This is a repository copy of *Refuting misconceptions: computer tutors for fraction arithmetic*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/206221/

Version: Published Version

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Empirical Research

Refuting Misconceptions: Computer Tutors for Fraction Arithmetic

Sabrina M. Di Lonardo Burr*[a], Heather Douglas[a], Maria Vorobeva[a], Kasia Muldner[a]

[a] Department of Cognitive Science, Carleton University, Ottawa, ON, Canada.

Abstract

Fractions, known to be difficult for both children and adults, are especially prone to misconceptions and erroneous strategy selection. The present study investigated whether a computer tutor improves fraction arithmetic performance in adults and if supplementing problem solving with erroneous examples is more beneficial than problem solving alone. Seventy-five undergraduates solved fraction arithmetic problems using a computer tutoring system we designed. In a between-subjects design, 39 participants worked with a problem-solving tutor that was supplemented with erroneous examples and 36 participants worked with a traditional problem-solving tutor. Both tutors provided hints and feedback. Overall, participants improved after the tutoring interventions, but there were no significant differences in gains made by the two conditions. For students with low prior knowledge about fraction arithmetic, the numerical gains were higher in the erroneous-example group than the problem-solving group, but this effect was not significant. Thus, computer tutors are useful tools for improving fraction knowledge. While erroneous examples may be particularly beneficial for students with low prior knowledge who may hold more misconceptions, more research is needed to make this conclusion.

Keywords: fractions, fraction misconceptions, educational technology, example-tracing tutor

Fraction skills are integral in science-based professions as well as less math-intensive jobs such as home care (calculating dosages) and retail sales (calculating discounts). Sixty-eight percent of adults report using fractions in their workplace (Handel, 2016). Fractions are also used during everyday activities like cooking (e.g., doubling or halving recipes that involve fractional amounts such as ¾ cups of flour) and helping children with schoolwork. Importantly, fraction skills are needed to critically evaluate statistics and probabilities in the media. For example, fraction knowledge is needed to recognize that 3 out of 4 dentists recommending a toothpaste is a stronger endorsement than 7 out of 10 dentists. Because fractions are used at work and at home, fraction competency is important for adults.

Although adults have been exposed to fractions at school, they continue to struggle with this challenging domain (DeWolf et al., 2014; Schneider & Siegler, 2010; Tan, 2020). Notably, fraction arithmetic is an area of difficulty. For example, in studies examining fraction knowledge amongst college students, students answered over 20% of the fraction arithmetic questions incorrectly (Siegler & Lortie-Forgues, 2015; Stigler et al., 2010; Tan, 2020). Thus, one goal of this work was to develop a computer tutor that provided guided practice for fraction problems, with the aim of improving fraction arithmetic skills. Previous studies have found that computer

tutors are an effective approach for delivering mathematics instruction in both middle- and high-school curricula (e.g., Ritter, Anderson, et al., 2007; Corbett et al., 2000). Koedinger et al. (2000) report that compared to traditional mathematics courses, students in courses using the Cognitive Tutor (Algebra I, Algebra II, and Algebra III) performed better on assessments of complex mathematical problem solving, standardized assessments of basic mathematical skills, and were more likely to pass their university mathematics courses. At the university level, computer tutors are beneficial for various mathematics concepts, including algebra (Ritter, Kulikowich, et al., 2007; Stewart et al., 2005) and calculus (Ayub et al., 2010).

The present study investigates whether a fraction computer tutor improves fraction performance. A second goal was to compare two types of instructional activities in terms of their impact on fraction performance. A traditional activity corresponds to problem solving, where students are given problems to work on and guidance for that process. However, problem solving alone induces a high cognitive load, which interferes with learning (Cooper & Sweller, 1987; Sweller & Cooper, 1985). One way to address this is by supplementing problem solving with worked examples (e.g., Atkinson et al., 2000; Chi et al., 1989; Große & Renkl, 2007; Sweller & Cooper, 1985). A standard worked example consists of a problem statement and a step-by-step correct derivation of the problem solution (Atkinson et al., 2000). Because the solution illustrated in the example is correct, students can learn how the solution is derived, which will help them solve a similar problem (Cooper & Sweller, 1987; Sweller & Cooper, 1985). A second less studied type of example is called an *erroneous example*. Erroneous examples include incorrect steps (i.e., errors) in the solutions, which students are tasked with finding and correcting (Adams et al., 2014).

A proposed benefit of erroneous examples relates to misconception refutation. Students at the beginning and intermediate stages of cognitive skill acquisition have various misconceptions because their knowledge is not yet complete or refined (VanLehn, 1999). Erroneous examples make the misconceptions explicit because they are illustrated in their solutions – since students are asked to correct the error(s), their own misconceptions are refuted in the process. In general, erroneous examples illustrate what *not* to do, providing an opportunity to extend knowledge. Indeed, some studies show that erroneous examples improve understanding and problem solving, while also promoting critical thinking (Borasi, 1996), reflection on misconceptions (Bransford et al., 2000), and constructive behaviours like the generation of explanations for why a solution is incorrect (Durkin & Rittle-Johnson, 2012). However, some educators have expressed concerns that including errors in examples will make students more likely to make errors themselves (Tsamir & Tirosh, 2005). Furthermore, learning through erroneous examples may place high demands on working memory (Große & Renkl, 2007). Working memory is loaded because students must retrieve and hold the correct solution steps in working memory, identify the incorrect step, provide an appropriate explanation for why that step is incorrect, and correct the incorrect step. Thus, other studies have found no evidence that erroneous examples improve learning (Chen et al., 2016; Huang, 2017; McLaren et al., 2016) or that erroneous examples are only beneficial for students with particular levels of prior knowledge (e.g., Große & Renkl, 2007; Heemsoth & Heinze, 2014; Huang et al., 2008).

In sum, the goals of the present study were to investigate if a computer tutor for fraction arithmetic improved adult fraction performance and to analyze whether supplementing problem solving with erroneous examples improved learning over problem solving alone for students with low and high prior knowledge. Erroneous examples and traditional problems were presented by a computer tutor we designed and implemented. The tutor provided immediate feedback for correctness on solution entries and interactively prompted users to

self-explain and correct example errors. We describe this tutor after describing the target domain, namely fraction arithmetic.

The Target Domain: Fraction Arithmetic

Fraction arithmetic errors are well documented. In the current study we focused on two sources of arithmetic errors: Errors attributed to *natural number bias* (a common fraction misconception; DeWolf & Vosniadou, 2015; Ni & Zhou, 2005; Stafylidou & Vosniadou, 2004; Van Dooren et al., 2015) and errors based on gaps in prior knowledge (Braithwaite et al., 2017; Braithwaite & Siegler, 2018b).

Errors Reflecting Natural Number Bias

One commonly observed source of errors in fraction arithmetic is *natural number bias,* which involves incorrectly applying natural number reasoning to solve fractions problems (DeWolf & Vosniadou, 2015; Meert et al., 2009; Ni & Zhou, 2005). For example, when adding or subtracting fraction quantities, students may view the numerator and denominator independently and apply natural number reasoning to solve the problem. As a result, students may add the numerators and denominators together and thus think $\frac{2}{3} + \frac{3}{4}$ could equal $\frac{5}{7}$ or even 12 (Braithwaite et al., 2018; Bruce et al., 2013). Another example of misapplying natural number reasoning involves multiplying fraction quantities. Students may incorrectly assume that the product will always be greater than the factors (Vamvakoussi & Vosniadou, 2004). Thus, when multiplying $\frac{2}{3} \times \frac{4}{5}$ students may consider the solution $\frac{8}{15}$ incorrect because it is smaller than either factor. Evidence indicates that although natural number bias diminishes as fraction skills develop (Braithwaite & Siegler, 2018b; Rinne et al., 2017), many adults continue to incorrectly apply natural number reasoning when comparing fraction quantities and when solving fraction procedures (DeWolf & Vosniadou, 2015; Obersteiner et al., 2013; Vamvakoussi et al., 2012).

Two theoretical views have been proposed to account for natural number bias. The *conceptual change* view posits that biased responses are related to weak conceptual knowledge thus improvement in fraction arithmetic requires improving conceptual knowledge (DeWolf & Vosniadou, 2015; Ni & Zhou, 2005). The second view, *dual-processing,* posits that natural number reasoning evokes an intuitive response to a fraction problem (Vamvakoussi et al., 2012). Since intuitive responses are often incorrect, generating correct solutions requires inhibition of intuitive natural number reasoning (for a discussion see Obersteiner et al., 2019). Consequently, reflection on fraction concepts can help inhibit the intuitive response (Obersteiner et al., 2013; Vamvakoussi et al., 2012). Accordingly, adults will benefit from reflecting on their responses. Importantly, a fraction tutor can provide opportunities to learn (strengthen existing knowledge and/or form new knowledge), review, and reflect. Thus, to reduce natural number bias we asked students to refute errors that corresponded to mistakes produced by misleading natural number reasoning (see Table 1, Cases 1 and 3).

Errors Reflecting Prior Fraction Knowledge

In addition to errors resulting from natural number bias, fraction errors may be related to gaps in prior knowledge. These gaps are the result of poorly encoded fraction concepts in the first place and/or due to forgetting previously taught content. The degree of forgetting is related to the distribution and amount of practice, with procedures that are not practiced frequently being more likely to be forgotten (Bahrick & Hall, 1991). To estimate how much practice students experienced, researchers have analyzed math textbooks. In American math textbooks (Grades 3 to 8), there are fewer division questions compared to other operations. Further, addition and subtraction of fractions include a similar number of questions involving equal and unequal

denominators whereas multiplication and division of fractions disproportionality involve questions with unequal denominators (Braithwaite et al., 2017; Braithwaite & Siegler, 2018a). Braithwaite et al. (2017) developed a computational model of fraction arithmetic and tested it with a distribution of fraction problems that mirrored the problem distributions in textbooks. Errors were categorized as execution- or strategy-based. The highest error rate (amongst operations) was observed with division and characterized by an execution error where the first fraction (instead of the second) was inverted and the two fractions multiplied (e.g., $\frac{2}{3} \div \frac{4}{5} = \frac{3}{2} \times \frac{4}{5}$). Common strategy errors included applying addition strategies for multiplication and division problems with equal denominators (e.g. $\frac{1}{4} \times \frac{3}{4} = \frac{3}{4}$; $\frac{6}{8} \div \frac{2}{8} = \frac{3}{8}$) and choosing the larger denominator when adding unit fractions with unequal denominators (e.g., $\frac{1}{4} + \frac{1}{8} = \frac{2}{8}$). Thus, in the current study we targeted less-practiced types of fraction problems and problems where students make common strategy errors, including division in general as well as addition and multiplication problems involving equal denominators (see Table 1, Cases 2, 4, 5, and 6).

In summary, fraction arithmetic is a suitable domain to assess both the effectiveness of a computer tutor and the effectiveness of erroneous examples for learning and reviewing of mathematical content. Fractions skills are important, but there are well-documented errors in adults' fraction arithmetic. Some of these errors are based on fraction misconceptions (i.e., natural number bias) and gaps in prior knowledge. However, adults have prior experience with fraction procedures, thus they come with a basic foundation, allowing them to reflect on their mistakes. Erroneous examples are designed to address fraction misconceptions and provide an opportunity for reflection. Fraction arithmetic thus provides an appropriate testing ground for our work.

Erroneous Examples

The format of an erroneous example can vary, but a common approach outlined in Adams et al. (2014) proposes the following guidelines: (1) the error in the example solution should be produced by a hypothetical student to avoid embarrassing any student, (2) the error process should be interactive and engaging, prompting students for explanations and feedback, and (3) the erroneous example should be sufficiently structured to minimize cognitive load. An erroneous example following these guidelines is illustrated in Figure 1.

Table 1

Common Fraction Errors by Operation

Case	Operation	Denominator	Error	
1	Addition	Common	Adding both the numerators and denominators	$\frac{1}{4} + \frac{1}{4} = \frac{2}{8}$
2	Addition	Different	Adding the numerators, choosing the larger denominator	$\frac{1}{4} + \frac{1}{8} = \frac{2}{8}$
3	Subtraction	Common & Different	Subtracting both the numerators and denominators	$\frac{4}{5} - \frac{1}{3} = \frac{3}{2}$
4	Multiplication	Common	Multiplying the numerators, keeping the denominator	$\frac{2}{7} \times \frac{3}{7} = \frac{6}{7}$
5	Division	Common	Dividing the numerators, keeping the denominator	$\frac{6}{7} \div \frac{3}{7} = \frac{2}{7}$
6	Division[a]	Common & Different	Inverting the first fraction and then multiplying	$\frac{2}{5} \div \frac{3}{7} = \frac{5}{2} \times \frac{3}{7} = \frac{15}{14}$

[a]Error specific to adults.

Figure 1. An erroneous example from our computer tutor illustrating the use of (i) a hypothetical student, (ii) an interactive and engaging prompt, and (iii) a structure that minimizes cognitive load.

Note. The section labels are added here but are not part of the tutor interface.

Erroneous examples have been used successfully to refute misconceptions in various areas of mathematics with students in middle and high school (Adams et al., 2014; Barbieri & Booth, 2016; Booth et al., 2013; Durkin & Rittle-Johnson, 2012; McLaren et al., 2015; Tsovaltzi et al., 2012). To illustrate, Adams and colleagues (2014) examined the effectiveness of an intelligent tutoring system that included erroneous examples designed to address common decimal misconceptions. Sixth grade middle school children who were given erroneous examples performed significantly better on a delayed post-test and were more accurate at judging whether their post-test answers were correct than those who were given traditional examples. When students were divided into low- and high-prior-knowledge groups, both groups had greater delayed post-test gains when they were presented with erroneous examples compared to traditional examples, suggesting that erroneous examples are effective for learners of varying prior-knowledge levels. Adams et al. concluded that by having students of all knowledge levels identify, explain, and correct errors, students can gain a deeper level of understanding of decimals.

In a replication of the Adams and colleagues (2014) study with a larger sample size, McLaren et al. (2015) had middle school students (11-13 years of age) learn about decimals through an interactive computer tutor. Students were either presented with problem solving supplemented with erroneous examples or standard problems. Specifically, in the erroneous examples condition, students were presented with a worked-out solution that contained errors and had to identify and correct the errors and self-explain the correct solution. In the standard condition, students were asked to solve problems by generating their solutions and self-explain the correct solutions. Students in the erroneous-example group performed significantly better on the delayed post-test, but not the immediate post-test, than students in the traditional problem-solving group. Mirroring prior findings, there was no significant difference in the effectiveness of erroneous examples for low- and high-prior-knowledge students.

Other studies have found more nuanced effects of erroneous examples. In particular, contrary to the findings of Adams et al. (2014) and McLaren et al. (2015), some have reported that erroneous examples are only beneficial for students with sufficient prior knowledge. For example, Heemsoth and Heinze (2014) taught Grade

6 students how to multiply and divide fractions over the course of 11 lessons, either through standard correct worked examples or erroneous examples. Though the erroneous-example group had a better understanding of incorrect strategies and concepts, with respect to fraction knowledge, only students with high prior knowledge benefited from erroneous examples. In contrast, low-prior-knowledge students benefitted more from standard correct examples. Similar results were reported in a study with university students (Große & Renkl, 2007). Specifically, university students were asked to explain both correct and erroneous examples presented via paper-and-pencil in the domain of probability. High-prior-knowledge students showed increased learning from the erroneous examples, while low-prior-knowledge students only learned from erroneous examples when the errors were explicitly highlighted. Thus, the results from these studies suggest that erroneous examples support learning only if students have a certain foundation of prior knowledge so that they can adequately reflect on errors embedded in the example solutions and their own misconceptions.

It would, however, be premature to conclude that erroneous examples are only beneficial for high-prior-knowledge-students, as the next study illustrates. Huang et al. (2008) designed a computer tutor for decimal concepts. In their study, Grade 6 students either worked with traditional paper-and-pencil test sheets or with a computer tutor, where they were asked to identify cognitive conflicts associated with their error. Similar to erroneous examples, when students made an error, the tutor presented them with a cognitive conflict (i.e., a wrong idea is presented to the user to prompt him/her to examine the reasonableness of his/her answer). The cognitive conflict was designed to help identify their error and clarify misconceptions. The intervention was successful: The tutor group had greater gains from pre-test to post-test than the paper-and-pencil group. Interestingly, within the computer tutor group, low-prior-knowledge students made greater gains than the high-prior-knowledge students, which is the opposite pattern to what was found by the other studies described above. In a study of decimal magnitudes paper-and-pencil tutors (Grades 4 and 5), Durkin and Rittle-Johnson (2012) suggested that erroneous examples can be beneficial to low-prior-knowledge students if students are presented with both correct and erroneous examples and asked to compare or contrast the examples. This was the case in a computer study by Stark et al. (2011) with undergraduate medical students' diagnostic competence. In this study, low-prior-knowledge students benefitted from both erroneous and correct examples when the examples were accompanied by elaborate feedback (Stark et al., 2011). Of note is that, with the exception of Stark et al., studies reporting an effect of prior knowledge involved paper and pencil materials rather than a computer tutor and so more work is needed to see if this effect transfers to a computer tutor context.

A potential downside of erroneous examples relates to affect. Learning from erroneous examples may be a more confusing and frustrating process, with higher confusion and frustration levels linked to poorer learning outcomes (Richey, Andres-Bray, et al., 2019; Richey, McLaren, et al., 2019). Thus, the varied results obtained in studies may reflect the difficulty of studying erroneous examples because many factors, including knowledge, affect, and how the error is presented can influence the usefulness of erroneous examples.

In summary, there is some evidence that erroneous examples promote learning, but there is a lack of consensus as to who will benefit from erroneous examples in comparison to traditional examples (i.e., low- vs. high-prior-knowledge students). In general, factors like the target domain and the age of the participants may influence the results and so more work is needed for a full understanding of when and how erroneous examples impact learning. In the present study we explore the utility of erroneous examples to supplement problem-solving with a computer tutor in the domain of fraction arithmetic.

The Fraction Computer Tutor

The instructional materials used in this study were administered by two versions of a computer tutor we built using the Cognitive Tutor Authoring Tool (CTAT; Aleven et al., 2009). This authoring tool supports the construction of so-called example-tracing tutors, used in the present study. The construction of example-tracing tutors does not require prior programming experience, making it accessible for both educators and researchers. One advantage of this type of tutor is that it can provide personalized step-by-step feedback and help. It accomplishes this by comparing student solutions against pre-stored correct and incorrect responses that are created by a human author when the tutor is designed (Aleven et al., 2009). There are three key components to the example-tracing tutor CTAT architecture: a) the front-end tutor interface that the student interacts with, for instance to view the erroneous example or solve the target problem; b) the authoring interface used by a human author to specify the problem steps, hints, and feedback (this is done during the tutor construction phase, and not seen by the student interacting with the tutor); c) an online platform where all actions within the tutor are stored and can be downloaded for subsequent analysis (Aleven et al., 2016). Both tutors created for this study included problems for users to solve and provided feedback and hints; one of the tutors additionally included erroneous examples.

We piloted initial versions of the tutor with both adults and children in order to obtain a range of feedback on aesthetics, wording, and difficulty. Though the tutor was not tested with children in the present study, it was designed so that it could be used by people of all ages who are learning or reviewing fraction concepts and procedures. First, the tutor was piloted with five numerical cognition experts. Based on their feedback, the examples were simplified, and the order of the example solution steps was made more explicit. Once these changes were implemented, the tutor was piloted with three children between the ages of 8 and 10. Based on their feedback, questions were made less wordy and aesthetics, such as the size of pictures, were adjusted.

Using the feedback from the pilots, we implemented two versions of the fraction tutor: An erroneous-example (EE) tutor that supplemented problem solving with erroneous examples, and a problem-solving (PS) tutor that did not include the erroneous example component. Both tutors contained the same six fraction word problems, namely two single-digit addition fraction problems, one single-digit subtraction problem, one single digit multiplication problem, and two single-digit division problems (see Figure 2 and Figure 3 for an example). However, the presentation of the problems differed slightly because the EE tutor included erroneous example components.

The erroneous example components for the EE tutor were based on the common fraction errors observed in adults (Braithwaite et al., 2017; Tan, 2020) as previously discussed. Table 1 shows six common fraction operation errors included in the six corresponding EE tutor activities - all were based on misconceptions reported in the literature, such as treating numerators and denominators as separate entities and applying addition/subtraction procedures to multiplication problems.

The EE tutor interface consisted of five sections and a Feedback Centre. As shown in Figure 2, Section 1 presented the problem description, while Section 2 corresponded to the first part of the erroneous example component, conceptualized as an incorrect response to the problem made by a fictional character. The second part of the erroneous example, shown in Section 3, asked users to identify the fictional character's error by choosing from a list of three options. This design aimed to promote active processing by asking users to identify the misconception – by stating it was incorrect, this also refuted that concept. Section 4 asked users to solve

the same fraction problem as was incorrectly solved in Section 2 by the fictional character. This gave users the opportunity to apply the correct procedure, thereby further refuting the misconception. Finally, Section 5, the Advice Centre, asked users to provide advice to the fictional character so that they could avoid making the same mistake in the future. This step aimed to encourage users to transfer their knowledge for the correct procedure to similar problems in the future. An additional component of the tutor, the Feedback Centre, will be described after we present the PS tutor, as both tutors included this component.

The PS tutor was populated with the same fraction problems as the EE tutor, but the interface did not include the erroneous example components (i.e., Sections 2 and 3, see Figure 3 for an example). In the Advice Centre, instead of asking users to help the fictional character avoid the mistake in the future, users were asked to identify the correct way to solve the problem. Thus, users of both tutors had the opportunity to identify the correct fraction procedure necessary to solve similar future problems with respect to the underlying fraction concept.

Both the EE tutor and the PS tutor provided immediate feedback on each user response by colouring correct responses green and incorrect responses red. Both tutors included a Feedback Centre that was always visible in the right panel on the screen (see Figure 2 and Figure 3), with on demand hints that users could access by clicking the corresponding button. Hints were displayed in the Feedback Centre, as were system generated messages, and the contents of these was the same in the two tutors (see Figure 4). The hints were designed to address difficulties solvers might have at certain steps in their problem-solving, such as help finding common denominators or providing brief explanations of procedures. Both versions of the tutor also occasionally generated encouraging messages, such as, "*Good job! Now let's give some advice to Samantha that she can use in the future!*"

Figure 2. Sample problem used in the Erroneous-Example Tutor, including (1) problem statement, (2) incorrect response by fictional character, (3) identification of the character's error, (4) problem-solving work area used to generate the correct solution, (5) Advice Centre to solidify the general procedure, and (5) the Feedback Centre.

Note. The section labels are added here but are not part of the tutor interface.

Figure 3. Sample problem used in the Problem-Solving Tutor.

Note. The interface is the same as for the EE tutor shown in Figure 2 except that the erroneous example components are absent here (i.e., Sections 2 and 3 in Figure 2).

Figure 4. Examples of hints provided to participants in the EE and PS Tutors.

Both the EE tutor and the PS tutor required that all items in a given exercise be answered correctly before the user could move on to the next exercise (example and subsequent problem in the EE tutor; problem in the PS tutor); all questions were completed in a set order in both tutors and the order was the same for both tutors. To

move from one item to the next, participants pressed the "Done" button located in the Feedback Centre (see Figure 2, Figure 3, and Figure 4). Users were prompted to move to the next question with a message in the Feedback Centre that stated, "*Click the Done button.*" If users tried to move on to the next question prior to producing a complete correct solution, they received the message "*I'm sorry, but you are not done yet. Please continue working.*"

Present Study

The present study compared learning outcomes from the two versions of the computer tutor described above – namely one that supplemented traditional problem solving with erroneous examples (E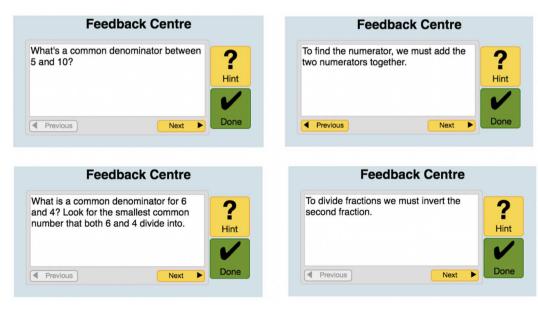E tutor) and one that consisted of only traditional problem solving (PS tutor). In both tutor versions participants received hints and feedback. The target population was undergraduate students who had not taken any university-level math courses. A prior study with undergraduate students found that erroneous examples were most effective for students with some prior knowledge of the topic (Große & Renkl, 2007). Since our population had exposure to fractions, erroneous examples are appropriate for this population - but whether it is a more beneficial learning strategy over traditional problem solving is an open question investigated in the present work. We also analysed the effect of erroneous examples for low- and high-prior-knowledge learners. Because there is a lack of consensus surrounding the effectiveness of erroneous examples, we did not make directional hypotheses. Instead, we asked the following questions:

Question I. Is a computer tutor an effective tool for improving fraction arithmetic in adults?

Question II. Does problem solving supplemented with erroneous examples lead to greater gains from pre-test to post-test than problem solving without erroneous examples?

Question II. Does the effect of erroneous examples depend on prior knowledge level?

Method

Participants

Eighty-seven undergraduate students from a Canadian university participated in the study (M_{age} = 21.06 years, *SD* = 5.15; 65.5% female). All participants spoke fluent English, with 61% identifying English as their first language. The most common first languages other than English were Chinese (10%), French (7%), and Arabic (5%). To avoid ceiling effects participants were not eligible to participate if they had current or previous enrolment in any post-secondary level mathematics courses. The most common majors were Cognitive Science and Psychology (46% of participants); other majors were predominantly Social Science programs. Most participants were in Year 1 (41%) or Year 2 (31%) of university. The study was approved by the Carleton University Research Ethics Board. Bonus course credit was provided as compensation for participation in the study.

Measures

Demographics

Basic demographic information was collected, including participant age, gender, and program of study.

Fraction Operations

To measure learning, participants completed two fraction operation tests: A pre-test and a post-test. The pre-test and post-test were paper-and-pencil tests designed for the present study. Both the pre- and post-test consisted of 18 unique items: Four addition (e.g., $\frac{1}{5} + \frac{2}{5}$), four subtraction (e.g., $\frac{7}{8} - \frac{4}{7}$), four division (e.g., $\frac{3}{5} \div \frac{1}{4}$), and four multiplication (e.g., $\frac{5}{6} \times \frac{3}{4}$) questions, as well as two word problems that required the sequential combination of operations. All questions consisted of single-digit numerators and denominators; total score was the sum of correct responses. The pre-test and post-test used different numbers but were analogous because they had the same problem structure. Participants had 20 minutes to complete the pre- and post-tests. The maximum possible score on each test was 18. The pre-test, post-test, and images of each of the problems presented in the fraction tutors can be found in the Supplementary Materials.

As part of a larger study, several other brief questionnaires about math attitudes and beliefs were administered (i.e., Math Confidence Scale (Hendy et al., 2014), the Short Grit Scale (Duckworth & Quinn, 2009), the Abbreviated Math Anxiety Scale (Hopko et al., 2003), the Mathematics Self-Concept, Self-Efficacy, and Anxiety Scale (Lee, 2009)) but because we do not report on their analysis, we do not describe them here.

Design and Procedure

A between-subjects design was used with the two computer tutors (EE, PS) serving as the two experimental conditions. Participants were assigned to one of the two conditions, alternating between EE and PS assignment to ensure an even number of participants were in each condition. The procedure for both conditions was the same.

The study took place in a laboratory; most experimental sessions included one participant, but several sessions included two (who were seated at opposite ends of a large room). Prior to entering the lab participants were unaware the study involved fractions (the title of the study on the online recruiting site was "Education with AI" and participants were told the study required completing a series of problem-solving tasks). After informed consent was obtained, participants were given 20 minutes to complete the fraction operation pre-test to assess their initial fraction performance; they were not allowed to use any assistive tools (i.e., calculators) and could not ask for help. Following this, participants completed several questionnaires (not reported in this study). They then had 20 minutes to work through the six questions using either the EE or PS tutor. After finishing these questions, participants were given 20 minutes to complete the fraction post-test. Finally, participants completed a post-tutor attitudes and beliefs questionnaire (not reported). The entire study took between one and one and a half hours to complete.

Results

Of the 87 participants recruited for the study, 75 were included in the analyses. Ten participants were excluded because they obtained a perfect score on the pre-test and thus were at ceiling performance. Two other two participants were excluded from the analysis (one had difficulty navigating the computer tutor and was unable to finish the tutor intervention and the other participant's data was lost). In total, there were 39 participants in the erroneous-example group and 36 participants in the problem-solving group.

The analyses related to our research questions were conducted using ANOVAs and *t*-tests. In addition to reporting frequentist statistics, Bayes factors are also reported, allowing for an evaluation of the fit of the data under the null and alternative hypotheses. Bayesian statistics offer advantages in terms of mathematical power and in contrast to frequentist statistics, allow researchers to make claims about the likelihood of both the null and alternative hypothesis given the evidence (Jarosz & Wiley, 2014). The Bayes factor, BF_{01}, is "a ratio that contrasts the likelihood of the data fitting under the null hypothesis with the likelihood of fitting under the alternative hypothesis" (Jarosz & Wiley, 2014, p. 3). The inverse, BF_{10}, puts the ratio in terms of the alternative hypothesis. In the present study, when the Bayes factor is in favour of the null hypothesis, the BF_{01} is reported; when in favour of the alternative hypothesis, the BF_{10} is reported. The interpretation of the strength of the evidence for the null or alternative hypothesis is in accordance with Jeffreys' (1961) guidelines (see Table 4 of Jarosz & Wiley, 2014).

Effect of Tutors on Learning

The descriptive statistics for the pre-test, post-test, and gain scores are in Figure 5. We first verified that there were no group differences on pre-test scores between the erroneous-example (EE) group and problem-solving (PS) group. This was the case, $t(73) = 0.81$, $p = .42$, $d = 0.19$, and thus we proceeded with the primary analyses. The pre-test scores were around 10/18 for both conditions, indicating that participants' a priori knowledge of fractions was modest.

Figure 5. Mean fraction pre-test, post-test, and gain scores for the EE and PS groups (Max Score = 18).

Note. Error bars represent standard error of the mean.

To measure learning, we used the standard method of calculating gain scores (i.e., post-test score – pre-test score). Collapsing across tutor groups, participants gained significantly from pre-test to post-test, $t(74) = 7.73$, $p < .001$, $d = 0.89$, $BF_{10} = 2.19e+08$. A one-way ANOVA found no significant difference in gain scores between the EE and PS groups, $F(1, 73) = 0.92$, $p = .34$, and the effect was very small, $\eta^2 = .01$. The estimated Bayes factor, $BF_{01} = 2.82$ indicates anecdotal support for the null hypothesis that the EE tutor did not provide additional benefits in comparison to the PS tutor. In summary, the computer tutor improved learning (Research Question I) but we did not find support that supplementing problem solving with erroneous examples resulted in higher learning than using a traditional problem-solving tutor (Research Question II).

Effect of Prior Knowledge

Previous studies have suggested that a student's prior knowledge can influence how much they gain from erroneous examples (e.g., Booth et al., 2013; Chen et al., 2016; Tsovaltzi et al., 2012). To test if this was the case in our data, participants were divided using a median split of the pre-test scores (*Mdn* = 10). We excluded from analysis participants with the median score and with scores right below and right above the median scores (as these scores are close to the median they have the potential to obscure results and thus removing them is advocated (Fulcher, 2005)): low-level-knowledge group (*n* = 32) and a high-level-knowledge group (*n* = 35). Table 2 displays mean pre- and post-test scores, separated by tutor group (i.e., EE and PS) and prior knowledge level (i.e., low and high).

Table 2

Mean Test Scores for Low-Level- and High-Level-Knowledge Learners, Separated by Tutor

	Low-Level				High-Level			
	EE (*n* = 17)		PS (*n* = 15)		EE (*n* = 16)		PS (*n* = 19)	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pre-test	4.88	2.76	5.40	2.47	14.88	1.78	14.79	1.65
Post-test	10.47	4.89	8.87	4.79	15.88	2.45	16.79	1.48
Gains	5.59	4.45	3.47	4.16	1.00	2.10	2.00	1.80

There was no significant difference between the EE and PS groups on pre-test scores for low-level-knowledge participants, $t(30) = 0.55$, $p = .58$, $d = 0.20$ or for the high-level-knowledge participants, $t(33) = 0.15$, $p = .88$, $d = 0.05$.

To assess learning, gain scores (post-test – pre-test) were analyzed in a 2(knowledge level: low, high) by 2(tutor: erroneous example, problem solving) between-subjects ANOVA. Of primary interest is the interaction between tutor type and knowledge level, since it informs on whether the effect of erroneous examples depends on knowledge level (see Figure 6). The low-level-knowledge students made greater gains than the high-level-knowledge students in the EE group, but this interaction was not significant, $F(1, 63) = 3.74$, $p = .058$, $\eta^2 = .05$. The Bayes factor, $BF_{10} = 1.44$, indicates anecdotal evidence for the inclusion of the interaction in the model. As shown in Figure 6, participants with low prior knowledge had higher mean gains from problem solving supplemented with erroneous examples than from problem solving alone, while the opposite pattern was true for participants with high prior knowledge and the difference in learning was smaller. However, we acknowledge that this difference is not statistically significant. Thus, with respect to Research Question III, we did not find evidence that for individuals with low prior knowledge, erroneous examples were more beneficial than problem solving alone.

There was also a significant main effect of knowledge level, $F(1, 63) = 14.08$, $p < .001$, $\eta^2 = .18$, which indicates that low-level-knowledge participants made greater learning gains than high-level-knowledge participants. The Bayes factor, $BF_{10} = 66.60$, indicates very strong support for the model that includes the effect of knowledge level. There was no main effect of tutor, $F(1, 63) = 0.48$, $p = .49$, $\eta^2 = .00$, and the Bayes factor, $BF_{01} = 3.04$, indicated substantial support for the null hypothesis that collapsed across knowledge levels, the two tutor groups did not differ in learning gains.

Error Type Analysis

Both versions of the tutor (EE, PS) were designed to provide instruction on procedures for fraction arithmetic. An exploratory analysis was conducted to identify the types of errors participants made, to see if the tutors reduced errors and to analyze if the tutor version reduced specific types of errors.
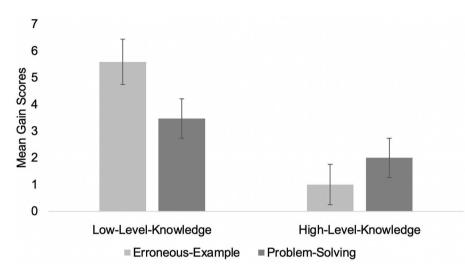

Figure 6. Mean fraction test gain scores for low-level and high-level fraction knowledge, separated by tutor condition.

Note. Error bars represent standard error of the mean.

We first flagged incorrect pre-test and post-test responses and then classified the errors based on the common misconceptions (see Table 1) and a review of errors by two numerical cognition experts. Table 3 shows the error classifications, broken down by tutor type (EE, PS) and operation. Classifiable errors fell into three main categories: Conceptual errors, arithmetic errors, and blank responses (no answer provided). Conceptual errors were the result of applying the incorrect procedure when trying to solve fraction arithmetic problems. For example, participants may have added both the numerator and denominator for an addition problem instead of finding a common denominator and only adding the numerators. Similarly, participants may have inverted and multiplied for a multiplication problem because this is the correct procedure for a division problem. Arithmetic errors were the result of incorrectly adding, subtracting, multiplying or dividing; they were not the direct result of a fraction misconception. For example, a participant may have known to invert the second fraction and multiply for a division problem, but when they multiplied, they obtained the incorrect response. Included in this category were reduction errors. For example, if a participant obtained a correct response of $\frac{9}{12}$ but incorrectly reduced the response to $\frac{2}{3}$.

The total number of conceptual errors did not significantly differ between the EE and PS groups on the pre-test (109 vs. 107) or on the post-test (72 vs. 47), $\chi^2(1, N = 335) = 3.11$, $p = .08$. The number of arithmetic errors also did not significantly differ for the EE and PS groups on the pre-test (48 vs. 78) or on the post-test (53 vs. 85), $\chi^2(1, N = 264) = 0.003$, $p = .96$. Furthermore, the number of blank responses where participants did not provide an answer at all did not significantly differ for the erroneous-example and problem-solving tutor groups on the pre-test (96 vs. 60) or on the post-test (18 vs. 22), $\chi^2(1, N = 196) = 3.58$, $p = .06$.

Table 3

Summary of Common Pre-Test and Post-Test Errors Across Each Operation for the Erroneous-Example and Problem-Solving Tutors

	Addition				Subtraction				Multiplication				Division			
	Pre-test		Post-test		Pre-test		Post-test		Pre-test		Post-test		Pre-test		Post-test	
Error	EE	PS	EE	PS	EE	PS	EE	PS	EE	PS	EE	PS	EE	PS	EE	PS
Conceptual Error																
Added/subtracted numerator and denominator	30	16	12	8	16	13	0	2	–	–	–	–	–	–	–	–
Only multiplied denominator when finding common denominator	0	4	3	5	0	3	3	7	–	–	–	–	–	–	–	–
Used bigger denominator as denominator	1	1	0	0	3	1	2	0	–	–	–	–	–	–	–	–
Found common denominator then multiplied/added/subtracted/divided numerators	–	–	–	–	–	–	–	–	15	4	11	11	14	10	6	11
Cross multiplied	–	–	–	–	–	–	–	–	7	29	1	5	–	–	–	–
Inverted and multiplied	–	–	–	–	–	–	–	–	6	6	0	12	–	–	–	–
Added numerator and multiplied denominator	–	–	–	–	–	–	–	–	2	1	0	6	–	–	–	–
Inverted first fraction	–	–	–	–	–	–	–	–	–	–	–	–	2	4	0	4
Cross divided/ inverted and divided	–	–	–	–	–	–	–	–	–	–	–	–	7	12	9	0
Multiplied instead of divided	–	–	–	–	–	–	–	–	–	–	–	–	4	5	0	1
Total	**31**	**21**	**15**	**13**	**19**	**17**	**5**	**9**	**30**	**40**	**12**	**34**	**27**	**31**	**15**	**16**
Arithmetic Error																
Reduction error	2	2	2	4	0	4	2	1	1	1	6	2	1	6	1	4
Arithmetic error	6	5	9	13	14	8	9	6	8	2	8	5	5	1	7	8
Miscellaneous	8	3	6	2	8	5	14	1	9	3	10	1	16	8	11	6
Total	**16**	**10**	**17**	**19**	**22**	**17**	**25**	**8**	**18**	**6**	**24**	**8**	**22**	**15**	**19**	**18**
Blank Response																
Blank or "I don't know"	0	4	3	0	6	9	6	6	0	28	4	3	54	55	9	9

Note. EE = erroneous-example tutor; PS = problem-solving tutor.

Importantly, when we totaled errors across both groups from pre-test to post-test, the number of conceptual errors decreased (216 vs. 119), whereas the number of arithmetic errors did not (126 vs. 138), $\chi^2(1, N = 598) = 16.67$, $p < .001$. Given that the tutor was designed to provide instruction and practice for fraction arithmetic procedures, not arithmetic in general, we would not expect to see arithmetic improvements. The reduction in conceptual errors suggests the computer tutor was effective at reducing fraction arithmetic misconceptions. Along these lines, the number of questions that were unanswered (i.e., left blank or participants wrote, "I don't know") significantly differed from pre-test to post-test, $\chi^2(1, N = 186) = 60.41$, $p < .001$. Thus, in addition to making fewer conceptual errors at post-test, participants were also more likely to attempt a problem that they previously did not know how to approach. Overall, the error analyses are informative as they show that both versions of the tutor were effective in improving conceptual fraction knowledge.

Discussion

The goal of the present study was to investigate the utility of a computer tutor and erroneous examples in the domain of fraction arithmetic. Fraction arithmetic was chosen because adults frequently require the use of fractions in the workplace (Handel, 2016) and because both children and adults struggle with this challenging

domain (DeWolf et al., 2014; Schneider & Siegler, 2010; Tan, 2020). Research with children and adults shows that they make similar conceptual and procedural errors (Braithwaite et al., 2017; Tan, 2020), suggesting that adults may not have correctly encoded fraction concepts. The study pre-test scores confirmed that the adults in our sample find fraction arithmetic difficult, as the pre-test mean was just above 50%. Thus, although participants were university students who all had learned fractions in their elementary school years, many forgot and/or never mastered some fraction concepts.

With the goal of improving fraction understanding, we built two versions of a fraction tutor using CTAT, a platform designed to support the construction of computer tutors. Unlike textbooks or worksheets, CTAT tutors are capable of sophisticated tutoring behaviours (Aleven et al., 2009), including step-by-step assistance, real-time feedback, and hints tailored to specific steps in the problem-solving solution. The motivation for including these types of support comes from studies on human tutoring indicating that such tailored support may be beneficial for supporting learning (although as reviewed in VanLehn (2011), it is still an open question as to the exact effect of each type of support). In the past, such support was only available from human tutors. However, human tutors are not always a viable option as there has been an increase in the number of students seeking private after-school tutoring in recent years and private tutoring can be costly (Hart & Kempf, 2015). Computer tutors address this limitation as they can be freely disseminated, particularly if they are online as is the case for our tutor.

Although participants learned from their interaction with the fraction tutor, we did not find evidence that supplementing problem solving with erroneous examples improved learning over traditional problem solving. We used Bayesian analyses in addition to frequentist statistics to triangulate results from alternative analysis methods. Bayesian statistics provided definitive evidence that the computer tutor led to significant learning gains. However, neither frequentist nor Bayesian statistics provided strong evidence in favour of erroneous examples. When all participants regardless of prior knowledge were considered, the effect size of erroneous examples was small and the Bayes Factor in favour of the null model was 2.8 (i.e., the null model was 2.8 times as likely as the alternative model that included erroneous examples). A Bayes Factor (BF) of 1-3 is considered anecdotal evidence for a given model (null or alternative), while a BF of 3-10 provides substantial evidence. Our results are close to the boundary between these two thresholds. Even if the Bayesian analysis provided substantial evidence, however, the small sample effect size suggests erroneous examples are not adding much.

While to the best of our knowledge our work is the first to incorporate Bayesian statistics into analyses about erroneous examples, in general our findings are in line with the findings of Adams et al. (2014) and McLaren et al. (2015) who did not find immediate benefits of erroneous examples. They did, however, find that erroneous examples led to greater learning gains on a delayed post-test. While we aimed to implement a delayed post-test phase, very few participants returned for this phase, making it impossible to analyze delayed post-test data. Another possible explanation for why we did not find erroneous examples to be more beneficial than problem solving alone relates to the support delivered by the CTAT tutor. Both groups received feedback and hints from the tutor. Feedback is generally beneficial, increasing learning gains (Shute, 2008) and has been found to be beneficial for both erroneous and correct examples (Stark et al., 2011). Thus, providing feedback and hints to both the EE and PS groups might have overshadowed any effects of erroneous examples, accounting for why we did not find significant differences in learning.

Some have reported that erroneous examples increased learning only for students with certain prior knowledge, although conflicting patterns have been reported. Some studies report that erroneous examples are only beneficial for students with high prior knowledge (Große & Renkl, 2007; Heemsoth & Heinze, 2014), while other studies found the opposite, namely that erroneous examples are more beneficial for students with low prior knowledge (Huang et al., 2008; Stark et al., 2011). In the present study, low-prior-knowledge students made greater gains than the high-prior-knowledge students in the EE group, but the difference was not statistically significant and the Bayes statistics provided only anecdotal evidence for the interaction term, with a BF of 1.44. It may be that erroneous examples are only beneficial for low-prior-knowledge students if the errors in the examples are explicitly highlighted (Große & Renkl, 2007). In the present study, participants were required to identify the error prior to moving to the problem-solving portion of the intervention, effectively highlighting the error, so this does not explain the lack of a significant effect. We also did not find evidence that erroneous examples were beneficial for participants with high prior knowledge. These individuals are less likely to have fraction misconceptions, and thus erroneous examples did not provide benefits.

The exploratory error analysis focused on identifying fraction misconceptions and whether interacting with the tutor reduced their frequency. Comparing the EE and PS versions of the fraction tutor, there were no significant differences in the types of errors made on either the pre-test or post-test. The latter result was somewhat surprising as the EE tutor explicitly illustrated conceptual errors and gave learners an opportunity to correct them by identifying the proper procedure. For both groups, the number of conceptual errors decreased in the post-test. Thus, while we did not find evidence that erroneous examples reduced errors more than standard problem solving, interacting with the fraction tutor did reduce fraction arithmetic misconceptions. Furthermore, the reduction in blank responses from pre-test to post-test means participants were more willing to try and devise a solution. This is a positive result as in educational settings, a blank response often receives no marks and cannot receive feedback to correct errors since no errors were recorded.

Educational Applications and Implications

The present study involved the domain of fraction arithmetic, something both children and adults struggle with (Braithwaite et al., 2017; DeWolf et al., 2014; Schneider & Siegler, 2010; Tan, 2020). As with other problem-solving domains, fraction skills can be improved with practice. However, traditional worksheets and textbook problems do not provide tailored feedback and support. Our results show that a fraction computer tutor helps promote learning of fraction concepts in adult participants. Since the tutor is online, it is easily accessible by individuals who wish to review fraction concepts and can be used to supplement remedial mathematics education in traditional classrooms. Since adults and children have similar fraction misconceptions (Braithwaite et al., 2017; Tan, 2020), the tutor may also be beneficial to younger students, but this requires empirical validation.

One important consideration is the feasibility of designing a computer tutor. Building a computer tutor traditionally has required technical expertise and many hours of programming. For the present study, we used CTAT (Aleven et al., 2009) to build both versions of the fraction tutor. With CTAT, no programming experience is required, making it feasible for teachers to design and implement their own computer tutor, tailored to the instruction and needs of their students.

Limitations and Future Work

In the present study, one test was used as a pre-test and a second test was used as a post-test. While the tests were carefully designed so that similar magnitude numerators and denominators were selected, these were not counterbalanced. The study also did not include a delayed post-test. Some previous studies have found that the benefits of erroneous examples only appear on delayed post-tests (Adams et al., 2014; McLaren et al., 2015). It is possible that the erroneous examples would lead to increased learning gains because they encourage deeper processing of the material during learning (Adams et al., 2014). A related consideration is the length of the intervention, which was relatively brief. Previous studies that found benefits of erroneous examples on either immediate or delayed post-tests had longer interventions that were between 75-120 minutes in length (e.g., Große & Renkl, 2007; McLaren et al., 2012, 2015).

Another limitation relates to the design of the tutor. The erroneous-example group was not required to generate free-form responses to identify and correct the error in the example, instead selecting items from a multiple-choice question. The latter strategy is standard in computer tutors (Conati & VanLehn, 2000) given the challenge of automatically parsing free-form typed input. However, the effect of erroneous examples may be strengthened when responses are generated rather than selected. With multiple-choice questions, there is a risk that students will click responses until they obtain the correct answer (i.e., receive the green highlight) and may not take the time to process the misconception. In future studies, open-ended questions in which participants generate a response should be considered as this may improve learning.

Yet another consideration relates to the inclusion of additional conditions. Future studies could compare the computer tutor conditions to paper-and-pencil erroneous example materials as well as paper-and-pencil problem-solving, to separate benefits of a computer tutor from benefits of the type of example. Future studies could also include a "pure" control condition, in which participants would not receive an intervention, instead just completing the pre-test, a non-math related task, and the post-test, to see if any gains occur simply from the pre-test priming participants' memory about fraction procedures.

Conclusion

The present study investigated the pedagogical value of a computer tutor for fraction arithmetic as well as the utility of erroneous examples to supplement problem solving. Both versions of the fraction tutor were successful as participants improved significantly from pre-test to post-test. In addition to standard frequentist statistics, our results were supplemented with effect sizes and Bayesian statistics. Despite all these methods, we did not find evidence that supplementing problem-solving activities with erroneous examples produced higher learning compared to problem solving without examples. Additionally, although low-prior-knowledge students in the erroneous-example condition had higher learning gains than in the problem-solving condition, this difference was not significant. Given that to date there are relatively few studies on erroneous examples, and even fewer with adults, there is a clear need for more work in this area.

Funding

Competing Interests

Acknowledgments

Supplementary Materials

The Supplementary Materials contain the pre-test, post-test, and images of each of the problems presented in the fraction tutors (for access see Index of Supplementary Materials below).

Index of Supplementary Materials

Di Lonardo Burr, S. M., Douglas, H., Vorobeva, M., & Muldner, K. (2020). *Supplementary materials to "Refuting misconceptions: Computer tutors for fraction arithmetic"* [Additional information]. OSF. https://osf.io/bg7j9/

References

Adams, D. M., 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, 36*, 401-411. https://doi.org/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*, 105-154.

Aleven, V., McLaren, B. M., Sewall, J., van Velsen, M., Popescu, O., Demi, S., . . . Koedinger, K. R. (2016). Example-tracing tutors: Intelligent tutor development for non-programmers. *International Journal of Artificial Intelligence in Education, 26*(1), 224-269. https://doi.org/10.1007/s40593-015-0088-2

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181-214. https://doi.org/10.3102/00346543070002181

Ayub, A. F. M., Mokhtar, M. Z., Luan, W. S., & Tarmizi, R. A. (2010). A comparison of two different technologies tools in tutoring Calculus. *Procedia: Social and Behavioral Sciences, 2*(2), 481-486. https://doi.org/10.1016/j.sbspro.2010.03.048

Bahrick, H. P., & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General, 120*(1), 20-33. https://doi.org/10.1037/0096-3445.120.1.20

Barbieri, C., & Booth, J. L. (2016). Support for struggling students in algebra: Contributions of incorrect worked examples. *Learning and Individual Differences, 48*, 36-44. https://doi.org/10.1016/j.lindif.2016.04.001

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. https://doi.org/10.1016/j.learninstruc.2012.11.002

PsychOpen GOLD

Borasi, R. (1996). *Reconceiving mathematics instruction: A focus on errors*. Norwood, NJ, USA: Ablex Pub.

Braithwaite, D. W., Pyke, A. A., & Siegler, R. S. (2017). A computational model of fraction arithmetic. *Psychological Review, 124*(5), 603-625. https://doi.org/10.1037/rev0000072

Braithwaite, D. W., & Siegler, R. S. (2018a). Children learn spurious associations in their math textbooks: Examples from fraction arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(11), 1765-1777. https://doi.org/10.1037/xlm0000546

Braithwaite, D. W., & Siegler, R. S. (2018b). Developmental changes in the whole number bias. *Developmental Science, 21*(2), Article e12541. https://doi.org/10.1111/desc.12541

Braithwaite, D. W., Tian, J., & Siegler, R. S. (2018). Do children understand fraction addition? *Developmental Science, 21*(4), Article e12601. https://doi.org/10.1111/desc.12601

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC, USA: National Academies Press. https://doi.org/10.17226/9853

Bruce, C., Chang, D., & Flynn, T. (2013). *Foundations to learning and teaching fractions: Addition and subtraction: Literature review* (Submitted to Curriculum and Assessment Branch, Ontario Ministry of Education). Retrieved from https://knaer-recrae.ca/images/resources/Yearley-Steeles/Foundations%20to%20Learning%20and%20Teaching%20Fractions%20Lit%20Review.pdf

Chen, X., Mitrovic, A., & Mathews, M. (2016). Do erroneous examples improve learning in addition to problem solving and worked examples? In A. Micarelli, J. Stamper, & K. Panourgia (Eds.), *Intelligent Tutoring Systems* (Vol. 9684, pp. 13–22). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-39583-8_2

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182. https://doi.org/10.1207/s15516709cog1302_1

Conati, C., & VanLehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education, 11*, 389-415.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*(4), 347-362. https://doi.org/10.1037/0022-0663.79.4.347

Corbett, A., McLaughlin, M., & Scarpinatto, K. C. (2000). Modeling student knowledge: Cognitive tutors in high school and college. *User Modeling and User-Adapted Interaction, 10*(2), 81-108. https://doi.org/10.1023/A:1026505626690

DeWolf, M., Grounds, M. A., Bassok, M., & Holyoak, K. J. (2014). Magnitude comparison with different types of rational numbers. *Journal of Experimental Psychology: Human Perception and Performance, 40*(1), 71-82. https://doi.org/10.1037/a0032916

DeWolf, M., & Vosniadou, S. (2015). The representation of fraction magnitudes and the whole number bias reconsidered. *Learning and Instruction, 37*, 39-49. https://doi.org/10.1016/j.learninstruc.2014.07.002

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (Grit–S). *Journal of Personality Assessment, 91*(2), 166-174. https://doi.org/10.1080/00223890802634290

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*(3), 206-214. https://doi.org/10.1016/j.learninstruc.2011.11.001

Fulcher, E. (2005). *A guide to coursework in psychology.* Hove, United Kingdom: Psychology Press.

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction, 17*(6), 612-634. https://doi.org/10.1016/j.learninstruc.2007.09.008

Handel, M. J. (2016). What do people do at work? A profile of U.S. jobs from the survey of workplace Skills, Technology, and Management Practices (STAMP). *Journal for Labour Market Research, 49*(2), 177-197. https://doi.org/10.1007/s12651-016-0213-1

Hart, D., & Kempf, A. (2015). *Public attitudes toward education in Ontario 2015: The 19th OISE survey of educational issues*. Toronto, ON, Canada: University of Toronto. Retrieved from https://www.oise.utoronto.ca/oise/UserFiles/Media/Media_Relations/Final_Report_-_19th_OISE_Survey_on_Educational_Issues_2015.pdf

Heemsoth, T., & Heinze, A. (2014). The impact of incorrect examples on learning fractions: A field experiment with 6th grade students. *Instructional Science, 42*(4), 659-660. https://doi.org/10.1007/s11251-014-9313-x

Hendy, H. M., Schorschinsky, N., & Wade, B. (2014). Measurement of math beliefs and their associations with math behaviors in college students. *Psychological Assessment, 26*(4), 1225-1234. https://doi.org/10.1037/a0037688

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The Abbreviated Math Anxiety Scale (AMAS): Construction, validity, and reliability. *Assessment, 10*(2), 178-182. https://doi.org/10.1177/1073191103010002008

Huang, X. (2017). Example-based learning: Effects of different types of examples on student performance, cognitive load and self-efficacy in a statistical learning task. *Interactive Learning Environments, 25*(3), 283-294. https://doi.org/10.1080/10494820.2015.1121154

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*(1), 61-76. https://doi.org/10.1016/j.compedu.2006.03.007

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving, 7*(1), Article 2. https://doi.org/10.7771/1932-6246.1167

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, United Kingdom: Oxford University Press.

Koedinger, K. R., Corbett, A., Ritter, S., & Shapiro, L. (2000). *Carnegie Learning's Cognitive Tutor[TM]: Summary research results*. Pittsburgh, PA, USA: Carnegie Learning. Retrieved from https://cdn.carnegielearning.com/assets/research/Carnegie_Mellon_Carnegie_Learning_Cognitive_Tutor_-_Summary_Research_Results.pdf

Lee, J. (2009). Mathematics Self-Concept, Self-Efficacy and Anxiety Scale [Data set]. *APA PsycTests.* https://doi.org/10.1037/t31473-000

McLaren, B. M., Adams, D., Durkin, K., Goguadze, G., Mayer, R. E., Rittle-Johnson, B., … van Velsen, M. (2012). To err is human, to explain and correct is divine: A study of interactive erroneous examples with middle school math students. In

A. Ravenscroft, S. Lindstaedt, C. D. Kloos, & D. Hernández-Leo (Eds.), *21st century learning for 21st century skills* (Vol. 7563, pp. 222–235). Heidelberg, Germany: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33263-0_18

McLaren, B. M., Adams, D. M., & Mayer, R. E. (2015). Delayed learning effects with erroneous examples: A study of learning decimals with a web-based tutor. *International Journal of Artificial Intelligence in Education, 25*(4), 520-542. https://doi.org/10.1007/s40593-015-0064-x

McLaren, B. M., van Gog, T., Ganoe, C., Karabinos, M., & Yaron, D. (2016). The efficiency of worked examples compared to erroneous examples, tutored problem solving, and problem solving in computer-based learning environments. *Computers in Human Behavior, 55*, 87-99. https://doi.org/10.1016/j.chb.2015.08.038

Meert, G., Grégoire, J., & Noël, M.-P. (2009). Rational numbers: Componential versus holistic representation of fractions in a magnitude comparison task. *Quarterly Journal of Experimental Psychology, 62*(8), 1598-1616. https://doi.org/10.1080/17470210802511162

Ni, Y., & Zhou, Y.-D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist, 40*(1), 27-52. https://doi.org/10.1207/s15326985ep4001_3

Obersteiner, A., Dresler, T., Bieck, S. M., & Moeller, K. (2019). Understanding fractions: Integrating results from mathematics education, cognitive psychology, and neuroscience. In A. Norton & M. W. Alibali (Eds.), *Constructing number* (pp. 135–162). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-00491-0_7

Obersteiner, A., Van Dooren, W., Van Hoof, J., & Verschaffel, L. (2013). The natural number bias and magnitude representation in fraction comparison by expert mathematicians. *Learning and Instruction, 28*, 64-72. https://doi.org/10.1016/j.learninstruc.2013.05.003

Richey, J. E., Andres-Bray, J. M. L., Mogessie, M., Scruggs, R., Andres, J. M. A. L., Star, J. R., . . . McLaren, B. M. (2019). More confusion and frustration, better learning: The impact of erroneous examples. *Computers & Education, 139*, 173-190. https://doi.org/10.1016/j.compedu.2019.05.012

Richey, J. E., McLaren, B. M., Andres-Bray, M., Mogessie, M., Scruggs, R., Baker, R., & Star, J. (2019). Confrustion in learning from erroneous examples: Does type of prompted self-explanation make a difference? In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Artificial Intelligence in Education* (Lecture Notes in Computer Science, Vol. 11625, pp. 445–457). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-23204-7_37

Rinne, L. F., Ye, A., & Jordan, N. C. (2017). Development of fraction comparison strategies: A latent transition analysis. *Developmental Psychology, 53*(4), 713-730. https://doi.org/10.1037/dev0000275

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review, 14*(2), 249-255. https://doi.org/10.3758/BF03194060

Ritter, S., Kulikowich, J., Lei, P., McGuire, C. L., & Morgan, P. (2007). What evidence matters? A randomized field trial of Cognitive Tutor Algebra I. In T. Hirashima, U. Hoppe, & S. Shwu-Ching Young (Eds.), *Supporting learning flow through integrative technologies* (pp. 12–20). Amsterdam, The Netherlands: IOS Press.

Schneider, M., & Siegler, R. S. (2010). Representations of the magnitudes of fractions. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1227-1238. https://doi.org/10.1037/a0018170

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.
https://doi.org/10.3102/0034654307313795

Siegler, R. S., & Lortie-Forgues, H. (2015). Conceptual knowledge of fraction arithmetic. *Journal of Educational Psychology, 107*(3), 909-918. https://doi.org/10.1037/edu0000025

Stafylidou, S., & Vosniadou, S. (2004). The development of students' understanding of the numerical value of fractions. *Learning and Instruction, 14*(5), 503-518. https://doi.org/10.1016/j.learninstruc.2004.06.015

Stark, R., Kopp, V., & Fischer, M. R. (2011). Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. *Learning and Instruction, 21*(1), 22-33. https://doi.org/10.1016/j.learninstruc.2009.10.001

Stewart, S., Thomas, M. O. J., & Hannah, J. (2005). Towards student instrumentation of computer-based algebra systems in university courses. *International Journal of Mathematical Education in Science and Technology, 36*(7), 741-749. https://doi.org/10.1080/00207390500271651

Stigler, J. W., Givvin, K. B., & Thompson, B. J. (2010). What community college developmental mathematics students understand about mathematics. *MathAMATYC Educator, 1*(3), 4-16.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59-89. https://doi.org/10.1207/s1532690xci0201_3

Tan, S. (2020). *Reduction procedures in fraction arithmetic* [Unpublished doctoral dissertation]. Carleton University, Ottawa, ON, Canada.

Tsamir, P., & Tirosh, D. (2005). In-service elementary mathematics teachers' views of errors in the classroom. *Focus on Learning Problems in Mathematics, 27*(3), 30-42.

Tsovaltzi, D., McLaren, B. M., Melis, E., & Meyer, A.-K. (2012). Erroneous examples: Effects on learning fractions in a web-based setting. *International Journal of Technology Enhanced Learning, 4*(3-4), 191-230. https://doi.org/10.1504/IJTEL.2012.051583

Vamvakoussi, X., Van Dooren, W., & Verschaffel, L. (2012). Naturally biased? In search for reaction time evidence for a natural number bias in adults. *The Journal of Mathematical Behavior, 31*(3), 344-355. https://doi.org/10.1016/j.jmathb.2012.02.001

Vamvakoussi, X., & Vosniadou, S. (2004). Understanding the structure of the set of rational numbers: A conceptual change approach. *Learning and Instruction, 14*(5), 453-467. https://doi.org/10.1016/j.learninstruc.2004.06.013

Van Dooren, W., Lehtinen, E., & Verschaffel, L. (2015). Unraveling the gap between natural and rational numbers. *Learning and Instruction, 37*, 1-4. https://doi.org/10.1016/j.learninstruc.2015.01.001

VanLehn, K. (1999). Rule-learning events in the acquisition of a complex skill: An evaluation of cascade. *Journal of the Learning Sciences, 8*(1), 71-125. https://doi.org/10.1207/s15327809jls0801_3

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197-221. https://doi.org/10.1080/00461520.2011.611369