International Journal of Innovation in Science and Mathematics Education, 32(3), 52-64, 2024

Analysing The Impact of Erroneous Examples on Third-Grade Students' Problem-Solving Proficiency

Rebeca García Olivares[a], Javier del Olmo-Muñoz[a*], José Antonio González-Calero[a] and David Arnau[b]

*Corresponding author: Javier del Olmo-Muñoz (Javier.Olmo@uclm.es)

Affiliation

[a] LabinTic. Lab for the integration of technology in the classroom. School of Education of Albacete, University of Castilla-La Mancha, Albacete, Spain.
[b] Department of Didactics of Mathematics, University of Valencia, Valencia, Spain.

ORCID

Rebeca García Olivares: https://orcid.org/0000-0003-4306-0372
Javier del Olmo-Muñoz: https://orcid.org/0000-0001-8754-0648
José Antonio González-Calero: https://orcid.org/0000-0003-0842-8151
David Arnau: https://orcid.org/0000-0002-2849-8248

Abstract

The potential benefits of introducing errors in problem solving has awakened interest in research into this understudied field. Here, we report the results of a quasi-experimental study with 85 third-grade students which examines whether erroneous examples might enhance students' problem-solving proficiency more effectively than worked ones. In the study, two conditions were established: a worked-example condition, where correct examples were presented before the students solved word problems, and an erroneous-example condition, where erroneous examples preceded word-problem solving. Our results demonstrate that post-test scores, after controlling the students' prior level, are significantly greater for the erroneous-example condition than the worked-example condition. Therefore, the erroneous-example approach seems to be more effective in developing a learner's problem-solving proficiency compared to a worked-example approach.

Keywords: worked examples, erroneous examples, word problems, problem solving, mathematics

Introduction

The use of examples as an instructional means is a common pedagogical technique and has been the subject of many studies in cognitive and educational psychology (e.g, Richey et al., 2019). Within this approach, which supports learning through the study of examples, there is a body of inconclusive findings among researchers between the use of worked (Renkl, 2017) and erroneous examples (Heemsoth & Heinze, 2014). While the efficacy of worked examples in mathematics education has been widely explored (Renkl, 2017; Richey et al., 2019; van Gog, Kester, & Paas, 2011), the potential benefits and drawbacks of erroneous examples, particularly in the domain of word problem solving, remain under-examined.

For the purpose of this research, "word problem-solving" is defined as the process of interpreting and solving "verbal descriptions of problem situations, presented within a scholastic setting, wherein one or more questions are raised the answer to which can be obtained by the application of mathematical operations to numerical data available in the problem statement or on numerical data derived from them" (Verschaffel, Schukajlow, Star, & Van Dooren, 2020, p. 1). It becomes evident that one cannot effectively address a given word problem unless they possess a firm grasp of the underlying concepts and a thorough understanding of the presented situation (Nicolas & Emata, 2018).

In this context, "problem-solving proficiency" refers to the ability to effectively and efficiently employ mathematical concepts and strategies to solve these types of problems. This proficiency is typically assessed based on the student's ability to understand, strategize, and execute steps to arrive at a correct solution (Karyotaki & Drigas, 2016). Importantly, problem-solving proficiency is not merely about obtaining a final correct answer, but it also involves the process and strategies employed to reach that solution. For complex, multi-step problems, each step may be evaluated separately to gain a more comprehensive understanding of a student's problem-solving skills. This approach provides a nuanced view of a student's capabilities, shedding light on their abilities to navigate problems of varying complexity, and offers valuable insights into their strengths and potential areas for improvement.

Errors are intrinsically linked to mathematical learning and, although their occurrence in a mathematics class may have positive effects on student achievement (see e.g., Heemsoth & Heinze, 2014), conventional instruction tends to avoid them, since there is a traditional view that committing errors would make them more salient and fix them in the memory and operational processes of the individual who makes them (Metcalfe, 2017). Some researchers oppose this conception of errors, as they consider them to be a valuable resource for learning and teaching. For instance, Borasi (1994) argued that a discussion concerning errors could benefit mathematics education as it encourages critical thinking about mathematical concepts, provides new problem-solving opportunities and motivates reflection and inquiry.

Given this background, the primary purpose of this study is to examine and compare the effectiveness of instruction based on erroneous examples with that of instruction based on worked examples, specifically in enhancing third-grade students' problem-solving proficiency. This investigation aims to contribute to the existing body of knowledge by providing insights into the potential benefits and limitations of using erroneous examples in mathematics education, especially in the context of word problem solving.

Literature review

Word problem solving in elementary education

In the realm of elementary education, word problem solving is a critical skill that intertwines various facets of mathematical understanding and real-world applications. Verschaffel and De Corte (1997) elaborate on the nuanced classifications of one-step arithmetic word problems. This classification highlights the often-problematic relationship between these problems and the real-world situations they are intended to represent. Moreover, the interaction between different types of knowledge and skills in solving these problems is complex and multifaceted, involving more than just mathematical calculation.

Identifying common and specific difficulties in problem solving is essential, as students frequently encounter obstacles such as deciphering key words, isolating necessary information and avoiding the inclination to guess answers without adequate reasoning (Phonapichat, Wongwanich, & Sujiva, 2014). This body of research, including the ideas of Fuson (1992) and those developed by Verschaffel and De Corte (1997), sheds light on the challenges faced by learners and also serves as a basis for developing teaching strategies that promote a more authentic and complete understanding of mathematics through the context of word problems. This approach is crucial for fostering authentic problem-solving skills in younger learners, enabling them to navigate and make sense of the mathematical aspects of their everyday lives.

Worked examples

Worked examples are defined as step-by-step demonstrations of how to solve a problem. Learning from worked examples means that correct examples are given to students before they try to solve problems on their own, which is a very effective method for the initial acquisition of cognitive skills (e.g., Große & Renkl, 2007; Hefter, Vom Hofe, & Berthold, 2022). On the one hand, a worked example usually consists of the statement of a problem, the problem-solving steps and the final solution itself. On the other hand, erroneous examples are worked examples that intentionally include one or more errors which the students are asked to detect, explain and/or correct, before moving on to a similar word problem to acquire a concept and develop metacognitive skills (Barbieri, Booth, Begolli, & McCann, 2021; Jaeger, Marzano, & Shipley, 2020).

In the field of problem solving, instruction based on worked examples has been the object of study of different researchers (for an overview see Barbieri et al., 2021). This approach seems to reduce problem-solving errors and learning time, as well as being helpful for the acquisition of initial cognitive skills (Renkl, 2017), and novel solving strategies (Vollman, 2021). However, recent studies show that worked examples seem to be a superficial form of instruction unless they are accompanied by self-explanation prompts, which facilitate a deeper learning process (e.g., Jaeger et al., 2020). Moreover, Vollman (2021) highlights that they are beneficial for learning only if corrective feedback is included.

Erroneous examples

Regarding the benefits of using incorrect examples, Heemsoth and Heinze (2014) found that reflecting on errors, and especially on erroneous examples, has a positive effect on knowledge and skill acquisition. Some researchers posit that learning from erroneous examples is more effective than learning from worked examples because errors create an impasse and are a stimulus to produce more self-explanations and, as a result, learning outcomes are better (e.g., Yang et al., 2016). Moreover, what differentiates erroneous examples from worked examples is that the former may encourage students to see errors as something which is not necessarily negative, which may help them to benefit from them and develop a constructive attitude towards errors (Yang et al., 2016).

Based on different researchers, Barbieri and Booth (2020) affirm that erroneous examples alone, or in combination with correct examples, might be considered beneficial for mathematical learning. Metcalfe (2017) clarify that, seeing one’s own errors followed by corrective feedback seems to be advantageous for learning; however, perceiving other people’s errors could impair the learning of a person who had not made those errors.

To help students in this process, and reduce the demands on prior knowledge and cognitive load from erroneous examples, there are different strategies such as highlighting errors (Fitzsimmons, Morehead, Thompson, Buerke, & Dunlosky, 2021; Große & Renkl, 2007), promoting self-explanations (Atkinson , Renkl, & Merril, 2003) and using scaffold comparison of worked and erroneous examples (Durkin and Rittle-Johnson, 2012). These strategies can make the use of erroneous examples beneficial to all learners (Adams et al., 2014; McLaren, van Gog, Ganoe, Karabinos & Yaron, 2012).

Previous studies

Although erroneous examples can be beneficial when integrated into mathematics education, research has shown variable results across different age groups and study contexts. For example, Adams et al. (2014) found that erroneous examples made students from middle school (11-13 years old) aware of their lack of knowledge, as they realized that their previous conceptions were incorrect and they felt motivated to accommodate the new information, which improved and strengthened their knowledge concerning decimals. Accordingly, Kapur (2014) stated that 14-15-year-old students, who solved ill-structured problems on the math concept of standard deviation, outperformed their counterparts who only solved well-structured problems. Nevertheless, Richey et al. (2019) analysed whether reflection on erroneous examples led sixth-grade students (with a mean age of 11.75 years) to greater confusion and frustration compared to solving practice problems, or if this process reinforced their understanding of decimals. The results did not show positive learning outcomes from erroneous examples because of the great confusion and frustration experienced by the learners.

Concerning the combined use of worked and erroneous examples as an instructional tool, Booth et al. (2013) carried out a study with eighth-grade students (demographic data not available, but US 8th graders are typically 13-14 years old) to test the impact of the different types of examples in algebraic instruction. The results suggest that the students who were in the combined condition (correct and erroneous examples) performed slightly better than those who worked only with erroneous examples, and significantly better than those who worked only with correct examples. In a similar study with middle-school students (age not provided, but they were also US 8th graders), Barbieri et al. (2021) also found that combining worked and erroneous examples improves procedural and conceptual algebraic knowledge more than regular problem solving.

On balance, the results from the aforementioned studies suggest that there is no complete agreement on the eventual benefits and drawbacks of these approaches. While studies such as Adams et al. (2014) and Kapur (2014) highlight the potential benefits of erroneous examples in improving students' understanding and problem-solving skills, others such as Richey et al. (2019) suggest potential drawbacks due to increased confusion and frustration. Hence, more research needs to be carried out in this field. It is also necessary to determine whether worked or erroneous examples may be more effective when combined with guided practice, such as corrective feedback, promoting self-explanations or comparing worked and erroneous examples. The debate over how –and whether– erroneous examples are valuable in the learning process is still ongoing, given the mixed evidence concerning their benefits (Fitzsimmons et al., 2021). The review by Darabi, Arrington, and Sayilir (2018) tried to shed light on the effectiveness of the failure-based instructional strategy, selecting, after the screening process, only twelve experimental interventions, of which none were conducted at elementary school grade levels. This scarcity of research led the authors to highlight the lack of empirical studies dealing with productive failure, which prevents a better understanding of its impact as an instructional strategy. Beyond that, few studies have explored the impact of these approaches on arithmetical word-problem solving.

Purpose of the study and research question

In view of the significant gaps and unresolved issues revealed by the literature, our study aims to contribute to the field by investigating the impact of error-based learning in elementary education. This focus on a younger student population and in the specific context of word problem solving represents a novel approach, potentially offering insights into the effectiveness of these instructional strategies at an earlier stage of mathematics education. In doing so, our research aims to enrich the existing body of knowledge and provide practical guidance for educators in optimising instructional designs to support young learners' mathematical development. Therefore, the present study seeks to give answer to the following research question: "Does instruction based on erroneous examples lead to higher scores on third-grade students' word problem-solving proficiency compared to instruction based on worked examples?"

Method

Participants

The participants were 85 third graders, whose ages ranged from 8 to 9 years old. They were in four different classes in two state schools, with a middle-class socioeconomic level in a city in Spain. Each class was randomly assigned to one of the two conditions; two classes were assigned to the control group, or worked-example condition (WEC), with 45 students (18 girls and 27 boys), and the other two classes were assigned to the experimental group, or erroneous-example condition (EEC), with 40 students (19 girls and 21 boys). All participants provided their consent to participate prior to the initiation of the study.

Design

The study is based on a quasi-experimental design which consisted of three phases for both conditions. In the first phase, a pre-test was given to the students to assess their level and prior knowledge in problem solving in order to: i) evaluate whether both groups are comparable; ii) design the following instruction sequence according to their level of performance and understanding; and iii) be used in combination with the subsequent post-test to assess the learning gains after the experiment. The pre-test lasted 40 minutes.

Afterwards, a second phase took place. In this stage, students from both groups individually completed an instruction sequence based on paper-and-pencil problem solving over a six-week period where we alternated the presentation of examples (worked or erroneous) with problems to be solved. To this effect, six instruction sessions were designed with different types of word problems, i.e., with additive steps (combination, change, and comparison) and with multiplicative steps (in particular, isomorphism of measurements). Unlike the first phase, the control group and the experimental group encountered different conditions. While for the WEC worked examples were presented before asking the students to solve word problems on their own, students in the EEC worked with erroneous examples before solving problems. All the sessions of this phase were conducted once a week and lasted 45 minutes for both conditions. After each session, students from both groups received corrective feedback to check their answers.

Finally, in the third phase, the students completed a post-test which consisted of isomorphic problems to those of the pre-test, i.e., with the same mathematical structure, but with different statements.

Materials

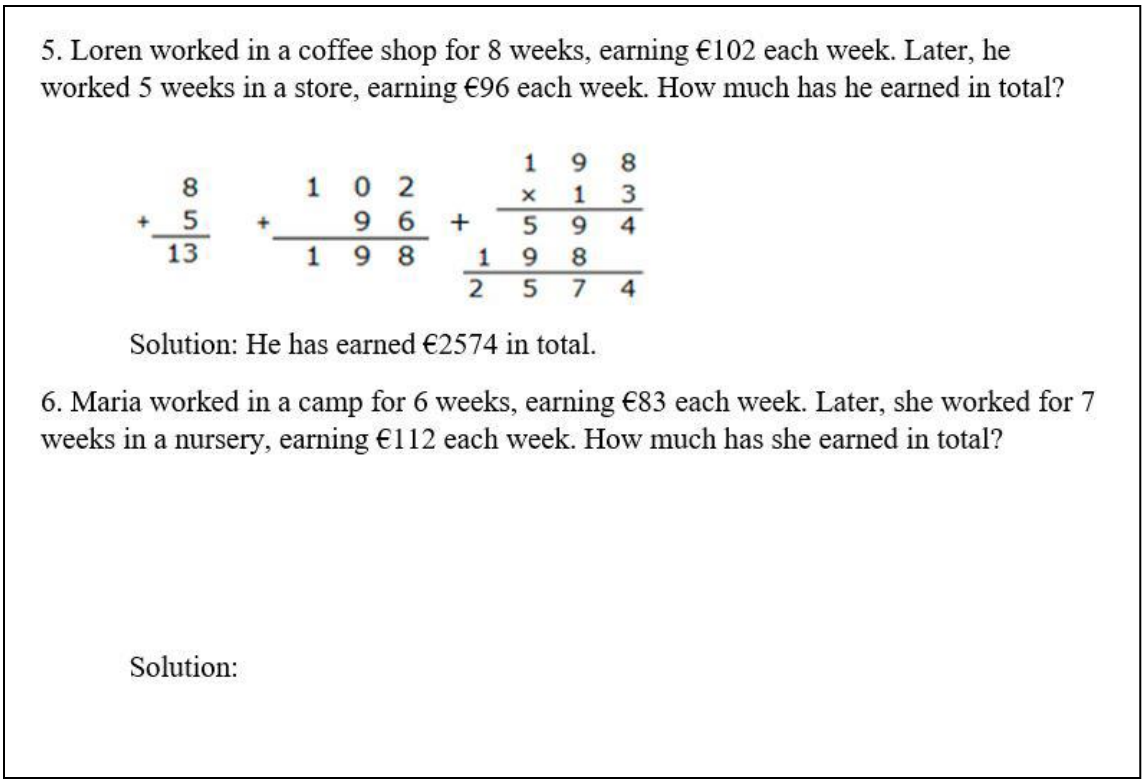
Pre-test and post-test

In order to assess the effectiveness of the intervention, a pre-test and a post-test were conducted. Before the instructional phase started, all the participants had to complete the same pre-test, which consisted of 14 word problems. The difficulty of the selected problems was adjusted to the curricular requirements for this age. In fact, they were selected from third grade textbooks. The tasks were one-step, two-step and three-step word problems with additive steps (combination, change, and comparison) or multiplicative steps (in particular, isomorphism of measurements). Most of the problems were hybrid in nature, combining two or more categories. With regard to the scoring of the answers, each problem was worth 1 point, with 14 points being the maximum score in the test. When the problems included multiple steps, each step was scored separately. For example, if a student in a three-step problem failed in the last step, then two out of three points were assigned to that problem.

After the instructional phase, which lasted six weeks, a post-test with 14 problems was presented to the students. Following the criteria of the pre-test, we used the same test for both conditions. The problems had the same mathematical structures as in the pre-test, but some surface elements were modified. The aim was to compare the students' answers before and after the instruction, and to check if there were improvements, or not, in relation to the results obtained in the pre-test. As in the pre-test, the total score was 14 points, 1 point for each problem.

Worksheets

For the instructional sessions, both conditions, EEC and WEC completed six worksheets based on problem solving over six weeks, one session per week. The instruction in the EEC included five pairs of problems, where the first problem in each pair was an example for the second problem. Students in this condition were informed that the examples may or may not contain errors. The reason for including worked examples in this condition was merely as a distractor for the students so they would not assume that all the problems they faced contained errors. Firstly, the students had to find and fix the erroneous examples (first problem), and afterwards they had to solve the corresponding paired problem (second problem). Figure 1 shows a pair of problems from the EEC.


5. Loren worked in a coffee shop for 8 weeks, earning €102 each week. Later, he worked 5 weeks in a store, earning €96 each week. How much has he earned in total?

8 + 5 = 13; 102 + 96 = 198; 198 × 13: 594 + 1980 = 2574

Solution: He has earned €2574 in total.

6. Maria worked in a camp for 6 weeks, earning €83 each week. Later, she worked for 7 weeks in a nursery, earning €112 each week. How much has she earned in total?

Solution:

Figure 1. Pair of problems from the EEC

The tests in the WEC included five correct examples and five problems. As in the EEC, the problems were paired, and the first problem of the pair was an example for the second one. Students in this condition were informed that all the examples were correctly solved and they were asked to solve the paired problems.

Feedback – sheets

After each session (except for the pre-test and post-test), feedback was provided to both groups in the following session. They were given a solution-sheet that included all the problems from the previous session. The solution-sheet presented the complete way to solve each problem, including all the steps and the final answer. In addition, they could compare the solution with their attempt as they also received their worksheets from the previous lesson. The students were asked to assess their responses in order to identify possible errors. In the case of EEC, the sheet included the answers to the problems that the students had to solve, as well as the answers to the erroneous examples. Errors were highlighted within the erroneous examples and the students were prompted to explain, in these examples, ‘what they had learned’. It is important to highlight at this point that even though the demands were different, both experimental conditions devoted the same amount of time to the feedback sheets.

The design of these materials was based on previous studies which dealt with erroneous and worked examples (Atkinson, Derry, & Renkl, 2000; Heemsoth & Heinze, 2014; McLaren et al., 2016). The type of problems, the contents and the statements of all the problems used in the study were taken from third-grade textbooks. Following Phonapichat et al. (2014), the steps in which students usually face more difficulties were considered to design the erroneous examples.

Procedure

The students worked alone and received no help throughout the completion of the instruction. One researcher was in the classroom and avoided any intervention that could influence the students' performance. The only interventions were devoted to answering questions about what they were expected to do. In addition, the students were encouraged to keep working even if they claimed that the tasks were too demanding, or that they did not know how to solve a problem. The teachers did not participate during the interventions; therefore, no instruction was needed for them, as one researcher oversaw the whole development of the sessions.

During the first session, students from both conditions completed the pre-test in a 40-minute session. In this session, the students were given a time to solve the problems. The researcher informed the students about the specific time they had to solve each problem. The students were notified when to start each problem, and they were also warned when 10 seconds were left and when the time expired. Once the time expired, all the students moved onto the next problem. In this way, all the learners worked at the same time and under the same conditions. Similarly, if they finished before the time expired, they could not move onto the next problem. In order to speed up their performance, the students were allowed to write and indicate the operations they needed to solve the problem without doing the calculations. This was only suggested for the cases in which they realized that they did not have enough time.

The experimental phase took place during the following six sessions. All the problems were the same for both conditions (the same structure and cover stories), but the difference resided in the errors presented in the EEC. At the beginning of these sessions (except for the first one), the students were asked to check their solutions from the previous session with a tick or a cross, comparing them with the feedback sheet. Afterwards, the next worksheet was given to the students. These sessions lasted 45 minutes each in both conditions.

Finally, in the last session, the post-test took place. As with the pre-test, this included 14 problems to be solved in 40 minutes. Once the students had completed the tests, we proceeded to the data coding to carry out the analysis of the results.

Data analysis

To address the research question at hand, we utilized an analysis of covariance (ANCOVA), a statistical method that allows for the examination of differences in post-test scores between experimental and control groups while accounting for pre-test measures (Ary, Jacobs, Razavieh, Sorensen, & Walker, 2014). Specifically, an ANCOVA was performed on the post-test scores, with the pre-test scores serving as a covariate. Both pre-test and post-test scores were standardized to a scale ranging from 0 to 10. No exclusion of any experimental subject from the data analysis occurred. Partial eta-squared values were utilized as a measure of effect size (Field, Miles, & Field, 2012). All statistical analyses were carried out using the R software (R Core Team, 2020).

Results

In order to analyse the results, Table 1 summarizes descriptive statistics for the two conditions, including the mean, adjusted means, standard deviation and sample sizes.

Table 1. Summary of pre-test and post-test results.

Group		Pre-test		Post-test		Adjusted Post-test	
	n	M	SD	M	SD	M	SD
Worked examples (WEC)	45	5.65	1.07	5.72	1.07	5.55	1.56
Erroneous examples (EEC)	40	5.14	0.87	6.22	0.87	6.42	1.56

Due to the study design, the participants were nested in classes. To estimate the clustering effect (the variability in achievement between classes), the intraclass correlation coefficient (ICC=.0035) and the design effect (Deff=1.07) were calculated. The ICC < 0.1 and Deff < 2.0, so, according to Heck and Thomas (2020), multilevel analysis is considered unnecessary. As a consequence, to evaluate differences on the post-test controlling the students' prior level, a one-way ANCOVA on the post-test scores with the pre-test scores as a covariate was conducted. A preliminary analysis, aimed to assess the homogeneity of regression assumption, indicated that the relationship between pre-test scores and post-test scores did not differ significantly across conditions, $F(1, 81) \approx 0.00$, $p = 0.996$. Concerning the covariate pre-test scores, it was found that the post-test scores were significantly related to the students' level before starting the instruction, $F(1, 82) = 118.70$, $p < .0001$, $r = .75$. There was also a statistically significant effect of the type of instruction after controlling the students' prior level in favour of the erroneous-examples condition ($F(1, 82) = 6.59$, $p = .0120$, partial eta-squared = .07). According to Cohen (1969), this effect can be considered as medium-sized.

Discussion and conclusion

The aim of this study was to examine the effectiveness of mathematics instruction based on erroneous examples compared to worked examples. The results reveal that those students who worked with erroneous examples achieved a statistically significant greater proficiency in word problem-solving scores —after controlling the students' prior level— than those who worked exclusively with worked examples. A possible explanation for these results could be related to the fact that the WEC received instruction with correct examples before they solved their corresponding paired problems, so students could replicate the same operations without needing to make sense of the problems. Thus, participants from the WEC were not challenged as much as those from the EEC to solve the problems, because they could just look for the operations that would enable them to pass from the current problem state to the final goal state (Sweller, 1988). Meanwhile, the students from the EEC needed to identify the incorrect steps, make sense of the errors and look up the correct solutions, which could provide a challenge to the existing mental models of the learner (Darabi et al., 2018), and mean reorganizing and integrating the new material with prior knowledge (Adams et al., 2014).

The fact that the EEC significantly outperformed the WEC in our study reveals that, although students in the experimental group may have felt confusion or frustration because they faced more challenging situations than their counterparts, these affective experiences did not prevent learning, contrary to what Richey et al. (2019) suggested. This finding is aligned with previous studies that revealed positive outcomes from the use of erroneous examples in learning (Adams et al., 2014; Heemsoth & Heinze, 2016; Kapur, 2014).

The current study was specifically oriented to promote word problem-solving abilities. This domain was selected due to the multiple difficulties, errors and anxiety that elementary students usually experience in this field, due to lack of knowledge and skills (Tambychik & Meerah, 2010). According to VanLehn (1988), the use of errors as a pedagogic tool may be especially beneficial in case of a lack of knowledge, skills, uncertainty or because of the difficulties of the tasks. Our results show the suitability of instruction based on errors to tackle students' difficulties in problem solving and to improve their problem-solving skills.

We can assume, as Durkin and Rittle-Johnson (2012) or Adams et al. (2014) reported, that participants in the EEC obtained better results because erroneous examples helped them be aware of their weaknesses in knowledge and their errors in previous conceptions. Consequently, they felt motivated to accommodate the new information to improve their existing mental models and to strengthen their knowledge. Likewise, the use of strategies such as corrective feedback to reduce the cognitive load could have improved the students' performance, as Metcalfe (2017) suggested. Also, self-explanations were prompted in the feedback, as proposed in several studies (e.g., Barbieri & Booth, 2020; Booth et al., 2013; Yang et al., 2016). In our study, these strategies were not implemented immediately after each problem was completed, but rather at the beginning of the next session. Still, these strategies may have helped the students to develop (or enhance) their critical thinking, error detection and error awareness skills (Tsovaltzi et al., 2010).

Our findings underscore the transformative potential of integrating erroneous examples into mathematics teaching, not only as a novel instructional tactic, but as a fundamental shift toward the acceptance of errors as fundamental learning moments. Consequently, our study extends the paradigm of problem-solving research by empirically validating the cognitive benefits of this approach, thereby advocating for broader pedagogical adoption.

We acknowledge that this study has some limitations that lead us to recommend future research. Firstly, although in this study we applied some strategies to reduce the cognitive load, we did not measure the students' cognitive load when facing erroneous examples, something which could be a useful variable to take into account in further research. In addition, a component of the study design that could explain the results is the reflection prompted in the experimental group about what they had learned after the feedback. This leads us to suggest future experimental research in which students in both conditions work with erroneous examples and only those in one condition are asked for self-explanations, so that eventual differences linked to metacognition in error-based learning can be evaluated. Finally, it should be noted that a non-standardised instrument was used in this study, something that is often the case with instruments consisting of word problems. In this respect, it has proved difficult in the past to make progress in the standardisation of an instrument which, if it were to be used, would probably be lengthy and hardly applicable. Nevertheless, it would be interesting for the research community to make efforts aimed at designing standardised instruments to strengthen the reliability of future research.

Word problems have long been a significant element of school mathematics across the world, and they have a prominent place in the curriculum since they provide practice for everyday situations in which students can apply the things they have previously learned (Verschaffel et al., 2020). Beyond the aforementioned limitations, the contribution of this study to mathematical research and education relies on how erroneous examples, paired with corrective feedback, look promising as an effective method for promoting student understanding and performance in arithmetical word problem solving, making it an approach worthy of consideration by educators and policymakers.

Acknowledgments

This work was supported by the Spanish Ministry of Science, Innovation and Universities through project PGC2018-096463-B-I00 financed by MCIN/AEI/10.13039/501100011033 and FEDER Una manera de hacer Europa; by Valencian Regional Government (Spain) under project AICO/2021/019; by the Castilla-La Mancha Regional Administration under Project SBPLY/19/180501/000278; and by the Ministry of Education, Culture and Sports of Spain under Grant FPU19/03857.

Declaration of interest statement

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics statement

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the University of Castilla-La Mancha (No. CEIS-632764-N671).

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

Ary, D., Jacobs, L. C., Razavieh, A., Sorensen, C. K., & Walker, D. A. (2014). *Introduction to research in education* (9th ed.). Cengage Learning.

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

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology*, *95*(4), 774–783. https://doi.org/10.1037/0022-0663.95.4.774

Barbieri, C. A., & Booth, J. L. (2020). Mistakes on display: Incorrect examples refine equation solving and algebraic feature knowledge. *Applied Cognitive Psychology*, *34*(4), 862–878. https://doi.org/10.1002/acp.3663

Barbieri, C. A., Booth, J. L., Begolli, K. N., & McCann, N. (2021). The effect of worked examples on student learning and error anticipation in algebra. *Instructional Science*, *49*(4), 419–439. https://doi.org/10.1007/s11251-021-09545-6

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

Borasi, R. (1994). Capitalizing on errors as “springboards for inquiry”: A teaching experiment. *Journal for Research in Mathematics Education*, *25*(2), 166. https://doi.org/10.2307/749507

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. Academic Press.

Darabi, A., Arrington, T. L., & Sayilir, E. (2018). Learning from failure: a meta-analysis of the empirical studies. *Educational Technology Research and Development, 66*(5), 1101–1118. https://doi.org/10.1007/s11423-018-9579-9

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

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications Ltd.

Fitzsimmons, C. J., Morehead, K., Thompson, C. A., Buerke, M., & Dunlosky, J. (2023). Can feedback, correct,

and incorrect worked examples improve numerical magnitude estimation precision? *The Journal of Experimental Education, 91*(1), 20–45. https://doi.org/10.1080/00220973.2021.1891009
Fuson, K. C. (1992). Research on Whole Number Addition and Subtraction. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning: A Project of the National Council of Teachers of Mathematics* (pp. 243–275). Macmillan Publishing Co, Inc.
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
Heck, R. H., & Thomas, S. L. (2020). An introduction to multilevel modeling techniques. In *An Introduction to Multilevel Modeling Techniques*. Routledge. https://doi.org/10.4324/9780429060274
Heemsoth, T., & Heinze, A. (2014). The impact of incorrect examples on learning fractions: A field experiment with 6th grade students. *Instructional Science*, *42*(4), 639–657. https://doi.org/10.1007/s11251-013-9302-5
Heemsoth, T., & Heinze, A. (2016). Secondary school students learning from reflections on the rationale behind self-made errors: A field experiment. *The Journal of Experimental Education*, *84*(1), 98–118. https://doi.org/10.1080/00220973.2014.963215
Hefter, M. H., Vom Hofe, R., & Berthold, K. (2022). Effects of a digital math training intervention on self-efficacy: Can clipart explainers support learners? *International Journal of Innovation in Science and Mathematics Education, 30*(4). https://doi.org/10.30722/IJISME.30.04.003
Jaeger, A. J., Marzano, J. A., & Shipley, T. F. (2020). When seeing what's wrong makes you right: The effect of erroneous examples on 3D diagram learning. *Applied Cognitive Psychology*, *34*(4), 844–861. https://doi.org/10.1002/acp.3671
Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, *38*(5), 1008–1022. https://doi.org/10.1111/cogs.12107
Karyotaki, M., & Drigas, A. (2016). Latest trends in problem solving assessment. *International Journal of Recent Contributions from Engineering, Science & IT (IJES)*, *4*(2), 4. https://doi.org/10.3991/ijes.v4i2.5800
McLaren, B. M., Adams, D., Durkin, K., Goguadze, G., Mayer, R. E., Rittle-Johnson, B., Sosnovsky, S., Isotani, S., & 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.), *EC-TEL 2012: 21st Century Learning for 21st Century Skills* (pp. 222–235). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33263-0_18
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
Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, *68*, 465–489. https://doi.org/10.1146/annurev-psych-010416-044022
Nicolas, C. A. T., & Emata, C. Y. (2018). An integrative approach through reading comprehension to enhance problem solving skills of Grade 7 mathematics students. *International Journal of Innovation in Science and Mathematics Education, 26*(3), 40–64.
Phonapichat, P., Wongwanich, S., & Sujiva, S. (2014). An analysis of elementary school students' difficulties in mathematical problem solving. *Procedia - Social and Behavioral Sciences*, *116*(2012), 3169–3174. https://doi.org/10.1016/j.sbspro.2014.01.728
R Core Team. (2020). *R: A language and environment for statistical computing*.
Renkl, A. (2017). Learning from worked-examples in mathematics: students relate procedures to principles. *ZDM - Mathematics Education*, *49*(4), 571–584. https://doi.org/10.1007/s11858-017-0859-3
Richey, J. E., Andres-Bray, J. M. L., Mogessie, M., Scruggs, R., Andres, J. M. A. L., Star, J. R., Baker, R. S., & 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
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7
Tambychik, T., & Meerah, T. S. M. (2010). Students' difficulties in mathematics problem-solving: What do they say? *Procedia - Social and Behavioral Sciences*, *8*(5), 142–151. https://doi.org/10.1016/j.sbspro.2010.12.020
Tsovaltzi, D., Melis, E., McLaren, B. M., Meyer, A.-K., Dietrich, M., & Goguadze, G. (2010). Learning from Erroneous Examples: When and How Do Students Benefit from Them? In M. Wolpers, P. A. Kirschner, M. Scheffel, S. Lindstaedt, & V. Dimitrova (Eds.), *Lecture Notes in Computer Science* (Vol. 6383, pp. 357–373). Springer-Verlag. https://doi.org/10.1007/978-3-642-16020-2_24
Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-

example pairs on novices' learning. *Contemporary Educational Psychology*, *36*(3), 212–218. https://doi.org/10.1016/j.cedpsych.2010.10.004
Vanlehn, K. (1988). *Toward a Theory of Impasse-Driven Learning*.
Verschaffel, L., & De Corte, E. (1997). Word Problems: A vehicle for promoting authentic mathematical understanding and problem solving in the primary school? In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics* (pp. 69–97). Psychology Press.
Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: a survey. *ZDM*, *52*(1), 1–16. https://doi.org/10.1007/s11858-020-01130-4
Vollman, E. P. (2021). *Learning beyond accuracy: Evidence for worked examples as a support for students' proportional reasoning gains* [Doctoral dissertation, The University of Chicago]. https://doi.org/10.6082/uchicago.2991
Yang, Z., Wang, M., Cheng, H. N. H., Liu, S., Liu, L., & Chan, T.-W. (2016). The effects of learning from correct and erroneous examples in individual and collaborative settings. *The Asia-Pacific Education Researcher*, *25*(2), 219–227. https://doi.org/10.1007/s40299-015-0253-2