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 Wolpers, M., Kirschner P.A., Scheffel, M., Lindstaedt, S. & Dimitrova, V. (eds), *Proceedings of the 5th European Conference on Technology Enhanced Learning, Sustaining TEL: From Innovation to Learning and Practice* (EC-TEL 2010), LNCS 6383, September/October, Barcelona, Spain. (pp. 357-373). Springer-Verlag Berlin Heidelberg.

Learning from Erroneous Examples: When and How do Students Benefit from them?

Dimitra Tsovaltzi, Erica Melis, Bruce M. McLaren, Ann-Kristin Meyer, Michael Dietrich, George Goguadze

DFKI GmbH, German Research Centre for Artificial Intelligence, Stuhlsatzenhausweg 3 (Building D3 2) D-66123 Saarbrücken Germany
Dimitra.Tsovaltzi@dfki.de

Abstract. We investigate whether *erroneous examples* in the domain of fractions can help students learn from common errors of other students presented in a computer-based system. Presenting the errors of others could spare students the embarrassment and demotivation of confronting their own errors. We conducted lab and school studies with students of different grade levels to measure the effects of learning with erroneous examples. We report results that compare the learning gains of three conditions: a control condition, an experimental condition in which students were presented with erroneous examples without help, and an experimental condition in which students were provided with additional error detection and correction help. Our results indicate significant metacognitive learning gains of erroneous examples for lower-grade students, as well as cognitive and conceptual learning gains for higher-grade students when additional help is provided with the erroneous examples, but not for middle-grade students.

Keywords. Erroneous examples, empirical studies, fractions misconceptions, adaptive learning, metacognition

1 Introduction

Erroneous examples are worked solutions that include one or more errors that the student is asked to detect and/or correct. Technology opens new possibilities of instruction with erroneous examples. For example, erroneous examples can be presented in a variety of ways, with different kinds of feedback, diverse tutorial strategies and sequencing of learning material. Adapting such features to the needs of individual students can contribute to better learning. A domain that seems to need alternative ways of instruction is fractions. There is evidence that students', but also preservice teachers' understanding of fractions is frequently underdeveloped [1] and misconceptions lead to poor performance in fraction problems [2].

However, erroneous examples have been scarcely investigated or used in mathematics teaching, particularly not in the context of technology enhanced learning (TEL). On the contrary, *correct* worked examples have been shown by studies in science learning to benefit students in learning mathematics and science problem solving [3], [4], [5], and [6]. A reason behind the reluctance to use erroneous examples comes from the fact that many mathematics teachers are sceptical about discussing errors in the classroom [7]. Teachers are cautious of exposing students to errors in fear that it could lead to incorrect solutions being assimilated by students, and often believe any discussion of errors should be avoided. As a consequence, the questions remain open on (i) if erroneous examples are beneficial for learning and (ii) how they should be presented.

The little theoretical work and empirical research on erroneous examples in mathematics has provided some evidence that studying errors can promote student learning of mathematics [8], [9], [10], [11], [12]. For example, Borasi argues that mathematics education could benefit from the discussion of errors by encouraging critical thinking about mathematical concepts, by providing new problem solving opportunities, and by motivating reflection and inquiry. Moreover, the highly-publicised TIMSS studies [13] showed that Japanese math students outperformed their counterparts in most of the western world. The key curriculum difference cited was that Japanese educators present and discuss incorrect solutions and ask students to locate and correct errors.

Siegler and colleagues conducted a controlled comparison of correct and incorrect examples [14], [15]. They investigated whether self-explaining both correct and incorrect examples is more beneficial than self-explaining correct examples only. They found that when students studied and self-explained *both* correct and incorrect examples they learned better. They hypothesised that self-explanation of correct and erroneous examples strengthened correct strategies and weakened incorrect problem solving strategies, respectively. Grosse and Renkl studied whether explaining both correct and incorrect examples makes a difference to learning and whether highlighting errors helps students learn from those errors [16]. Their empirical studies (in which no help or feedback was provided) showed some learning benefit of erroneous examples, but unlike the less ambiguous Siegler et al results, the benefit they uncovered was only for learners with strong prior knowledge and for far transfer. Research in other domains, such as medical education, has demonstrated the benefits of erroneous examples in combination with elaborate feedback in the acquisition of problem-solving schemata [17].

2 Theoretical Background and Design

We take the earlier controlled studies further by investigating erroneous examples in the context of Technology-Enhanced learning with ActiveMath, a web-based system for mathematics [18]. Our ultimate goal is to develop micro and macroadaptation for the presentation of erroneous examples for individual students since the benefit of erroneous examples may depend on individual skills, grade level, etc. Microadaptation refers to the teaching strategy, or step-by-step feedback, inside an erroneous example based on the student's performance. Macroadaptation refers to the choice of task for the student, as well as the frequency and sequence of the presentation of erroneous examples. In this paper, we focus on the empirical results that inform our work on the adaptive technology. In contrast to the Siegler studies, we are interested in the interaction of students' study of erroneous examples and the situational and learner characteristics. Extending the work of Grosse and Renkl, we investigate erroneous examples with and adaptive error-detection and error-correction help in the context of TEL. This novel design relies on the intelligent technology of ActiveMath. Our primary rationale for including error detection and correction help in the empirical studies is that students are not accustomed to working with and learning from erroneous examples in mathematics. Thus, they may not have the required skills to review, analyse, and reflect upon such examples. Taking this strand and providing additional help, we also extend the work of Kopp and colleagues in medical education [17] for mathematics education.

We believe that learning from errors can help students develop (or enhance) their critical thinking, error detection, and error awareness skills. This conjecture stems from various foundational elements, most of which are supported, at least partially, by past research. To begin with, a student can learn important error detection and correction skills by studying erroneous examples, something that is not possible with correct examples and difficult with unsupported problem solving. Moreover, erroneous examples may weaken students' incorrect strategies, as opposed to worked examples that strengthen correct strategies [14]. Additionally, similar to worked examples, erroneous examples do not ask a student to perform as in problem solving, but provide a worked-out solution that includes one or more errors. They could, in effect, reduce extraneous cognitive load in comparison to problem solving [19] while increasing germane cognitive load in the sense of creating cognitive conflict situations. Erroneous examples may, further, guide the learner toward a learning orientation rather than a performance orientation, especially in combination with help that increases student's understanding and, hence, their involvement in the learning process [14].

With regard to the possible drawbacks of erroneous examples, we hypothesise that a student is less likely to exhibit the feared 'conditioned response' of behaviourist theory [20] when studying errors that the student has not made him/herself and thus has not (necessarily) internalised. On the contrary, students may benefit from erroneous examples when they encounter them at the right time and in the right way. For example, rewarding the student for error detection may lead to annotations of these errors in memory such that they will be avoided in subsequent retrieval. At the same time, a student is unlikely to be demotivated by studying common errors in the domain, made by others, as when emphasising errors the student has made him/herself. Some of our preliminary work has already demonstrated the motivational potential of erroneous examples [21]. In the course of our investigation of erroneous examples, we aim to answer the following research questions:

When:

1. Do advanced students, in terms of grade level, gain more from erroneous examples than less advanced students?

How:

2. Can students' cognitive skills, conceptual understanding, and transfer abilities improve through the study of erroneous examples?
3. Does work with erroneous examples help to improve the metacognitive competencies of error detection, error awareness and error correction?

Based on this, our hypothesis is that presenting erroneous examples to students will give them the opportunity to find and react to errors in a way that will lead to deeper, more conceptual learning and better error-detection (i.e., metacogntive) skills. This, in turn, will help them improve their cognitive skills and will promote transfer. We further hypothesise that the effect of erroneous examples depends on whether students are supported in finding and correcting the error and on when and how they are introduced to the students. We present our design of erroneous examples and the studies we conducted to address these questions in the context of TEL.

2.1 Design of Erroneous Examples in ActiveMath

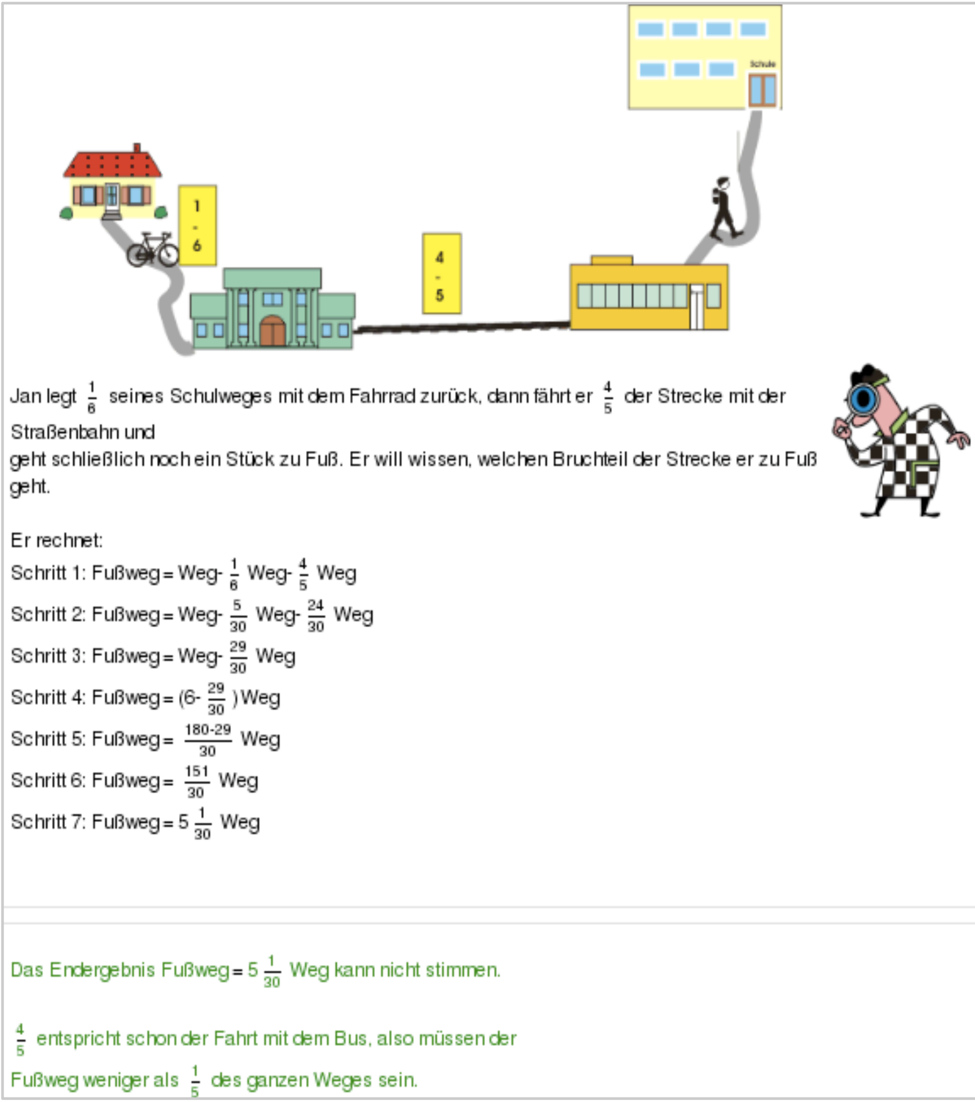

Fig. 1 An Online Erroneous Example and the error-awareness and detection (EAD) feedback (bottom).

Erroneous examples, as well as standard fraction, exercises are online in ActiveMath. Their presentation is through a tutorial strategy, which defines when and how to provide help, signal correct and wrong answers, give answers away, show previous steps of the students, etc. Erroneous examples in ActiveMath include instances of typical errors students make, which address standard problems students face with rule-application, or errors that target common misconceptions and deal with more fundamental conceptual understanding in fractions.

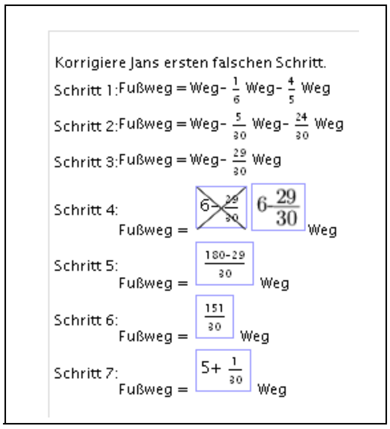

Fig. 2. Error-correction Phase

Erroneous examples consist of two phases: error detection and error correction. Fig. 1 displays the task presented in the first phase. Its translation is: "Jan rides his bike for 1/6 of the path to school, then drives with the tram 4/5 of the path and finally walks the rest of the path. He wants to know what fraction of the path he walks. He calculates:…" The steps of the erroneous solutions are presented as choices in a multiple-choice question (MCQ) and students have to select the erroneous step. After completing this phase, the student is prompted to correct the error, as shown in Fig. 2 ("Correct Jan's first wrong step").

2.2 Feedback Design

"Why is the 4th step wrong?"

- because the length is 6*path (due to 1/6)
- because the entire length is 5*6=30*path
- because the entire length is 1*path
- I don't know.

"How can you represent the whole path?"

- With 100
- With 1*path
- I don't know.

Fig. 3. "Why" and "How" MCQs

Based on pilot studies [23], we designed feedback for helping students understand and correct the errors. There are four types of unsolicited feedback: minimal feedback, error-awareness and detection (EAD) feedback, self-explanation feedback and error-correction scaffolds. (i) Minimal feedback consists of flag feedback (checks for correct and crosses for incorrect answers) along with a text indication. (ii) EAD feedback (bottom of Fig. 1) focuses on supporting the metacognitive skills of error detection and awareness and appears on the screen after the student has indicated having read the task. As an example, the English translation of the EAD feedback in Fig. 1 is: "The result, walking distance=5 1/30, cannot be correct. Travel with the bus is already 4/5 of the total distance, so the walking distance must be less than 1/5." (iii) Self-explanation feedback (Fig. 3) is presented in the form of MCQs. It aims to help students understand and reason about the error through "why" and "how" questions. (iv) Error-correction scaffolds prepare the student for correcting the error in the second phase and also have the form of MCQs. The choices in the MCQs are related to different misconceptions or typical errors. By addressing such misconceptions and errors, MCQs are meant to prepare the students for correcting the error in Phase 2. Students receive minimal feedback on their choices, and eventually the correct answer. MCQs are nested (2 to 5 layers). If a student chooses the right answer at the two top-level MCQs (the "why" and "how" questions), then the next levels, the error-correction MCQs, will be skipped.

In the second, error-correction, phase the chosen step is crossed out, and an additional editable box is provided for correcting the error (cf. Fig. 2). After that error-specific feedback is provided, e.g., "*You forgot to expand the numerators"*, along with the correct solution. Here, we allow students only one attempt in order to avoid overlap with problem-solving exercises and to be able to assign learning effects to erroneous examples.

3 Studies

To assess the learning effects of erroneous examples at different grade levels and settings, we conducted lab studies with 6th, 7th and 8th-graders and school studies with 9th and 10th-graders. The participants came from both urban and suburban German schools. We first explain the general aspects of our experimental design that are common to all studies and discuss the particulars of each of the studies and the results in the following sections.

General Design. All studies used three conditions. The control condition, No-Erroneous-Examples (NOEE), included standard fraction problems of the form 3/4+5/7 with minimal feedback and the correct solution, but no erroneous examples. The experimental condition Erroneous-Examples-With-Help (EEWH) included standard exercises, and erroneous examples with provision of help (EAD, error detection/correction MCQs, and error-specific help). The condition Erroneous-Examples-Without-Help (EEWOH) included standard exercises, and erroneous examples but without additional help. The design included a pre-questionnaire, a pretest, a familiarisation, an intervention, a posttest and a post-questionnaire. The pre and posttests were counterbalanced.

3.1 6th-Grade Lab Study

Twenty-three paid volunteers in the 6th-grade participated in this study, distributed as follows: EEWH=8, EEWOH=7, NOEE=8. They had just completed a course on fractions at school. The mean of their term-grade in mathematics across conditions was 2.04 (*SD*=.88) (best=1 vs. fail=6), so the participants were generally good students. There was no significant difference in the means of the pretest among conditions (*p*=.8). The experiment was completed in one session with breaks.

3.1.1 Materials

In the intervention, both groups solved six sequences of three items: standard exercise - standard exercise - erroneous example. These sequences trained skills that are typical fraction topics taught at school, e.g. fraction addition/subtraction with like denominators and with unlike denominators, addition of whole numbers with fractions, as well as word problems but no modelling, which would require students to use fraction operators to represent the word problems.

The posttest consisted of similar problems, and a transfer problem (which was a four-fraction addition, as opposed to the maximum of three in the intervention). Finally, three erroneous examples which included three open conceptual questions each on error detection and awareness were part of the posttest. The questions were of the kind “Why can Oliver’s solution not be correct?”, “What mistake did Oliver make?”, “Why did Oliver make this mistake? What did he not understand about fractions?” These questions were designed to test students’ error detection skills as well as their understanding of basic fraction principles. For example, the mistake Oliver made was that he added the denominators 6 and 8 in the exercise 7/6+5/8. The answer to the question about what Oliver did not understand would be “That if one adds the denominators 6 and 8, one gets 14ths, to which 6ths and 8ths cannot be transformed.”, which addresses the basic concept of common denominators.

3.1.2 Results: 6th-Grade Lab Study

Table 1. Descriptive Statistics: Lab Study 6th-Grade

	Condition	**EEWH**	**EEWOH**	**NOEE**
Score	*Subscore*	*mean(sd)%*	*mean(sd)%*	*mean(sd)%*
Cognitive Skills	**Pretest**	80.2(26.7)	85.7(17.8)	86.5(12.5)
	Post-pre-diff	-2.1(33.6)	1.2(21.7)^	2.1(23.9)+
Metacognitive Skills (EE)	**EE-find**	91.7(15.4)+	76.2(31.7)^	66.5(35.6)
	EE-correct	80.2(12.5)+	75.0(21.0)^	68.7(25.9)
	EE-ConQuest*	64.6(25.5)+	60.2(33.3)^	41.7(21.2)
	EE-total	75.3(16.8)+	67.9(27.5)^	54.7(23.0)
	Total-time-on-postEE	16.9(6.2)^	13.8(5.5)+	18.0(5.1)
Transfer	**Transfer**	75.0(46.2)+	71.4(48.8)	75.0(46.3)^

Note: +=best, ^=middle learning gains, *= also conceptual skill

The results for the erroneous examples scores follow our hypotheses, although they were mostly insignificant. The EEWH condition scored highest in almost all scores. For all these scores, EEWOH came second, followed by NOEE. The big variances between conditions were only significant for correcting the error (EE-correct) in the erroneous examples. Nevertheless, we ran an ANOVA for that score, since the group size is almost the same across conditions, which means that ANOVA can produce robust results. The condition showed no significant effect in the ANOVA, there was a significant difference when comparing EEWH and NOEE for finding the error in the planned tests (Helmert contrasts) (p=.044, d=1.39). Another quite big difference was between EEWH and NOEE for the total erroneous example score (p=.065, d=1.20), which includes correcting the error and answering conceptual questions. These learning gains related to erroneous examples did not transfer to the cognitive skills where the differences between pretest and posttest are minimal in either direction for all conditions and there was a ceiling effect both in the pretest (M =84.1, SD=19.3) and the posttest (M =84.4, SD=15.8). This might be due to the high prior knowledge level of the participants.

As we did not have access to the term grades of the participants before the experiment, the conditions were not balanced in that respect. Therefore, we analysed the data with the term-grade but also with the pretest score as covariates, to capture the possible influence of previous math and fraction knowledge, respectively, on the learning effects. With this analysis, there is a main effect for erroneous examples in answering conceptual questions (p=.036, d=1.01), and in the total erroneous examples score (p=.028, d=3.3), when comparing the two erroneous example conditions with the control. The same scores were also significantly higher for EEWH vs. NOEE (p=.022, d=1.11/ p=.008, d=1.32) respectively). Additionally, the difference for finding the error was significantly higher for EEWH vs. NOEE (p=.029, d=1.06).

Discussion: 6th Grade. We found significant differences in the scores for erroneous examples, which means that erroneous examples, in general, and the additional help, in particular, supported better the metacognitive skills. The higher performance in the conceptual questions related to understanding the error also indicates better conceptual understanding for the erroneous examples conditions and for the help condition. However, we had no evidence that studying erroneous examples had an effect on standard cognitive skills, where the students already had a very high level.

3.2 Lab Study: 7th and 8th Grade

Twenty-four students in the 7th and 8th-grade took part, eight in each of the three conditions. 7th and 8th-graders are similarly advanced beyond 6th-graders in their understanding of fractions, according to our expert teachers. They have had more opportunity to practice, but often retain their misconceptions in fractions. The mean of their term grade in mathematics was a little lower (M=2.8, SD=1.2) compared to the 6th grade, but still at the upper-level of the grading scale. The pretest mean difference was not significant between conditions (p=.8). Consistent with the judgments of the expert teachers, there was no significant difference in the scores of the 7th compared to the 8th grade (p=.896). Participants completed all experimental phases in one session with breaks in between.

3.2.1 Materials

7th and 8th-graders worked with the same materials as the 6th-gradres but also solved modelling exercises that were not used in the 6th-grade, since such exercises are not typically encountered in German schools in this grade. An example of a modelling exercise is: “Eva invited her friends to her birthday party. They drank 8 3/7 bottles of apple juice as well as 1 5/6 bottles of lemonade. How many bottles did they drink all together?” The expected modelling in this exercise is: 8 3/7 + 1 5/6. In total, there were seven sequences of exercises in this study. A modelling exercise also testing transfer was added to the posttest. By including such fraction modelling, we aimed to induce and measure conceptual understanding.

3.2.2 Results: 7th-8th-Grade Lab Study

As a whole, the results do not support our hypotheses for the 7th and 8th-grade. All differences in scores are small and not significant. NOEE scored better in almost all scores, apart from the conceptual questions, where EEWOH did best. EEWOH was also second best in finding the error and in the total erroneous examples score. EEWH came second in the cognitive skills, correcting the error, transfer exercises, and modelling. The standard deviation for all scores except for improvement on cognitive skills was highest for EEWH. Since the group size is the same across conditions, the results of the ANOVA can be considered robust although Levene's test was significant for finding the error (p=.018), conceptual questions (p=.000) and for the total score on erroneous examples (p=.000). The only statistically significant score in the ANOVA test was the time spent on the posttest erroneous examples ($F(2, 21)$ =5.59, p=.011, n^2=.22), where NOEE spent significantly more time than the erroneous-examples conditions together (p=.009, d=-1.18) and EEWH alone (p=.003, d=-1.45). Moreover, the term-grade is a significant covariate on answering conceptual questions ($F(1,21)$=4.49, p=.047, n^2=.18) and quite big for the total erroneous examples score ($F(1,21)$=4.03, p=.059, n^2=.06) which in both cases decreases the difference between the control and the erroneous example conditions that originally scored worse. There is also a significant effect of the condition for the time spent on erroneous examples ($F(2,21)$=5.28, p=.014, n^2=.35) when term-grade is considered as covariate.

Table 2. Descriptive Statistics: Lab Study 7th-8th-Grade

	Condition	**EEWH**	**EEWOH**	**NOEE**
Score	***Subscore***	***mean(sd)%***	***mean(sd)%***	***mean(sd)%***
Cognitive Skills	**Pretest**	73.7(26.7)	71.2(19.7)	77.9(12.4)
	Post-pre-diff	2.4(24.4)^	-4.3(26.6)	6.9 (17.9)+
Metacognitive Skills (EE)	**EE-find**	68.7(34.7)	75.0(13.4)^	90.6(12.9)+
	EE-correct	57.8(26.7)^	54.7(21.1)	65.6(20.8)+
	EE-ConQuest*	55.2(46.5)	62.5(12.6)+	61.5(19.4)^
	EE-total	59.3(37.1)	63.7(11.9)^	69.8(15.0)+
	Total-time-on-postEE	8.1(4.3)+	11.5(4.2)^	15.5(4.8)
Transfer	**Transfer**	45.2(45.8)^	38.0(36.0)	67.3(28.5)+
Conc. Underst.	**Modelling**	36.4(42.2)^	19.8(35.0)	40.8(48.6)+

Note: +=best, ^=middle learning gains, *= also conceptual skill

An important result in this study is a significant difference in the scores for finding and correcting the error in ($t(23)$=4.89, p<.001, r=.71). The standard deviation for the two metacognitive competencies is comparable, but the mean for correcting is more than 0.5 point lower than for finding the error (M=3.12, SD= .95 for finding, M=2.54, SD=.99 for correction), which means that a significant number of participants was able to find the error but not correct it. This is also true when comparing separate conditions, where the difference between finding and correcting the error in EEWOH and EEWH is significant, but it is a stronger effect in EEWOH ($t(7)$=6.56, p<.001, r=.92), who could correct about 1/3 of the mistakes they found, and much less in the EEWH ($t(7)$=4.19, p=.001, r=.85), who could correct 1/2 of the mistakes they found. The same phenomenon occurred even with students who could solve exercises, but could not correct errors of the same type. For example, most students could add fractions with unlike denominators, but could not correct related errors.

Discussion: 7th-8th-Grade. An explanation for the fact that the erroneous examples conditions, and especially the EEWH condition, did not perform better in the metacognitive skills of erroneous examples, for which they were trained, is the little time they spent on erroneous examples in the posttest. Moreover, the long session might have overloaded the students and especially the ones in the EEWH condition whose sessions last long (two and a half hours) because of the help provided. The possible resulting fatigue might be the reason why they did not spent more time on erroneous examples in the posttest.

A plausible interpretation for the fact that the term grade is a significant covariate for answering conceptual questions, but not for cognitive skills is that a higher level of prior math knowledge is required to process new conceptual knowledge. This high-level knowledge is not necessary to deal with trained cognitive skills, which can be done by using well-practice solutions steps (algorithmically). The difference between finding and correcting the error may mean that although students know the correct rules for performing operations on fractions and can recognise errors that violate these rules, they still have knowledge gaps that surface when asked to correct the error.

3.3 School Study: 9^{th}-10^{th}-Grade

In order to test the use of erroneous examples outside the lab we conducted school studies. The school studies tested students from two different schools, one urban and one suburban, of yet a higher level (9^{th} and 10^{th}-grade). Our expert teachers advised that these students typically still exhibit common fractions misconceptions. However, 9^{th} and 10^{th}-graders have, on average, an overall higher math knowledge. Since we found that math knowledge (term grade) has a covariating effect on conceptual understanding, we wanted to see if erroneous examples would have a better effect with these higher grade students.

Forty-three students completed the study, distributed as follows: EEWH=14, EEWOH=18, NOEE=11. The difference in the pretest was not significant either between 9^{th} and 10^{th} grade (p=.12), or between conditions (p=.7). They completed the experiment on two days separated by a week. Many students did not attend school on the second experimental day, which led to the unbalanced conditions.

3.3.1 Materials

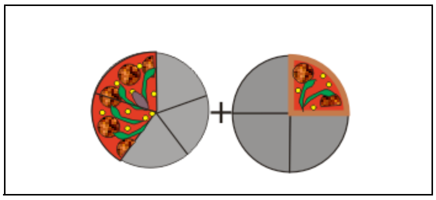

Fig. 4. Pizza representation of fraction problem 3/5+1/4

To emphasise conceptual learning and avoid a ceiling effect in this study we shifted from the traditional school fraction curriculum and included more conceptual exercises to address the basic principles for fractions. For instance, we used the principles of "addition as increasing", "subtraction as decreasing", and "part of a whole" [22]. In effect, we reorganised the materials to reflect this shift and also added one sequence to train the basic concept "part of a whole". Thus, we incorporated conceptual errors on top of the rule-application errors which were the focus of the previous lab studies. We also changed the order of presentation of the erroneous examples in the intervention; a sequence here consisted of standard exercise – erroneous examples – standard exercise, to test whether giving students the opportunity to train a bit after the erroneous examples would make a difference. Furthermore, we adjusted the pretest and posttest exercises to test these concepts and added a transfer exercise for fraction subtraction and two that asked students to transform a fraction operation represented with pizzas into a numerical fraction representation. For example, the task in Fig. 4 had to be represented as 3/5+1/4, a standard kind of exercise done at schools.

3.3.2 Results: School Study 9th- and 10th-Grade

The results of the school studies supported our hypotheses. The participants in the EEWH condition scored higher in most scores used as measures of learning. There is no clear second place that varies between EEWOH and NOEE. Although the variances tend to be high, they are mostly comparable between conditions, allowing for analysis of variance. The biggest differences were in favour of EEWH for the modelling exercises in total ($F(2,39)=3.11$, $p=.056$, $n^2=.14$), and for modelling the basic concept "part of a whole" alone which just missed significance ($F(2,39)=3.20$, $p=.051$, $n^2=.14$), and the transfer exercises (cf. Table 3) which was not significant. Conceptual understanding gave results with very high variance. Still, EEWH scored significantly higher than NOEE ($p=.016$, $d=1.28$) in modelling "part of a whole", which was trained, and even lower than in the pretest for the same concept. However, NOEE scored higher for modelling "relative part of" that was not trained during intervention, but was meant to test transfer from the more general concept "part of a whole". EEWOH scored very low in both concepts.

Table 3. Descriptive Statistics School Studies 9th, 10th-Grade

	Condition	**EEWH**	**EEWOH**	**NOEE**
Score	***Subscore***	***mean(sd)%***	***mean(sd)%***	***mean(sd)%***
Cognitive Skills	**Pretest**	75.99(11.9)	64.3(20.6)	61.3(12.7)
	Diff-post-pre-total	9.3(13.8)+	2.0(24.5)	6.3(22.5)^
	Transform-total	21.5(22.9)+	8.23(32.2)^	1.6(48.7)
Metacognitive Skills (EE)	**EE-find**	53.6(27.5)+	47.2(27.0)	52.3(28.4)^
	EE-correct	30.4(28.0)+	15.3(24.5)	27.3(32.5)^
	EE-ConQuest*	46.4(20.3)^	48.2(25.3)+	41.7(29.3)
	EE-total	44.6(20.9)+	41.4(22.9)	40.9(26.8)^
	Total-time-on-EE	7.6(7.9)	5.8(3.3)+	6.2(3.5)^
Transfer	**Add-subtr-total**	30.4(36.9)+	16.7(34.3)	18.2(25.2)^
	Conc-transf-total*	49.4(36.8)+	31.9(30.0)	34.9(32.0)^
	Transfer-total	39.88(22.6)+	24.3(26.8)	26.5(28.6)^
Conceptual Understanding	**Part-of-whole**	21.4(42.9)+	-5.6(62.8)^	-30.7(38.5)
	Addition-as-incr	69.6(41.8)+	62.5(47.2)^	45.5(52.2)
	Subtr-as-decreas	50.0(51.9)+	25.0(42.9)	27.3(46.7)^
	Rel-part-of	28.6(46.9)^	8.3(25.7)	31.8(46.2)+
	Modelling-total	56.7(33.0)+	32.6(25.6)	33.5(29.9)^

Note: +=best, ^=middle learning gains, *=also conceptual skill

Moreover, the cognitive skills in the taught exercises increased more for EEWH that had a variance of about 10% lower than the other two conditions. EEWH reached the mean of 85.2 ($SD=14.9$) in the posttest and surpassed the other two conditions by almost 20%. This difference in the posttest was also significant, as measured by an ANOVA ($F(2,39)= 5.42$, $p=.008$, $n^2=.13$). Similarly, for the transformation exercises, EEWH scored a mean of 88.0 (SD=20.99) in the posttest and significantly better than the other conditions ($F(2,39)= 3.48$, $p=.04$, $n^2=.15$). The planned Helmert contrasts indicated significant main effects of erroneous examples (EEWH and EEWOH conditions) for modelling the concept "part of a whole" ($p=.038$, $d=0.63$). For the conditions EEWH and NOEE, there was a significant higher performance of EEWH in modelling "part of a whole" ($p=.016$, $d=1.28$) and almost for modelling in general ($p=.056$, $d=0.74$). The condition EEWH did significantly better than EEWOH in modelling in general ($p=.026$, $d=0.81$).

We also tested the possible covariating effect of the pretest-total. The pretest score was meant to indicate significant differences based on the prior fraction knowledge. Interestingly, the pretest scores have a covariating effect on learning for the taught cognitive skills ($F(1,39)=21.56$, $p=.000$, $n^2=.36$) and for the transformation skills ($F(1,39)=9.52$, $p=.004$, $n^2=.22$) separately. Taking this effect into account turns both of these scores into significant effects (taught cognitive skills, $F(2,39)=3.46$, $p=.041$, $n^2=.14$ and transformation skills $F(2,39)=3.45$, $p=.042$, $n^2=.13$), as well as for modelling the concept "part of whole" ($F(2,39)=3.96$, $p=.027$, $n^2=.17$). Through a closer look with help of the planned contrasts, a main effect of erroneous examples can be seen (EEWH and EEWOH conditions) for modelling the concept "part of a whole" ($p=.04$, $d=0.67$). For the conditions EEWH and NOEE, there was a significantly higher performance of EEWH in the transformation exercises ($p=.015$, $d=0.8$) and modelling "part of a whole" ($p=.007$, $d=0.9$) and almost for taught cognitive skills ($p=.058$, $d=0.62$). The condition EEWH only did significantly better than EEWOH in the taught cognitive skills ($p=.019$, $d=0.77$), but differences were also quite big in the transformation exercises ($p=.085$, $d=0.56$), and in modelling in general ($p=.061$, $d=0.61$). An additional interesting result is that EEWH condition spent significantly more time-on-task ($p=.037$, $d=0.68$) than any of the other two, but even more than NOEE ($p=.016$, $d=0.79$).

Moreover, although we did not find any significant difference between conditions in metacognitive skills, we again found that significantly more students across conditions could find the error in the posttest erroneous examples than correct it $t(42)=8.84$, $p=.000<.001$, $r=.81$.

The above results for the 9th and 10th-grade are by and large robust, however the variance was significantly big for transformation ($p=.003$), where EEWH has a much smaller variance than any of the other two conditions. This difference, in combination with the different sample sizes, might make the results for this score unreliable.

Discussion: 9th- and 10th-Grade School Study. The most striking result is that erroneous examples with help had a significant effect for the taught cognitive skills over erroneous examples without help and an almost significant effect over no erroneous examples. The source of this might be the higher math knowledge of the more advanced students (9th- and 10th-grade). Confronting the students with more conceptual exercises on fractions may have contributed to this result, too.

Moreover, our results indicated a beneficial use of erroneous examples as a whole for the conceptual exercises dealing with modelling. The high variance in modelling the basic concepts tested in this experiment indicates that some students could understand the principle behind them and had no problems applying them, whereas others were just confused. This effect is particularly high for the EEWOH in modelling "part of a whole", as well as not for modelling "relative part of" that was not taught at all during intervention, but was meant to test transfer from the more general concept "part of a whole". Both of these concepts seem to have been particularly confusing for EEWOH. The higher variance and the negative learning effect in "part of a whole" for EEWOH may mean that this condition was confused by being asked to represent the difficult concept "part of a whole" explicitly and conceptually (as opposed to the standard school algorithmic approach). Since they received no help, they could not recover from the confusion at all, unlike EEWH, and scored badly both in this trained concept ("part of a whole"), and in the transfer concept ("relative part of"). A possible explanation for the high learning effect of EEWH but also the high variance is that they had the chance to practice the basic concept "part of a whole" in the way taught by the erroneous examples with additional support. This resulted in scoring better at the relevant exercise, and relatively high in transferring from this concept to "relative part of", but might have still confused students who rely on purely procedural/algorithmic solutions. NOEE managed to score better than EEWOH as they just used the standard algorithmic strategy taught at school without necessarily having a deeper understanding of the concept.

4 General Discussion

In our studies with different levels of students we had some consistent results but also some results that reveal a difference in how erroneous examples with error detection and correction help can influence mathematics learning.

We found that more advanced students (9^{th} and 10^{th}-grade) benefit from erroneous examples with help in terms of cognitive skills (including standard problem solving) in general, as opposed to erroneous examples without help or no use of erroneous examples at all. Although this was not the case for either of the two less advanced levels that we tested, it might have been an artefact of the very high prior fraction knowledge of the particular participants (6^{th}, 7^{th}, and 8^{th}-grade). Moreover, we had some sources of evidence that deep conceptual understanding is influenced by erroneous examples with additional error detection and correction help. Such evidence includes the better performance of the EEWH over the NOEE condition at the conceptual questions in the 6^{th}-grade, as well as the higher scores in modelling "part of a whole" for EEWH vs. NOEE and in modelling as a whole for EEWH vs. EEWOH (both for the 9^{th} and 10^{th}-grade). The higher grades (9^{th}, 10^{th}) also received more intervention materials aiming at conceptual understanding. The difference in conceptual understanding between EEWH and EEWOH for the same grade levels might have also instigated the respective difference in cognitive skills. Additionally, we found that erroneous examples can also influence the metacognitive skill of error detection for lower-grade (6^{th}-grade) but highly competent students. There is a possible twofold explanation for this. First, these students, who have just learned fractions can handle the demanding erroneous examples because the cognitive skills and domain knowledge that erroneous examples presuppose is readily available to them. Second, there is room for improving their error-detection significantly as they have not yet applied much what they have learned to make errors themselves and, hence, practice error-detection on their own errors.

Overall, we found some main effects for erroneous examples for the less advanced 6^{th}-grade (metacognitive skills) and the more advanced 9^{th} and 10^{th}-grade (conceptual understanding), but most effects were for erroneous examples with help. This is consistent with the results of Kopp and colleagues [17] in the medical domain in terms of the benefit of erroneous examples with help, although the domains differ a lot and a comparison is tenuous. We also found that the use of erroneous examples without help might be worse than no use of erroneous examples for conceptual and transfer skills, which is not reliably true for metacognitive skills. As a whole, the inconsistent performance observed in the school study with regard to the modelling might mean that there was a conflict between the standard procedural way fractions are taught at school and the conceptual way our erroneous examples deal with them. This effect might be stronger for EEWOH who are left confused, due to the lack of guidance. More familiarity with erroneous examples and the conceptual strategy might counter-balance this confusion, especially when combined with provision of help. Siegler [14] has suggested that requests for explanation of correct and incorrect strategies lead to a period of "cognitive ferment" (p. 51) and only later do they cause the development of correct strategies and the ability to self-explain. He attributes this delay to a state of increased uncertainty and variability. Another possible explanation for the lack of transfer between the taught concept ("part of a whole") and the untaught concept ("relative part of") might be that the theoretically subordinate category of "relative part of" is actually not cognitively subordinate, and hence there is no transfer between the two basic concepts in terms of learning.

Our results do not support the use of erroneous examples with or without help for medium-advanced students (7^{th} and 8^{th}-grade), where prior knowledge seems to play a crucial role. A reason for that might be the combination of the high grade level but also the high competence (term grade and pretest scores) which the participants had and was not the case for the 9^{th} and 10^{th}-grade.

An interesting mismatch between the competencies of finding and correcting the error across conditions is evident in our results. This mismatch persisted in all our studies independent of student level or material design, and it was significant in our studies with the two higher grade students (7^{th}-8^{th} and 9^{th}-10^{th}-grades). Ohlsson [24] has described this phenomenon as dissociation between declarative and practical knowledge. It is intriguing that in our school studies with 9^{th} and 10^{th}-graders, students' cognitive skills did improve through erroneous examples, despite the fact that their ability to find errors developed significantly more than that of correcting errors. This might show that the competence of correcting typical errors is not necessary for monitoring, correcting, or avoiding one's own errors. That is consistent with Ohlsson's argument, that when the competency for finding errors is active, it functions as a self-correction mechanism that, given enough learning opportunities, can lead to a reduction of performance errors. However, it is a new finding in comparison to previous research in erroneous examples that has not differentiated between the competencies of finding and correcting errors. In our case, it seems like erroneous examples with error-detection and error-correction help that specifically train finding errors and explaining them might offer the required learning opportunities without the need to develop the error-correction skill. The contribution of such help is also in line with the theoretical work by van Gog and her colleagues [25] who have advocated its use in the context of worked examples as a way for promoting conceptual understanding.

Regarding the presentation of erroneous examples, we have at least a first indication that they are more beneficial when presented after the students have been confronted with standard exercises and followed again by standard exercises, since we only found a significant improvement at tasks other than erroneous examples when this order of presentation was used. A potential explanation is that this gives students the opportunity to review the material before working with erroneous examples that might also increase the perceived relevance of erroneous examples, as well as to practice what they have learned after the presentation of the erroneous examples.

In general, although our results show room for further improvement in the students' understanding of fractions, they still reveal a good trend for erroneous examples as an instructional method that can help students in this demanding domain, in particular for more advanced students (higher grade) who have had fractions in the past, but also for students just having learned fractions. They also indicate that previous results on the benefits of self-explaining correct and incorrect examples by Siegler and colleagues in water displacement and mathematical equality problems [14], [15] and Grosse and Renkl in probability problems [16] are transferable to online erroneous examples in the fraction domain.

4.1 Open questions

A question that remains open is how less and medium advanced students (6th, 7th, and 8th-grade) can be helped to improve their cognitive skills using erroneous examples. A practical measure here, in terms of our design of online erroneous examples, may be to allow students to explicitly request more help. It is likely that they will use this extra feature if they feel uncertain about their answer, thus overcoming a possible shortcoming of our design of online erroneous examples which assumes that if students can answer the basic "why" and "how" process-oriented MCQs they do not need error detection and correction help; MCQs dealing with this issue are skipped in that case. However, this might be too coarse an indicator for when and how much help is needed. Moreover, it underestimates the difficulty students have with applying rules.

The materials and instructional design might also need adaptations. For instance, the results might be clearer if we enrich our conceptual exercises and test more whether errors that reveal lack of conceptual understanding are not used. We want to elaborate more on such conceptual exercises since the standard fraction exercises practiced at school might be too simple to be able to influence students' performance alone through process-oriented help, as we have observed in our studies with the less-advanced and medium-advanced students and is also hypothesised from a theoretical perspective by [25]. A good start would be to try to replicate our results for the advanced students (9th and 10th-grade) using the new more conceptual materials with the other grade levels. A more representative sample in terms of prior math and fraction knowledge is also a prerequisite for this test. Furthermore, the replication of the results would help rule out the possibility that the materials alone and not the level made the difference in our results. Moreover, due to the big variances in the study with 9th and 10th grades and the unequal group sizes we are already planning to collect additional data points to be able to draw even more reliable conclusions.

Finally, we want to test whether the order of presentation really plays a significant role, by using the more conceptual material and varying the order of presentation between different conditions.

Acknowledgments. This work was supported by the DFG - Deutsche Forschungsgemeinschaft under the ALoE project ME 1136/7.

5 References

1. Jones Newton, K.: An Extensive Analysis of Preservice Teachers' Knowledge of Fractions. American Educational Research Journal, 45(4), 1080--1110 (2008)
2. Stafylidou S., Vosniadou, S.: The development of students' understanding of the numerical values of fractions. Learning and Instruction, 14, 503—518 (2004)
3. McLaren, B.M., Lim, S.J., Koedinger, K.R.: When and how often should worked examples be given to students? New results and a summary of the current state of research. Love, B. 7. C., McRae, K., V.M.S., (eds.). Proceedings of the 30th Annual Conference of the Cognitive Science Society, pp. 2176--2181, Cognitive Science Society (2008)
4. Paas, F.G., van Merrienboerg, J.J.G.: Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. Journal of Educational Psychology, 86(1), 122--133 (1994).

5. Renkl, A.: Learning from worked-out examples: A study on individual differences. Cognitive Science 21, 1—29 (1997)
6. Trafton, J.G., Reiser, B.J. The contribution of studying examples and solving problems. In Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society (1993)
7. Tsamir, P., Tirosh, D.: In-service mathematics teachers' views or errors in the classroom. In International Symposium: Elementary Mathematics Teaching, Prague, August (2003)
8. Borasi, R.: Capitalising on errors as "springboards for inquiry": A teaching experiment. Journal for Research in Mathematics Education, 25(2), 166--208 (1994)
9. Müller, A.. Aus eignen und fremden Fehlern lernen. Praxis der Naturwissenschaften 52(1), 18--21 (2003).
10. Oser, F., Hascher, T.: Lernen aus Fehlern - Zur Psychologie des negativen Wissens. Schriftenreihe zum Projekt: Lernen Menschen aus Fehlern? Zur Entwicklung einer Fehlerkultur in der Schule, Pädagogisches Institut der Universität Freiburg, (1997).
11. T. Seidel, M. Prenzel. Mit Fehlern umgehen - Zum Lernen motivieren. Praxis der Naturwissenschaften 52(1), 30—34 (2003)
12. Strecker, Ch.: Aus Fehlern lernen und verwandte Themen. http://www.blk.mat.uni-bayreuth.de/material/db/33/fehler.pdf.
13. OECD: International report PISA plus (2001)
14. Siegler, R.S.: Microgenetic studies of self-explanation. In N. Granott and J. Parziale (eds.) Microdevelopment, Transition Processes in Development and Learning, pp. 31--58. Cambridge University Press (2002)
15. Siegler, R.S., Chen, Z.: Differentiation and integration: Guiding principles for analyzing cognitive change. Developmental Science 11, 433—448 (2008)
16. Grosse, C.S. and Renkl, A.: Finding and fixing errors in worked examples: Can this foster learning outcomes? Learning and Instruction 17, 612--634, (2007)
17. Kopp, V., Stark, R., Fischer, M. R.: Fostering diagnostic knowledge through computer-supported, case-based worked examples: effects of erroneous examples and feedback. Medical Education 42, 823--829 (2008)
18. Melis, E., Goguadze, G., Homik, M.,. Libbrecht, P., Ullrich, C., Winterstein, S.: Semantic-aware components and services in ActiveMath. British Journal of Educational Technology. Special Issue: Semantic Web for E-learning, 37(3), 405--423 (2006)
19. Paas, F.G., Renkl, A., Sweller, J.: Cognitive load theory and instructional design: Recent developments. Educational Psychologist 38, 1--4 (2003)
20. Skinner, B.F. The behavior of organisms: An experimental analysis. Appleton-Century, New York, US (1938)
21. Melis, E.: Design of erroneous examples for ActiveMath. B. Bredeweg Ch.-K. Looi, G. McCalla and J. Breuker (eds.), Artificial Intelligence in Education. Supporting Learning Through Intelligent and Socially Informed Technology. 12th International Conference (AIED 2005) 125, pp. 451--458. IOS Press, (2005)
22. Malle, G.: Grundvorstellungen zu Bruchzahlen. Mathematik Lehren 123, 4--8 (2004)
23. Tsovaltzi, D., Melis, E., McLaren, B.M., Dietrich, M., Goguadze, G., Meyer, A-K.: Erroneous Examples: A Preliminary Investigation into Learning Benefits_ In Cress, U., Dimitrova, V., Specht, M. (eds.), Proceedings of the Fourth EC-TEL 2009, LNCS 5794, pp. 688—693. Springer-Verlag, Berlin, Heidelberg (2009)
24. Ohlsson, S.: Learning from Performance Errors. Psychological Review 103(2), 241--262 (1996)
25. van Gog, T., Pass, F., van Merrienboerg, J.J.G.: Process-Oriented Worked Examples: Improving Transfer Performance Through Enhanced Understanding. Instructional Science 32, 83--98 (2004)