceptions, as we hypothesized, the finding that students in the ErrEx condition were more likely to make high confidence correct responses on the delayed posttest indicates that erroneous examples helped strengthen students' meta-cognitive awareness of their decimal knowledge somewhat more than problem solving.

Do the Groups Differ on Their Satisfaction with the Online Lesson?

For the evaluation survey, four categories, each of which entailed multiple questions as described previously, were created to assess different aspects of the lesson: "Lesson

Enjoyment", "Ease of Interface Use", "Feelings of Math Efficacy" and "Perceived Material Difficulty". The PS condition students were significantly more likely to report that they liked the lesson compared to the ErrEx students $F(1, 388)=4.29$, $MSE= 23.49$, $p=.04$, $d =. 21$. Although there were no significant differences between the conditions in terms of perceived lesson difficulty, $F(1, 388)=1.69$, $MSE= 4.96$, $p=.19$, $d=−.13$, participants in the PS condition found it significantly easier to interact with the tutor interface, $F(1, 388)=12.94$, $MSE= 124.97$, $p<.001$, $d=.37$ There were no significant differences between the two conditions in terms of reporting that the lesson led to more positive feelings about math, $F(1, 388)=2.08$, $MSE= 9.66$, $p=.15$, $d=.15$. The higher satisfaction ratings of the PS group on two key measures is another major finding of this study[8]

Do the Groups Differ on Time on Task?

We also wanted to see how much time students in the two groups spent doing the lesson. The erroneous examples students may have performed better on the delayed posttest, but did the extra steps and additional time in the instructional phase contribute to this benefit? On average, students in the ErrEx condition took 71.43 ($SD =21.98$) minutes to complete the lesson while students in the PS condition took 51.09 ($SD= 20.40$) minutes. An independent samples t-test revealed this difference to be significant; participants in the ErrEx condition took significantly longer to complete the lesson, $t(388)=9.48$, $p<.001$. In addition to t-tests, regression analyses for gains between pretest and the immediate and delayed posttest were run with condition and time-on-task entered at Step 1 and the interaction term entered at Step 2. Although there was a non-significant effect of time-on-task on pretest-to-delayed-posttest gains, $β=.10$, $t= 1.88$, $p=.06$, there was no significant interaction between duration and condition on delayed posttest performance, $β=.02$, $t= .32$, $p=.75$. There were no significant effects or interactions with duration for pretest to immediate posttest gains. Overall, there is no evidence that time on task contributed more to one group than the other.

Discussion

Empirical Findings

Overall, students liked the lesson significantly better when they only engaged in traditional problem solving ($d=.21$ for liking rating) and the problem solving students found the user interface easier to interact with ($d=.37$), yet students who learned with erroneous examples showed higher learning gains as measured on a delayed posttest

---

[8] Looking at differences between the two schools, there was only a significant difference for perceived material difficulty, in which students from school B ($M=5.24$, $SD=1.68$) reported finding the instructional materials more difficult than students from school B ($M=5.75$, $SD=1.71$), $F(1, 386)=8.04$, $MSE=23.16$, $p=.01$, $d=.29$ There was only one marginally significant interaction between school and treatment group, concerning the question about making students feel good about math, $F(1, 386)=3.83$, $MSE=17.69$, $p=.05$, reflecting a pattern in which participants in the ErrEx group were less likely to report that the intervention made them feel good about math at school B while there were no significant differences between the two groups at school A.

(d=.33). In other words, students liked the lesson better when they could engage in problem solving, but they learned better when they were asked to tackle and learn with erroneous examples, consistent with the admonishment that "liking is not learning". This point was further supported by there being no significant correlations between students liking ratings and pre-to-post learning gains, r (390)=−.05, p=.29, or liking ratings and pre-to-delayed learning gains, r (390)=−.01, p=.80. In addition, a hierarchical regression analysis showed that there was no significant interaction between liking and the two conditions in terms of increasing learning gains either for the immediate, β=.08, t=1.08, p=.28, or delayed posttest, β=.04, t=.50, p=.62.

The results of this study replicate the pattern of findings in a previous study in which the erroneous examples group outperformed the problem-solving group on a delayed posttest but not an immediate posttest (Adams et al. 2014). In other words, these new results add support to the emergent finding that erroneous examples lead to a delayed, but not immediate, learning effect. This pattern of significant differences on delayed tests rather than immediate tests is also consistent with research on other generative learning activities such as self-testing (Dunlowsky et al. 2013; Fiorella and Mayer 2015).

Theoretical Implications

Asking learners to identify and self-explain errors in someone else's worked-out solutions to mathematics problems can prime deeper cognitive processing during learning than simply asking a learner to solve the problems on his or her own. This is the theoretical rationale for presenting erroneous examples. In addition, asking students to analyze erroneous examples, with feedback, is intended to help learners develop metacognitive skills, particularly, monitoring and evaluating steps in a problem-solving plan that can persist over time.

A possible explanation for the longer-term retention of erroneous examples is that erroneous example study, which involves elements of both example study *and* problem solving (i.e., fixing the erroneous solutions and solving practice problems), may provide and strengthen "don't do X" knowledge and/or more general declarative/conceptual knowledge, in addition to supporting procedural knowledge. Put another way, the erroneous example students may be developing multiple cognitive paths such that "don't do X" (or conceptual knowledge) compensates for weakness in "do X" procedural knowledge. This explanation is in line with Bob Siegler's theory (Siegler 2002) in which students saw and explained both correct and incorrect examples and that group performed better than the one that saw and explained correct examples only. In essence, he theorized that the erroneous example / worked example treatment strengthened both the "do X" and "don't do X" knowledge of students.

Learning from erroneous examples can be seen as similar to a *desirable difficulty* (Yue et al. 2013), in which making a learning task more difficult can result in deeper and longer-lasting learning than making the learning task very straightforward. A possible explanation for how erroneous examples are similar to desirable difficulties comes from cognitive load theory (Moreno and Park 2010). In order to update long-term memory and make it flexibly accessible, students must be prompted to engage in deeper processing (also called generative or germane processing) of the instructional material. Traditional instructional approaches, such as presenting students with consecutive problems on the same topic, may ease working memory and intrinsic processing,

but may not promote the generative/germane processing that leads to long-term memory benefits like erroneous examples do.

Practical Implications

Although the present results suggest the potential of erroneous examples to aid learning, an important practical issue concerns the proper balance of direct instruction, problem solving and erroneous examples. In the present study, students in the erroneous examples group received a combination of erroneous example and problem solving items.

    Another important practical issue concerns the role of feedback in erroneous examples, because without feedback, students run the risk of learning the incorrect way to solve problems. In the present study, students could not move forward until they had corrected errors and produced a correct solution strategy.

    We expected that, like the Groβe and Renkl study (2007), higher prior knowledge students would benefit more from erroneous examples than lower prior knowledge students in this study. However, we did not find a difference between high and low prior knowledge students, indicating that students of any level could benefit from erroneous examples. Perhaps our materials, unlike those of the Groβe and Renkl study, were designed so that even lower prior knowledge students could easily follow, interact with, and learn from the examples without incurring excessive cognitive load. The Groβe and Renkl work was also different in that it focused on errors related to confusing problem types instead of deeply entrenched misconceptions, which is what our study focused on. In other words, erroneous examples may be more helpful for students with low prior knowledge when they involve common misconceptions.

Limitations

This was a study conducted over five class periods that focused on just a single topic within the U.S. middle-school mathematics curriculum. In addition, many of our decimal problems are single-step problems, unlike the more complex, multi-step problems in studies like that of Große and Renkl (2007). More research is clearly needed to determine whether and how erroneous examples can make a difference to learning across the mathematics curriculum and in topics of varying difficulty and complexity.

    Another possible limitation is that students were prompted to give procedural, rather than conceptual, explanations to the incorrect and correct solutions. One might expect that conceptual explanations would help students more effectively overcome their misconceptions and lead to deeper learning. Conceptual explanations of decimal content and problems, expressed succinctly and simply enough for middle school students to understand, were exceedingly difficult to write, so we used procedural explanations. Yet, interestingly, even with procedural explanations, students in the erroneous examples condition learned more deeply than those in the problem solving condition. Left for future research is experimenting with the effect of conceptual self-explanations.

    Finally, it could be argued that the two comparison groups, erroneous examples and problem solving, differ on more than a single variable. The erroneous examples group was prompted to self explain both the error that was observed and the correct way to solve the problem. In the problem-solving group, on the other hand, students were

prompted to self explain only the correct solution. It goes to the different nature of these instructional material types that they differ on this aspect, yet the fact is the erroneous examples condition received more self-explanation prompting than the problem-solving condition. It is possible that that difference in the design contributed to the delayed effect found in this study.

Conclusion

This paper has presented a study that provides evidence that erroneous examples may lead to deeper and longer-lasting learning as compared to supported problem solving. The study described here is a replication of an earlier study (Adams et al 2014), and the results are in line with that study. Furthermore, the study provides strong support for the notion that "liking is not learning", since students in the erroneous examples group liked the materials less and found the user interface harder to work with than the problem solving group, yet they learned the material more deeply.

References

Adams, D., McLaren, B. M., Durkin, K., Mayer, R.E., Rittle-Johnson, B., Isotani, S., & Van Velsen, M. (2014). Using erroneous examples to improve mathematics learning with a web-based tutoring system. *Computers in Human Behavior, 36C (2014),* 401–411. Elsevier. doi: 10.1016/j.chb.2014.03.053.

Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: example-tracing tutors. *International Journal of Artificial Intelligence in Education, 19*(2), 105–154.

Aleven, V., Myers, E., Easterday, M., & Ogan, A. (2010). Toward a framework for the analysis and design of educational games. In: *Proceedings of the 2010 I.E. International Conference on Digital Game and Intelligent Toy Enhanced Learning.* (pp. 69–76). doi:10.1109/DIGITEL.2010.55.

Anthony, L. (2008). *Developing handwriting-based Intelligent Tutors to enhance mathematics learning.* Unpublished doctoral dissertation, Carnegie Mellon University, USA.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: combining fading with prompting fosters learning. *Journal of Educational Psychology, 95,* 774–783.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: differentiating between correct and incorrect examples. *Learning and Instruction, 25,* 24–34.

Borasi, R. (1996). *Reconceiving mathematics instruction: A focus on errors*. Ablex Publishing Corporation.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington: National Academy Press.

Brueckner, L. J. (1928). Analysis of difficulties in decimals. *Elementary School Journal, 29,* 32–41.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General, 127*(4), 355–376.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah: Lawrence Erlbaum Associates, Inc.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, R., & Glaser, R. (1989). Self explanations: how students study and used examples in learning to solve problems. *Cognitive Science, 13*, 145–182.

Chi, M. T. H., DeLeeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 25*(4), 471–533.

College Board (2015). Identifying sentence errors. From the College Board PSAT/NMSQT website: https://www.collegeboard.org/psat-nmsqt/preparation/writing-skills/sentence-errors.

Dunlowsky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58.

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*, 206–214.

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that improve understanding*. New York: Cambridge University Press.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy* (1st ed.). New York: Palgrave Macmillan.

Glasgow, R., Ragan, G., Fields, W. M., Reys, R., & Wasman, D. (2000). The decimal dilemma. *Teaching Children Mathematics, 7*(2), 89–93.

Graeber, A., & Tirosh, D. (1988). Multiplication and division involving decimals: preservice elementary teachers' performance and beliefs. *Journal of Mathematics Behavior, 7*, 263–280.

Groβe, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: can this foster learning outcomes? *Learning and Instruction, 17*(6), 612–634.

Gunderman, R. B., & Burdick, E. J. (2007). Error and opportunity. *American Journal of Roentgenology, 188*(4), 901–903.

Guthrie, E. R. (1952). *The psychology of learning*. New York: Harper & Brothers.

Hausmann, R. G. M., & Chi, M. T. H. (2002). Can a computer interface support self-explanation? *International Journal of Cognitive Technology, 7*, 4–14.

Hiebert, J. (1992). Mathematical, cognitive, and instructional analyses of decimal fractions. Chapter 5 in *Analysis of arithmetic for mathematics teaching*, pp 283–322. Lawrence Erlbaum.

Hiebert, J., & Wearne, D. (1985). A model of students' decimal computation procedures. *Cognition and Instruction, 2*, 175–205.

Huang, T.-H., Liu, Y.-C., & Shiu, C.-Y. (2008). Construction of an online learning system for decimal numbers through the use of cognitive conflict strategy. *Computers & Education, 50*, 61–76.

Hull, C. L. (1952). *A behavior system: An introduction to behavior theory concerning the individual organism*. New Haven: Yale University Press.

Irwin, K. C. (2001). Using everyday knowledge of decimals to enhance understanding. *Journal for Research in Mathematics Education, 32*(4), 399–420.

Isotani, S., McLaren, B. M., & Altman, M. (2010). Towards intelligent tutoring with erroneous examples: A taxonomy of decimal misconceptions. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Proceedings of the 10th International Conference on Intelligent Tutoring Systems (ITS-10), Lecture Notes in Computer Science, 6094* (pp. 346–348). Berlin: Springer.

Johnson, C. I., & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. *Computers in Human Behavior, 26*, 1246–1252.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579–588.

King, A. (1994). Guiding knowledge construction in the classroom: effects of teaching children how to question and how to explain. *American Educational Research Journal, 31*(2), 338–368.

Lomas, J. D., Patel, K., Forlizzi, J., & Koedinger, K. (2013). *Optimizing challenge in an educational game using large-scale design experiments. Proceedings of CHI2013*. New York: ACM Press.

Mayer, R. E., & Johnson, C. I. (2010). Adding instructional features that promote learning in a game-like environment. *Journal of Educational Computing Research, 42*, 241–265.

McLaren, B. M., Lim, S., & Koedinger, K. R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2176–2181). Austin: Cognitive Science Society.

Moreno, R., & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory*. Cambridge: Cambridge University Press.

National Health Care, U.K. (2013), Intrathecal injection error video: https://www.youtube.com/watch?v=cipFuDxiF2Y.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington: U.S. Department of Education.

Paas, F., & van Merriënboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122–133.

Putt, I. J. (1995). Preservice teachers ordering of decimal numbers: when more is smaller and less is larger! *Focus on Learning Problems in Mathematics, 17*(3), 1–15.

Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self explanation. *Learning and Instruction, 12*, 529–556.

Renkl, A. (2014). The worked examples principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 391–412). New York: Cambridge University Press.

Renkl, A., & Atkinson, R. K. (2010). Learning from worked-out examples and problem solving. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory*. Cambridge: Cambridge University Press.

Resnick, L. B., Nesher, P., Leonard, F., Magone, M., Omanson, S., & Peled, I. (1989). Conceptual bases of arithmetic errors: the case of decimal fractions. *Journal for Research in Mathematics Education, 20*(1), 8–27.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: an iterative process. *Journal of Educational Psychology, 93*, 346–362.

Sackur-Grisvard, C., & Léonard, F. (1985). Intermediate cognitive organizations in the process of learning a mathematical concept: the order of positive decimal numbers. *Cognition and Instruction, 2*, 157–174.

Salden, R. J. C. M., Aleven, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science, 38*, 289–307.

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: not an artefact of lousy control conditions. *Computers in Human Behavior, 25*(2009), 258–266.

Shoebottom, P. (2015). Error correction. From the Frankfurt International School website: http://esl.fis.edu/grammar/correctText/.

Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Granott & J. Parziale (Eds.), *Microdevelopment, Transition Processes in Development and Learning* (pp. 31–58). Cambridge: Cambridge University Press.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. New York: Appleton-Century.

Stacey, K. (2005). Travelling the road to expertise: A longitudinal study of learning. In. H. Chick & J. Vincent (Eds.), *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (vol 1, pp.19–36). University of Melbourne: PME.

Stacey, K., Helme, S., & Steinle, V. (2001). Confusions between decimals, fractions and negative numbers: A consequence of the mirror as a conceptual metaphor in three different ways. In M. v. d. Heuvel-Panhuizen (Ed.), *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 217–224). Utrecht: PME.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*, 59–89.

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

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Swigger, K. M., & Wallace, L. F. (1988). A discussion of past programming errors and their effect on learning Assembly language. *The Journal of Systems and Software, 8*, 395–399.

The Doctor's Company (2013). Video on learning from errors. https://www.youtube.com/watch?v=-ol5jM7YHH0.

Tsamir, P. & Tirosh, D. (2003). In-service mathematics teachers' views of errors in the classroom. In International Symposium: Elementary Mathematics Teaching, Prague.

Tsovaltzi, D., Melis, E., & McLaren, B. M. (2012). Erroneous examples: Effects on learning fractions in a web-based setting. *International Journal of Technology Enhanced Learning (IJTEL)*. V4 N3/4 2012 pp 191–230.

WHO (2014). World Health Organization (WHO): "Learning from Errors to Prevent Harm" workshop. http://www.who.int/patientsafety/education/curriculum/PSP_mpc_topic-05.pdf.

Widjaja, W., Stacey, K., & Steinle, V. (2011). Locating Negative Decimals on the Number Line: Insights into the Thinking of Pre-service Primary Teachers. *Journal of Mathematical Behavior*. 30, 80–91. http://dx.doi.org/10.1016/j.jmathb.2010.11.004.

Wylie, R., & Chi, M. T. H. (2014). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 413–432). New York: Cambridge University Press.

Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: an undesired desirable difficulty. *Journal of Educational Psychology, 105*, 266–277.

Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction, 4*(3), 137–66.