



# Learner-generated graphic representations for word problems: an intervention and evaluation study in grade 3

Barbara Ott<sup>1</sup> 

Published online: 3 September 2020

© The Author(s) 2020

## Abstract

The use of written or graphic representations is essential in mathematics. Graphic representations are mainly used and researched as instruments for problem solving. There is a gap in research for interventions that use learner-generated graphic representations as documents for reflection processes for promoting the development of children's graphic representation competences. This is the focus of the study presented here. The study examines to what extent such an intervention has an effect on how the children take into account a mathematical structure in their self-generated graphic representations, how they ensure a mathematical matching with the word problem, and what degree of abstraction they choose. Additionally, the effect on the solution rates is investigated. The results show that children in the intervention group more frequently pay attention to a mathematically appropriate structure, compared with children in the control groups. This result is statistically significant. At the same time, children keep the degree of abstraction relatively constant. Solution rates improve continuously, but the difference is not significant.

**Keywords** Inscriptions · Graphic representations · Word problems · Reflection · Intervention study

## 1 Introduction

In the practice of mathematics, the use of inscriptions, i.e., representations that exist in material form (Roth & McGinn, 1998), is essential. Without such material representations, it is virtually impossible to acquire a mathematical understanding (Dörfler, 2008; Goldin & Shteingold, 2001). Accordingly, representation is now normatively set as a mathematical competence in the standards and curricula of many countries (e.g., NCTM, 2000). This competence is often

---

✉ Barbara Ott  
barbara.ott@phsg.ch

<sup>1</sup> St. Gallen University of Teacher Education, Notkerstrasse 27, 9000 St. Gallen, Switzerland

associated with problem solving. Furthermore, in research, the flexible and adaptive use of different representations is considered essential for mathematical problem solving (Heinze, Star, & Verschaffel, 2009). The generation of a graphic representation is regarded as an important heuristic (Hembree, 1992). At the same time, it is often reported that many learners rarely use graphic representations as heuristics (Fagnant & Vlassis, 2013; Lopez Real & Veloo, 1993). This is particularly evident at the primary school level, where interventions involving graphic representations for problem solving often have almost no positive effects (Hembree, 1992).

Thus, a tension between mathematically and didactically motivated ideas on the one hand and the use by learners at the primary school level on the other hand becomes apparent. In intervention studies, graphic representations are mostly examined as instruments for problem solving (e.g., Van Essen & Hamaker, 1990). There is a gap in the research on the question as to how learner-generated graphic representations develop when they are created as documents for others and reflected upon in class. The study described here makes a contribution in this direction.

## 2 Theoretical background

### 2.1 Graphic representations for word problems

Various forms of representations can be distinguished (Goldin & Shteingold, 2001). A distinction between descriptive and depictive representations (Schnotz, 2002) seems to be useful in the case of graphic representation for word problems. We will see below that word problems are a particular example of descriptive representations, while graphic representations form a subset of depictive representations.

Descriptive representations consist of symbols and are associated with the content they represent by means of a convention. Descriptive representations contain relational characters for structural mapping (Schnotz, 2001). For example, texts or mathematical equations are descriptive representations. For describing something in texts, nouns are put by verbs and prepositions in relation to each other. In contrast, the signs used in depictive representations “are associated with the content they represent through common structural features” (Schnotz, 2002, p. 103). Depictive representations have structural properties that correspond to properties of the facts to be presented. In contrast to descriptive representations, information can be taken directly. Thus, they enable efficient conclusion processes and are particularly well suited to producing inferences in the observer, which is important for problem solving (Larkin & Simon, 1987; Schnotz, 2001).

Word problems are forms of descriptive representations. They can be understood as tasks presented in text form, where the content is largely meaningless and interchangeable (Schipper, 2009). The focus is on the verbally described mathematical relationships. Figure 1 shows an example of a word problem. Semantically different word problems can describe the same mathematical relationships and lead to the same mathematical operations (Verschaffel, Greer, & De Corte, 2000). Examples used in this study can be found in Table 6 in the Appendix. A change of the linguistic surface structure can influence the degree of difficulty. If, for example, the information in the text is mentioned in the order necessary for processing, this has a positive effect on the solution process (Stern, 1998). Word problems are often criticized (Verschaffel et al., 2000): One

A spruce grows about 15 cm per year. In the garden there is an 83 cm high spruce. How old is it?

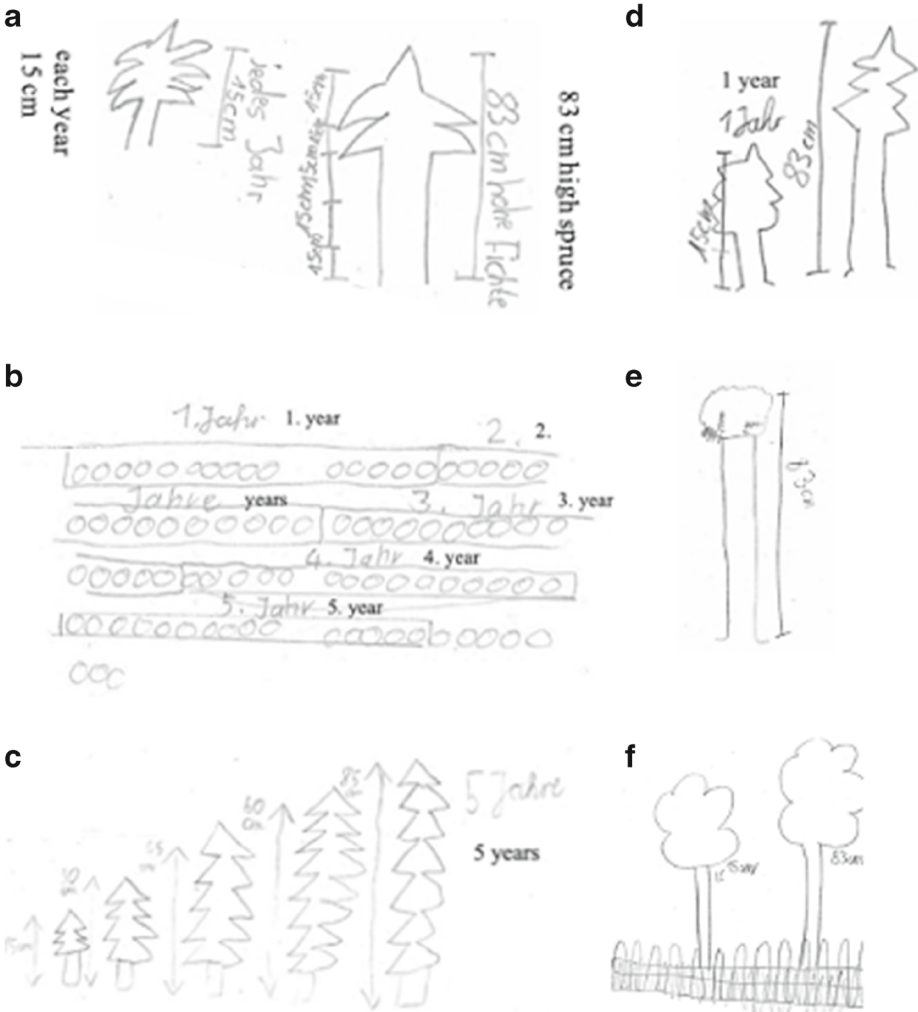


Fig. 1 Graphic representations (with translation) and the associated word problem (Ott, 2016)

main criticism is that word problems that are used in classroom practices often do not have any genuine references to reality, but are artificial problems. At the same time, it is reported that students often “solve these problems in a stereotyped and artificial way” (Verschaffel et al., 2000, p. 12). However, depending on how they are used in teaching, word problems also have the potential to develop several mathematical competences (Verschaffel et al., 2000). For example, they can be used to “develop new mathematical structures, notations, etc in the course of exploring the modeling of phenomena” (Verschaffel et al., 2000, p. 173).

Graphic representations are forms of depictive representations. The term *graphic* refers to representations consisting of lines and dashes (Cox, 1999). Sketches and drawings can be understood as graphic representations. In this study, *graphic representation* refers to paper and pencil representations. Graphic representations are characterized by aspects of space that are

mapped onto content elements (Stern, Aprea, & Ebner, 2003). In graphic representations, therefore, due to the position of the individual elements, relationships become directly apparent (Larkin & Simon, 1987).

In a preliminary study (Ott, 2016), three key features of graphic representations for word problems were identified: *mathematical structure*, *mathematical matching*, and *degree of abstraction*. A *mathematical structure* may be defined based on set theory (Rinkens, 1973): Relationships between amorphous elements of a set can be determined by defining linkages on the set. A structure is thereby imposed on the set. In word problems, information is presented with quantities and nouns which are related to each other by verbs and prepositions. The word problem is thus given a mathematical structure. For a graphic representation, it is necessary to invent signs for objects, e.g., quantities and nouns, that shall be represented. Relationships between these signs are determined by the arrangement of the signs on the sheet. Since the mathematical structure is defined on the signs of these objects, they are called structurally relevant objects. Figure 1 shows six examples of graphic representations. While in a, c, d, e, and f trees or rulers are structurally relevant object signs, in b, these are circles as elements of a set. a, b, and c show the mathematical structure of a linear equation in the arrangement of the signs for the structurally relevant objects. d only shows the relationship between 15 cm and 1 year. e and f only show structurally relevant objects, e.g., for the quantity 83 cm, without an arrangement. Graphic representations with a mathematical structure can be understood as signs with a relational character, whose perceptible basis is an inscription. Therefore, they take on the character of diagrams (Dörfler, 2008). There is mathematical matching between a word problem and a graphic representation if both are “informationally equivalent” (Palmer, 1978). This is the case if both of the following conditions are satisfied: Firstly, there is a match regarding the objects. This means there is a match between the quantities that can be split into measured value (e.g., 83) and measuring unit (e.g., cm) (Kirsch, 1997) on the text side, and the signs for the structurally relevant objects on the graphic side. Secondly, there is a match regarding the operations between the verbs and prepositions on the text side and the arrangement of the signs for the structurally relevant objects on the graphic side. In a, b, d, and f of Fig. 1, there is a complete matching regarding the measured values. In these examples, all measured values given in the task (15 cm, 83 cm) can be identified in the graphic representation. In c and e, this matching is partial, since only the 15 cm (c) or the 83 cm (e) is visible, while in b the measuring unit is not considered; this is the case in the other examples. Regarding operations, there is a complete matching in a, b, and c, a partial matching in d, and no matching in e and f. According to Peschek (1988), the degree of abstraction can be characterized as being the degree of focusing on the representation of the word problem’s mathematical aspects. Two indicators are identified: a focus on the structurally relevant objects (indicator 1) and a focus on the mathematically relevant qualities of the structurally relevant objects (indicator 2). In a, b, c, d, and e of Fig. 1, indicator 1 can be considered high since no other objects are drawn. In f, it is low, because a fence is also drawn. Indicator 2 can be considered high in b and low in the other examples, because detailed trees are drawn. The key features can be analyzed independently of each other. However, the prerequisite for determining the mathematical matching and the degree of abstraction is that structurally relevant objects can be identified in the graphic representation.

## 2.2 Types of graphic representations for word problems

In the literature, different types of graphic representations are distinguished. A distinction is often made between schematic and pictorial representations (Hegarty & Kozhevnikov, 1999; Presmeg, 1986): In schematic representations, the focus is on the spatial relations described in a problem; in pictorial representations, the focus is on the visual appearance of the objects described in a problem (Hegarty & Kozhevnikov 1999). Rellensmann, Schukajlow, and Leopold (2017) make a similar distinction regarding the process of mathematical modeling by distinguishing between mathematical drawings and situational drawings: In mathematical drawings, the focus is on the mathematical model described in the problem; in a situational drawing, the situation described in the problem is pictorially depicted.

If the learners are responsible for both the process of generation of the graphic representation and the final product, we speak of learner-generated graphic representations (Van Meter & Garner, 2005). With regard to learner-generated graphic representations, a study by Sherin (2000) indicates that these graphic representations can differ in even more diverse ways. In a preliminary study (Ott, 2016, 2017), the following learner-generated graphic representations for word problems can be distinguished at the primary school level:

- *Off the text*: there is no link to the text with regard to the content.
- *Illustrative*: there is a link to the text, but no structurally relevant objects are represented.
- *Object-related*: there is a link to the text and structurally relevant objects are represented although relations between them are not identifiable in the arrangement.
- *Diagrammatic*: there is a link to the text, structurally relevant objects are represented, and relations between them are identifiable in the arrangement.

Figure 1 shows four examples of graphic representations that are diagrammatic (a, b, c, d) and two examples that are object-related (e, f). Figure 2 shows a graphic representation for the word problem that is off the text (a) and illustrative (b).

Graphic representations that represent parts of the mathematical structure (*object-related, diagrammatic*) can differ in their degree of mathematical matching and the degree of abstraction (Ott, 2016, 2017). The mathematical matching may be *complete, partial, or nonexistent* in terms of measured values, measuring units, or operations. The degree of abstraction can be either *high* or *low* with regard to the two indicators.

A spruce grows about 15 cm per year. In the garden there is an 83 cm high spruce. How old is it?



a



b

Fig. 2 Graphic representations for the given word problem that are off the text (a) and illustrative (b)

### 2.3 Teaching graphic representations for word problems

With regard to studies on generating graphic representations for word problems, two types can be distinguished (Fagnant & Vlassis, 2013): encouraging students to use specified diagram types (e.g., Diezmann, 2002) and encouraging students to generate their own graphic representations (e.g., Van Dijk, Van Oers, & Terwel, 2003a). The results are inconsistent in terms of positive benefits. Some studies show that children often find it difficult to use predefined representations (Fagnant & Vlassis, 2013; Pantziara, Gagatsis, & Elia, 2009). Pantziara et al. (2009) suggest that the diagrams did not fit the learners' personal preferences and mental models and conclude that the interpretation of diagrams is also essential for improving diagram competence. They report that students often tried to transform the given diagrams into pictorial representations to interpret them. Van Dijk et al. (2003a) and Van Dijk, Van Oers, Terwel, and Van den Eeden (2003b) compared the two approaches in the fifth grade and found better results when learners generated their own graphic representations. They conclude "that designing models in co-construction may lead to a deeper insight into the meaning and use of models and consequently make possible a more flexible approach in problem solving" (Van Dijk et al., 2003b). What is so far unclear is the effect of teaching that combines the generation of one's own graphic representation with the interpretation of given graphic representations.

### 2.4 Learner-generated graphic representations and problem solving

Learner-generated graphic representations can fulfill two functions (Selter, 1993): as instruments, they function as an aid for problem solving in the sense of a private representation; as documents, they record the results and the solutions. In the latter function, they are public and related to an addressee. Such graphic representations are more complete, more richly inscribed, and more conventional than private representations (Cox, 1999).

The type of graphic representation employed by the student seems to have an influence on the success of problem solving: While there is a positive connection between schematic representation and successful problem solving, the generation of pictorial representations is negatively connected with problem solving success (Hegarty & Kozevnikov, 1999). These findings are relativized by results regarding modeling tasks: According to Rellensmann et al. (2017), both situational and mathematical drawings are related to modeling performance but in different ways. While the accuracy of mathematical drawings is directly related to modeling performance, the accuracy of the situational drawing is indirectly related, mediated by the accuracy of mathematical matching.

Learners hardly use sketches for problem solving (Fagnant & Vlassis, 2013; Lopez Real & Veloo, 1993; Van Essen & Hamaker, 1990). Lopez Real and Veloo (1993) report that the request to draw a sketch improves solution rates. In most studies, however, this improvement does not occur (Hembree, 1992). Hembree's meta-analysis showed that the use of graphic representations in problem solving can be trained. Comparing different instructional methods, training in drawing diagrams offered the largest improvement in problem solving. However, these positive effects do not yet occur during primary school age (Hembree, 1992). Accordingly, intervention studies by Van Essen and Hamaker (1990) showed that fifth graders could benefit from training in generating graphic representations for problem solving, while first and second graders could not. Similarly, Hembree (1992) concludes "that earlier grades should focus on problem representation instead of stressing solutions" (p. 269). Studies show that primary school children often find it difficult to represent mathematical relationships, preferring instead to illustrate the content of the

task (Hasemann, 2006; Ott, 2016). When learners generate graphic representations that contain a great deal of mathematically irrelevant information, this can be problematic because it can make it difficult for them to recognize mathematical structures (Presmeg, 1986). To improve children's representation skills, an as-yet unexplored approach can be seen in generating graphic representations, not as instruments for problem solving but as documents for later reflection processes in class.

### 3 Research questions

The following research questions will be examined:

Does an intervention in the 3rd grade based on reflective discussions about children's graphic representations for word problems

1. Have a positive effect on the attention paid to the key features of graphic representations (mathematical structure, mathematical matching, degree of abstraction) in the learner-generated graphic representations?
2. Have a positive effect on the solution rates?

The questions focus on measurable changes based on the children's documents. Studies show that generating graphic representations in co-construction leads to learners' representations becoming less realistic and more focused on mathematical relationships (Van Dijk et al., 2003a). With regard to problem solving, studies show positive effects of interventions that encourage reflection on different approaches in problem solving (Sturm, 2018). Since the intervention in the study presented here combines the generation of one's own graphic representation with the reflections on graphic representations, it is deduced that the intervention evaluated here has a positive effect on the attention paid to the key features of graphic representations by the 3rd graders. It is therefore assumed that the students generate mathematically correct, abstract graphic representations that are appropriate to the problem. Since there is a positive connection between schematic representation and successful problem solving (Hegarty & Kozhevnikov, 1999), it is deduced that the intervention evaluated here has a positive effect on the solution rates.

The following hypotheses will be tested:

- H1: After the intervention, children in the intervention group pay more attention to the key features of graphic representations, i.e.,
- (a) The mathematical structure
  - (b) The mathematical matching
  - (c) A high degree of abstraction

in their documents than children in the control groups.

- H2: Children in the intervention group solve word problems more often correctly after the intervention than children in the control groups.

## 4 Method

### 4.1 Design

The study is designed with three groups (intervention group and two control groups) and three waves of measurement (pretest, posttest, follow-up test). A total of nine intervention units were carried out. The teaching in the intervention group was carried out by the author. In the control groups, the usual mathematics lessons were continued by the class teachers. The teachers in control group 1 were additionally given the word problems from the intervention to use them in their lessons. They received no further training.

### 4.2 Participants

The study was conducted in the 3rd grade. Two classes from each of three primary schools took part and formed one of the three groups. All schools are from a suburban location. The participants from the intervention group comprised 35 children (18 boys, 17 girls). The average age at the first wave of measurement was 8 years and 4 months (youngest child: 7 years and 5 months; oldest child: 9 years). Control group 1 consisted of a total of 43 children (21 boys, 22 girls). The average age was 8 years and 4 months (youngest child: 7 years and 8 months; oldest child: 9 years and 10 months). Control group 2 consisted of a total of 46 children (11 male and 35 female). The average age was 8 years and 4 months (youngest child: 7 years and 6 months; oldest child: 10 years and 2 months). A total of 33 children in each group took part during all six test days.

### 4.3 Items

Both the intervention items and the test items were developed as word problems based on schoolbook tasks according to the curriculum. Main criterion for the development was the extent to which the formulation of the text suggests the graphic representation of the mathematical structure. Three types can be distinguished (see Table 1): Word problems of type A are characterized by the fact that the structurally relevant objects can be directly drawn, and their arrangement is described in the text. In contrast, in word problems of type C, the structurally relevant objects are not directly drawable by their physical properties. For a graphic representation, signs for these objects and their arrangement must be invented. For type B tasks, some but not all structurally relevant objects can be drawn directly.

Another criterion was to develop challenging tasks for the learners through the arithmetic or semantic structures of the tasks. For example, tasks for comparison situations or linear equations were developed. An overview of all items used in the intervention and the paper pencil test can be found in Table 6 in the Appendix.

**Table 1** Examples for three types of word problems

Type A	Type B	Type C
A picture is 20 cm wide and 30 cm high. It is glued onto a large piece of paper in such a way that a 15-cm-wide margin remains on all sides. How wide and how high must the large paper be?	A spruce grows about 15 cm per year. In the garden there is an 83 cm high spruce. How old is it?	In a race, Tina takes 74 s. Lara needs 12 s less than Tina. Pia takes 84 s. How many seconds does Lara need? How many seconds does Pia take longer than Lara?



## 4.4 Intervention

In the intervention, reflections (Freudenthal, 1991) should enable learners to develop their competences in graphic representation. The intervention was divided into two phases, each taking place on a weekly basis (Ott, 2018). In phase 1, both classes of the intervention group received a letter from the author informing them of the word problem of the week. Each student was asked to generate a graphic representation for this given word problem containing everything that is important for him or her to understand and solve the word problem. In order for the students to make their graphic representations as complete as possible (Cox, 1999), the instruction was to generate a graphic representation that is understandable for others. In case the students were able to solve the problem, they were also required to note the solution. The students generated their graphic representations on their own without further assistance during the free-work periods, which took place every morning in the classes. Each student dropped the document with his or her own graphic representation into a class mailbox, which was emptied by the author at the end of the week. From these learner-generated graphic representations, a maximum of three graphic representations were selected by the author for phase 2 of the intervention. The selection was made so that the examples differed as much as possible in the way the drawing was done. Depending on which graphic representations were to be found in the children's documents, an attempt was made to select examples with different mathematical structure, mathematical matching, or degree of abstraction.

In phase 2, these selected graphic representations formed the basis of the reflection discussions with the whole class. Enlarged copies of the selected children's documents were attached one after the other to the blackboard. The whole class sat down in front of the blackboard. The aim of the reflection was to collectively explain the shown graphic representation and thus to try to understand the point of view of the author of the graphic representation. The author of the graphic representation was allowed, if he or she so wished, to comment on his or her graphic representation at the end of the reflection. Otherwise, the representation remained anonymous. Since the interpretation of the graphic representations is a challenge for the children, each graphic representation was first examined and analyzed individually. The reflection process was supported and stimulated by the author, in order to encourage a change in perspective (Freudenthal, 1991; Schülke, 2013): "What was the child probably thinking while making the math drawing? ... What do you particularly like about the math drawing? Why? ... What do you suppose: Why did the child draw the things in the math drawing, the relationships between things, and the labels like this?" In the course of reflection, the students also sometimes wished to improve on the graphic representations and discussed different possibilities. This was the case, for example, if the graphic representations were mathematically incomplete. Once all the graphic representations on the blackboard had been individually reflected in this way one after the other, the children were asked to compare them with each other and work out similarities and differences. The comparison in turn could relate to the way the drawing was done, how the mathematical structure was depicted, the mathematical matching or the degree of abstraction. The comparison also promotes a change in perspective and hence initiates a reflection process (Freudenthal, 1991, Schülke, 2013). The children also discussed why one or the other graphic representation fits the given word problem particularly well. In the discussions, they referred to both mathematical and content aspects of the word problem. Furthermore, they discussed the possibility to see the solution of the word problem in the graphic representation as well as the importance of the content-related references in the graphic representations. Some children preferred more abstract graphic

representations and focused on the mathematical aspects. Others preferred more concrete graphic representations as the content of the word problem was important to them.

#### 4.5 Paper and pencil test

The same paper and pencil test was performed at each test time. The test consisted of eight word problems. Three of them are shown in Table 1. The mathematical structures of six test items are similar to those of the intervention items. Two items show another mathematical structure. Four test items are formulated similarly to intervention items and differ only in terms of content, and the other four differ in both respects (see Table 6 in the Appendix). The instruction was the same as in the intervention. Testing took place on two successive days, with four test items each day. The tests were carried out by trained test leaders and the author.

#### 4.6 Analysis

In a preliminary study (Ott, 2016, 2017), an analysis tool for graphic representations for word problems was developed. The analysis tool makes it possible to clearly assign each learner-generated graphic representation to a category according to mathematical structure, mathematical matching, and the degree of abstraction (see Section 2.3). The good interrater reliability of  $K=0.81$  (mathematical structure),  $K=0.99$  (mathematical matching with regard to the measured values),  $K=0.96$  (mathematical matching with regard to the measuring units),  $K=0.99$  (mathematical matching with regard to the operations), and  $K=0.90$  (degree of abstraction) allows this analysis tool to be used in the study presented here. The objectivity of the evaluation is guaranteed by the standardized procedure specified in an analysis guideline (Ott, 2016, 2017).

### 5 Results

The 2780 documents produced by the children were encoded by two raters. The interrater reliabilities at the three test times varied between  $K=0.95$  and  $K=0.98$  and are thus to be regarded as very good. Three graphic representations in Fig. 1 show the development of the graphic representations of a particular student from pretest (e) via posttest (d) to follow-up test (a). Different qualitative developments can be seen in Ott (2016, 2020). The focus here is on the quantitative part of the study. In order to stay conservative, all statistical hypotheses were tested against an alpha level of 5% (two-tailed). Therefore, the reduced sample ( $N=33$  in each group) is used. The criterion of a sample size  $>30$  in each group, as a prerequisite for performing a two-way analysis of variance (ANOVA), is thus fulfilled (Bortz & Schuster, 2010). A factorial ANOVA was conducted for each of the three key features of graphic representations and for the solution rates to compare the main effects of *time of testing* and *group* and the interaction effect between *time of testing* and *group*. The characteristic values of interest are, if present, coded with 1, or 0 if they are not present; total values are formed for all eight items in a test time. Table 2 shows the main and interaction effects.

**Table 2** Main effects of time of testing (TZP) and group (G) and interaction effects (TZP  $\times$  G)

Key feature		Factor	<i>F</i> -value	df	<i>p</i> value	Partial eta-squared ( $\eta_p^2$ )
Mathematical structure	Structurally relevant objects and their relationships	TZP	30.020	1.845	< .001	.238
		G	6.059	2	.003	.112
		TZP $\times$ G	11.591	3.690	< .001	.195
	Relationships only	TZP	94.364	2	< .001	.496
		G	11.805	2	< .001	.197
		TZP $\times$ G	14.870	4	< .001	.237
Mathematical matching	Measured value	TZP	57.400	2	< .001	.374
		G	2.433	2	.093	.048
		TZP $\times$ G	12.267	4	< .001	.204
	Measuring unit	TZP	79.441	2	< .001	.453
		G	4.423	2	.015	.084
		TZP $\times$ G	24.353	4	< .001	.337
	Operations	TZP	82.114	2	< .001	.461
		G	11.102	2	< .001	.188
		TZP $\times$ G	14.556	4	< .001	.233
Degree of abstraction	Indicator 1	TZP	28.168	1.779	< .001	.227
		G	1.886	2	.157	.038
		TZP $\times$ G	3.060	3.559	.023	.060
	Indicator 2	TZP	13.119	2	< .001	.120
		G	0.392	2	.677	.008
		TZP $\times$ G	0.309	4	.872	.006
Solution rates	TZP	41.41	2	< .001	.301	
	G	1.46	2	.237	.030	
	TZP $\times$ G	1.47	4	.213	.030	

## 5.1 Mathematical structure

In a first step, the representation of the signs for structurally relevant objects and their arrangement is tested. In a second step, only the representation of the relationships is examined more specifically. Table 3 shows the development of the mean values of the groups in comparison. Since the Mauchly test proves that sphericity is violated, the Greenhouse-Geisser correction was utilized.

### 5.1.1 Representation of structurally relevant objects and their relationships

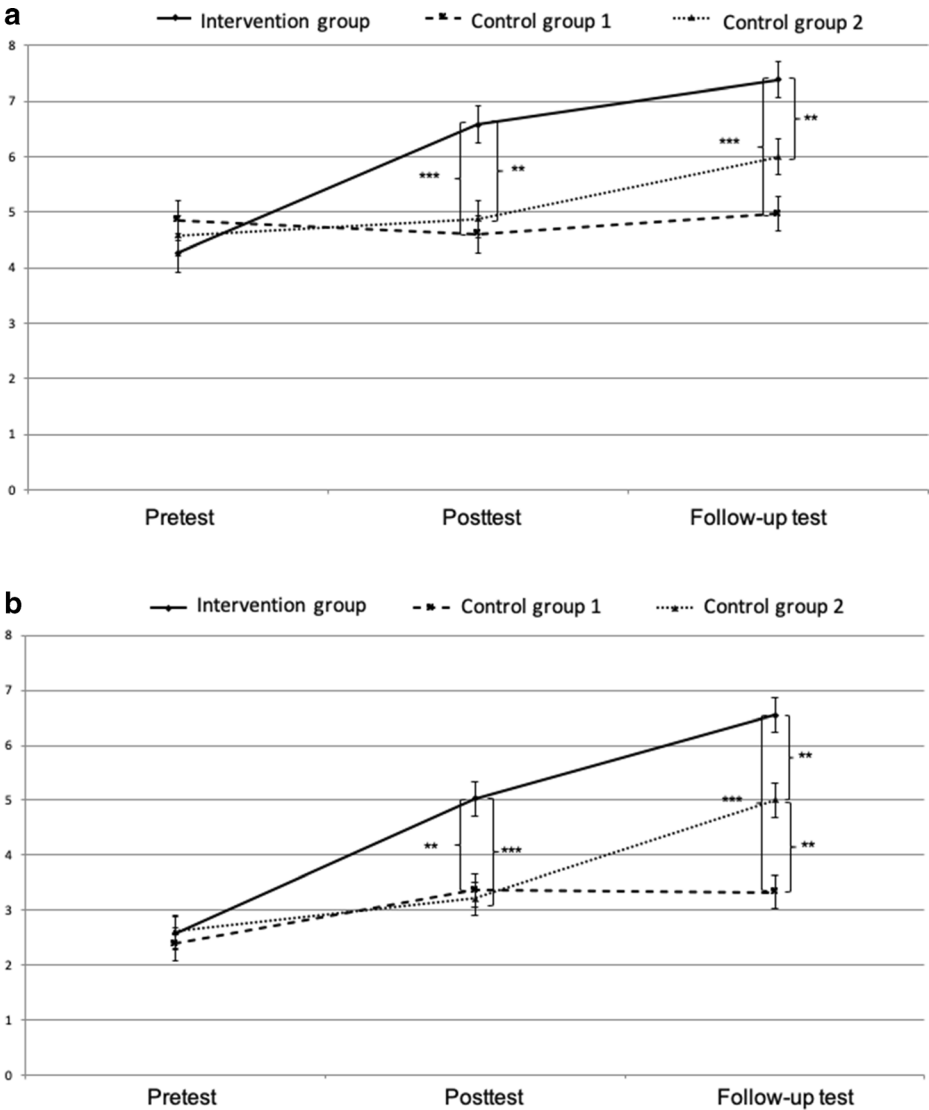
The main effects of time of testing and group are statistically significant at the .05 significance level (see Table 2). The interaction effect is significant ( $p < .001$ ) too. The effect size of the interaction ( $\eta_p^2 = .195$ ) is classified as strong (Cohen, 1988). The pairwise comparisons with the Bonferroni

**Table 3** Mean values and standard deviations with regard to the mathematical structure ( $N = 33$  in each group)

		Pretest	Posttest	Follow-up test
Structurally relevant objects and their relationships	Intervention group	<i>M</i> 4.27 ( <i>SD</i> 1.91)	<i>M</i> 6.58 ( <i>SD</i> 1.09)	<i>M</i> 7.39 ( <i>SD</i> 0.79)
	Control group 1	<i>M</i> 4.85 ( <i>SD</i> 2.12)	<i>M</i> 4.61 ( <i>SD</i> 2.28)	<i>M</i> 4.97 ( <i>SD</i> 2.37)
	Control group 2	<i>M</i> 4.58 ( <i>SD</i> 2.02)	<i>M</i> 4.88 ( <i>SD</i> 2.19)	<i>M</i> 6.00 ( <i>SD</i> 1.90)
Relationships only	Intervention group	<i>M</i> 2.58 ( <i>SD</i> 1.58)	<i>M</i> 5.03 ( <i>SD</i> 1.53)	<i>M</i> 6.55 ( <i>SD</i> 1.06)
	Control group 1	<i>M</i> 2.39 ( <i>SD</i> 1.52)	<i>M</i> 3.36 ( <i>SD</i> 1.71)	<i>M</i> 3.33 ( <i>SD</i> 1.96)
	Control group 2	<i>M</i> 2.61 ( <i>SD</i> 2.03)	<i>M</i> 3.21 ( <i>SD</i> 2.06)	<i>M</i> 5.00 ( <i>SD</i> 2.11)

correction show no significant differences ( $ps > .05$ ) at the pretest time. At the posttest time, there were significant differences between the intervention group and control group 1 ( $p < .001$ ) and between the intervention group and control group 2 ( $p = .002$ ). Significant differences between the intervention group and control group 1 ( $p < .001$ ), as well as between the intervention group and control group 2 ( $p = .007$ ), are also observed in the follow-up test. The remaining pair comparisons show no significant differences ( $ps > .05$ ) (see Fig. 3a).

Conclusively, hypothesis H1(a) can be confirmed. After the intervention, the intervention group more frequently represents structurally relevant objects and relationships in the documents than the control groups.



**Fig. 3** Group interaction with regard to the structurally relevant objects and their relationships (a) and with regard to the relationships only (b) ( $N = 33$  each group;  $**p < .01$ ,  $***p < .001$ )

## 5.1.2 Representation of relationships only

Both main effects are statistically significant at the .05 significance level too (see Table 2). The interaction effect is also significant ( $p < .001$ ). The effect size of the interaction ( $\eta_p^2 = .237$ ) is classified as strong (Cohen, 1988). At the pretest time, pairwise comparisons with the Bonferroni correction reveal no significant differences ( $ps > .05$ ). At the posttest time, there are significant differences between the intervention group and control group 1 ( $p = .001$ ) and between the intervention group and control group 2 ( $p < .001$ ). Significant differences between the intervention group and control group 1 ( $p < .001$ ), as well as the intervention group and control group 2 ( $p = .002$ ), are also found in the follow-up test. At the follow-up test time, a significant difference occurs between control group 2 and control group 1 ( $p = .001$ ). The remaining pair comparisons show no significant differences ( $ps > .05$ ) (see Fig. 3b). Hypothesis H1(a) can also be confirmed for the specification on relationships. After the intervention, the intervention group more frequently represents relationships in the documents than the control groups.

## 5.2 Mathematical matching

The mathematical matching is tested separately for measured values, measuring units, and operations. Table 4 shows the development of the mean values of the groups in comparison. The Mauchly test proves that sphericity is fulfilled.

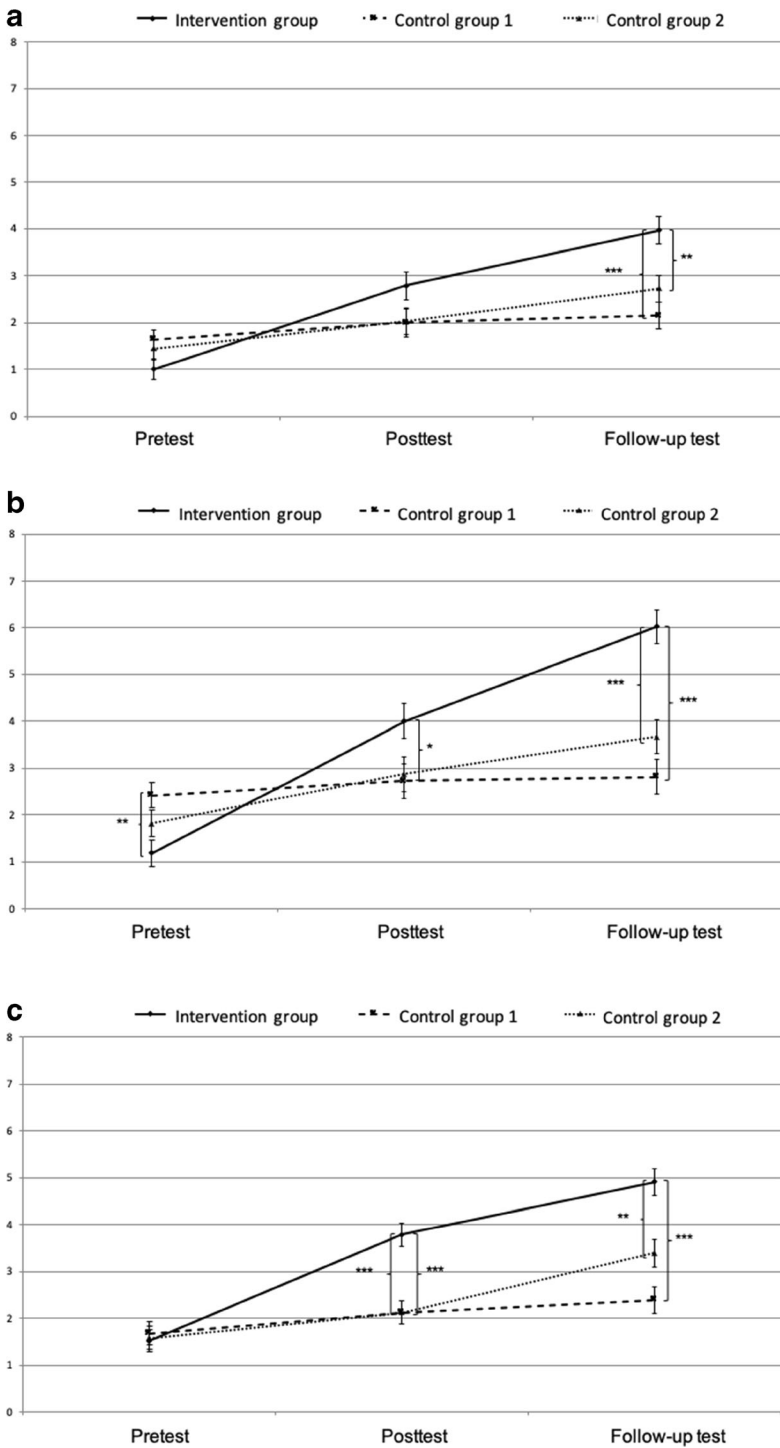
### 5.2.1 Measured values

The main effect of time of testing is significant ( $p < .001$ ) and the main effect of group is not significant ( $p = .093$ ) (see Table 2). The interaction effect is significant ( $p < .001$ ). The effect size of the interaction ( $\eta_p^2 = .204$ ) is to be classified as strong (Cohen, 1988). The pairwise comparisons with the Bonferroni correction do not show significant differences ( $ps > .05$ ) at the pretest and posttest time. The follow-up test shows a significant difference between the intervention group and control group 1 ( $p < .001$ ) and between the intervention group and control group 2 ( $p = .009$ ). There is no significant difference between the two control groups ( $p > .05$ ) (see Fig. 4a).

Hypothesis H1(b) can be confirmed for the follow-up test time. Three months after the intervention, the intervention group more frequently observes the mathematical matching of the measured values in the documents than the control groups.

**Table 4** Mean values and standard deviations with regard to the mathematical matching ( $N = 33$  in each group)

		Pretest	Posttest	Follow-up test
Measured values	Intervention group	<i>M</i> 1.00 ( <i>SD</i> .79)	<i>M</i> 2.79 ( <i>SD</i> 1.64)	<i>M</i> 3.97 ( <i>SD</i> 1.38)
	Control group 1	<i>M</i> 1.64 ( <i>SD</i> 1.48)	<i>M</i> 2.00 ( <i>SD</i> 1.82)	<i>M</i> 2.15 ( <i>SD</i> 1.86)
	Control group 2	<i>M</i> 1.45 ( <i>SD</i> 1.25)	<i>M</i> 2.03 ( <i>SD</i> 1.61)	<i>M</i> 2.73 ( <i>SD</i> 1.68)
Measuring units	Intervention group	<i>M</i> 1.18 ( <i>SD</i> 1.21)	<i>M</i> 4.00 ( <i>SD</i> 2.11)	<i>M</i> 6.03 ( <i>SD</i> 1.51)
	Control group 1	<i>M</i> 2.42 ( <i>SD</i> 2.05)	<i>M</i> 2.73 ( <i>SD</i> 2.28)	<i>M</i> 2.82 ( <i>SD</i> 2.44)
	Control group 2	<i>M</i> 1.82 ( <i>SD</i> 1.38)	<i>M</i> 2.88 ( <i>SD</i> 1.95)	<i>M</i> 3.67 ( <i>SD</i> 2.16)
Operations	Intervention group	<i>M</i> 1.52 ( <i>SD</i> 1.35)	<i>M</i> 3.79 ( <i>SD</i> 1.19)	<i>M</i> 4.91 ( <i>SD</i> 1.63)
	Control group 1	<i>M</i> 1.67 ( <i>SD</i> 1.45)	<i>M</i> 2.12 ( <i>SD</i> 1.22)	<i>M</i> 2.39 ( <i>SD</i> 1.50)
	Control group 2	<i>M</i> 1.58 ( <i>SD</i> 1.44)	<i>M</i> 2.12 ( <i>SD</i> 1.75)	<i>M</i> 3.39 ( <i>SD</i> 1.89)



**Fig. 4** Group interaction with regard to the appropriate measured values (a), the appropriate measuring units (b) and the appropriate operations (c) ( $N = 33$  each group; \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ )

## 5.2.2 Measuring units

Both main effects are statistically significant at the .05 significance level (see Table 2). The interaction effect is significant ( $p < .001$ ) too. The effect size of the interaction ( $\eta_p^2 = .337$ ) is to be classified as strong (Cohen, 1988). The pairwise comparisons with the Bonferroni correction reveal a significant difference between the intervention group and control group 1 at the pretest time ( $p = .006$ ). The documents of the children in the intervention group have the lowest values for the matching of measuring units. There is also a significant difference between the intervention group and control group 1 at the posttest time ( $p = .049$ ). The disadvantage for the intervention group observed during the pretest turns into an advantage at the posttest time. The follow-up test reveals significant differences between the intervention group and control group 1 ( $p < .001$ ) and between the intervention group and control group 2 ( $p < .001$ ). The remaining pair comparisons show no significant differences ( $ps > .05$ ) (see Fig. 4b). Hypothesis H1(b) can be confirmed for the follow-up test time. Three months after the intervention, the intervention group more frequently observes the mathematical matching of the measuring units in the documents than the control groups.

## 5.2.3 Operations

Regarding the operations, both main effects are statistically significant at the .05 significance level (see Table 2). The interaction effect is significant ( $p < .001$ ) too. The effect size of the interaction ( $\eta_p^2 = .233$ ) is classified as strong (Cohen, 1988). The pairwise comparisons with the Bonferroni correction reveal significant differences for the posttest time between the intervention group and both control groups ( $ps < .001$ ). There are also significant differences between the intervention group and control group 1 ( $p < .001$ ) and between the intervention group and control group 2 ( $p = .001$ ) at the follow-up test time. The remaining pair comparisons show no significant differences ( $ps > .05$ ) (see Fig. 4c). Hypothesis H1(b) can be confirmed. After the intervention, the intervention group more frequently observes the mathematical matching of the operations in the documents than the control groups.

## 5.3 Degree of abstraction

The two indicators of the degree of abstraction are tested individually. Table 5 shows the development of the means of the groups in comparison. Since the Mauchly test proves that sphericity is violated, the Greenhouse-Geisser correction was utilized.

### 5.3.1 Indicator 1

The main effect of testing time is significant ( $p < .001$ ) and the main effect of group is not significant ( $p = .157$ ). The interaction effect is significant ( $p = .023$ ) (see Table 2). The effect size of the interaction ( $\eta_p^2 = .060$ ) is to be classified as moderate (Cohen, 1988). The pairwise comparisons

**Table 5** Mean values and standard deviations with regard to the degree of abstraction ( $N = 33$  in each group)

		Pretest	Posttest	Follow-up test
Indicator 1 (Focus on the structurally relevant objects)	Intervention group	$M 2.85 (SD 1.66)$	$M 4.39 (SD 1.30)$	$M 4.88 (SD 1.08)$
	Control group 1	$M 3.03 (SD 1.74)$	$M 3.39 (SD 1.98)$	$M 3.82 (SD 1.98)$
	Control group 2	$M 3.21 (SD 1.64)$	$M 3.58 (SD 1.82)$	$M 4.58 (SD 1.86)$
Indicator 2 (Focus on their mathematically relevant qualities)	Intervention group	$M 2.00 (SD 1.54)$	$M 2.73 (SD 1.46)$	$M 2.94 (SD 1.17)$
	Control group 1	$M 1.91 (SD 1.42)$	$M 2.30 (SD 1.49)$	$M 2.64 (SD 1.56)$
	Control group 2	$M 1.88 (SD 1.58)$	$M 2.67 (SD 1.95)$	$M 2.70 (SD 1.94)$

with the Bonferroni correction do not show significant differences ( $ps > .05$ ) at the pretest and posttest times. The follow-up test time shows a significant difference between the intervention group and control group 1 ( $p = .036$ ). The remaining pair comparisons at the follow-up point show no significant differences ( $ps > .05$ ). Hypothesis H1(c) has to be partially rejected for indicator 1. After the intervention, the intervention group does not more frequently focus on the structurally objects in the graphic representations compared with the control groups, with the exception of control group 1 at the follow-up test.

### 5.3.2 Indicator 2

All effects are not statistically significant at the 0.5 significance level, except for the main effect of testing time (see Table 2). Hypothesis H1(c) has to be rejected for indicator 2. After the intervention, the intervention group does not more frequently focus on the mathematically relevant qualities of the structurally relevant objects compared with the control groups. Descriptive results suggest that it is the items that are decisive for the degree of abstraction, but not the mathematical structure (Ott, 2016).

## 5.4 Solution rates

The Mauchly test proves that sphericity is fulfilled. All effects are not statistically significant at the .05 significance level, except for the main effect of testing time (see Table 2). Hypothesis H2 has to be rejected. The intervention group does not correctly solve word problems more often after the intervention than the control groups.

Descriptive results (Ott, 2016) show that in the intervention group, the continuous increase in correct solutions is accompanied by a decrease in the proportion of false solutions and in the proportion of documents in which no solution is specified. This picture does not appear in the control groups.

## 6 Discussion

### 6.1 Summary of the results

The results show a differentiated picture with regard to the key features of graphic representations. Since the intervention group did not experience any other support in problem solving or graphic representation than the intervention itself during the period of the study, the results likely suggest that the findings are a result of the intervention.

#### 6.1.1 Mathematical structure and mathematical matching

At the time of the pretest, the mean values for the representation of mathematical relationships in all three groups are low. This finding confirms earlier observations (Hasemann, 2006; Ott, 2016). After the intervention, the intervention group more often generates object-related and, above all, more diagrammatic graphic representations. The findings on mathematical matching with regard to the operations show that the intervention group after the intervention not only more frequently represents mathematical relationships to a noteworthy extent than the control groups but also pays significantly more attention to mathematical matching concerning the operations given in the word problems. The learners also pay more attention to the depiction of the given measured values and measuring units.



Here, too, they differ significantly from the control groups. Since the interaction effect is significant, these developments can be attributed to the intervention. The word problems alone cannot be regarded as decisive for the development, due to the significant difference to control group 1. The level of the interaction effect shows that the differences between the groups are also practically significant. Since the development does not stop after the intervention, it can be regarded as sustainable.

These findings are in line with those of Van Essen and Hamaker (1990), who report that learner-generated drawings are richer and more oriented towards mathematical relationships after a short period of training. In the study presented here, the richness is reflected in an increased representation of the measured values and measuring units. These results also complement findings by Van Dijk et al. (2003b): a combination of the self-generation of graphic representations and processes of reflection on them in class may promote the graphic representation of mathematical structures with mathematical matching to the given word problem.

### 6.1.2 Degree of abstraction

Regarding the degree of abstraction, the findings do not meet expectations. The intervention group does not pay attention to a higher level of abstraction after the intervention than the control groups. A low degree of abstraction is a typical way of drawing among children (Sherin, 2000). However, the result is in contrast with the findings of other studies (Lopez Real & Veloo, 1993; Van Dijk et al., 2003a; Van Essen & Hamaker, 1990) in which the graphic representations become more formalized and schematic. The results must be interpreted against the background of the intervention: the children were asked to record everything of importance to them for solving the word problem. The graphic representation should be understandable for someone else. When representations are made for others, they tend to be richer (Cox, 1999). The children seem to consider realistic drawings to be more understandable for others. This complements the findings of Pantziara et al. (2009) who found that learners tried to transform schematic representations into pictorial representations to interpret them. No statement can be made concerning the degree of abstraction if the graphic representations had been created as heuristics.

However, with a relatively constant degree of abstraction, the children in the intervention group paid more attention to the mathematical structure and matching after the intervention. It can be interpreted in such a way that children do not neglect the content of a word problem but establish more flexible mathematical relationships. The results suggest that the development of the degree of abstraction is independent of the representation of the mathematical structure and matching. This should be examined in further investigations. Nevertheless, it must be taken into account that mathematically irrelevant information can make it difficult to recognize mathematical structures and therefore poses an obstacle for problem solving (Presmeg, 1986). This could provide an explanation of the results with regard to the solution rates.

### 6.1.3 Solution rates

The intervention group does not correctly solve the word problems significantly more often than the control groups. This is in contrast with the results of Van Dijk et al. (2003a), in whose study designing models using a co-construction approach had a positive effect. However, in the study presented here, the graphic representations were generated as documents for others and not as instruments. In addition, in the test processing and in the intervention, the focus was not on problem solving. This study shows that generating a mathematically matching diagrammatic representation does not guarantee that the problem will be solved correctly. This is in line with results of Van Essen and Hamaker (1990) and

could be interpreted in that the children regard the graphic representation and the task solution as independent from each other. One reason for this can be seen in the instruction, which does not use graphic representations as a heuristic. A further cause can be seen in the intervention, in which the focus was not on the connection between graphic representation and correct solutions. Further research with a modified instruction, which includes problem solving to a greater extent, is needed. According to Hembree (1992), another reason could be the children's age: positive effects on problem solving through interventions in graphic representation do not yet occur at primary school age.

## 6.2 Limitations

No study comes without limitations. First, the sample of the study is not globally representative, but can be regarded as characteristically representative with regard to graphic representations in mathematics lessons (Bortz & Döring, 2006). Second, for the development of representation skills, the time frame set for practical reasons with nine intervention units is relatively short. The multi-group design makes it possible to attribute effects to the intervention and to control for possible disturbance variables, such as age-related development, habituation to the test, the influence of the intervention tasks themselves, and class composition. Nevertheless, the possibility of interference cannot be avoided in a study under almost real conditions. The development of control group 1 could be influenced by such effects. To minimize external influences, the intervention group was taught by the author. It is known that studies of this kind show slightly higher effects (De Boer, Donker, & Van der Werf, 2014). In the future, an independent replication should show whether the same effects occur when teachers carry out the intervention themselves. Here, a follow-up test was carried out to analyze long-term effects. In addition, it would make sense to examine the children's class discussions more closely for gaining a better understanding of the development processes. The qualitative part of the study presented here provides some information in this regard (Ott, 2016, 2020). Furthermore, in the results, ceiling effects are partly observable, as also expressed in the change to the scatter over time. This is due to the fact that the test does not differentiate any further at the upper end. Finally, no statement can be made as to the extent to which the children use graphic representations as heuristics for problem solving.

## 7 Conclusion

In the intervention presented here, learners of the 3rd grade first generated their own graphic representations for a given word problem as documents for others and then reflected on some of these learner-generated graphic representations in class. The aim of the reflection was to collectively explain the shown graphic representation and thus to try to understand the point of view of the author of the graphic representation. The results suggest that learners are enabled by this intervention to more frequently pay attention to a mathematically appropriate structure in their graphic representations for word problems. The results show that this effect is also sustainable. Even if the solution rates did not improve to a statistically significant extent, the basis for the use of graphic representations as an instrument for problem solving can still be laid. All in all, teaching that combines the self-generation of graphic representations and reflection processes on them in class seems to be positive for the development of graphic representation competences.

**Funding** Open access funding provided by St.Gallen University of Teacher Education.

# Appendix

**Table 6** Items of the paper pencil test (italic) and the intervention (normal) ordered by mathematical structure, type of word problem, and similarity of formulation

Type of word problem		Type A	Type B	Type C
Mathematical structure	Two-dimensional figure	<p>Similar formulation</p> <p><i>A picture is 20 cm wide and 30 cm high. It is glued onto a large piece of paper in such a way that a 15 cm wide margin remains on all sides. How wide and how high must the large paper be?</i></p> <p>From a square meadow each side is 50 m long. It is divided by a fence into four squares of equal size. How long are the sides of the small squares? How many meters of fence are needed to fence in all the small squares?</p> <p><i>Mama is sewing on a jacket from the bottom to the top new buttons to close. She needs nine buttons. Between the places where she sews the buttons on, she always leaves a gap of 5 cm. How long is the zipper on the jacket?</i></p>	<p>A flower bed is 40 cm wide and 60 cm long. Small stones are scattered around the bed to create a 20 cm wide path. How wide and how long is the bed with the path?</p>	
	(a-1)•b-problem			
	Comparison	<p>Steffi has 24 fishes. She has 8 fishes more than Thomas. How many fishes does Thomas have?</p>	<p>Max is 78 cm tall. If he climbs up on a 20 cm high box, he's the same height as Klaus. Paul is 15 cm smaller than Klaus. How tall is Paul?</p>	<p>Similar formulation</p> <p><i>In a race, Tina takes 74 s. Lara needs 12 s less than Tina. Pia takes 84 s How many seconds does Sara needs 57 min for her homework. Jule needs 26 min less than Sara. Andi needs 69 min.</i></p>

Table 6 (continued)

Type of word problem		Type B	Type C
Type A			
Linear equation		<p>Similar formulation</p> <p>A spruce grows about 15 cm per year. In the garden there is an 83 cm high spruce. How old is it?</p> <p>Workers build a high-rise building. Every week, they make 3 m. The high-rise building is now 23 m high. How long have the workers been building it?</p>	<p>Lara need? How many seconds does Pia take longer than Lara?</p> <p>does Jule need? How many minutes does Andi take longer than Sara?</p> <p>Lisa's selling at the flea market. She makes twice as much money on Saturday as she does on Sunday. In total, she earns 21 €. How much money does she earn on Sunday?</p>
System of linear equations	<i>The children put their reading folders on the shelf. 5 folders next to</i>	<p>Bastian tinkers 7 easter bunnies. Anna tinkers twice as many. Franziska tinkers four times as many Easter bunnies as Bastian. How many easter bunnies do they tinker altogether?</p> <p>Sabine is knitting a scarf. In the morning she knits 7 cm. In the afternoon she knits three times as much. In the evening she knits another 2 cm. How long is her scarf at the end of the day?</p>	<p>Christian is practising swimming. The first day he swims 6 min. Then every day he swims 5 min longer than the day before. How long does he swim the fourth day?</p>

Table 6 (continued)

Type of word problem		Type A	Type B	Type C
Linear equation with two unknowns	Similar formulation	<p><i>each other are 35 cm wide. How wide are 8 folders?</i></p> <p><i>The kit contains 24 wheels. With these, cars and tricycles are to be built. How many cars and tricycles can be built, so that there's no wheel left?</i></p>	<p>In the stable I count 22 legs. They belong to pigs and chickens. How many pigs and chickens can be in the stable?</p>	

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* [Research methods and evaluation for human and social scientists] (4ed.). Heidelberg, Germany: Springer.
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* [Statistics for human and social scientists] (7ed.). Berlin, Heidelberg: Springer.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2ed.). USA: Lawrence Erlbaum Associates.
- Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction*, 9, 343–363.
- De Boer, H., Donker, A. S., & Van der Werf, M. P. C. (2014). Effects of the attributes of educational interventions on students' academic performance: A meta-analysis. *Review of Educational Research*, 84(4), 509–545.
- Diezmann, C. (2002). Enhancing students' problem solving through diagram use. *Australian Primary Mathematics Classroom*, 7(3), 4–8.
- Dörfler, W. (2008). Mathematical reasoning: Mental activity or practice with diagrams. In M. Niss (Ed.), *ICME 10 proceedings, regular lectures, CD-Rom* (p. 17). Roskilde, Denmark: Roskilde University and IMFUFA.
- Fagnant, A., & Vlassis, J. (2013). Schematic representations in arithmetical problem solving: Analysis of their impact on grade 4 students. *Educational Studies in Mathematics*, 84, 149–168.
- Freudenthal, H. (1991). *Revisiting mathematics education*. Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Goldin, G. A., & Shteingold, N. (2001). Systems of representations and the development of mathematical concepts. In A. Cuoco & F. R. Curcio (Eds.), *The roles of representation in school mathematics* (pp. 1–23). Reston, VA: National Council of Teachers of Mathematics.
- Hasemann, K. (2006). Rechengeschichten und Textaufgaben - Vorgehensweisen, Darstellungsformen und Einsichten von Kindern am Ende des 2. Schuljahres [Arithmetic stories and word problems - procedures, forms of representation and insights of children at the end of the 2nd school year]. In E. Rathgeb-Schnierer & U. Roos (Eds.), *Wie rechnen Matheprofis? Ideen und Erfahrungen zum offenen Mathematikunterricht* (pp. 15–26). Munich, Germany: Oldenbourg.
- Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology*, 91(4), 684–689.
- Heinze, A., Star, J. R., & Verschaffel, L. (2009). Flexible and adaptive use of strategies and representations in mathematics education. *Zentralblatt für Didaktik der Mathematik*, 41, 535–540.
- Hembree, R. (1992). Experiments and relational studies in problem solving: A meta-analysis. *Journal for Research in Mathematics Education*, 23(3), 242–273.
- Kirsch, A. (1997). *Mathematik wirklich verstehen* [Really understand mathematics] (3ed.). Cologne, Germany: Aulis-Verlag Deubner.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65–99.
- Lopez Real, F., & Veloo, P. K. (1993). Children's use of diagrams as a problem-solving strategy. In N. Hirabayashi, & L. Shigematsu (Eds.), *Proceedings of the Seventeenth International Conference for the Psychology of Mathematics Education* (pp. 169–176). Tsukuba, Japan: University of Tsukuba.
- NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Ott, B. (2016). *Textaufgaben grafisch darstellen: Entwicklung eines Analyseinstruments und Evaluation einer Interventionsmaßnahme* [Graphic representations for word problems: Development of an analysis instrument and evaluation of an intervention measure.]. Münster, Germany: Waxmann.

- Ott, B. (2017). Children's drawings for word problems – Design of a theory and an analysis tool. In T. Dooley, & G. Gueudet (Eds.), *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education* (CERME10, February 1–5, 2017) (pp. 3984–3991). Dublin, Ireland: DCU Institute of Education and ERME.
- Ott, B. (2018). 'Mach doch einfach mal 'ne Skizze!' Grafisches Darstellen im Sachrechnen [Just make a sketch!] Graphic representation for word problems]. In E. Plackner & J. Postupa (Eds.), *MaMut – Veranschaulichen* (pp. 7–30). Hildesheim, Germany: Franzbecker.
- Ott, B. (2020). Über Darstellungen reflektieren: Darstellungswechsel in der Primarstufe fördern [Reflecting on representations: Encourage a change of representation at the primary level]. In G. Kadunz (Ed.), *Zeichen und Sprache im Mathematikunterricht* (pp. 113–147). Berlin, Germany: Springer Spektrum.
- Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 259–303). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pantziara, M., Gagatsis, A., & Elia, I. (2009). Using diagrams as tool for the solution of non-routine mathematical problems. *Educational Studies in Mathematics*, 72, 39–60.
- Peschek, W. (1988). Untersuchungen zur Abstraktion und Verallgemeinerung [Studies on abstraction and generalization]. In W. Dörfler (Ed.), *Kognitive Aspekte mathematischer Begriffsentwicklung* (pp. 127–190). Vienna, Austria: Hölder-Pichler-Tempsky.
- Premeg, N. (1986). Visualisation and mathematical giftedness. *Educational Studies in Mathematics*, 17(3), 297–311.
- Rellensmann, J., Schukajlow, S., & Leopold, C. (2017). Make a drawing. Effects of strategic knowledge, drawing accuracy, and type of drawing on students' mathematical modelling performance. *Educational Studies in Mathematics*, 95, 53–78.
- Rinkens, H. D. (1973). *Abstraktion und Struktur: Grundbegriffe der Mathematikdidaktik* [Abstraction and structure: Basic concepts of mathematics didactics]. Ratingen, Germany: Henn.
- Roth, W.-M., & McGinn, M. K. (1998). Inscriptions: Toward a theory of representing as social practice. *Review of Educational Research*, 68(1), 35–59.
- Schipper, W. (2009). *Handbuch für den Mathematikunterricht an Grundschulen [Manual for the teaching of mathematics in primary schools]*. Braunschweig, Germany: Schroedel.
- Schnotz, W. (2001). Wissenserwerb mit Multimedia [Knowledge acquisition with multimedia]. *Unterrichtswissenschaft*, 29(4), 292–318.
- Schnotz, W. (2002). Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, 14(1), 101–120.
- Schülke, C. (2013). *Mathematische Reflexion in der Interaktion von Grundschulkindern* [Mathematical reflection in the interaction of primary school children]. Münster, Germany: Waxmann.
- Selter, C. (1993). *Eigenproduktionen im Arithmetikunterricht der Primarstufe* [Going one's own way in arithmetic lessons at primary level]. Wiesbaden: DUV.
- Sherin, B. L. (2000). How students invent representation of motion: A genetic account. *The Journal of Mathematical Behavior*, 19(4), 399–441.
- Stern, E. (1998). *Die Entwicklung des mathematischen Verständnisses im Kindesalter* [The development of mathematical understanding in childhood]. Lengerich, Germany: Pabst.
- Stern, E., Aprea, C., & Ebner, H. G. (2003). Improving cross-content transfer in text processing by means of active graphical representations. *Learning and Instruction*, 13(2), 191–203.
- Sturm, N. (2018). *Problemhaltige Textaufgaben lösen* [Solving problematic word problems]. Wiesbaden, Germany: Springer Spektrum.
- Van Dijk, I. M. A. W., Van Oers, B., & Terwel, J. (2003a). Providing or designing? Constructing models in primary maths education. *Learning and Instruction*, 13, 53–72.
- Van Dijk, I. M. A. W., Van Oers, B., Terwel, J., & Van den Eeden, P. (2003b). Strategic learning in primary mathematics education: Effects of an experimental program in modelling. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 9(2), 161–187.
- Van Essen, G., & Hamaker, C. (1990). Using self-generated drawings to solve arithmetic word problems. *Journal of Educational Research*, 83(6), 301–312.
- Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review*, 17(4), 285–325.
- Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse, the Netherlands: Swets & Zeitlinger Publishers.