



Effects of multiple choice options in mathematics learning

Cornelia S. Große

FB3 Mathematics / Computer Science, University of Bremen, Bremen, Germany

For correspondence: cornelia.grosse@uni-bremen.de

Abstract

In recent years, the ability to use mathematics flexibly in everyday situations is increasingly accentuated, and mathematics education gives more and more attention to problems with realistic contexts. However, it is reported very often that learners have substantial difficulties in connecting mathematics to real situations. In order to solve complex and authentic problems embedded in real situations it is necessary to find translations from verbal descriptions to mathematical notations and to interpret mathematical results with respect to the real situation. In the present work, it was tested experimentally whether translation and interpretation competencies are fostered by presenting multiple choice options from which the learners have to select the correct one. While benefits of translation answer choices emerged, interpretation answer choices did not support learning. However, when the learners were asked to rate how much they liked the problems, opposed results were obtained, indicating that objective learning outcomes and subjective scores did not correspond. Possible reasons for the inconsistent and to some extent unexpected learning outcomes and for the divergence between objective results and subjective statements are discussed.

Keywords: mathematics learning, word problems, modeling problems, multiple choice options.

Introduction

In recent years, the need for using mathematics in everyday life is increasingly accentuated, and mathematics education gives more and more attention to problems with realistic contexts. While this is a very appealing approach at first glance, a more detailed look quickly reveals the many problems learners are confronted with when trying to solve these problems. A first obstacle to overcome refers to finding a transition from the verbal problem statement to a mathematical notation; and after performing mathematical operations the mathematical solution has to be interpreted in view of the described situation, which again is a great challenge. Still, in order to flexibly apply mathematics in “out-of-mathematical” situations, these competencies are of major importance.

Within the conceptual and theoretical outline of the PISA study (e.g., OECD, 2003) “mathematical literacy” highlights the importance of connecting mathematics to real situations. Mathematizing authentic situations is called “modeling” (e.g., Yerushalmi, 1997); accordingly, modeling problems are complex and authentic problems embedded in real situations (e.g., Blum and Borromeo Ferri, 2009). It is a characteristic feature of modeling problems that on the one hand they do not need to give all relevant information explicitly, and on the other hand, irrelevant information may be presented. Thus, the learner first has to distinguish between relevant and irrelevant information; and second, missing information has to be estimated or otherwise obtained. After selecting a matching mathematical model and conducting intra-mathematical operations, the learner has to choose an appropriate interpretation of the mathematical results.

However, the distinction between word problems and modeling problems is not always clear and consistent. For example, Palm (2008) defines the scope of different types of “word problems”; however, on closer inspection, these “word problems” are very close to the “modeling problems”

described above. In his classification, “category 1 word problems” require an interpretation of a remainder, thus, the mathematical result has to be rounded up or down in order to get a result matching reality. In order to solve “category 2 word problems” in a realistic way, it is necessary to recognize that given quantities cannot enter the calculation as a whole because parts of these quantities could not be used in practice. In contrast, in order to successfully solve “category 3 word problems”, relevant information is missing, so that a realistic result can only be reached via estimations.

There is general consent that mathematics education should focus on modeling problems (e.g., Borromeo Ferri, 2009; see also Yerushalmi, 1997). In order to enhance understanding and transfer, real contexts are highlighted by several theories; specifically, constructivist approaches (for an overview see Reinmann and Mandl, 2006) emphasize the importance of situated learning (e.g., Lave and Wenger, 1991) in concrete contexts. Accordingly, it is recommended to connect learning processes with authentic situations, in order to acquire knowledge from the perspective of application (see also Vye et al., 1997).

While being appealing in theoretical papers, this approach leads to several problems in practice. A large amount of research confirms that when solving mathematics problems, learners frequently do not consider reality; occasionally, children take a “shortcut” by just choosing a mathematical operation, inserting numbers from the problem formulation, carrying out the calculation, and writing down the result without checking for plausibility (e.g., Greer, 1997; Reusser and Stebler, 1997; Verschaffel et al., 1994, 1997, 1999).

A small number of studies explicitly addresses the acquisition of mathematical modeling competencies; accordingly, it is actually possible to foster modeling competencies (e.g., Galbraith and Clatworthy, 1990). In the study of Zöttl et al. (2010), the effectivity of several measures was analyzed, among them heuristic worked examples, prompts for self-explanations, self-tests, and the possibility to write a learning journal. Overall, their learning environment supported learning outcomes considerably; however, it is not possible to conclude which specific measure(s) in fact triggered the improvement. According to Schukajlow and Blum (2011; see also Blum and Borromeo Ferri, 2009) cognitively activating education which concentrates on individual work fosters modeling competencies. Schukajlow et al. (2015) examined effects of asking students to find multiple solutions to modeling problems; while they did not find direct effects on modeling performance, they found indirect effects of multiple solutions via the number of solutions the learners developed and their experience of competence.

In contrast, improving skills in less complex word problem solving is addressed by many studies; for example, with respect to the effectivity of trainings which concentrate on metacognitive strategies (e.g., de Kock and Harskamp, 2014; Mevarech et al., 2010), and with regard to effects of schema-based instruction (for an overview see Powell, 2011). More specifically, Mwangi and Sweller (1998) examined the effectivity of worked examples for the acquisition of word problem solving competencies; they found positive effects of worked examples, specifically in an integrated format, on word problem solving.

In domains such as mathematics, worked examples – consisting of a problem statement, solution steps, and the final result – are frequently used, and their effectivity is documented convincingly. Many effective example characteristics and learners’ activities have been identified (for an overview see Atkinson et al., 2000; Renkl, 2014 provides a theory of example-based learning which integrates theoretical considerations and results from observational learning and analogical reasoning). The effectiveness of learning with worked examples is often explained by the cognitive load theory (e.g., Paas et al., 2003); it is claimed that worked examples reduce element interactivity and thereby reduce extraneous cognitive load (Sweller, 2010), which, in turn, provides cognitive resources for constructing and automating a cognitive schema (cf. van Gog et al., 2004, 2006, 2008). Yet, Renkl

(1997) showed that a large number of learners processes worked examples only in a passive way, with substantially reduced learning outcomes.

Research has shown that connecting examples and problems fosters learning (e.g., Stark et al., 2000; Trafton and Reiser, 1993); specifically, a “fading procedure” is recommended (e.g., Renkl et al., 2004) which means that a sequence of worked examples is used, but progressively more and more steps of the solutions are skipped until only the problem statement is presented to the learners. It is an important benefit of this approach that the requirements for the students increase only gradually; however, even if one knows that a similar step as presented now has to be solved by oneself afterwards, it might still be difficult to concentrate one’s attention on the important details of the solution step. A feature which forces the learners to concentrate on relevant details in a worked-out solution is using incorrect steps.

An increasing number of studies examines the effectivity of learning with incorrect examples (e.g., Adams et al., 2014; Booth et al., 2013; Durkin and Rittle-Johnson, 2012; Große and Renkl, 2007; Heemsoth and Heinze, 2014; McLaren et al., 2012); accordingly, learning with incorrect worked examples can be very effective. However, taking into account the length and intricacy of solutions to complex word problems, finding errors in an incorrectly solved problem would be a very challenging task. If the learners are not given any hint where exactly the errors could be located, this could demand too much of the learners. In addition, for example with respect to estimations, it might be very difficult for learners to decide whether they are “a little bit unlikely” (which might still be rated as correct) or “definitely unrealistic” (which then would have to be marked as an error). Thus, in consequence, the effectivity of implementing learning with incorrect examples in order to foster transition competencies seems at least to be debatable.

In a nutshell, empirical results suggest that it is indeed possible to foster transition competencies. However, it seems to be a very challenging task, and the question remains still open whether the very demanding translation and interpretation processes can be supported by presenting learners worked solutions without the risk of being processed only superficially.

In addition to effects of specific learning methods on learning outcomes, effects on subjective variables should be taken into account as well, as a rich body of research documents the influence of interest and enjoyment on comprehension, learning, and performance (e.g., Giannakos, 2013; Schukajlow and Rakoczy, 2016; for an overview see Pekrun, 2006). In the context of the present work, especially self-efficacy is of particular importance, as motivation and achievement are significantly influenced by self-efficacy (for an overview, see Bandura and Locke, 2003). Self-efficacy can be influenced by the experimental variation of the design of learning materials; for example, it can be fostered by lowering the requirements for the learners: if learners can meet the expectations, their confidence that they will henceforth be able to keep up will increase (Bandura, 1977). It is assumed that expectations of personal efficacy determine how much effort is taken and how long it is held up in the face of challenges and difficulties (Bandura, 1977). Overall, a positive attitude toward a learning method forms a prerequisite for effective learning and a solid foundation for application in practice.

In the present experiment, the idea of using incorrect examples in order to foster transitions between reality and mathematics was taken up; however, in order to avoid demanding too much of the learners, and in order to design the decision process unambiguously for the learners, they were given three multiple choice options from which they could select the right solution step. In this way, it was ensured that the learners knew exactly what they were intended to do, and they knew exactly where to look for possible errors. They were told that from three multiple choice options, only one was right, and after selecting one multiple choice option they could then proceed with the learning materials. Große (2014) showed that, compared to detailed word problems, focused word problems were substantially more acknowledged by the learners; hence, it can be supposed that reducing the demands of word problems in the beginning, and supporting the students to meet those demands,

can reduce the individually perceived difficulty of the tasks, which in turn can foster the confidence of the learners that they can succeed, leading to positive effects on motivation and achievement (Bandura, 1977; Bandura and Locke, 2003). With respect to mathematics education, it is not only important that a learning method is effective with respect to the acquisition of skills, but it is also important that the learners express a positive attitude toward this method. Only then can be assumed that learners will succeed motivated and confident; thus, in the present study, subjective ratings were examined as well.

The following research questions were raised:

1. Is it effective to present learners multiple choice options for the translation part of word problems?

Hypothesis 1: Learning results – especially with respect to translation competencies – are enhanced if learners have the opportunity to select a translation from a select list.

2. Is it effective to present learners multiple choice options for the interpretation part of word problems?

Hypothesis 2: Learning results – especially with respect to interpretation competencies – are enhanced if learners have the opportunity to select an interpretation from a select list.

3. Does the presentation of multiple choice options foster a positive attitude toward the problems?

Hypothesis 3: Reducing the requirements for the learners by presenting multiple choice options fosters a positive attitude toward the problems.

Methods

Sample and design.

In this study, $n = 147$ 8th grade students (mean age: 13.98 years, $SD = .55$ (4 missing values); 79 female) of a German “Gymnasium” (secondary school) participated. The experiment took place within regular mathematics courses; the learners within each class were randomly assigned to the experimental groups. A two-factorial design with four experimental conditions was implemented, with the factors “translation choice” (with versus without) and “interpretation choice” (with versus without). In the groups with translation choice, the learners were given three multiple choice options for the translation step from the verbal description to a mathematical notation; in the groups with interpretation choice, the learners were given three multiple choice options for the interpretation step from the mathematical result to an interpretation with respect to the real situation. All other steps had to be performed without support. Thus, in the group “with translation and interpretation choice” ($n = 37$), the learners first had to select the right transition to a mathematical notation, then they had to perform the calculation on their own, and finally they had to select the right interpretation of the mathematical result. In the group “translation choice only” ($n = 37$), the participants were only given multiple choice options for the translation part and had to perform the rest on their own; in the “interpretation choice only” group ($n = 37$), the participants started solving the problem on their own but were given multiple choice options for the interpretation part. In the group “without choice” ($n = 36$), the participants had to solve the problems without support.

Materials.

As learning domain, word problems which can be solved with linear equation systems were chosen. In order to vary the difficulty of the translation and interpretation steps, some problems used in the present study included estimations in the translation step (“category 3” word problems according to the classification of Palm, 2008), and some required the learners to interpret a remainder realistically (“category 1” word problems according to the classification of Palm, 2008).

In the learning phase, four problems were introduced, two included estimations and two the interpretation of a remainder. Depending on the experimental condition, the learners were presented multiple choice options for the translation step from the real situation to a mathematical notation and

/ or for the interpretation step from a mathematical solution to a real result. The learners did not receive general information on the role of translation or interpretation in modeling, and they did not receive general instruction on how to translate or interpret results. In order to avoid copying, learners sitting next to each other were assigned to different experimental conditions, and the multiple choice options were designed in a manner so that copying mindlessly would lead to wrong answers. Table 1 depicts one of the problems provided in the learning phase.

Instruments.

Pretest: Assessment of prior knowledge. A pretest following the “rapid testing” rationale (Kalyuga and Sweller, 2004; see also Kalyuga, 2008) examined the prior knowledge; in this type of test, learners are required to find a first move toward the solution of a problem in a very short period of time. Kalyuga and Sweller (2004) report a very high correlation with a traditional test and conclude that the rapid test has a very high concurrent validity. The present pretest contained 10 problems and the participants were given 20 seconds to find a first step toward the solution of each problem. Each problem was presented on a separate sheet, and after 20 seconds the learners were told to turn the page and begin the next problem. For four problems, the learners needed to come up with a first translation step from a verbally presented situation to a mathematical notation; three problems were intra-mathematical (indicate a first step toward solving an equation or an equation system); for three problems, the learners needed to interpret a mathematical solution with respect to an authentic situation. For each correct first step, the learners were given 1 point (partial credit for partially correct solutions); thus, the maximum score for the pretest was 10 points. Writing down more than the first solution step did not lead to a higher score.

Subjective evaluation. After the learning phase, the learners answered the item “I like problems such as those on the last four pages” on a 4-stage rating-scale (1 indicating “yes” and 4 indicating “no”).

Post-test: Assessment of learning outcomes. The post-test addressed the following aspects:

Translation. In this part of the post-test, two problems were presented to the learners where they had to find a translation from a verbal description of a real situation to a mathematical notation; e.g., *Sabine wants to sew 3 pillowcases. She buys cloth with blue stars for € 8.50. In addition, she buys 2 packages of decorative ribbon for € 2.50 each and 3 zippers which cost € 3.00 each. How much does she have to pay in total? Please write down an equation for solving this problem (you do not need to perform the calculation).* For each problem, a maximum of 1 point was given (partial credit for partially correct answers); thus, the maximum score for the post-test category “translation” was 2 points.

Interpretation. This category comprised two problems where the learners had to interpret given mathematical results with respect to a real situation; e.g., *Max wants to buy roses for his mother. He figures out that his money suffices for 7.72 roses. How many roses can he buy?* Each realistic answer was awarded with 1 point; hence, the maximum score for the post-test category “interpretation” was 2 points.

Complete problems. The learners were asked to solve two complete word problems. The first one included the interpretation of a remainder: *Lars is passionate about taking pictures and wants to buy exactly 20 photo albums for his next photo projects (they should suffice for the next 6 months). There are very nice albums for € 6.00 each and very simple ones for € 1.50 each. Lars likes the expensive albums much more so that he wants to buy as many of them as possible. However, he can only spend € 70.00. Thus, how many expensive and how many low-priced albums should he buy?* The second problem included estimation: *A chemistry teacher caused an explosion in his preparation for the next lesson, and materials were damaged which should have been distributed in two classes. The school secretary then ordered new materials. For the class 10a she paid € 115.00 for a preparation CD, 20 workbooks and an answer booklet; for the class 10b she paid € 140.00 for two preparation CDs and 24 workbooks. In addition to the exact amount of damage, the insurance wants an approximate breakdown of the loss occurred. The school secretary knows that answer booklets in chemistry cost*

Table 1. Example problem from the learning phase (translated from German).

Problem formulation (same for all conditions):

Yasmin wants to invite 12 friends to a pool party. Therefore, she wants to buy 15 lampions. In the store she finds very nice lampions which cost € 4.00 each and simple lampions for € 1.00 each. Yasmin likes the expensive lampions so much that she wants to buy as many as possible. She has only € 44.00 with her; thus, she also has to buy some cheap lampions. Hence, how many expensive and how many cheap lampions should she buy?

Translation and interpretation choice

For the first solution step, you are given three alternatives. Please find out how the solution should begin:

A Let x be the number of expensive lampions and y the number of cheap lampions. Then it must hold:
 $x + y = 15$
 $4 \cdot x + y = 44$

B Let x be the number of expensive lampions and y the number of cheap lampions. Then it must hold:
 $x + y = 15$
 $x + 4 \cdot y = 44$

C Let x be the price for an expensive lampion (in €) and y the price for a cheap lampion (in €). Then it must hold:
 $x + y = 15$
 $4 \cdot x + y = 44$

Which solution begin is correct? [free space for writing down]

Please find the following solution steps on your own. [free space for writing down]

For the answer, you are given three alternatives. Please find out what the answer should be:

A Let us assume that it results: $x = 9.67$ and $y = 5.33$.
 Thus, Yasmin should buy 10 expensive lampions and 5 cheap lampions.

B Let us assume that it results: $x = 9.67$ and $y = 5.33$.
 Thus, Yasmin should buy 9 expensive lampions and 6 cheap lampions.

C Let us assume that it results: $x = 9.67$ and $y = 5.33$.
 Thus, an expensive lampion costs € 9.67 and a cheap lampion costs € 5.33.

Which answer is correct? [free space for writing down]

Translation choice only

For the first solution step, you are given three alternatives. Please find out how the solution should begin:

A Let x be the price for an expensive lampion (in €) and y the price for a cheap lampion (in €). Then it must hold:
 $x + y = 15$
 $4 \cdot x + y = 44$

B Let x be the number of expensive lampions and y the number of cheap lampions. Then it must hold:
 $x + y = 15$
 $4 \cdot x + y = 44$

C Let x be the number of expensive lampions and y the number of cheap lampions. Then it must hold:
 $x + y = 15$
 $x + 4 \cdot y = 44$

Which solution begin is correct? [free space for writing down]

Please find the following solution steps on your own. [free space for writing down]

Interpretation choice only

Please find the solution on your own. [free space for writing down]

For the answer, you are given three alternatives. Please find out what the answer should be:

A Let us assume that it results: $x = 9.67$ and $y = 5.33$.
 Thus, Yasmin should buy 9 expensive lampions and 6 cheap lampions.

B Let us assume that it results: $x = 9.67$ and $y = 5.33$.
 Thus, Yasmin should buy 10 expensive lampions and 5 cheap lampions.

C Let us assume that it results: $x = 9.67$ and $y = 5.33$.
 Thus, an expensive lampion costs € 9.67 and a cheap lampion costs € 5.33.

Which answer is correct? [free space for writing down]

No choice

Please find the solution on your own. [free space for writing down]

about the same as in physics and biology – last week, she ordered one for € 4.90 and one for € 5.10. Thus, she can estimate the costs for the answer booklets, but she does not know the prices for the workbooks and the CDs. As she cannot reach anyone in the mail order by phone, she estimates the price for an answer booklet and then calculates the prices for the CDs and the workbooks. Show that you can do that as well!

For each problem, the learners could get a maximum of one point for the translation part, and a maximum of one point for the interpretation part, respectively (partial credit for partially correct answers).

Procedure.

First, the participants worked on the pretest (experimenter-paced; 20 seconds per problem). Subsequently, in the learning phase, the experimental variation took place. In order to avoid learners getting stuck in the first problem and having no time left for the following three problems, the learners first got 9 minutes for problems 1 and 2 (and were allowed to move back and forth between them); afterwards, they got again 9 minutes for problems 3 and 4 (and were allowed to move back and forth between them). Thus, it was ensured that each learner at least tried one problem of the “estimation” type and one problem of the “interpret a remainder” type. Afterwards, the learners were asked to give a subjective evaluation. Subsequently, they worked on the post-test (20 minutes). Lastly, the participants indicated their age and gender. The individual learning time and the time for working on the tests was kept constant. In total, the experiment took about 45 minutes.

Results

With respect to the pretest, no significant differences between the experimental groups were found, main effect “translation choice”: $F(1, 143) = .62, p = .433$, main effect “interpretation choice”: $F(1, 143) = .62, p = .433$, interaction effect: $F(1, 143) = .19, p = .667$. Hence, the learners in all experimental groups had approximately the same level of prior knowledge.

Table 2 shows the means (and standard deviations) of the prior knowledge and post-test scores in the experimental groups. No heterogeneous slopes from the post-test measures to the pretest score were found, hence, the pretest score entered GLM models as a continuous covariate. The analyses were computed as two-factorial analyses of variance, with “translation choice” and “interpretation choice” as independent variables and the pretest score as covariate.

Table 2. Means and standard deviations (in parentheses) of the pretest and post-test scores.

	With translation choice		Without translation choice	
	With interpretation choice	Without interpretation choice	With interpretation choice	Without interpretation choice
Pretest (max: 10)	6.20 (1.95)	6.54 (1.52)	6.54 (1.77)	6.64 (1.44)
Translation				
Separate items (max: 2)	.80 (.73)	1.00 (.79)	1.05 (.74)	1.03 (.84)
Complete problem “interpretation of a remainder” (max: 1)	.23 (.37)	.25 (.35)	.11 (.27)	.11 (.24)
Complete problem “estimation” (max: 1)	.26 (.39)	.20 (.36)	.21 (.36)	.39 (.42)
Interpretation				
Separate items (max: 2)	1.57 (.60)	1.76 (.55)	1.65 (.63)	1.83 (.45)
Complete problem “interpretation of a remainder” (max: 1)	.08 (.28)	.08 (.25)	.19 (.40)	.25 (.42)
Complete problem “estimation” (max: 1)	.07 (.21)	.05 (.20)	.03 (.11)	.08 (.19)

Translation.

With respect to the separate post-test “translation” items, neither the main effect “translation choice” nor the main effect “interpretation choice” reached the level of significance, $F(1, 142) = .72, p = .398$,

and $F(1, 142) = .16, p = .694$, respectively. The interaction effect was not significant, $F(1, 142) = .58, p = .449$. As expected, the influence of prior knowledge was significant, $F(1, 142) = 23.45, p < .001, \text{partial } \eta^2 = .142$, large effect. Concerning the complete problems, it should be noted that the scores in both complete problems were low in general, indicating that they were indeed very challenging for the learners. Both problems were analyzed separately as they addressed different aspects (interpreting a remainder realistically versus estimate missing information realistically). Concerning the translation part of the “interpretation of a remainder” problem, the main effect “translation choice” was significant, $F(1, 142) = 8.99, p = .003, \text{partial } \eta^2 = .060$, medium effect; indicating that learning with translation answer choices was more effective than learning without translation answer choices. However, neither the main effect “interpretation choice” nor the interaction effect reached the level of significance, $F(1, 142) = .00, p = .986$, and $F(1, 142) = .01, p = .930$, respectively. The influence of prior knowledge was again significant, $F(1, 142) = 20.07, p < .001, \text{partial } \eta^2 = .124$, medium effect.

With respect to the translation part of the “estimation” problem, the influence of prior knowledge was significant, $F(1, 142) = 38.82, p < .001, \text{partial } \eta^2 = .215$, large effect. Neither the main effect “translation choice” nor the main effect “interpretation choice” reached the level of significance, $F(1, 142) = .74, p = .392$, and $F(1, 142) = .38, p = .537$, respectively. However, the interaction effect was significant, $F(1, 142) = 5.75, p = .018, \text{partial } \eta^2 = .039$, small effect. As displayed in Figure 1, if translation answer choices were presented, it made only little difference whether interpretation answer choices were presented as well. In contrast, in the case of no translation options, it was much more beneficial to also omit the interpretation options; considerably lower results were obtained in the group with interpretation options only compared to the “no choice” group. According to a simple effects test, the difference between these two groups was actually significant, $p = .035$. In the case of “without interpretation choice”, the difference between the groups with and without translation choice was also significant, $p = .023$.

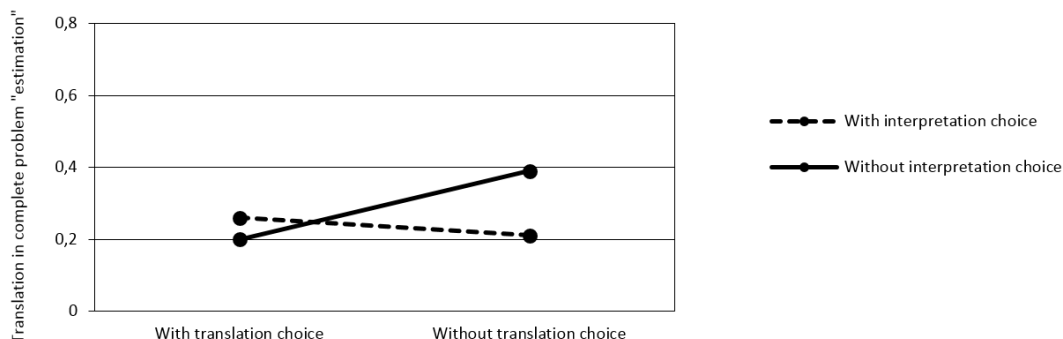


Figure 1. Post-test complete problem “estimation”: translation performance in the experimental groups.

Interpretation.

With respect to the separate post-test “interpretation” items, the main effect “translation choice” was not significant, $F(1, 142) = .30, p = .584$. However, the main effect “interpretation choice” just barely missed the level of significance, $F(1, 142) = 3.42, p = .067, \text{partial } \eta^2 = .023$, small effect; learning without interpretation choice led to descriptively better interpretation results than learning with interpretation answer choices. The interaction effect was not significant, $F(1, 142) = .04, p = .848$. Again, the influence of prior knowledge was large, $F(1, 142) = 36.81, p < .001, \text{partial } \eta^2 = .206$, large effect.

Concerning the interpretation part of the “interpretation of a remainder” problem, a significant main effect “translation choice” emerged, $F(1, 142) = 5.78, p = .018, \text{partial } \eta^2 = .039$, small effect; indicating that learning without translation choice options was more beneficial than learning with translation

choice options. In contrast, neither the main effect “interpretation choice” nor the interaction effect reached the level of significance, $F(1, 142) = .26, p = .611$, and $F(1, 142) = .30, p = .586$, respectively. The influence of prior knowledge was not significant, $F(1, 142) = .13, p = .721$.

Concerning the interpretation part of the “estimation” problem, apart from a significant influence of prior knowledge, $F(1, 142) = 4.27, p = .041, \text{partial } \eta^2 = .029$, small effect, no significant effects were obtained, $F(1, 142) = .11, p = .744$ for “translation choice”, $F(1, 142) = .12, p = .725$ for “interpretation choice”, and $F(1, 142) = 1.55, p = .216$ for the interaction effect, respectively.

Complementing the results concerning objective learning outcomes, subjective statements were analyzed. On a four-stage rating scale, the learners rated the statement “I like problems such as those on the last four pages” (1 indicating “yes” and 4 indicating “no”). The group “interpretation choice only” reached the most positive scores ($M = 2.96, SD = .77$), followed by the “no choice” ($M = 3.01, SD = .67$; one missing value), “translation and interpretation choice” ($M = 3.22, SD = .92$), and “translation choice only” ($M = 3.38, SD = .72$) groups. The main effect “translation choice” reached the level of significance, $F(1, 141) = 5.37, p = .022, \text{partial } \eta^2 = .037$, small effect. Neither the main effect “interpretation choice” nor the interaction effect were significant, $F(1, 141) = 1.01, p = .316$, and $F(1, 141) = .30, p = .585$, respectively. The influence of prior knowledge was significant, $F(1, 141) = 6.82, p = .010, \text{partial } \eta^2 = .046$, small effect.

Discussion

Translation and interpretation competencies of the participants were measured on the one hand with separate items addressing only one specific aspect; on the other hand, the learners had to solve two complete word problems addressing translation and interpretation competencies. These two complete problems were analyzed separately, as they implemented “translation” and “interpretation” differently. Whereas the “interpretation of a remainder problem” was quite straightforward with respect to the translation requirements, the interpretation part was quite demanding. In contrast, the “estimation” problem began with a challenging translation part, whereas the interpretation part was less complicated.

With respect to translation competencies as measured separately by specific “translation” problems in the post-test, no significant group differences emerged. Thus, contrary to expectation, providing multiple answer choices for the translation steps did not enhance learning. However, with respect to the translation part of the “interpretation of a remainder” problem, a significant main effect “translation choice” emerged, indicating that in case of rather straightforward translation requirements, learning with translation answer choices was indeed more effective than learning without translation answer choices. With respect to the translation performance in the “estimation” problem – addressing a much more complicated translation – , a significant interaction effect emerged, indicating that if translation answer choices were presented, it made only little difference whether interpretation answer choices were presented as well, whereas in the case of no translation options, it was much more beneficial to also omit the interpretation options; “translation choice only” and “interpretation choice only” were both significantly less effective than “no choice”. This means that a demanding translation in form of estimation seems to be supported best when the students are trained to solve problems completely on their own; in contrast, a mix of multiple choice support in one problem part with no support in the other problem part is not beneficial. However, it has to be acknowledged that the errors in the choice options were difficult to detect, so it is possible that several learners had difficulties in identifying the correct response option – which, in addition to the fact that no feedback was provided, might have led to the effect that unguided problem solving was more effective than an extensive search for an error.

In the context of the translation options it should be noted that the learners could use them in different ways. To begin with, learners could try to find their own translation from the situation to a mathematical model, and identify and tick the closest related translation in the answer choices. They also could use the answer choices in order to get an idea what a translation could look like and to narrow down the search space. However, the answer options provided yet another opportunity: The learners could take them and try to validate whether they fulfill the requirements given in the problem description. It is a limitation of the present study that it cannot be traced back how the learners proceeded and which strategy – or combination of strategies – they chose.

With respect to the acquisition of interpretation competencies as measured separately by specific “interpretation” problems in the post-test, against expectation, presenting multiple choice options for the interpretation part in the learning phase was not effective; in fact, the effect just barely missed the level of significance, but displayed a small effect in “unexpected” direction: learning without interpretation choice led to better interpretation results than learning with interpretation answer choices. However, the scores in the separate interpretation items were very high across all groups, so a ceiling effect cannot be excluded with certainty.

With respect to the interpretation part of the “interpretation of a remainder” problem, the same effect appeared and actually reached the level of significance. What at first sight seems implausible may, however, be explained by the consideration that possibly multiple choice options make the learners lazy and “wait and see”, lowering commitment and dedication. In addition, it should be noted that the errors in the multiple choice options were anything but obvious, and just upon close inspection was it possible to determine the right option. This might have irritated the learners, leading to lower learning outcomes. Concerning the interpretation part of the “estimation” problem, no significant effects were found; indicating that with respect to generating a quite straightforward interpretation, translation and interpretation answer choices do not support learning.

The impression arises that having the possibility to just select a solution part might elicit some sort of expectations or even lethargy with respect to the rest of the solution process, possibly leading to a lower extent of personal engagement and initiative when it comes to “traditional” problem solving without support. This, in turn, might explain the surprisingly low or inconsistent effectivity of the presented multiple choice options.

Another explanation can be seen in interesting evidence from a different field of research. According to McDaniel et al. (2007), for subsequent memory performance, recall tests are more advantageous compared to recognition tests; this indicates that working on multiple choice tests enhances learning less than working on free recall tests. For the present context this means that even if translation and interpretation answer choices were intended to simplify and facilitate solving complex word problems, the skill acquisition might as well be fostered effectively by just working on problems like in a “free recall” condition.

In sum, it can be concluded that effects of translation and interpretation answer choices depend on the concrete requirements of the presented problems; results cannot be generalized across different levels of difficulty. With respect to the acquisition of translation skills, the results are inconsistent, translation answer choices being helpful when it comes to rather straightforward translation requirements in complete problems. With respect to the acquisition of interpretation competencies it can be concluded that they are fostered especially effectively – at least in most cases – when no interpretation answer choices are presented.

However, even if objective learning outcomes are without any doubt very important, subjective statements of the learners are critical as well. A learning method may be very effective in a laboratory situation; however, that does not mean a lot when it comes to mathematics learning in “real” settings: When learners do not like what they are intended to do, then the resulting lack of motivation could

easily even out the advantages of the learning method. Thus, subjective attitudes of the learners toward the problems presented in the learning phase were assessed as well. The presentation of “interpretation choice only” scored best; however, the factor “interpretation choice” was not significant. In contrast, translation answer choices actually reduced the acceptance. Apparently, learners prefer to start problems on their own, compared to being required to work on multiple choice options. The response options concerning the translation from the verbal description to a mathematical notation differed only with respect to details, so that the differences between them might have been difficult to detect, which in turn might have been confusing and unnerving for the learners. In contrast, the response options concerning the interpretation of the mathematical solution with regard to the real situation differed in a more eye-catching way. Thus, if the differences between the multiple choice options for the translation step had been more eye-catching, more positive reactions might have been provoked.

With respect to “real” mathematics education, it can be concluded that presenting translation answer choices may enhance translation competencies; however, at least if they are designed as in the present experiment, learners do not appreciate them very much. In contrast, interpretation choice options are appreciated by the learners – at least when presented exclusively – but they do not support learning. Overall, the effectivity of translation and interpretation options is falling short of expectations, and subjective results are not consistent with objective scores. The present experiment shows that with respect to fostering transitions between reality and mathematics, objective learning outcomes and subjective scores do not necessarily correspond. This aspect is of major importance and demonstrates the importance of putting the discussion on fostering transition competencies in mathematics education on a firmer footing.

Acknowledgements

I would like to thank the German Research Foundation (DFG; contract number GR2706/4-1) and the Central Research Development Fund (CRDF) of the University of Bremen for their support. And special thanks go to Bàra W. Grollius for her assistance.

References

- Adams, D. M., McLaren, B. M., Durkin, K., Mayer, R. E., Rittle-Johnson, B., Isotani, S., and van Velsen, M. (2014). Using erroneous examples to improve mathematics learning with a web-based tutoring system. *Computers in Human Behavior*, 36, 401-411. doi: 10.1016/j.chb.2014.03.053
- Atkinson, R. K., Derry, S. J., Renkl, A., and Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of educational research*, 70, 181-214.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Bandura, A., and Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology*, 88, 87-99. doi: 10.1037/0021-9010.88.1.87.
- Blum, W., and Borromeo Ferri, R. (2009). Mathematical modelling: Can it be taught and learnt? *Journal of Mathematical Modelling and Application*, 1, 45-58.
- Booth, J. L., Lange, K. E., Koedinger, K. R., and 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. doi: 10.1016/j.learninstruc.2012.11.002.
- Borromeo Ferri, R. (2009). Zur Entwicklung des Verständnisses von Modellierung bei Studierenden [On the development of students' understanding of modeling]. In M. Neubrand (Hrsg.), *Beiträge zum Mathematikunterricht 2009* (S. 139-142). Münster: WTM.
- De Kock, W. D., and Harskamp, E. G. (2014). Can teachers in primary education implement a metacognitive computer programme for word problem solving in their mathematics classes? *Educational Research and Evaluation*, 20, 231-250.
- Durkin, K., and Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction*, 22, 206-214. doi: 10.1016/j.learninstruc.2011.11.001.
- Galbraith, P. L., and Clatworthy, N. J. (1990). Beyond standard models – Meeting the challenge of modelling. *Educational Studies in Mathematics*, 21, 137-163.
- Giannakos, M. N. (2013). Enjoy and learn with educational games: Examining factors affecting learning performance. *Computers & Education*, 68, 429-439. doi: 10.1016/j.compedu.2013.06.005

- Greer, B. (1997). Modelling reality in mathematics classrooms: The case of word problems. *Learning and Instruction*, 7, 293-307.
- Große, C. S. (2014). Learning to solve story problems – supporting transitions between reality and mathematics. *European Journal of Psychology of Education*, 29, 619–634. doi: 10.1007/s10212-014-0217-6.
- Große, C. S., and Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, 17, 612-634. doi: 10.1016/j.learninstruc.2007.09.008.
- Heemsoth, T., and Heinze, A. (2014). The impact of incorrect examples on learning fractions: A field experiment with 6th grade students. *Instructional Science*, 42, 639-657. doi: 10.1007/s11251-013-9302-5.
- Kalyuga, S. (2008). When less is more in cognitive diagnosis: A rapid online method for diagnosing learner task-specific expertise. *Journal of Educational Psychology*, 100, 603-612. doi: 10.1037/0022-0663.100.3.603.
- Kalyuga, S., and Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, 96, 558–568. doi: 10.1037/0022-0663.96.3.558.
- Lave, J., and Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., and Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19, 494-513. doi: 10.1080/09541440701326154.
- McLaren, B. M., Adams, D., Durkin, K., Gogvadze, G., Mayer, R. E., Rittle-Johnson, B., Sosnovsky, S., Isotani, S., and 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 *21st Century Learning for 21st Century Skills* (pp. 222-235). Berlin: Springer.
- Mevarech, Z. R., Terkieltaub, S., Vinberger, T., and Nevet, V. (2010). The effects of meta-cognitive instruction on third and sixth graders solving word problems. *ZDM Mathematics Education*, 42, 195-203. doi: 10.1007/s11858-010-0244-y.
- Mwangi, W., and Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction*, 16, 173-199. doi: 10.1207/s1532690xci1602_2.
- OECD (2003). *The PISA 2003 assessment framework – Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.
- Paas, F., Renkl, A., and Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38, 1-4.
- Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics*, 67, 37-58. doi: 10.1007/s10649-007-9083-3.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational psychology review*, 18, 315-341. doi: 10.1007/s10648-006-9029-9
- Powell, S. R. (2011). Solving word problems using schemas: A review of the literature. *Learning Disabilities Research & Practice*, 26, 94-108.
- Reinmann, G., and Mandl, H. (2006). Unterrichten und Lernumgebungen gestalten [Teaching and designing learning environments]. In A. Krapp and B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (pp. 613-658). Weinheim: Beltz PVU.
- Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, 21, 1-29.
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38, 1-37. doi: 10.1111/cogs.12086
- Renkl, A., Atkinson, R. K., and Große, C. S. (2004). How fading worked solution steps works - A cognitive load perspective. *Instructional Science*, 32, 59-82.
- Reusser, K., and Stebler, R. (1997). Every word problem has a solution—the social rationality of mathematical modeling in schools. *Learning and Instruction*, 7, 309-327.
- Schukajlow, S., and Blum, W. (2011). Zum Einfluss der Klassengröße auf Modellierungskompetenz, Selbst- und Unterrichtswahrnehmungen von Schülern in selbstständigkeitsorientierten Lehr-Lernformen [On the effect of class size on modeling competency and self-reported perceptions of students in self-regulated learning environments]. *Journal für Mathematik-Didaktik*, 32, 133-151. doi: 10.1007/s13138-011-0025-3.
- Schukajlow, S., Krug, A., and Rakoczy, K. (2015). Effects of prompting multiple solutions for modelling problems on students' performance. *Educational Studies in Mathematics*, 89, 393-417. doi: 10.1007/s10649-015-9608-0
- Schukajlow, S. and Rakoczy, K. (2016). The power of emotions: Can enjoyment and boredom explain the impact of individual preconditions and teaching methods on interest and performance in mathematics? *Learning and Instruction*, 44, 117-127.
- Stark, R., Gruber, H., Renkl, A., and Mandl, H. (2000). Instruktionale Effekte einer kombinierten Lernmethode: Zahlt sich die Kombination von Lösungsbeispielen und Problemlöseaufgaben aus? [Instructional effects of a combined learning method: How effective is the combination of worked examples and problems to-be-solved?]. *Zeitschrift für Pädagogische Psychologie*, 14, 206-218.
- Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review*, 22, 123-138. doi: 10.1007/s10648-010-9128-5.
- Trafton, J. G., and Reiser, B. J. (1993). The contributions of studying examples and solving problems to skill acquisition. In M. Polson (Ed.), *Proceedings of the 15th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Van Gog, T., Paas, F., and van Merriënboer, J. J. G. (2004). Process-oriented worked examples: improving transfer performance through enhanced understanding. *Instructional Science*, 32, 83-98.
- Van Gog, T., Paas, F., and van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction*, 16, 154-164. doi: 10.1016/j.learninstruc.2006.02.003.
- Van Gog, T., Paas, F., and van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction*, 18, 211-222. doi: 10.1016/j.learninstruc.2007.03.003.

- Verschaffel, L., De Corte, E., and Borghart, I. (1997). Pre-service teachers' conceptions and beliefs about the role of real-world knowledge in mathematical modelling of school word problems. *Learning and Instruction*, 7, 339-359.
- Verschaffel, L., De Corte, E., and Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction*, 4, 273-294.
- Verschaffel, L., De Corte, E., and Vierstraete, H. (1999). Upper elementary school pupils' difficulties in modeling and solving nonstandard additive word problems involving ordinal numbers. *Journal for Research in Mathematics Education*, 30, 265-285.
- Vye, N. J., Goldman, S. R., Voss, J. F., Hmelo, C., Williams, S., and Cognition and Technology Group at Vanderbilt (1997). Complex mathematical problem solving by individuals and dyads. *Cognition and Instruction*, 15, 435-484. doi: 10.1207/s1532690xci1504_1.
- Yerushalmy, M. (1997). Mathematizing verbal descriptions of situations: A language to support modeling. *Cognition and Instruction*, 15, 207-264. doi: 10.1207/s1532690xci1502_3.
- Zöttl, L., Ufer, S., and Reiss, K. (2010). Modelling with heuristic worked examples in the KOMMA learning environment. *Journal für Mathematik-Didaktik*, 31, 143-165. doi: 10.1007/s13138-010-0008-9.